

Katherine MUNRO
Stefan PAPP
Zoltan TOTH
Wolfgang WEIDINGER
Danko NIKOLIĆ

3. Auflage

HANDBUCH DATA SCIENCE UND KI

Mit Machine Learning
und Datenanalyse
Wert aus Daten generieren

HANSER

Munro / Papp / Toth / Weidinger / Nikolić
Antosova Vesela / Bruckmüller / Cadonna / Eder / Gorzala / Hahn / Langs /
Licandro / Mata / McIntyre / Meir-Huber / Móra / Pasieska / Rugli / Wazir / Zauner

Handbuch Data Science und KI



Bleiben Sie auf dem Laufenden!

Der Hanser Computerbuch-Newsletter informiert Sie regelmäßig über neue Bücher und Termine aus den verschiedenen Bereichen der IT. Profitieren Sie auch von Gewinnspielen und exklusiven Leseproben. Gleich anmelden unter

www.hanser-fachbuch.de/newsletter

Katherine Munro, Stefan Papp, Zoltan C. Toth,
Wolfgang Weidinger, Danko Nikolić,
Barbora Antosova Vesela, Karin Bruckmüller,
Annalisa Cadonna, Jana Eder, Jeannette Gorzala,
Gerald A. Hahn, Georg Langs, Roxane Licandro,
Christian Mata, Sean McIntyre, Mario Meir-Huber,
György Móra, Manuel Pasieska, Victoria Rugli,
Rania Wazir, Günther Zauner

Handbuch Data Science und KI

Mit Machine Learning und Datenanalyse Wert aus
Daten generieren

3., aktualisierte und erweiterte Auflage

HANSER



Print-ISBN: 978-3-446-47937-1

E-Book-ISBN: 978-3-446-48072-8

Epub-ISBN: 978-3-446-48357-6

Alle in diesem Werk enthaltenen Informationen, Verfahren und Darstellungen wurden zum Zeitpunkt der Veröffentlichung nach bestem Wissen zusammengestellt. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund sind die im vorliegenden Werk enthaltenen Informationen für Autor:innen, Herausgeber:innen und Verlag mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor:innen, Herausgeber:innen und Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Weise aus der Benutzung dieser Informationen – oder Teilen davon – entsteht. Ebenso wenig übernehmen Autor:innen, Herausgeber:innen und Verlag die Gewähr dafür, dass die beschriebenen Verfahren usw. frei von Schutzrechten Dritter sind. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt also auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benützt werden dürften.

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers (m/w/d) verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

Die endgültige Entscheidung über die Eignung der Informationen für die vorgesehene Verwendung in einer bestimmten Anwendung liegt in der alleinigen Verantwortung des Nutzers.

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.d-nb.de> abrufbar.

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdruckes und der Vervielfältigung des Werkes, oder Teilen daraus, vorbehalten. Kein Teil des Werkes darf ohne schriftliche Einwilligung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder einem anderen Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung – mit Ausnahme der in den §§ 53, 54 UrhG genannten Sonderfälle –, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Wir behalten uns auch eine Nutzung des Werks für Zwecke des Text- und Data Mining nach § 44b UrhG ausdrücklich vor.

© 2025 Carl Hanser Verlag GmbH & Co. KG, München
Kolbergerstraße 22 | 81679 München | info@hanser.de
www.hanser-fachbuch.de

Lektorat: Sylvia Hasselbach

Copy editing: Jürgen Dubau, Freiburg/Elbe

Coverkonzept: Marc Müller-Bremer, www.rebranding.de, München

Covergestaltung: Tom West

Titelmotiv: [gettyimages/ValeryBrozhinsky](https://www.gettyimages.com)

Herstellung: le-tex publishing services GmbH, Leipzig

Satz: Eberl & Koesel Studio, Kempten

Druck: CPI books GmbH, Leck

Printed in Germany

Inhalt

Vorwort	XXIII
Danksagungen	XXV
1 Einführung	1
<i>Stefan Papp</i>	
1.1 Über dieses Buch	1
1.2 Die Halford Group	2
1.2.1 Alice Halford – Vorsitzende	3
1.2.2 Analysten	5
1.2.3 „CDO“	6
1.2.4 Vertrieb	8
1.2.5 IT	9
1.2.6 Sicherheit	11
1.2.7 Leiter der Produktion	11
1.2.8 Kundenbetreuung	13
1.2.9 HR	14
1.2.10 CEO	15
1.3 Kurz und bündig	16
2 Das A und O der KI	19
<i>Stefan Papp</i>	
2.1 Die Datenverwendungszwecke	20
2.1.1 Bias	20
2.1.2 Datenkompetenz	23

2.2	Kulturschock	25
2.3	Ideenfindung	29
2.4	Entwurfsprozessmodelle	32
2.4.1	Design Thinking	32
2.4.2	Double Diamond	33
2.4.3	Durchführung von Workshops	34
3	Cloud-Dienste	44
	<i>Stefan Papp</i>	
3.1	Einführung	45
3.2	Cloud-Essentials	45
3.2.1	XaaS	48
3.2.2	Cloud-Anbieter	49
3.2.3	Native Cloud-Dienste	51
3.2.4	Cloud-native Paradigmen	54
3.3	Infrastructure as a Service	56
3.3.1	Hardware	57
3.3.2	Verteilte Systeme	59
3.3.3	Linux Essentials für Datenexperten	63
3.3.4	Infrastructure as Code	70
3.4	Platform as a Service	75
3.4.1	Cloud Native PaaS-Lösungen	76
3.4.2	Externe Lösungen	81
3.5	Software as a Service	84
3.6	Kurz und bündig	85
4	Datenarchitektur	87
	<i>Zoltan C. Toth und Sean McIntyre</i>	
4.1	Übersicht	87
4.1.1	Maslowsche Bedürfnishierarchie für Daten	88
4.1.2	Anforderungen an die Datenarchitektur	89
4.1.3	Die Struktur einer typischen Datenarchitektur	90
4.1.4	ETL (Extrahieren, Transformieren, Laden)	95
4.1.5	ELT (Extrahieren, Laden, Transformieren)	96
4.1.6	ETLT	97
4.2	Datenerfassung und -integration	97

4.2.1	Datenquellen	98
4.2.2	Traditionelle Dateiformate	100
4.2.3	Moderne Dateiformate	102
4.2.4	Welche Speicheroption soll ich wählen?	105
4.3	Data Warehouses, Data Lakes und Lakehouses	105
4.3.1	Data Warehouses	106
4.3.2	Data Lakes und Cloud-Datenplattformen	110
4.4	Data Transformation	114
4.4.1	SQL	116
4.4.2	Big Data & Apache Spark	127
4.4.3	Cloud-Datenplattformen für Apache Spark	136
4.5	Workflow-Orchestrierung	138
4.5.1	Dagster und der Modern Data Stack	141
4.6	Ein Anwendungsfall und seine Datenarchitektur	142
4.7	Kurz und bündig	148
5	Data Engineering	150
	<i>Stefan Papp</i>	
5.1	Abgrenzung zum Software Engineering	151
5.2	Programmiersprachen	153
5.2.1	Code oder kein Code?	153
5.2.2	Auswahl der Programmiersprache	155
5.2.3	Python	156
5.2.4	Scala	160
5.3	Software-Engineering-Prozesse für Daten	162
5.3.1	Konfigurationsmanagement	163
5.3.2	CI/CD	164
5.4	Datenpipelines	166
5.4.1	Gemeinsame Merkmale einer Datenpipeline	167
5.4.2	Datenpipelines in der Unified Data Architecture	167
5.5	Speicheroptionen	172
5.5.1	Datei-Ära	172
5.5.2	Datenbank-Ära	173
5.5.3	Data-Lake-Ära	175
5.5.4	Serverless-Ära	176

5.5.5	Polyglotte Speicherung	177
5.5.6	Data-Mesh-Ära	178
5.6	Tooling	180
5.6.1	Batch: Airflow	180
5.6.2	Streaming: Kafka	182
5.6.3	Transformation: Databricks Notebooks	187
5.7	Gemeinsame Herausforderungen	189
5.7.1	Datenqualität und unterschiedliche Standards	189
5.7.2	Skewed Data	191
5.7.3	Überlastete operationelle Systeme	192
5.7.4	Operative Legacy-Systeme	193
5.7.5	Plattform- und Informationssicherheit	193
5.8	Kurz und bündig	194
6	Data Governance	195
	<i>Victoria Rugli, Mario Meir-Huber</i>	
6.1	Warum brauchen wir Data Governance?	195
6.1.1	Beispiel 1: Mit Data Governance Klarheit schaffen	197
6.1.2	Beispiel 2: Die (negativen) Auswirkungen einer mangelhaften Data Governance	198
6.2	Die Bausteine der Data Governance	199
6.2.1	Data Governance erklärt	200
6.3	Menschen	202
6.3.1	Data Ownership	203
6.3.2	Data Stewards	206
6.3.3	Data Governance Board	208
6.3.4	Change Management	210
6.4	Prozess	212
6.4.1	Verwaltung von Metadaten	213
6.4.2	Management der Datenqualität	217
6.4.3	Datensicherheit und Datenschutz	221
6.4.4	Stammdatenmanagement	225
6.4.5	Datenzugang und Suche	229
6.5	Technologie (Datenmanagement-Tools)	232
6.5.1	Open-Source-Tools	232

6.5.2	Cloud-basierte Data-Governance-Tools	239
6.6	Kurz und bündig	244
7	Machine Learning Operations (ML Ops)	245
	<i>Zoltan C. Toth, György Móra</i>	
7.1	Übersicht	245
7.1.1	Umfang von MLOps	246
7.1.2	Datenerhebung und Exploration	247
7.1.3	Feature Engineering	247
7.1.4	Modelltraining	248
7.1.5	In der Produktion eingesetzte Modelle	249
7.1.6	Bewertung des Modells	249
7.1.7	Model Understanding	250
7.1.8	Modellversionierung	250
7.1.9	Modellüberwachung	251
7.2	MLOps in einer Organisation	251
7.2.1	Die wichtigsten Vorteile von MLOps	252
7.2.2	Erforderliche Fähigkeiten für MLOps	252
7.3	Verschiedene gängige Szenarien im MLOps-Bereich	253
7.3.1	Integration von Notebooks	253
7.3.2	Features in der Produktion	255
7.3.3	Modelleinsatz	257
7.3.4	Modellformate	258
7.4	MLOps-Werkzeuge und MLflow	259
7.4.1	MLflow	260
7.5	Kurz und bündig	263
8	Cybersicherheit von Machine-Learning-Systemen	264
	<i>Manuel Pasiëka</i>	
8.1	Einführung in die Cybersicherheit	265
8.2	Angriffsfläche	267
8.3	Angriffsmethoden	268
8.3.1	Model Stealing	268
8.3.2	Datenextraktion	271
8.3.3	Data Poisoning	273
8.3.4	Adversariale Angriffe	276

8.3.5	Backdoor-Angriffe	279
8.4	Sicherheit von großen Sprachmodellen durch maschinelles Lernen	282
8.4.1	Datenextraktion	283
8.4.2	Jailbreaking	284
8.4.3	Prompt Injection	286
8.5	KI-Bedrohungsmodellierung	289
8.6	Verordnungen	291
8.7	Wie geht es jetzt weiter?	293
8.8	Zusammenfassung	295
8.9	Kurz und bündig	296
9	Mathematik	297
	<i>Annalisa Cadonna</i>	
9.1	Lineare Algebra	298
9.1.1	Vektoren und Matrizen	299
9.1.2	Operationen zwischen Vektoren und Matrizen	303
9.1.3	Lineare Transformationen	305
9.1.4	Eigenwerte, Eigenvektoren und Eigendekomposition	306
9.1.5	Andere Matrixzerlegungen	308
9.2	Kalkulus und Optimierung	310
9.2.1	Ableitung	311
9.2.2	Gradient und Hessian	313
9.2.3	Gradientenabstieg	315
9.2.4	Eingeschränkte Optimierung	317
9.3	Wahrscheinlichkeitsrechnung	318
9.3.1	Diskrete und kontinuierliche Zufallsvariablen	319
9.3.2	Erwartungswert, Varianz und Kovarianz	323
9.3.3	Unabhängigkeit, bedingte Verteilungen und Bayes-Theorem	325
9.4	Kurz und bündig	327
10	Statistik – Grundlagen	329
	<i>Rania Wazir, Georg Langs, Annalisa Cadonna</i>	
10.1	Daten	331
10.2	Einfache lineare Regression	332
10.3	Multiple lineare Regression	341
10.4	Logistische Regression	343

10.5	Wie gut ist unser Modell?	352
10.6	Kurz und bündig	353
11	Business Intelligence (BI)	355
	<i>Christian Mata</i>	
11.1	Einführung in Business Intelligence	358
11.1.1	Definition von Business Intelligence	358
11.1.2	Rolle in Organisationen	359
11.1.3	Entwicklung von Business Intelligence	360
11.1.4	Data Science und KI im Kontext von BI	362
11.1.5	Daten für die Entscheidungsfindung	366
11.1.6	Verstehen des geschäftlichen Kontextes	368
11.1.7	Business-Intelligence-Aktivitäten	371
11.2	Grundlagen des Datenmanagements	373
11.2.1	Was sind Datenmanagement, Datenintegration und Data Warehousing?	374
11.2.2	Datenbeladung – Der Fall von ETL oder ELT	375
11.2.3	Datenmodellierung	377
11.3	Reporting und Datenanalyse	385
11.3.1	Reporting	385
11.3.2	Berichtsarten	389
11.3.3	Datenanalyse	390
11.3.4	Visuelle Datenanalyse	392
11.3.5	Trends in Reporting und Datenanalyse	394
11.4	BI-Technologien und Werkzeuge	396
11.4.1	Relevante BI-Technologien	396
11.4.2	Verbreitete BI-Werkzeuge (BI-Tools)	401
11.5	BI und Data Science: Ergänzende Disziplinen	405
11.5.1	Unterschiede zwischen BI und DS	405
11.5.2	Gemeinsamkeiten von BI und DS	406
11.5.3	Synergien bei BI und DS	406
11.6	Ausblick für Business Intelligence	408
11.6.1	Erwartungen an die Entwicklung von BI	409
11.7	Kurz und bündig	411

12	Maschinelles Lernen	413
	<i>Georg Langs, Katherine Munro, Rania Wazir</i>	
12.1	Einführung	413
12.2	Grundlegendes: Feature Spaces	415
12.3	Klassifizierungsmodelle	419
12.3.1	K-Nearest-Neighbor-Klassifikator	419
12.3.2	Support Vector Machine	420
12.3.3	Entscheidungsbäume	421
12.4	Ensemble-Methoden	423
12.4.1	Bias und Varianz	424
12.4.2	Bagging: Random Forests	426
12.4.3	Boosten: AdaBoost	430
12.4.4	Die Grenzen der Merkmalskonstruktion und -auswahl	431
12.5	Unüberwachtes Lernen: Lernen ohne Etiketten	432
12.5.1	Clustering	432
12.5.2	Manifold Learning	433
12.5.3	Generative Modelle	434
12.6	Künstliche neuronale Netze und Deep Learning	436
12.6.1	Das Perzeptron	436
12.6.2	Künstliche neuronale Netze	437
12.6.3	Deep Learning	439
12.6.4	Convolutional Neural Networks	440
12.6.5	Training von Convolutional Neural Networks	441
12.6.6	Rekurrente neuronale Netze	444
12.6.7	Long Short-Term Memory Networks	446
12.6.8	Autoencoder und U-Nets	447
12.6.9	Adversariales Training	449
12.6.10	Generative Adversarial Networks	451
12.6.11	Cycle-GANs und Style-GANs	453
12.7	Transformer-Modelle und Aufmerksamkeitsmechanismen	455
12.7.1	Die Transformer-Architektur	455
12.7.2	Was der Aufmerksamkeitsmechanismus leistet	457
12.7.3	Anwendungen von Transformer-Modellen	458
12.8	Reinforcement Learning	459

12.9	Andere Architekturen und Lernstrategien	463
12.10	Validierungsstrategien für Techniken des maschinellen Lernens	463
12.11	Schlussfolgerung	465
12.12	Kurz und bündig	466
13	Großartige künstliche Intelligenz erschaffen	467
	<i>Danko Nikolić</i>	
13.1	Wie KI mit Data Science und maschinellem Lernen zusammenhängt	468
13.2	Eine kurze Geschichte der KI	472
13.3	Fünf Empfehlungen für die Entwicklung einer KI-Lösung	475
13.3.1	Empfehlung Nummer eins: Seien Sie pragmatisch	475
13.3.2	Empfehlung Nummer zwei: Erleichtern Sie Maschinen das Lernen – schaffen Sie induktive Verzerrungen	478
13.3.3	Empfehlung Nummer drei: Analysen durchführen	484
13.3.4	Empfehlung Nummer vier: Hüten Sie sich vor der Skalierungsfalle	487
13.3.5	Empfehlung Nummer fünf: Hüten Sie sich vor der Verallgemeinerungsfalle (so etwas wie ein kostenloses Mittagessen gibt es nicht)	499
13.4	Intelligenz auf menschlicher Ebene	505
13.5	Kurz und bündig	508
14	Signalverarbeitung	510
	<i>Jana Eder</i>	
14.1	Einführung	511
14.2	Abtastung und Quantisierung	512
14.3	Frequenzbereichsanalyse	515
14.3.1	Fourier-Transformation	517
14.4	Rauschunterdrückung und Filtertechniken	523
14.4.1	Rauschunterdrückung mit einem Gaußschen Low-pass-Filter ...	525
14.5	Analyse des Zeitbereichs	527
14.5.1	Signalnormierung und Standardisierung	527
14.5.2	Signaltransformation und Merkmalsextraktion	528
14.5.3	Techniken zur Zerlegung von Zeitreihen	531
14.5.4	Autokorrelation: Verstehen der Signalähnlichkeit über die Zeit ..	534
14.6	Analyse des Zeit-/Frequenzbereichs	537

14.6.1	Kurzzeit-Fourier-Transformation und Spektrogramm	537
14.6.2	Diskrete Wavelet-Transformation	538
14.6.3	Gramian Angular Field	539
14.7	Die Beziehung zwischen Signalverarbeitung und maschinellem Lernen	542
14.7.1	Techniken für das Feature Engineering	542
14.7.2	Vorbereitung auf maschinelles Lernen	543
14.8	Praktische Anwendungen	544
14.9	Kurz und bündig	546
15	Basismodelle	547
	<i>Danko Nikolić</i>	
15.1	Die Idee eines Basismodells	547
15.2	Wie trainiert man ein Basismodell?	551
15.3	Wie verwenden wir Basismodelle?	553
15.4	Ein Durchbruch: Das Lernen hat kein Ende	562
15.5	Kurz und bündig	564
16	Generative KI und große Sprachmodelle	566
	<i>Katherine Munro, Gerald A. Hahn, Danko Nikolić</i>	
16.1	Einführung in die „Gen-KI“	566
16.2	Generative KI-Modalitäten	568
16.2.1	Methoden für das Training generativer Modelle	569
16.3	Große Sprachmodelle	570
16.3.1	Was sind LLMs?	570
16.3.2	Wie wird so etwas wie ChatGPT trainiert?	572
16.3.3	Methoden zur direkten Verwendung von LLMs	574
16.3.4	Methoden zur Anpassung von LLMs	587
16.4	Schwachstellen und Grenzen von Gen-KI-Modellen	598
16.4.1	Einführung	598
16.4.2	Prompt Injection und Jailbreaking-Angriffe	598
16.4.3	Halluzinationen, Konfabulationen und Begründungsirrtümer ..	603
16.4.4	Urheberrechtliche Bedenken	605
16.4.5	Bias	609
16.5	Erstellen robuster, effektiver Gen-KI-Anwendungen	612
16.5.1	Kontrollstrategien während der gesamten Entwicklung und Nutzung	612

16.5.2	Guardrails	615
16.5.3	Sicherer und erfolgreicher Einsatz generativer KI	615
16.6	Kurz und bündig	618
17	Natürliche Sprachverarbeitung (NLP)	622
	<i>Katherine Munro</i>	
17.1	Was ist NLP und warum ist es so wertvoll?	623
17.2	Warum „traditionelles“ NLP im „Zeitalter der großen Sprachmodelle“ lernen?	624
17.3	NLP-Datenaufbereitungstechniken	626
17.3.1	Die NLP-Pipeline	626
17.3.2	Konvertierung des Eingabeformats für Machine Learning	634
17.4	NLP-Aufgaben und -Methoden	636
17.4.1	Regelbasiertes (symbolisches) NLP	637
17.4.2	Machine-Learning-Ansätze	641
17.4.3	Neurales NLP	651
17.4.4	Transfer Learning	659
17.5	Das Wichtigste in Kürze	673
18	Computer Vision	676
	<i>Roxane Licandro</i>	
18.1	Was ist Computer Vision?	676
18.2	Ein Bild sagt mehr als tausend Worte	678
18.2.1	Das menschliche Auge	679
18.2.2	Das Bildaufnahmeprinzip	681
18.2.3	Digitale Dateiformate	687
18.2.4	Bildkomprimierung	688
18.3	Ich sehe was, was du nicht siehst	690
18.3.1	Computergestützte Fotografie und Bildmanipulation	693
18.4	Computer-Vision-Anwendungen und zukünftige Richtungen	697
18.4.1	Image-Retrieval-Systeme	698
18.4.2	Objekterkennung, Klassifizierung und Verfolgung	701
18.4.3	Medizinische Computer Vision	702
18.5	Menschen sehen lassen	707
18.6	Kurz und bündig	709

19	Modellierung und Simulation – Erstellen Sie Ihre eigenen Modelle	711
	<i>Günther Zauner, Wolfgang Weidinger</i>	
19.1	Einführung	712
19.2	Allgemeine Aspekte	713
19.3	Modellierung zur Beantwortung von Fragen	714
19.4	Reproduzierbarkeit und Lebenszyklus des Modells	716
19.4.1	Der Lebenszyklus einer Modellierungs- und Simulationsfrage	718
19.4.2	Parameter- und Output-Definition	720
19.4.3	Dokumentation	723
19.4.4	Verifizierung und Validierung	724
19.5	Methoden	729
19.5.1	Gewöhnliche Differentialgleichungen (ODEs)	729
19.5.2	Systemdynamik (SD)	731
19.5.3	Diskrete Ereignissimulation	734
19.5.4	Agentenbasierte Modellierung	738
19.6	Beispiele für Modellierung und Simulation	741
19.6.1	Dynamische Modellierung von Eisenbahnnetzen zur optimalen Wegfindung mit agentenbasierten Methoden und Reinforcement Learning	742
19.6.2	Strategien zur agentenbasierten Covid-Modellierung	744
19.6.3	Deep-Reinforcement-Learning-Ansatz für eine optimale Nachschubpolitik in einer VMI-Umgebung	751
19.6.4	Valide Lösungen für ein ressourcenbeschränktes Projektplanungsproblem mithilfe von bestärkendem Lernen und Bewertung der Robustheit mit diskreter Ereignissimulation	754
19.6.5	Zusammenfassung und Lessons Learned	759
19.7	Kurz und bündig	760
20	Visualisierung von Daten	764
	<i>Barbora Antosova Vesela</i>	
20.1	Geschichte	766
20.2	Welche Tools Sie verwenden sollten	772
20.3	Arten von Datenvisualisierungen	775
20.3.1	Streudiagramm	775
20.3.2	Liniendiagramm	776

20.3.3	Säulen- und Balkendiagramme	777
20.3.4	Histogramm	778
20.3.5	Tortendiagramm	779
20.3.6	Box Plot	780
20.3.7	Heat Map	780
20.3.8	Baumdiagramm	781
20.3.9	Andere Arten von Visualisierungen	782
20.4	Wählen Sie die richtige Datenvisualisierung	783
20.5	Tipps und Tricks	786
20.6	Präsentation der Datenvisualisierung	791
20.7	Kurz und bündig	791
21	Datengetriebene Unternehmen	794
	<i>Mario Meir-Huber, Stefan Papp</i>	
21.1	Die drei Ebenen eines datengesteuerten Unternehmens	795
21.2	Kultur	795
21.2.1	Unternehmensstrategie für Daten	796
21.2.2	Die Analyse des aktuellen Stands	799
21.2.3	Unternehmenskultur und Organisation einer erfolgreichen Datenorganisation	801
21.2.4	Kernproblem: der Fachkräftemangel	810
21.3	Technologie	812
21.3.1	Die Auswirkungen von Open Source	813
21.3.2	Cloud	813
21.3.3	Auswahl des Anbieters	814
21.3.4	Data Lake aus der Unternehmensperspektive	814
21.3.5	Die Rolle der IT	815
21.3.6	Data Science Labs	816
21.3.7	Revolution in der Architektur: das Data Mesh	817
21.4	Business	818
21.4.1	Daten kaufen und teilen	819
21.4.2	Implementierung des analytischen Anwendungsfalls	820
21.4.3	Self-Service Analytics	821
21.5	Kurz und bündig	821

22	Leistungsstarke Teams schaffen	822
	<i>Stefan Papp</i>	
22.1	Neue Teams, neues Glück	823
22.2	Storming	823
22.2.1	Szenario: 50 Shades of Red	823
22.2.2	Szenario: Retrospektive	828
22.3	Norming	831
22.3.1	Change Management und Transition	831
22.3.2	RACI-Matrix	834
22.3.3	SMART	836
22.3.4	Agile Prozesse	838
22.3.5	Kommunikationskultur	840
22.3.6	DataOps	842
22.4	Forming	847
22.4.1	Szenario: Eine neue Morgendämmerung	847
22.4.2	Wachstumsgedanken	849
22.5	Kurz und bündig	852
23	Gesetz über künstliche Intelligenz	854
	<i>Jeannette Gorzala, Karin Bruckmüller</i>	
23.1	Einführung	855
23.2	Definition von KI-Systemen	857
23.3	Anwendungsbereich und Zweck des KI-Gesetzes	859
23.3.1	Der risikobasierte Ansatz	860
23.3.2	Unannehmbare Risiken und verbotene KI-Praktiken	862
23.3.3	Hochriskante KI-Systeme und Compliance	865
23.3.4	Mittleres Risiko und Transparenzverpflichtungen	867
23.3.5	Geringes Risiko und freiwillige Selbstverpflichtungen	868
23.4	KI-Modelle mit allgemeinem Verwendungszweck	869
23.5	Zeitplan und Anwendbarkeit	872
23.6	Sanktionen	872
23.7	KI und zivilrechtliche Haftung	873
23.8	KI und strafrechtliche Haftung	873
23.9	Kurz und bündig	877

24	AI in verschiedenen Branchen	878
	<i>Stefan Papp, Mario Meir-Huber, Wolfgang Weidinger, Thomas Tremel, Marek Danis</i>	
24.1	Automobilindustrie	883
24.1.1	Vision	884
24.1.2	Daten	885
24.1.3	Anwendungsfälle	885
24.1.4	Herausforderungen	886
24.2	Luftfahrt	888
24.2.1	Vision	889
24.2.2	Daten	889
24.2.3	Anwendungsfälle	890
24.2.4	Herausforderungen	891
24.3	Energie	891
24.3.1	Vision	892
24.3.2	Daten	893
24.3.3	Anwendungsfälle	893
24.3.4	Herausforderungen	894
24.4	Finanzen	895
24.4.1	Vision	895
24.4.2	Daten	896
24.4.3	Anwendungsfälle	896
24.4.4	Herausforderungen	898
24.5	Gesundheit	899
24.5.1	Vision	899
24.5.2	Daten	900
24.5.3	Anwendungsfälle	901
24.5.4	Herausforderungen	901
24.6	Regierung	902
24.6.1	Vision	903
24.6.2	Daten	903
24.6.3	Anwendungsfälle	904
24.6.4	Herausforderungen	907
24.7	Kunst	908
24.7.1	Vision	909

24.7.2	Daten	909
24.7.3	Anwendungsfälle	910
24.7.4	Herausforderungen	910
24.8	Produktion	911
24.8.1	Vision	911
24.8.2	Daten	912
24.8.3	Anwendungsfälle	912
24.8.4	Herausforderungen	913
24.9	Öl und Gas	914
24.9.1	Vision	914
24.9.2	Daten	915
24.9.3	Anwendungsfälle	916
24.9.4	Herausforderungen	917
24.10	Einzelhandel	918
24.10.1	Vision	918
24.10.2	Daten	919
24.10.3	Anwendungsfälle	919
24.10.4	Herausforderungen	920
24.11	Anbieter von Telekommunikation	921
24.11.1	Vision	921
24.11.2	Daten	922
24.11.3	Anwendungsfälle	922
24.11.4	Herausforderungen	924
24.12	Transport	925
24.12.1	Vision	925
24.12.2	Daten	926
24.12.3	Anwendungsfälle	926
24.12.4	Herausforderungen	926
24.13	Lehre und Ausbildung	927
24.13.1	Vision	928
24.13.2	Daten	928
24.13.3	Anwendungsfälle	929
24.13.4	Herausforderungen	930
24.14	Die digitale Gesellschaft	930
24.15	Kurz und bündig	932

25	Klimawandel und KI	933
	<i>Stefan Papp</i>	
25.1	Einführung	933
25.2	KI – ein Klimaretter?	935
25.3	Messung und Verringerung von Emissionen	936
	25.3.1 Basislinie	936
	25.3.2 Datenanwendungsfälle	938
25.4	Sequestrierung	940
	25.4.1 Biologische Sequestrierung	941
	25.4.2 Geologische Sequestrierung	943
25.5	Vorbereiten auf die Auswirkungen	945
25.6	Geoengineering	946
25.7	Greenwashing	948
25.8	Ausblick	950
25.9	Kurz und bündig	952
26	Mindset und Community	953
	<i>Stefan Papp</i>	
26.1	Data Driven Mindset	954
26.2	Data-Science-Kultur	957
	26.2.1 Start-up oder Beratungsunternehmen?	957
	26.2.2 Labs statt Konzernpolitik	958
	26.2.3 Keiretsu statt Einzelkämpfertum	958
	26.2.4 Agile Softwareentwicklung	960
	26.2.5 Firmen- und Arbeitskultur	961
26.3	Antipatterns	964
	26.3.1 Abwertung von Fachwissen	964
	26.3.2 Die IT wird es schon richten	966
	26.3.3 Widerstand gegen Veränderungen	966
	26.3.4 Besserwisser-Mentalität	967
	26.3.5 Schwarzmalerei	968
	26.3.6 Pfennigfuchseriei	969
	26.3.7 Angstkultur	970
	26.3.8 Kontrolle über die Ressourcen	970
	26.3.9 Blindes Vertrauen in die Ressourcen	971

26.3.10	Das Schweizer Taschenmesser	972
26.3.11	Over-Engineering	973
26.4	Kurz und bündig	973
27	Vertrauenswürdige KI	974
	<i>Rania Wazir</i>	
27.1	Rechtlicher und Soft-Law-Rahmen	975
27.1.1	Normen	978
27.1.2	Verordnungen	979
27.2	KI-Stakeholder	981
27.3	Fairness in der KI	982
27.3.1	Bias	984
27.3.2	Fairness-Metriken	987
27.3.3	Unerwünschten Bias in KI-Systemen reduzieren	992
27.4	Transparenz von KI-Systemen	993
27.4.1	Dokumentieren der Daten	994
27.4.2	Dokumentieren des Modells	996
27.4.3	Explainability (Erklärbarkeit)	997
27.5	Schlussfolgerung	1000
27.6	Kurz und bündig	1000
28	Epilog	1001
	<i>Stefan Papp</i>	
28.1	Halford 2.0	1001
28.1.1	Umwelt, Soziales und Governance	1002
28.1.2	HR	1003
28.1.3	Kundenzufriedenheit	1005
28.1.4	Produktion	1007
28.1.5	IT	1008
28.1.6	Strategie	1010
28.2	Letzte Worte	1012
28.3	Kurz und bündig	1013
29	Die Autor:innen	1014
	Index	1023

Vorwort

Dieses Vorwort wurde NICHT von ChatGPT (oder Ähnlichem) geschrieben.

Während ich diese Aussage treffe, frage ich mich, wie oft sie in Zukunft für Texte oder andere Medienformen gelten wird. In den letzten zwei Jahren hat dieses KI-gestützte Tool enorme Popularität erlangt und Data Science und KI einen unglaublichen Bekanntheitsgrad verschafft. Infolgedessen sind die Erwartungen an Künstliche Intelligenz exponentiell gestiegen und haben solche Höhen erreicht, dass man sich fragen könnte, ob sie jemals erreicht werden können.

Das Thema KI folgt dem bekannten Hype-Zyklus. Einige dieser hohen Erwartungen sind wohlverdient: Diese leistungsstarke Technologie wird die Art und Weise, wie wir leben und arbeiten, in vielerlei Hinsicht verändern. Um ein Beispiel zu nennen: Einige Universitäten erwägen, von ihren Studenten keine Seminararbeiten mehr zu verlangen, da es nicht möglich ist zu überprüfen, ob sie von einem KI-Tool geschrieben wurden.

Aber wir müssen uns auch auf einige Enttäuschungen in der Zukunft gefasst machen, da die KI unweigerlich die überzogenen Erwartungen mancher Leute nicht erfüllen kann. Selbst wenn die Vorstellungen vernünftig sind, ist der Zeitrahmen, den diese Menschen und Organisationen für die Umsetzung von KI-Projekten im Sinn haben, oft nicht realistisch. Dies führt zu weiteren Enttäuschungen, wenn die erhoffte Wirkung und der erhoffte Wert nicht innerhalb des gewünschten Zeitrahmens erreicht werden können.

Die ersten Anzeichen dafür sind bereits zu erkennen, denn ChatGPT und ähnliche Tools liefern eine Fülle von wortgewandten und kohärenten – aber nicht korrekten – Informationen. Die neue Welle von „KI-Experten“, die immer haarsträubendere Versprechungen über die von ihnen oder ihren Unternehmen erfundenen Tools machen, die nur schwer zu halten sein werden, trägt nicht zu mehr Vertrauen bei. Im Grunde genommen verkaufen sie digitales „Schlangenöl“.

All dies erhöht den Druck auf die Data Scientists, mit diesen Erwartungen umzugehen und gleichzeitig weiterhin das gleiche Ziel zu erreichen, das sie seit Jahrzehnten verfolgen:

verständliche Antworten auf Fragen anhand von Daten zu geben.

Aus diesem Grund sind neutrale Organisationen wie die Vienna Data Science Group (VDSG, www.vdsg.at), die den interdisziplinären und internationalen Wissensaustausch zwischen Datenexperten fördert, so notwendig und wichtig. Wir engagieren uns nach wie vor stark für die Entwicklung des gesamten Data-Science- und KI-Ökosystems (Ausbildung, Zertifizierung, Standardisierung, Studien zu den gesellschaftlichen Auswirkungen usw.) in Europa und darüber hinaus. Dieses Buch ist nur eine unserer Bemühungen, um dieses Ziel zu erreichen. Denn trotz all des Hypes und der Übertreibungen in der KI- und Datenlandschaft bleibt Data Science dasselbe: eine interdisziplinäre Wissenschaft, die eine sehr heterogene Gruppe von Spezialisten versammelt. Sie setzt sich aus drei großen Strömungen zusammen, und wir sind stolz darauf, dass wir in jeder von ihnen Experten haben:

- Informatik und IT
- Mathematik und Statistik
- Fachwissen in der Branche oder dem Bereich, in dem Data Science und künstliche Intelligenz angewendet werden.

Die VDSG (www.vdsg.at) hat schon immer einen ganzheitlichen Ansatz für Data Science verfolgt, und das ist auch in diesem Buch nicht anders: Ab Kapitel 1 stellen wir ein fiktives Unternehmen vor, das datengetriebener werden möchte, und begleiten es im Laufe des Buches bis zum Ende seiner Datentransformation in Kapitel 28. Auf dem Weg dorthin gehen wir auf viele Herausforderungen ein und bieten Ihnen so praktische Einblicke, die nur dank des regen Austauschs in unserer großen Data-Science- und KI-Community möglich waren.

Das Ergebnis ist eine stark erweiterte Ausgabe unseres Data Science & AI-Handbuchs mit zehn neuen Kapiteln zu Themen wie Aufbau von KI-Lösungen (Kapitel 13), Foundation Models (Kapitel 15), Large Language Models und generative KI (Kapitel 16) sowie Klimawandel und KI (Kapitel 25). Ergänzend dazu werden auch die grundlegenden Themen Datenarchitektur, Engineering und Governance (Kapitel 4, 5 und 6) behandelt und mit Machine Learning Operations (MLOps, Kapitel 7) abgerundet, das sich zu einer eigenen, sehr wichtigen Disziplin entwickelt hat.

Um eine solide Grundlage zu schaffen, die Ihnen hilft, all dies zu verstehen, haben wir wieder eine Einführung in die zugrunde liegende Mathematik (Kapitel 9) und Statistik (Kapitel 10), die in Data Science verwendet werden, sowie Kapitel über die Theorie hinter Machine Learning, der Signalverarbeitung und der Computer Vision (Kapitel 12, 14 und 18) aufgenommen. Wir haben auch Themen behandelt, die mit der Wertschöpfung aus Daten zu tun haben, wie z. B. Business Intelligence (Kapitel 11) und Data Driven Enterprises (Kapitel 21), sowie wichtige Informationen, die Ihnen helfen,

Daten sicher zu nutzen, einschließlich Kapiteln über das neue EU-KI-Gesetz (Kapitel 23) und vertrauenswürdige KI (Kapitel 27).

Diese umfangreiche Erweiterung des Opus Magnum der VDSG dient vor allem einem Zweck:

ein realistisches und ganzheitliches Bild von Data Science und KI zu vermitteln.

Data Science und KI entwickeln sich derzeit in einem unglaublich schnellen Tempo, und das gilt auch für ihre Auswirkungen auf die Gesellschaft. Das bedeutet, dass die Verantwortung, die auf den Schultern der Data Scientists lastet, ebenfalls gewachsen ist, und damit auch die Notwendigkeit für Organisationen wie die VDSG, sich zu engagieren und diese Herausforderungen zu bewältigen.

Packen wir's an!

Sommer 2024

Wolfgang Weidinger

Danksagungen

Wir, die Autoren, möchten diese Gelegenheit nutzen, um unseren Familien und Freunden, die uns geholfen haben, unsere Gedanken und Einsichten in diesem Buch auszudrücken, unseren aufrichtigen Dank auszusprechen. Ohne ihre Unterstützung und Geduld wäre diese Arbeit nicht möglich gewesen.

Ein besonderer Dank aller Autoren geht an Katherine Munro, die viel zu diesem Buch beigetragen und viel Zeit und Mühe in die Bearbeitung unserer Manuskripte investiert hat.

Für meine Eltern, die immer gesagt haben, dass ich alles schaffen kann. Wir hätten nie erwartet, dass es so etwas sein würde.

Katherine Munro

Ich möchte mich bei meiner Frau und der Vienna Data Science Group für ihre kontinuierliche Unterstützung auf meinem beruflichen Weg bedanken.

Zoltan C. Toth

Wenn ich an die Menschen denke, die mich am meisten unterstützt haben, möchte ich mich bei meinen Eltern bedanken, die immer an mich geglaubt haben, egal, was passiert ist, und bei meiner Partnerin Verena, die in den letzten Monaten wieder sehr geduldig war, während ich an diesem Buch gearbeitet habe.

Darüber hinaus bin ich sehr dankbar für die Unterstützung und Motivation, die ich von den Menschen erhalten habe, die ich durch die Vienna Data Science Group kennengelernt habe.

Wolfgang Weidinger

1

Einführung

Stefan Papp

„Ich möchte CDO werden anstelle des CDOs.“

Iznogoud (angepasst)



Fragen, die in diesem Kapitel beantwortet werden:

- Wie könnte ein fiktives Unternehmen aussehen, das vor seiner Transformation in ein datengesteuertes Unternehmen steht?
- Welche Herausforderungen muss ein Unternehmen bewältigen, um datengesteuert zu werden?
- Wie werden die Kapitel in diesem Buch Ihnen, dem Leser, helfen, solche Herausforderungen in Ihrer eigenen Organisation zu erkennen und zu bewältigen?

1.1 Über dieses Buch

Dieses Buch bietet einen praktischen, erfahrungsbasierten Einblick in verschiedene Aspekte von Data Science und künstlicher Intelligenz. In dieser, unserer dritten, Auflage tauchen die Autoren auch tief in einige der aufregendsten und sich schnell entwickelnden Themen unserer Zeit ein, darunter Large Language Models und generative KI.

Das vorrangige Ziel der Autoren ist es, dem Leser einen ganzheitlichen Zugang zu diesem Bereich zu vermitteln. Aus diesem Grund ist dieses Buch nicht rein technisch: Die Reife von Data Science und KI hängt ebenso sehr von der Arbeitskultur ab, insbesondere vom kritischen Denken und evidenzbasierter Entscheidungsfindung, wie von den Kenntnissen in Mathematik, neuronalen Netzen, KI-Frameworks und Datenplattformen.

In den letzten Jahren sind sich die meisten Experten einig geworden, dass künstliche Intelligenz unsere Arbeits- und Lebensweise verändern wird. Für eine ganzheitliche Betrachtung müssen wir auch den Status quo betrachten, wenn wir verstehen wollen, was getan werden muss, um unsere vielfältigen Ambitionen mithilfe von KI zu erfüllen. Ein nützlicher Rahmen dafür ist es zu untersuchen, wie Menschen mit den Herausforderungen der Datentransformation aus einer organisatorischen Perspektive umgehen. Aus diesem Grund werden wir dem Leser ein fiktives Unternehmen vorstellen, das am Anfang seiner Reise steht, evidenzbasierte Entscheidungsfindung in seine Unternehmensidentität zu integrieren. Wir werden dieses fiktive Unternehmen, in dem vieles datenorientierter sein könnte, als Modell verwenden, um mögliche Herausforderungen zu skizzieren, denen Organisationen begegnen können, wenn sie datenorientierter werden wollen. Am Ende dieses Buches wird unser hypothetisches Unternehmen auch als Modell dafür dienen, wie ein datengesteuertes Unternehmen aussehen könnte. In den dazwischen liegenden Kapiteln gehen wir auf viele dieser Herausforderungen ein und geben praktische Ratschläge, wie man sie bewältigen kann.

Falls Sie als Leser lieber keine Prosa über ein fiktives Unternehmen lesen möchten, um etwas über solche typischen organisatorischen Herausforderungen zu erfahren, empfehlen wir Ihnen, dieses Kapitel zu überspringen und mit einem Kapitel zu beginnen, das Ihren Interessen entspricht. Als ganzheitliches Buch über dieses Gebiet behandeln die Autoren künstliche Intelligenz, maschinelles Lernen, generative KI, Modellierung, Verarbeitung natürlicher Sprache, Computer Vision und andere relevante Bereiche. Wir behandeln Engineer-Themen wie Datenarchitektur und Datenpipelines, die für die Umsetzung datengetriebener Projekte in die Produktion unerlässlich sind. Schließlich gehen wir auch auf kritische soziale und rechtliche Fragen im Zusammenhang mit der Nutzung von Daten ein. Jeder Autor geht sehr detailliert auf sein Fachgebiet ein, damit Sie einen großen Nutzen daraus ziehen können.

Wir bitten die Leser, sich direkt mit uns in Verbindung zu setzen und uns mitzuteilen, wie wir unser ehrgeiziges Ziel erreichen können, die Standardliteratur für einen ganzheitlichen Ansatz in diesem Bereich zu werden. Wenn Sie der Meinung sind, dass einige neue Inhalte in einer der nächsten Ausgaben behandelt werden sollten, können Sie die Autoren über berufliche Netzwerke wie LinkedIn finden.

In diesem Sinne, fangen wir an.

1.2 Die Halford Group

Bob betrat das Bürogebäude der Halford Group, einem Hersteller von Konsumgütern, zu denen auch die meistverkaufte Gummiente gehörte. Nachdem er die Bürotüren passiert hatte, fühlte er sich in die Achtzigerjahre zurückversetzt. Die Besucher mussten sich am Eingang anmelden, Formulare ausfüllen, um im Falle eines Unfalls haft-

bar zu werden, und versprechen, keine Fotos zu machen – und das war nur der erste Schritt. Als Bob den Aufzug mit seinen Messingknöpfen und dem glänzenden Mahagonidekor betrat, hätte er schwören können, dass er die Kulisse des Films „The Wolf of Wall Street“ betreten hatte.

Im Büro der Geschäftsführung sah es ähnlich aus. Den bräunlichen Teppichen sah man ihr Alter an, und die Tapeten sahen aus, als hätten sie den Rauch vieler Marlboro-Männer aus den Achtzigern eingeatmet. Die abgenutzten Ledersofas und der wuchtige Holzschreibtisch (wieder aus Mahagoni) schienen eine Erinnerung an eine große, aber ferne Vergangenheit zu sein. Bob konnte sich vorstellen, wie sein Vater – ein Mann, der immer stolz darauf gewesen war, im Verkauf tätig zu sein und den Lehren von Zig Ziglar zu folgen – in jungen Jahren mit dieser Firma Geschäfte gemacht hatte.

Dieses Bild in Bobs Vorstellung wurde unterbrochen, als eine junge Frau den Raum betrat, und Bob wurde sofort in die Gegenwart zurückgeworfen. Mit entschlossener Miene schritt sie auf Bob zu und griff nach seiner Hand. Etwas verblüfft betrachtete er den platinblonden Haarschopf und die Tätowierungen, die von ihrem maßgeschneiderten Anzug nicht ganz verdeckt wurden, und hob seine Hand als Antwort. Die Frau lächelte.

1.2.1 Alice Halford – Vorsitzende

„Ich bin Alice Halford“, sagte sie, „ich bin die Enkelin von Big Harry Halford, dem Gründer dieser Gruppe. Er hat sein Imperium von Grund auf aufgebaut.“

Bob hatte alle Legenden über den alten Halford-Boss gelesen. Aus jedem Artikel über ihn ging hervor, dass er nicht auf viele Leute hörte. Stattdessen war „Big Harry“ ein stolzer, entschlossener Kapitän, der den Kurs vorgab und von seinem Team absoluten Gehorsam verlangte. Soweit Bob wusste, wurde in den Wirtschaftsmagazinen noch nicht viel über Alice geschrieben. Allerdings hatte er in Vorbereitung auf dieses Treffen einen Artikel gelesen. Alice sei anders als der große alte Familienpatriarch, hieß es darin. Sie hatte sich in einem erbitterten Kampf gegen drei ehrgeizige Brüder durchgesetzt und war vom Vorstand zur Vorsitzenden gewählt worden, weil sie große Pläne hatte, die Firma in ein modernes Unternehmen umzuwandeln, das dem Zeitgeist des 21. Jahrhunderts entsprechen würde.

„Obwohl er erfolgreich war, würde die heutige Generation meinen Großvater als Dinosaurier bezeichnen, der nur genug Fußspuren hinterlassen wollte, damit die nächste Generation weiß, dass er da war“, sagte Alice. „Besonders in seinen letzten Lebensjahren stand er Veränderungen skeptisch gegenüber. Viele Chefberater angesehener Unternehmen kamen hochobersten Hauptes in unsere Büros und erklärten uns, dass unsere langfristige Existenz davon abhängen würde, ein datengesteuertes Unternehmen zu werden. Aber mein Großvater hatte immer ein Sprichwort: In

dem Moment, in dem ein Computer entscheidet und nicht mehr ein Gründer, der sich auskennt und seinem Bauchgefühl folgt, ist es vorbei. All die einst stolzen Berater und ihre Unterstützer aus dem Unternehmen dachten, sie könnten jede Führungskraft davon überzeugen, ihre Vorstellungen von einem modernen Unternehmen zu übernehmen, aber letztendlich zogen sie den Schwanz ein und gingen.“

Alice lächelte Bob an und fuhr fort: „Der Ruhestand meines Großvaters war längst überfällig, aber schließlich zwangen ihn seine exotischen kubanischen Zigarren und seine Gewohnheit, teuren Whiskey zu trinken, dazu, sein Arbeitsleben zu beenden. Ich habe den Vorsitz des Verwaltungsrats übernommen. Ich möchte alle Gerüche des letzten Jahrhunderts beseitigen. Als ich antrat, stellte ich fest, dass Teile des Unternehmens hochgradig toxisch waren. Meine Strategieberater wiesen mich darauf hin, dass jede große Organisation ein gewisses Maß an organisatorischer Arroganz und Ineffizienz aufweist. Sie ermahnten mich auch, meine Erwartungen niedrig zu halten. Während viele Enthusiasten behaupten, dass KI die Welt für immer verändern wird, ist jedes große Unternehmen wie ein lebender Organismus mit vielen verschiedenen Unterteilungen und Eigenschaften. Die Veränderung der Unternehmenskultur ist ein langer Prozess, und viele Unternehmen stehen vor ähnlichen Herausforderungen. Letztlich wird jedes Unternehmen von Menschen geführt, und niemand kann Menschen über Nacht ändern. Einige mögen mit Veränderungen einverstanden sein, einige wenige wünschen sich vielleicht sogar, dass sie zu schnell erfolgen, aber die meisten Menschen werden sich auf die eine oder andere Weise gegen Veränderungen sträuben.“

Gleichzeitig ist mir klar, dass uns die Zeit davonläuft. Wir haben gelernt, dass unsere Hauptkonkurrenten uns voraus sind, und wenn wir nicht aufholen, werden wir schließlich nicht mehr mithalten können. Unser derzeitiger CEO kommt aus dem Finanzbereich und braucht daher die Unterstützung eines Datenstrategen. Bob, Sie wurden uns als hervorragender Experte für die Umwandlung eines Unternehmens in ein datengesteuertes Unternehmen empfohlen, das die traditionellen Geschäftsmodelle auf den Kopf stellt. Sie können mit jedem reden, Sie haben alle Freiheiten, die Sie brauchen. Danach bin ich neugierig auf Ihre Ideen, das Unternehmen von Grund auf zu verändern.“

Bob nickte enthusiastisch. „Ich liebe Herausforderungen. Ihre Sekretärin hat mir bereits gesagt, dass ich am Nachmittag keine weiteren Termine haben sollte. Können Sie mich Ihrem Team vorstellen? Ich würde gerne mehr darüber erfahren, wie es arbeitet und welche Anforderungen es hat.“

„Ich dachte mir, dass Sie das tun wollen. Zuerst werden Sie die Analysten David und Anna kennenlernen. Dann treffen Sie Tom, den Vertriebsleiter. Es wäre am besten, wenn Sie auch mit dem IT-Manager Peter sprechen würden ...“ Alice unterbrach sich, seufzte und fuhr fort: „Schließlich habe ich für Sie ein Treffen mit unserem Produktionsleiter, der Reklamationsabteilung, unserem Sicherheitschef und schließlich mit unserer Personalabteilung arrangiert. Ich werde Ihnen unseren neuen CEO vorstel-

len, der heute einfliegt, um beim Abendessen Einzelheiten zu besprechen. Ich habe einen Tisch in einem guten Restaurant in der Nähe reserviert. Aber es macht Sinn, wenn man zuerst mit allen anderen Beteiligten spricht. Ich habe meine Kollegen gebeten, jeweils ein persönliches Gespräch mit Ihnen zu vereinbaren. Das wird ein anstrengender Nachmittag, Bob.“

1.2.2 Analysten

Als Alice den Raum verließ, erschienen in der Tür ein Mann, offenbar Mitte vierzig, und eine etwa gleichaltrige Frau. Das mussten die Analysten David und Anna sein. Als keiner der beiden bereit schien, den Raum zu betreten, winkte Bob sie herein. Er fühlte sich an ein Empowerment-Seminar erinnert, das er vor einigen Jahren besucht hatte: Der Trainer war wild entschlossen gewesen, aus allen Teilnehmern starke Führungspersönlichkeiten zu machen, hatte aber davor gewarnt, dass nur die Tatkräftigen die Welt beherrschen würden. Diese Analysten schienen das genaue Gegenteil zu sein. David lachte nervös, als er eintrat, und Anna hielt ihren Blick gesenkt, als sie sich auf den nächstgelegenen Platz begab. Beide schienen nicht gerade begeistert zu sein, dort zu sein; Bob wollte sich gar nicht ausmalen, wie sie in dem „Urschrei“-Test des Seminars abgeschnitten hätten.

David und Anna setzten sich, und Bob versuchte, mit Fragen über ihre Arbeit das Eis zu brechen. Es dauerte eine Weile, aber schließlich begannen sie zu reden.

„Nun, wir erstellen Berichte für die Geschäftsführung“, sagte David. „Wir bemühen uns um Genauigkeit und versuchen, unsere Berichte pünktlich einzureichen. Das hat sich zu einem gewissen Ruf entwickelt“, fügte er hinzu.

Bob erkannte, dass er, wenn er sie zum Reden bringen wollte, seine berühmte Rede halten musste, die er folgendermaßen zusammenfasste: „Ihre Aufgabe in dieser Besprechung ist es, über Ihr Problem zu sprechen. Meine Aufgabe ist es zuzuhören.“ Schließlich musste er das Unternehmen Halford in ein datengesteuertes Unternehmen umwandeln, und sie waren diejenigen, die am engsten mit den Daten des Unternehmens arbeiteten.

Bob beendete seine Rede mit Elan, aber Anna zuckte nur mit den Schultern. „Die Geschäftsleitung will viel wissen, aber unsere Möglichkeiten sind begrenzt.“

Bob bemühte sich, beiden in die Augen zu sehen, doch Anna wandte sich schnell ab. „Aber was hindert Sie daran, Ihre Arbeit uneingeschränkt zu tun?“

„Unsere größte Herausforderung ist der Batch-Prozess aus der Hölle“, meldete sich David plötzlich zu Wort. „Dieser berüchtigte tägliche Job läuft über Nacht und extrahiert alle Daten aus den operativen Datenbanken. Er ist ungeheuer komplex. Ich habe aufgehört zu zählen, wie oft dieser Job im Laufe der Zeit fehlgeschlagen ist.“

Ich habe sie, dachte Bob und nickte ermutigend.

„Und niemand weiß, warum dieser Job scheitert“, schaltete sich Anna ein. „Und wenn er scheitert, wissen wir nicht, ob die Daten korrekt sind. Bis jetzt gab es noch nie ein Problem, wenn wir einen Bericht mit fragwürdigen Zahlen abgaben. Aber das liegt wahrscheinlich daran, dass die meisten Manager die von uns gelieferten Fakten und Zahlen ohnehin ignorieren.“

„Genau!“ David hob dramatisch seine Hände. Bob begann zu befürchten, dass er in ein Wespennest gestochen hatte.

„Wenn ein Job scheitert, bin ich es, der zur IT-Abteilung gehen muss“, sagte David. „Ich kann es einfach nicht mehr hören, dass den Nerds der Speicherplatz ausgegangen ist und dass DevSecOps mal wieder einen Firewall-Port geschlossen hat. Alles, was ich will, sind die Daten, um meine Berichte zu erstellen. Ich kämpfe auch oft mit unserer Sicherheitsabteilung. Manchmal sind ihre Prozesse so streng, dass sie eine Innovation fast sabotieren. Gelegentlich habe ich den Eindruck, dass sie den Zugang zu Datenquellen absichtlich sperren, um uns zu ärgern.“

„Oft werden wir gefragt, ob wir etwas Anspruchsvolleres machen wollen“, sagt Anna und schüttelt frustriert den Kopf. „Es ist immer das gleiche Muster. Ein Manager besucht ein Seminar und kommt zu uns, um uns zu fragen, ob wir ‚KI‘ machen können. Wenn Sie mich ehrlich fragen, würde ich gerne etwas Anspruchsvolleres machen, aber wir haben Angst, dass das ganze System auseinanderbricht, wenn wir etwas ändern. Deshalb bin ich froh, wenn wir der Geschäftsleitung die Daten vom Vortag zur Verfügung stellen können.“

„Verstehen Sie uns nicht falsch, ML und KI wären fantastisch. Aber unser Unternehmen muss sich noch die Grundlagen erarbeiten. Ich glaube, die meisten unserer Manager haben keine Ahnung, was KI bewirkt und was wir damit machen könnten. Aber werden sie es zugeben? Auf keinen Fall.“

Anna lehnte sich verärgert zurück. Bob brauchte sie nicht zu fragen, um zu wissen, dass sich beide bereits bei anderen Unternehmen bewarben.

1.2.3 „CDO“

In der Mittagspause stürmte ein hagerer Mann in einem schwarzen Rollkragenpullover ins Büro. Er wirkte nervös, als ob jemand hinter ihm her wäre. Seine Augen huschten durch den Raum und vermieden den Blickkontakt. Sein ganzer Körper zapelte, und er konnte seine Hände nicht ruhig halten.

„Ich bin der CDO. Mein Name ist Cesario Antonio Ramirez Sanchez; nennen Sie mich Cesar“, stellte er sich mit spanischem Akzent vor.

Bob war überrascht, dass dieses Treffen nicht angekündigt worden war. Währenddessen näherte sich sein unerwarteter Besucher immer wieder einem Stuhl und entfernte sich wieder von ihm, als könne er sich nicht entscheiden, ob er sich setzen sollte oder nicht.

„CDO? Ich habe diese Position im Organigramm nicht gesehen“, antwortete Bob ruhig, „ich habe einen Cesario Antonio Rami gesehen ...“

„Nein, nein, nein ... Das ist nicht mein offizieller Titel. Es ist das, was ich tue“, sagte Cesar dramatisch. „Ich verändere das Unternehmen von Grund auf, verstehen Sie? Wie ein Guerillakrieg. Ohne Leute wie mich wäre diese Firma noch in der Steinzeit, verstehen Sie?“

„Ich bin an der Meinung aller interessiert“, antwortete Bob, „aber ich bin Alice unterstellt und kann mich nicht an Geheimoperationen beteiligen.“

„Nein, nein, nein ..., alles ist ganz einfach. In dieser Firma laufen nur Schwachköpfe herum ...“ Cesar hob den Finger, holte tief Luft, nickte zweimal und fuhr fort. „Ich weiß ... die Personalabteilung sagt mir immer, ich soll freundlich zu den Leuten sein und keine bösen Worte sagen. Aber wir haben in dieser Firma nur Data Warehouses. Nicht einmal einen Data Lake. Catastrófica! Es ist das 21. Jahrhundert, und diese Dinosaurier arbeiten wie in Lateinamerika hace veinte años. Increíble!“

Er holte noch einmal Luft und fuhr dann fort: „Modernisieren wir diese Bruchbude! Fangen wir bei null an. Es gibt so viel zu tun. Zuerst müssen wir diese alten Geräte auf den Müll werfen, verstehst du? Und sie durch streaming-fähige PLCs ersetzen. Dann streamen moderne Edge-Computing-Dienste alles mit Kafka an verschiedene Datenspeicher. Alle Probleme sind gelöst. Und dann haben wir eine Echtzeit-Analyseschicht auf einem Data Mesh.“

Bob starrte sein Gegenüber an, das nicht in der Lage zu sein schien, seine Augen oder seinen Körper länger als einen Moment ruhig zu halten. „Es tut mir leid, ich verstehe nicht.“

„Sie sind ein Experte, Sie haben einen Dokortitel, oder? Sie sollten verstehen: moderne Fabrik, IoT, Industrie 4.0, Factory of the Future.“

Bob beschloss, nicht zu antworten. Stattdessen zog er die Augenbrauen hoch und wartete darauf, was Cesar als Nächstes sagen würde.

„So viel Potenzial“, fuhr Cesar fort. „Und alles wird vergeudet. Warum redet die Personalabteilung immer über die Gefühle der Menschen? Alles ist so einfach. Diese alte Firma muss modern werden. Wir brauchen keine Künstler, wir brauchen Leute mit Verstand. Wenn ich Kunst will, höre ich mir Mariachi in Cancún an. Wenn die jetzigen Leute Schwachköpfe sind, stellt neue Leute ein. Kluge Leute, mit Dokortitel und Erfahrung. Meine alten Chefs in Lateinamerika, das können Sie sich nicht vorstellen, hätten alle gefeuert, auch die Personalabteilung. Lassen Sie uns später reden; ich bin in der IT-Abteilung en la cava.“

Bob hatte keine Zeit, zu antworten. Cesar verließ den Raum so schnell, wie er ihn betreten hatte.

1.2.4 Vertrieb

Ein großer, schlanker, grauhaariger Mann betrat den Raum, nahm am Ende des Tisches Platz, lehnte sich zurück und schenkte Bob ein Verkäuferlächeln, für das Colgate Millionen bezahlt hätte.

„Ich bin Tom Jenkins. Meine Freunde nennen mich ‚die Lawine‘. Denn wenn ich für eine Wiedervorlage zum Telefon greife, kann mich niemand mehr aufhalten. In den Neunzigern habe ich an einem einzigen Tag vier Verkäufe gemacht. Können Sie sich das vorstellen?“

Ich verstehe schon, du bist ein Held, dachte Bob. Halte mal den Ball flach.

„Mein Name ist Bob. Ich bin ein Berater, der diesem Unternehmen helfen soll, datenorientierter zu werden.“

Toms gewinnendes Lächeln verschwand, als Bob „Daten“ erwähnte.

„Ich habe zu viel von den Daten gehört“, sagte Tom. „Keine Analyse kann Bauchgefühl und Erfahrung ersetzen. Verstehen Sie mich nicht falsch. Ich liebe genaue Daten über meine Verkaufszahlen, aber man sollte einem erfahrenen Mann vertrauen, dass er seine eigenen Entscheidungen trifft. Kein Computer wird mir jemals sagen, welchen potenziellen Kunden ich anrufen soll. Wenn ich an meinem Schreibtisch sitze, weiß ich, was so richtig durchstarten wird.“

„Bei allem Respekt. Ich kann Ihnen viele Beispiele dafür zeigen, wie ein evidenzbasierter Ansatz Kunden zu mehr Einnahmen verholfen hat.“

„Haben Sie sich gerade selbst gehört?“ antwortete Tom, „evidenzbasiert ...! Man gewinnt Verkäufe nicht mit klugen Reden. Man muss an den Emotionen und Beziehungen der Menschen arbeiten. Kein Computer wird jemals bessere Geschäfte machen als ein Verkäufer mit einem gewinnenden Lächeln. Ich gebe Ihnen ein Beispiel: Eines Tages zeigten unsere Verkaufsdaten, dass wir in unserer wichtigsten Region weniger Produkte verkauften. Einige Datenanalysten erzählten mir etwas über demografische Veränderungen. Was für ein Blödsinn!

Also bin ich losgezogen und habe mit den Leuten gesprochen. Ich kenne meine Leute da draußen. Sie sind alle großartige Menschen. Alles tolle Leute! Sehr klug und hemdsärmelig. Ich liebe das. Wir aßen Steaks und tranken Bier, dann stellte ich unsere neue Produktlinie vor. Raten Sie mal, wer danach Verkäufer des Monats war?

Kein Computer muss mir sagen, wie ich auf meine Kunden zugehen soll. Solange die Verkaufsberichte stimmen und wir die Provisionen berechnen können, ist alles in Ordnung. Es ist der Verkäufer, nicht der Computer, der ein Geschäft abschließt.“

Damit erhob sich Tom, die „Lawine“, wieder. Er lud Bob in ein fantastisches Restaurant ein – „Ich kenne den Besitzer, und glauben Sie mir, er macht die besten Steaks, die Sie je probiert haben“ – und war weg.

1.2.5 IT

Zehn Minuten nach dem geplanten Beginn der Sitzung wartete Bob immer noch auf das Teammitglied, von dem er im Vorfeld am meisten gehört hatte: den IT-Leiter Peter. Sein Name war von verschiedenen Leuten mehrfach erwähnt worden, aber immer, wenn Bob nachgefragt hatte, wollten die Leute nicht antworten oder seufzten nur und sagten ihm: „Sie werden schon sehen.“

Schließlich stürmte Peter atemlos und schwitzend in den Raum. „Der Weg von meinem Büro im Keller in dieses Stockwerk ist ein Albtraum“, sagte er zwischen zwei Atemzügen. „Man trifft im Aufzug so viele Leute, die etwas wollen. Ich stehe ständig unter Stress, das können Sie sich nicht vorstellen! Hier, ich habe uns ein paar Sandwiches mitgebracht. Ich habe ein kleines Nebengeschäft in der Gastronomie. Man braucht so ein Hobby, um hier zu überleben. Ohne ein Hobby wird man in diesem Geschäft verrückt.“

Peter war ein untersetzter, rotgesichtiger Mann, der schon mit Halford zusammen war, als er noch viel jünger war und viel mehr Haare hatte. Er saß etwas zu bequem in seinem Stuhl, mit dem Selbstbewusstsein eines Mannes, der schon so lange dabei war, dass er praktisch zum Mobiliar gehörte.

An Selbstvertrauen mangelt es ihm nicht, das ist sicher, dachte Bob. Ich frage mich, wie viele schmutzige Geheimnisse dieser Mann im Laufe der Jahre erfahren hat, die nur er kennt.

„Okay, dann lassen Sie uns über IT reden“, seufzte Peter, nachdem Bob die Sandwiches abgelehnt hatte. „Meine Kollegen aus dem Vorstand und die Führungskräfte verstehen immer noch nicht, was sie täglich von mir verlangen. Wenn sie mich zu Besprechungen einladen, tauche ich oft nicht mehr auf. Wir sind ein großes Unternehmen, aber niemand will in die IT investieren. Ich bin unterbesetzt; wir schaffen es kaum, das Unternehmen am Laufen zu halten. Wollen wir eine Zigarette rauchen gehen?“

„Nein, danke“, sagte Bob, aber Peter hatte bereits ein Päckchen aus seiner Hosentasche gekramt. Er plapperte den ganzen Weg zum Raucherraum und sprang von einem Thema zum anderen. Bob erfuhr alles über Peter, von seinem Lieblingsessen über sein Privathaus bis hin zu seinem Leistenbruch, der offenbar immer schlimmer wurde. Nachdem Peter die erste Zigarette in den Mund gesteckt hatte, wandte er sich wieder dem Thema zu, das Bob wirklich interessierte.

„Die Schlipsträger wollen etwas, ohne die Konsequenzen zu kennen. Auf der einen Seite wollen sie, dass alles sicher ist, aber auf der anderen Seite wollen sie moderne Datenlösungen. Oft fragen sie mich an einem Tag nach einer Sache, und am nächsten Tag priorisieren sie etwas anderes. Um es ganz offen zu sagen: Ich hatte schon viele Gespräche mit diesen externen Beratern. Wenn ich ihnen erlauben würde, das zu tun, worum sie mich bitten, könnte ich gleich alle unsere Daten auf einen Dateiserver

legen und Hacker einladen, sie direkt herunterzuladen. Um die Dinge am Laufen zu halten, muss man das gesamte Unternehmen mit einer Firewall versehen!“ Peter drückte seine Zigarette aus und griff nach einer neuen.

Bob nutzte die Gelegenheit, um sich einzuklinken. „Können Sie mir mehr über Ihre IT-Abteilung erzählen? Ich war auf der Suche nach einer Dokumentation über die IT-Landschaft. Ich habe nicht viele Informationen über Ihre internen Dateifreigaben gefunden. Welchen Cloud-Anbieter nutzen Sie derzeit?“

Peter lachte und fing dann an zu husten. Mit Tränen in den Augen antwortete er: „Ich sagte doch, ich bin unterbesetzt. Glauben Sie wirklich, dass ich Zeit für eine Dokumentation habe?“ Er deutete auf seinen Kopf. „Keine Sorge, hier oben in den grauen Zellen ist alles gespeichert. Und wir haben eine No-Cloud-Strategie. Cloud ist nur eine Marketing-Sache, wenn Sie mich fragen. Wenn wir Lösungen selbst bauen, ist es sicherer, und wir haben alles unter Kontrolle. Wenn ich nur mehr Leute hätte ... Haben Sie schon einen meiner Jungs kennengelernt, Cesar? Er ist in Ordnung, wenn er nicht redet. Das kommt leider nur selten vor. Ich mag es nicht, wenn Leute denken, sie seien schlauer als ich. Er kennt die zwei Regeln von Peter noch nicht. Regel Nummer 1: Geh deinem Chef nicht auf die Nerven. Regel Nummer 2. Befolge Regel Nummer 1.“

Peter lachte, schnippte die zweite Zigarette auf den Boden und holte eine Tüte aus seiner anderen Tasche. Sie war voll mit Karamellbonbons: Peter steckte sich eine in den Mund und fuhr laut kauend fort. „Alice hat mich gefragt, ob ich Ihnen Bill, meinem leitenden Ingenieur, vorstellen darf, aber ich habe abgelehnt. Dieser Typ hat den Verstand eines Fuchses, aber die Kommunikationsfähigkeit eines Esels. Außerdem wird er nervös, wenn man ihm direkt in die Augen sieht. Ich mache mir immer Sorgen, dass er sich in die Hose macht – oder bin ich schon wieder zu politisch inkorrekt? Unsere Personalabteilung sagt mir ständig, dass ich freundlicher sein soll. Aber in dieser Klapsmühle lernt man, seinen Stress abzubauen, indem man sagt, was man denkt. Also entschuldigen Sie bitte meinen Sarkasmus. Ich bin die letzte Person, die zwischen dem Chaos und einer funktionierenden IT-Landschaft steht, das Management nervt mich ständig mit dummen Anfragen, und zu guter Letzt macht sich die Personalabteilung mehr Gedanken darüber, wie ich kommuniziere, als darüber, wie sie die Leute finden, die mir helfen könnten, unser Unternehmen am Laufen zu halten.“

Es bedurfte einiger Versuche, bis Bob sich endlich von Peters Gemecker lösen konnte, um zu seinem nächsten Treffen zu gehen. Sogar als er ging, forderte Peter Bob immer wieder auf, ihn einmal in seinem Lebensmittelgeschäft zu besuchen, wo sie unter vier Augen etwas trinken könnten und Peter seine „Kriegsgeschichten“ über Halford offener erzählen könnte.

1.2.6 Sicherheit

Während er auf den Vertreter der Personalabteilung wartete, erhielt Bob eine Sprachnachricht von Suzie Wong, der Leiterin der Datensicherheit. Als Bob sie abspielte, hörte er im Hintergrund Verkehrsgeräusche.

„Entschuldigung, dass ich nicht gekommen bin. Die Schule hat mich einbestellt, weil eines meiner Kinder krank geworden ist. Ich hoffe, eine Sprachnachricht ist in Ordnung. Ich bin Suzie Wong. Ich arbeite schon seit Jahren bei Halford. Man nennt mich die menschliche Firewall gegen Innovation. Ich fasse das als Kompliment auf, weil es irgendwie bedeutet, dass ich meine Arbeit gut mache. Könnte ein Unternehmen mit einer Sicherheitschefin zufrieden sein, die ihren Job leichtnimmt? Mein Vorgänger war viel entspannter als ich. Er war in den Fünfigern und hatte es sich ein wenig zu bequem gemacht, weil er dachte, er würde sich mit einem sicheren Job zur Ruhe setzen. Und dann eines Tages ... gab es diese Sicherheitslücke. Sein Kind ist immer noch in der Privatschule, er ist plötzlich ohne Job und, nun ja, ich erspare Ihnen die Details.

Die Leute denken oft, dass ich nur da bin, um ihre Anfragen zur Datenfreigabe abzunicken, aber meine eigentliche Aufgabe ist der Schutz der Privatsphäre unserer Kunden. Data Scientists müssen mir beweisen, dass die Daten unserer Kunden sicher sind, wenn sie mit ihnen herumspielen wollen. Leider nehmen das zu viele zu leicht.

Wenn der Antragsteller das Verfahren befolgt, könnte ein Privacy Impact Assessment innerhalb einer Woche durchgeführt werden. Ich werde Ihnen später einen Link zu unserem Sicherheitsportal schicken, damit Sie es überprüfen können. Sie werden selbst sehen, dass wir nichts Unmögliches verlangen.

Ich bin die letzte Verteidigungslinie, die dafür sorgt, dass wir keine saftigen Geldstrafen zahlen müssen, weil jemand dachte, es handele sich nur um Daten, mit denen er herumspielt. Manche Leute nennen mich auch scherzhaft ‚Frau Nein‘, denn das ist meine übliche Antwort, wenn sie nicht erklären können, warum ich ihnen Sicherheitsausnahmen gewähren oder Zugang zu Daten mit privaten Kundeninformationen geben sollte. Manche Leute beschwerten sich, dass es auf diese Weise Monate dauern kann, bis eine Sicherheitsgenehmigung erteilt wird. Aber solange Ingenieure und Data Scientists noch nicht wissen, wie man Sicherheitsfragen richtig angeht, ist es mir egal, ob es Jahre dauert, bis ich mein endgültiges OK gebe. Wie auch immer, entschuldige mich jetzt, ich bin in der Schule ...“

1.2.7 Leiter der Produktion

Bob hatte noch etwas Zeit bis zu seinem nächsten Treffen und suchte online nach seinem nächsten Gesprächspartner. Er entdeckte einen Mann mittleren Alters mit einer langen Geschichte in den sozialen Medien, einschließlich einiger fragwürdiger Fotos seines jüngeren Ichs in einem Che-Guevara-T-Shirt. Bob schmunzelte. Dieser

junge Mann konnte froh sein, dass ihr Gespräch nicht in den Zeiten des Kalten Krieges stattfand.

Schließlich betrat Bobs Gesprächspartner den Raum. Er war muskulös, und sein buschiger schwarzer Bart zeigte die ersten Anzeichen von Grauwerden.

„Mein Name ist Hank. Freut mich, Sie kennenzulernen“, sagte er mit tiefer Stimme.

„Ich habe gehört, Sie sind neu in Ihrer Position“, sagte Bob.

„Ja. Alice hat meinen Vorgänger gefeuert, weil er ein Tyrann war. Ich bin jetzt einer der ersten der, wie sie es nennt, ‚neuen Generation‘. Ich habe zugesagt, weil ich hier jetzt etwas verändern kann. Lassen Sie mich auf den Punkt kommen: Was haben Sie hier vor?“

Bob lächelte und sagte: „In Fabriken geht es oft darum, maschinelles Lernen für die Automatisierung einzusetzen. Denken Sie an Prozesse, bei denen Menschen die Qualität eines Artikels manuell prüfen. Stellen Sie sich vor, Sie könnten das alles automatisieren. Eine Kamera prüft jedes Teil, und fehlerhafte Teile – die wir ‚Ausschuss‘ nennen – werden automatisch herausgefiltert.“

Hank versteifte sich. „Meine Aufgabe ist es, Arbeitsplätze zu schützen, nicht ihre Beseitigung zu unterstützen. Einige unserer Fabriken befinden sich in Dörfern, wo sie die einzige Quelle für Arbeit sind.“

„Fast jedes Land durchläuft einen demografischen Wandel. Können Sie garantieren, dass Sie in der Lage sein werden, genügend Arbeitskräfte zu halten, um die Fabriken am Laufen zu halten? Wie wäre es, das Gleiche mit weniger Menschen zu tun?“

„Aber wenn man darum bemüht, ein paar Leute wegzuo optimieren, können in so einem Projekt auch viele Menschen arbeitslos gemacht werden“, sagte Hank. „Was ist, wenn man in ein paar Jahren überhaupt keine Arbeiter mehr braucht? Ich möchte nicht die Tür zu einem System öffnen, das die Bourgeoisie reicher macht und den einfachen Proletarier arbeitslos werden lässt.“

„Das ist sehr unwahrscheinlich“, sagte Bob. „Ich sehe, Sie sind solidarisch mit Ihren Mitarbeitern, Hank. Haben Sie daran gedacht, Anwendungsfälle zu erforschen, um sie zu schützen? Wir können zum Beispiel mithilfe von Computer Vision erkennen, ob Fabrikarbeiter Helme tragen.“

Hank schaute Bob tief in die Augen. Bob konnte nicht genau sagen, ob das ein gutes oder ein schlechtes Zeichen war, aber eines wurde ihm klar: Diesem Mann würde er nicht gerne auf einer dunklen, leeren Straße begegnen.

„Ich verstehe, dass es für meine Kollegen von Vorteil sein könnte“, sagte Hank. „Ich will nur kein trojanisches Pferd öffnen: Ich führe ein IT-System ein, um Unfälle zu vermeiden, und das nächste macht die Arbeiter überflüssig. Aber ich habe Alice versprochen, dass ich sie unterstützen werde. Sie ist ein guter Mensch. Ich werde mit meinen Kollegen sprechen. Ich muss sie mit ins Boot holen, aber eine Sache ist nicht verhandelbar: Wir werden niemals ein System tolerieren, das die Arbeitsplätze von Menschen vollständig ersetzt.“

1.2.8 Kundenbetreuung

Die nächste Gesprächspartnerin, eine ältere Frau mit perfekt glänzendem, silbernem Haar, betrat den Raum. Sie setzte sich und fuhr vorsichtig mit den Fingern über ihren klassischen Dutt, um sicherzugehen, dass kein einziges Haar fehl am Platz war.

„Ich bin Annie von der Beschwerdestelle“, sagte sie in einem etwas aristokratischen Ton. Sie schien sich mehr für ihre sauber manikürten Nägel zu interessieren als für Bob, als sie fortfuhr: „Ich weiß ehrlich gesagt nicht, warum Sie mit mir sprechen wollen.“

„Nun, Teil eines datengesteuerten Unternehmens ist oft auch eine kundenorientierte Strategie. Wir können die Kundenabwanderung und andere Metriken anhand von Daten messen. Die meisten meiner Kunden wollen Daten nutzen, um ihren Erfolg zu maximieren. Sie haben ihre Abteilungen sogar in ‚Abteilung für Kundenzufriedenheit‘ umbenannt, um dies zu unterstreichen.“

„Aha“, sagte Annie. Es herrschte eine unangenehme Stille, während sie mit ihrem anderen Ärmel das Zifferblatt ihrer antiken Uhr polierte.

Bob räusperte sich, um ihre Aufmerksamkeit zu gewinnen. „Wären Sie daran interessiert, anhand von Daten mehr über Ihre Kunden zu erfahren?“

„Warum sollte ich?“

„Um ihre Kunden besser zu betreuen?“

„Wir haben solide Produkte. Die meisten Beschwerden entbehren jeder Grundlage. Wir glauben, je weniger Geld wir für verwirrte Kunden ausgeben, desto mehr haben wir übrig, um unsere Produkte zu verbessern. Das nenne ich den echten Kundennutzen, den wir bieten.“

Aha. Bob erkannte das berühmte Argument gegen Investitionen in eine Domain, die nicht direkt zu Einnahmen führt. *Wahrscheinlich bekommt sie einen Bonus, wenn sie die jährlichen Kosten niedrig hält*, dachte er und sah eine Chance.

„Und wie halten Sie die Kosten im Moment gering?“

„Wir haben ein Offshore-Callcenter. Dort werden etwa 80 % der Anrufe bearbeitet, wonach viele dieser Kunden einfach aufgeben. Die verbleibenden 20 % werden an ein kleines Team von fortgeschrittenen Kundendienstmitarbeitern weitergeleitet. Ich weiß, das klingt hart, aber Sie können sich nicht vorstellen, wie viele verwirrte Menschen versuchen, uns anzurufen, ohne überhaupt ein Problem zu haben. Einige – so scheint es – rufen uns nur an, um zu reden.“

„Richtig. Und haben Sie an die Möglichkeit gedacht, Kosten zu senken, indem Sie Chatbots mit generativer KI entwickeln? Es gibt auch viele Möglichkeiten, Data Science zu nutzen, um Kundenbeschwerden zu filtern. Wenn sie richtig geschult sind, erhalten Ihre Kunden einen besseren Support, und Sie senken die Kosten.“

„Würde es ausreichen, das Offshore-Zentrum zu schließen?“

Erwischt. „Wenn man es richtig macht, ja.“

Zum Gefühlten ersten Mal sah Annie Bob direkt an. „Wie viel würde es kosten?“

„Im Moment ist es noch schwer abzuschätzen.“

Annie dachte eine Weile nach und stand dann auf, um zu gehen. An der Tür hielt sie inne. „Sobald Sie es wissen, rufen Sie mich sofort an.“

1.2.9 HR

„Ich bin ... ich bin Pratima“, sagte eine Frauenstimme an der Tür. Sie ging auf Bob zu, schaute ihn mit einem einladenden Lächeln an und fragte: „Wie kann ich Ihnen helfen, Bob?“

„Hallo, Pratima. Setzen wir uns doch. Wie Sie wissen, bin ich hier, um dieses Unternehmen in ein stärker datenorientiertes Unternehmen umzuwandeln. Ich habe auf LinkedIn gesehen, dass Sie zuvor für sehr moderne Unternehmen mit einer starken Datenkultur gearbeitet haben. Wie ist es jetzt, für ein Unternehmen zu arbeiten, das noch am Anfang seiner Reise steht?“

„Alice bat mich, Ihnen gegenüber offen zu sein. Ich habe diesen Job als Karriereschritt angenommen, um in eine Führungsposition aufzusteigen. Diese Entscheidung hat mich jedoch vor mehr Herausforderungen gestellt als erwartet. In meinem vorherigen Job hatten wir die richtige Atmosphäre, um neue Talente anzuziehen. Es war ein Umfeld, das für Spitzenleistungen prädestiniert war: schicke Büroräume, eine moderne Arbeitskultur mit flachen Hierarchien, coole Produkte, an denen man arbeiten konnte, und viele talentierte, vielfältige Kollegen. Die Rekrutierung war einfach, weil neue Kandidaten den Geist unserer Gemeinschaft spürten.“

Pratima seufzte.

„In diesem Unternehmen können wir aber nicht verbergen, dass wir beim Übergang erst am Anfang stehen. Die Bewerber haben in der Regel viele Angebote, aus denen sie wählen können. Manchmal müssen wir zusehen, wie perfekte Kandidaten weggehen, weil wir noch kein warmes und einladendes Umfeld für Datenexperten bieten. Wenn Manager über KI und die Umstellung auf Daten diskutieren, übersehen einige vielleicht den menschlichen Aspekt. Was ist, wenn Sie die perfekte Datenstrategie entwickeln, aber nicht genügend Talente finden? Viele Unternehmen sind mit diesem Problem konfrontiert, und immer steht ein Elefant im Raum. Um ein datengesteuertes Unternehmen zu werden, müssen Sie ein Umfeld schaffen, das Menschen anzieht, die anders denken, und das bedeutet, dass Sie Ihre Kultur ändern müssen.“

„Glauben Sie, dass die Unternehmensleitung Angst hat, zu viele Veränderungen voranzutreiben, weil sie fürchtet, alles zu verlieren?“

„Ich verstehe, dass einige erfahrene Mitarbeiter enttäuscht sein und sogar kündigen könnten, wenn sich ihre gewohnte Umgebung zu modernisieren beginnt. Aber wenn man sich überhaupt nicht verändert, steckt man im Schlamm fest, und die Konkurrenz wird einen überflüssig machen. Der Dalai Lama sagt, wir sollten die Veränderung sein, die wir uns wünschen.“

„Richtig. Und ich glaube, es war Seneca, der einmal sagte: ‚Nicht, weil die Dinge schwierig sind, wagen wir sie nicht. Sondern weil wir es nicht wagen, sind sie schwierig.‘“

„Stimmt! Aber ich muss jetzt gehen. Ich freue mich darauf, unser Gespräch fortzusetzen.“

1.2.10 CEO

Alice und Bob trafen sich am Abend in einem Fusion Restaurant in der Innenstadt. Alice stellte Bob Santiago vor, den langjährigen CFO, der zum neuen CEO wurde. Nach einem ausgezeichneten Essen bestellten sie den berühmten armenischen Cognac und begannen mit der eigentlichen Diskussion.

„Ich will ehrlich zu Ihnen sein, Bob“, begann Santiago. „Alle Ihre Ideen zur Umgestaltung von Halford klingen fantastisch, aber als Ökonom und Zahlenmensch ist meine erste Frage: Wie viel wird das alles kosten?“

Oh Mann. Bob war auf diese Frage vorbereitet, aber er wusste, dass Santiago die Antwort nicht gefallen würde. „Es kommt darauf an“, sagte er, und Santiago sah genauso unzufrieden aus, wie Bob es erwartet hatte.

„Ich verstehe, dass jeder auf die Kosten schaut“, fuhr Bob fort, „aber die Geschichte ist voll von Unternehmen, die es versäumt haben, innovativ zu sein, und in Konkurs gegangen sind, während ihre Konkurrenten voranschritten. Wenn man das gesamte Spektrum der künstlichen Intelligenz betrachtet, wird kaum ein Unternehmen am Ende so arbeiten wie bisher.“

„Einige Unternehmen empfehlen, dass wir mit Workshops zur Data Literacy beginnen, um Führungskräfte in die Lage zu versetzen, Daten und Zahlen effizient zu interpretieren. Literacy hört sich so an, als wollten sie uns wieder das Lesen und Schreiben beibringen – und das natürlich für viel Geld. Verstehen Sie mich bitte nicht falsch. Ich verstehe, dass wir innovativ sein müssen, aber wenn ich alles genehmige, was die Berater mir vorschlagen, sind wir bald pleite.“

„Aber wenn Ihr Führungsteam nicht ‚in Daten denken‘ kann“, sagte Bob und deutete dabei Anführungszeichen an, „wie soll es dann an unserem geplanten Strategie-Workshop teilnehmen, in dem wir spezifische Data-Science-Optionen für unsere Geschäftsziele erkunden?“

„Worin besteht der Unterschied?“

„In den Data-Literacy-Workshops wollen wir ein Verständnis dafür schaffen, wie man Daten interpretiert. Im Strategie-Workshop erstellen wir eine Liste von Anwendungsfällen zur Verbesserung der Prozesse in Ihrem Unternehmen und priorisieren diese, um neue Datenlösungen schrittweise zu integrieren.“

„Ich weiß, dass wir ein paar harte Nüsse zu knacken haben. Einige unserer Mitarbeiter glauben nicht daran, dass wir datengesteuert arbeiten wollen, und wir müssen möglicherweise massiv in Enablement investieren. Wir haben einmal externe Unternehmen gebeten, uns bei der Modernisierung unserer IT zu helfen. Kein Beratungsunternehmen hat mir ein Angebot mit einem Festpreis für ein Umstellungsprojekt gemacht. Sie sagten immer, wir stünden vor einem Loch ohne Boden. Führung ist der einzige Weg, um voranzukommen. Wenn das Führungsteam überzeugt ist und sich anpasst, kann sich diese Kultur ausbreiten.“

Ihre betriebliche IT muss allmählich reifen und modernisiert werden. Seien Sie sich jedoch bewusst, dass eine analytische Schicht auch außerhalb der Unternehmens-IT aufgebaut werden kann. Ein Risiko besteht darin, Daten zu einem reinen IT-Problem zu machen. Die IT ist ein Teil einer Transformation, aber ein datengesteuertes Unternehmen zu werden, ist weit mehr, als einigen Ingenieuren einen Job zu geben, um Plattformen zu bauen.“

„Für mich ist die Sache klar“, sagte Alice. „Entweder wir modernisieren uns, oder wir verschwinden allmählich von der Bildfläche. Bob, was brauchst du, um uns zu helfen?“

Bob schaute von einem zum anderen und überlegte sorgfältig, was er als Nächstes sagen sollte. „Datenorientiert zu werden bedeutet nicht, dass man einen Haufen Data Scientists einstellt, die ein bisschen zaubern, und plötzlich verdient das Unternehmen mit KI tonnenweise Geld. Wie gesagt, der erste Schritt besteht darin, die Beteiligten in Einklang zu bringen. Für mich ist das das A und O der KI: die Schaffung einer Datenkultur, die auf kritischem Denken und evidenzbasierten Entscheidungen beruht.“

„Großartig“, antwortete Alice. „Fangen wir damit an.“

1.3 Kurz und bündig



Erwartungsmanagement

Die meisten Unternehmen sehen die Notwendigkeit, datengesteuert zu werden, da sie wissen, dass Organisationen, die die technische Entwicklung ignorieren, meist scheitern.

Einige Mitarbeiter haben vielleicht unrealistische Erwartungen, wie schnell eine Umstellung vorantreiben kann. Wir weisen darauf hin, dass die Umstellung auf ein datengesteuertes Unternehmen nicht nur eine Änderung von Praktiken und Prozessen bedeutet, sondern oft auch eine kulturelle Überarbeitung der Art und Weise, wie das Unternehmen seine Geschäfte führt.

Viele Mitarbeiter befürchten, dass sie einen Teil ihrer Autonomie aufgeben müssen oder sogar ihren Arbeitsplatz ganz an Computer verlieren, wenn KI in ihrem Unternehmen eingeführt wird. Eine Organisation, die datengesteuert wird, muss sich damit auseinandersetzen.

Technologiefokus und fehlende Strategie

Einige Unternehmen versuchen, ein Patentrezept zu finden, das alle Probleme löst. „Wir setzen einfach diese Technologie ein, wenden KI auf diese oder jene Weise an, und schon sind alle unsere Probleme gelöst“, denken sie. Eine zu starke Fokussierung auf Technologie ist jedoch ein Antipattern, das die Entwicklung eines Unternehmens hin zu einer datengesteuerten Organisation behindern kann.

Bei Data Science und KI geht es um mehr als nur das Verständnis von Frameworks und Methoden

Es ist zwar unerlässlich, über ein Team von qualifizierten Data Scientists und KI-Ingenieuren zu verfügen, um die richtigen KI-Frameworks auszuwählen und komplexe KI-Systeme zu entwickeln, aber für große Unternehmen gibt es noch viele andere Überlegungen, die zu beachten sind. Es ist ein Risiko, nicht zu verstehen, welche Bedürfnisse ein Unternehmen hat und wo KI einen Unterschied machen kann. Mit dem falschen Ziel wird jede Strategie scheitern.

Zusammenarbeit zwischen Analysten und IT

In einigen Unternehmen stellt die IT-Abteilung die Plattformen zur Verfügung, die die Analysten nutzen müssen. Wenn diese Plattformen fehleranfällig oder veraltet sind, kann das für die Analysten frustrierend sein. In modernen Umgebungen müssen nicht alle analytischen Plattformen von einer zentralen IT-Abteilung verwaltet werden. Dies kann den Datenteams mehr Freiheit geben, selbstständig zu arbeiten.

IT

Vielen IT-Teams fehlen die Ressourcen, um die für Data-Science-Plattformen erforderlichen Datenpipelines zu erstellen. Oft klafft eine Lücke zwischen Anwendern und Ingenieuren, sodass sie nur schwer miteinander kommunizieren können.

Die IT-Abteilung, insbesondere die Betriebsabteilung, ist oft darauf bedacht, Probleme zu vermeiden. Daher sind viele bestrebt, ihre Systeme vor Veränderungen zu schützen. Sie wollen den Zugriff auf Daten erschweren, um die Plattformen sicher zu halten. Data Scientists hingegen möchten einfach auf Daten zugreifen, um schnell Fortschritte zu erzielen. Dies kann zu Reibungen zwischen beiden Teams führen.

Kosten

Die Einführung von maschinellem Lernen, Data Science und künstlicher Intelligenz kann teuer werden. Es ist selten einfach zu sagen, wie stark sich Data Science auf die Ergebnisse des Unternehmens auswirken wird: Das Verhältnis zwischen Input (wie Zeit, Aufwand und Ressourcen) und Output ist alles andere als deterministisch. Die Alternative ist jedoch noch düsterer. Ein Unternehmen, das nicht bereit ist, in Innovationen zu investieren, wird schließlich seine Wettbewerbsfähigkeit verlieren und den Konkurs riskieren.

Schutz von Daten und Privatsphäre

Bürokratie um Datenschutzrichtlinien kann einige Projekte verlangsamen und aufwendiger machen, aber sie sind absolut notwendig. Außerdem möchte wahrscheinlich niemand in einem System leben, in dem die Privatsphäre nicht respektiert wird. In der täglichen Arbeit ist der Schutz der Privatsphäre prozessorientiert: Es geht darum, diese Prozesse transparent und effizient zu gestalten.

Mitarbeitende gewinnen

Die Einführung von Data Science kann bedeutendere Veränderungen in der Unternehmensstruktur oder -kultur erfordern, die verborgene Konflikte und Herausforderungen ans Licht bringen könnten, über die niemand sprechen möchte.

Datenexperten sind eine seltene Spezies, und da es nur wenige von ihnen auf dem Arbeitsmarkt gibt, wie kann ein Unternehmen ohne die erforderlichen Fähigkeiten überhaupt an Veränderungen denken? Um Ingenieure und Data Scientists zu gewinnen, ist oft ein Angebot erforderlich, das über den berühmten Obstkorb im Büro hinausgeht.

2

Das A und O der KI

Stefan Papp

„Unzureichende Fakten bergen immer Gefahren.“

Spock

„Wir haben zwei Ohren und einen Mund, also sollten wir mehr zuhören als sagen.“

Zenon von Kitium



Fragen, die in diesem Kapitel beantwortet werden:

- Wie könnte das menschliche Ego Einzelner den Erfolg oder Misserfolg von KI- und Dateninitiativen beeinflussen?
- Was ist Bias (Voreingenommenheit), und warum ist es wichtig, dieses Thema zu adressieren?
- Wie finden Sie objektiv Ihre Anwendungsfälle und Ziele für Data Science und KI?
- Wie können Workshops durchgeführt werden, um datenorientierter zu werden, und wie können gemeinsame Herausforderungen angegangen werden?

Lassen Sie uns mit einer Hypothese beginnen: Jede Form von Data Science und schließlich auch der künstlichen Intelligenz entwickelt sich aus einer evidenzbasierten Entscheidungsfindung, die auf kritischem Denken beruht. Selbst wenn die außergewöhnlichsten Köpfe zusammenkämen und ihnen unendlich viele Ressourcen zur Verfügung stünden, würden sie scheitern, wenn ihre Arbeit auf falschen Schlussfolgerungen, Fakten und Annahmen beruhte.

In Kapitel 1 stellen wir Tom, den Verkäufer, vor. Als Technologieskeptiker will er Entscheidungen aus dem Bauch heraus treffen. Stellen wir uns vor, der Personalleiter von Halford stellt eine neue Verkäuferin, Cherry, ein, die sich auf Fakten und Logik

verlässt. Angenommen, sie ist ebenso qualifiziert und erfahren wie Tom, wird sie dann erfolgreicher sein als er?

In der Wissenschaft geht es um Beweise. Die wissenschaftliche Methode soll subjektive Eindrücke oder Intuition bei der Erforschung eines Themas eliminieren. Die Aufgabe des Data Scientists besteht darin, Beweise für die vom Unternehmen vorgeschlagenen Unternehmungen zu erbringen, und dann in einem zweiten Schritt diese Beweisführung durch statistische Schlussfolgerungen unter Verwendung von Datenpipelines und möglicherweise maschinellen Lernmodellen zu automatisieren. Wenn wir uns eine Organisation vorstellen, die vollständig datengesteuert und faktenbasiert ist, könnten wir uns eine Einheit vorstellen, die ihren Betrieb vollständig darauf ausrichtet, Hypothesen darüber zu sammeln, was gut oder schlecht für das Unternehmen ist, und die auf hochautomatisierte Weise Beweise und Widerlegungen liefert. Sich auf Bauchgefühl und Intuition zu verlassen, könnte als veraltet und als mittelalterlicher Aberglaube abgetan werden.

In späteren Kapiteln werden wir zeigen, dass generative KI nicht immer evidenzbasierte Informationen liefert und dass künstliche Intelligenz voreingenommen sein kann, genau wie Menschen. Bedeutet dies, dass Cherry, die Logik und Fakten liebt, genauso voreingenommen sein kann wie Tom, wenn sie sich auf KI verlässt?

Beginnen wir mit einer Geschichte, um diese Hypothese aus einem anderen Blickwinkel zu betrachten.

2.1 Die Datenverwendungszwecke

Der Personalleiter von Halford hat gerade zwei junge Data Scientists eingestellt: Mahsa und Rachid. Beide beginnen damit, potenzielle analytische Anwendungsfälle zu untersuchen, um einen ersten Anwendungsfall zu ermitteln, der dem gesamten Unternehmen den Wert von Daten aufzeigen kann. Da es sich um ein kleines Team handelt, teilen sie sich die Arbeit auf, befragen verschiedene Abteilungen getrennt und untersuchen sie auf Data-Science-Anwendungsfälle. Irgendwann treffen sich Rachid und Mahsa im Besprechungsraum, um ihre Zwischenergebnisse zu besprechen.

2.1.1 Bias

„Wir sind ein Produktionsunternehmen“, sagt Rachid. „Wenn ich mir all die Fälle ansehe, die ich untersucht habe, sehe ich ein enormes Potenzial, die Zahl der fehlerhaften Teile in unseren Produktionsprozessen zu verringern. Ich weiß, dass das Produktionsteam unserer Arbeit immer noch ablehnend gegenübersteht, aber sie müssen den Wert dessen erkennen, was wir hier tun.“

„Vielleicht ist diese Feindseligkeit ein guter Grund, unseren ersten Anwendungsfall woanders zu starten“, antwortet Mahsa. „Wenn wir unser erstes Projekt vermasseln und uns einen schlechten Ruf erarbeiten, wird es lange dauern, bis wir das Vertrauen der Belegschaft zurückgewinnen. Wir haben Verbündete in der Abteilung für Kundenbeschwerden; sie haben eine neue Chefin, die sich beweisen will, und sie hat mir gesagt, dass sie die erste Abteilung in Halford sein will, die vollständig datengesteuert ist. Wir sollten mit der Entwicklung eines KI-Chatbots beginnen.“

„Komm schon“, sagt Rachid, „Beschwerdemanagement ist nicht unser Kerngeschäft. Konzentrieren wir uns auf etwas, das Einnahmen bringt.“

„Wir verdienen Geld mit zufriedenen Kunden“, entgegnet Mahsa.

„Ja, und *meine* Data-Science-Fallstudien werden dafür sorgen, dass unsere Kunden bessere Produkte bekommen.“

„Du hast es in den ersten Tagen selbst gesagt: Wenn es um die Datenqualität geht, heißt es: ‚Garbage in, garbage out‘. Meinem ärgsten Feind würde ich die Fabrikdaten nicht geben wollen. Wir werden einen Monat brauchen, um die Werksdaten auf ein akzeptables Niveau zu bringen.“

„Aber kein langweiliger Chatbot löst ein Geschäftsproblem“, betont Rachid. „Ich habe schon mit Fabrikdaten gearbeitet. Ich weiß, wovon ich spreche. Soweit ich weiß, gehörst du zu den besten Absolventen deiner Universität, aber was hast du außer der Erforschung von Sprachmodellen in einem Universitätskurs gemacht? Hast du irgendwelche praktischen Erfahrungen?“

Mahsa schüttelt den Kopf und fügt laut hinzu: „Ich habe hart daran gearbeitet, diese Strategie zusammen mit dem neuen Leiter der Beschwerdeabteilung zu entwickeln. Wenn wir mit etwas anderem anfangen, ist meine ganze Mühe umsonst gewesen.“

* * *

Obwohl Mahsa und Rachid noch am Anfang ihrer Laufbahn stehen, kommt diese Art von Gespräch auch bei erfahreneren Fachleuten vor. Beide werben für ihre Ideen und ignorieren die Meinung des anderen. Beide versuchen, fachlich zu argumentieren, aber sie liefern keine schlüssigen Beweise für ihre Aussagen. Letztlich geht es ihnen darum, einen persönlichen Vorteil zu erlangen: Für Mahsa geht es darum, ihre gute Beziehung zum Leiter der Reklamationsabteilung zu pflegen, und für Rashid darum, seine Erfahrungen aus früheren Projekten zu nutzen. Die Diskussion wird mehr und mehr von Emotionen bestimmt. Sie könnten sich in einem Teufelskreis eines Confirmation Bias befinden.

Vielleicht schlafen die beiden über diesen Konflikt, und vielleicht wachen sie am nächsten Tag mit frischem Geist auf und ändern ihre Meinung. Mahsa könnte anerkennen, dass Rashid Recht hatte, als er sagte, dass Produktionsdaten für den Erfolg des Unternehmens entscheidend sind. Rashid könnte zustimmen, dass Mahsas Begründung, in einer anderen Abteilung anzufangen, der risikoärmere Ansatz sein könnte. Im schlimmsten Fall werden beide auf ihren Ansichten beharren, und ihre

Differenzen könnten zu einer Fehde führen, wenn sie nicht einen Manager mit ausreichenden Führungsqualitäten haben, der ihnen hilft, ihren Konflikt zu lösen.



Confirmation Bias

Als Confirmation Bias bezeichnet man das Phänomen, dass wir dazu neigen, Dinge zu ignorieren, die unseren Ansichten widersprechen, aber alle Beweise, die unsere Ansichten bestätigen, in Erinnerung behalten.



Bild 2.1 Confirmation Bias¹⁾

Bias wird oft bereits in den ersten Statistikvorlesungen für Studenten behandelt. Ohne zu lernen, Daten korrekt zu interpretieren, können wir keine fundierten Aussagen über unsere Umwelt machen.

Der Lebenszyklus von Data Science besteht in der Regel aus den folgenden Schritten:

1. Erstellen einer Hypothese darüber, was anhand von Daten geändert werden kann
2. Erstellen eines Konzeptnachweises (PoC)
3. Automatisieren einer Lösung

Allerdings kann es zu Diskussionen kommen, bis eine Lösung in Produktion ist, nachdem sie zuvor mehrere Runden positiver Bewertungen und Rückmeldungen erhalten hat. Nehmen wir an, Mahsa hat einen Weg gefunden, einen KI-Chatbot zur Verwaltung der Kundenkommunikation einzuführen, und sie hat gezeigt, wie Kunden mit diesem Chatbot kommunizieren können. Rashid könnte immer noch einwenden und behaupten, er könne mehr Wert schaffen als Mahsa, wenn er bis dahin das Budget für die ersten Anwendungsfälle von Fabrikdaten erhalten hätte.

¹ Copyright: <https://www.simplypsychology.org/confirmation-bias.html>

Selbst wenn die ersten Tests darauf hindeuten, dass Mahsas Innovation den Bedarf an Callcenter-Mitarbeitern verringern könnte, könnte es Monate dauern, um zu überprüfen, ob sich die allgemeine Kundenzufriedenheit dadurch verbessert. Vielleicht hat Mahsa Recht, und sie wird schließlich zur Heldin der Leiterin der Beschwerdeabteilung und all der zufriedenen Kunden, die sie dann bedienen können. Ein anderes mögliches Ergebnis ist, dass Rashids Hypothese richtig ist: Es spielt keine Rolle, ob Kunden mit einem Chatbot chatten oder Antworten von einem Menschen erhalten, der ihre Probleme nicht sofort versteht; was Kunden wirklich wollen, sind Produkte von besserer Qualität.



Übungen

- Fragen Sie einen Chatbot mit generativer KI (GenAI), z. B. Google Gemini oder ChatGPT, nach anderen Arten von Bias.
- Stellen Sie sich vor, wir würden einen Weg finden, die Selbsttäuschung in dieser Welt loszuwerden. Wäre die Welt dann ein besserer Ort? Warum ist die Selbsttäuschung in der Menschheit überhaupt so weit verbreitet?
- Wie stellen Sie sicher, dass Sie selbst nicht befangen sind?
- Was halten Sie von der folgenden Hypothese? „Voreingenommene Informationen, die von künstlicher Intelligenz oder Data-Science-Projekten stammen, sind schlimmer als menschliche Aussagen, die voreingenommen sind.“
- Nachdem Sie die Debatte zwischen Rashid und Mahsa gelesen haben, könnten Sie sich Strategien ausdenken, die ihre Debatten von vornherein verhindern würden?

2.1.2 Datenkompetenz

Eine Frage im Dialog in Abschnitt 2.1.1 beschäftigt sich mit dem tatsächlichen Wert für das Unternehmen. In diesem Beispiel haben wir über die Verringerung des Ausschusses in einer Fabrik gesprochen, aber was bedeutet das in Zahlen? Eine Aussage wie „Jeden Tag reduzieren wir die Anzahl der Ausschüsse um 30 %“ klingt hervorragend, aber was bedeutet das ohne jeden Bezug?

In unserem Beispiel ist die wirtschaftliche Optimierung das Ziel. Wir brauchen Daten zu Kennzahlen, die mit der Steigerung langfristiger Gewinne verbunden sind. Ein Beispiel für eine solche Finanzkennzahl ist die **Kapitalrendite** oder **Amortisierung**. Versuchen wir, dies anhand unseres Beispiels einer Fabrik zu verstehen. Mit einem neuen System können wir unsere Produktionspipeline verbessern und täglich etwa 1000 EUR einsparen, was uns jährlich stolze 365 000 EUR einbringt. Was aber, wenn die Investition 2 000 000 EUR kostet? Nun, in etwa fünfeinhalb Jahren wird sich diese Lösung amortisiert haben. Was aber, wenn die Produktionsleitung in fünf Jahren stillgelegt werden soll, weil eine neue Generation von Werksleitungen sie ersetzen wird?

Eine weitere Frage sind die Opportunitätskosten. Stellen Sie sich vor, dass ich für 2 000 000 EUR die Kosten in einem anderen Anwendungsfall um mehr als 365 000 EUR senken könnte.

Wie sicher sind wir, dass unsere Zahlen korrekt sind? Sind wir sicher, dass wir nicht etwas übersehen haben? Vielleicht sogar etwas, das alles andere überflüssig macht? Welche Risiken bestehen, wenn etwas Unerwartetes passiert?

Dies ist der Moment, in dem Data Science auf die Erwartungen der Unternehmen trifft. Das Verständnis komplexer Systeme, ihrer Interaktion und der Metriken, die sie beschreiben, ist mehr als nur eine Aufgabe für einen Business Analysten. Ein Arbeitsablauf für einen Data-Science-Fall von Anfang bis Ende könnte wie folgt aussehen:

1. **Identifizieren eines Unternehmensziels:** Die Unternehmensleitung möchte die Automatisierung der Fabrik erhöhen, um die Qualität zu steigern und den manuellen Aufwand zu verringern.
2. **Erstellen von Hypothesen:** Derzeit überprüfen Mitarbeiter den Output der Fabrikpipeline manuell auf Produkte mit Qualitätsmängeln. Wir stellen die Hypothese auf, dass „die Erkennung von Ausschuss durch Computer-Vision-Algorithmen die Kosten senken und die Qualität verbessern wird“.
3. **Angebot eines Werts (Value Proposition):** Wir fassen die stündlichen Kosten pro Linie für manuelle Qualitätskontrollen zusammen. Wir schätzen auch die Kosten für die Entwicklung eines Systems für automatisierte Qualitätskontrollen und dessen Einsatz auf allen Linien. Wir stellen fest, dass sich die Investition in drei Jahren amortisiert haben dürfte.
4. **Erstellen Sie einen Konzeptnachweis:** Wir erstellen einen Prototyp, und anhand von Trainingsdaten sind die ersten Ergebnisse vielversprechend. Auf der Grundlage unseres Modells und weiterer Erkenntnisse während der Konzeptionsphase kommen wir zu dem Schluss, dass das Unternehmen von unserer Value Proposition profitieren wird.
5. **Einsatz der ersten Testversion:** Wir installieren Videokameras in einer unserer kleineren Produktionspipelines, und Maschinenbauingenieure fügen einen Mechanismus hinzu, der eine Ware aus der Pipeline entfernt, wenn das Computer-Vision-Modell ihnen mitteilt, dass es Defizite aufweist.
6. **Kontinuierliche Verbesserung des Produkts bis zur Reife:** Unsere erste Testversion muss noch stark verbessert werden. Wir sammeln Daten, um die Effizienz unseres Algorithmus zu messen und unsere Lösung schrittweise zu verbessern. Wenn die Leistung unseres Algorithmus mit der menschlicher Arbeiter übereinstimmt, setzen wir die Lösung in verschiedenen Fabriken ein. Das computer-gestützte System wird weiter verbessert. Irgendwann wird das neue System die alte Lösung übertreffen und überall eingesetzt werden.
7. **Eine Retrospektive:** Nach dem Erfolg unseres Projekts treffen wir uns, um zu besprechen, wie wir in Zukunft noch erfolgreicher sein können.

Skeptiker könnten die Herausforderungen sehen, denen sich ein Unternehmen stellen könnte. Was ist, wenn das Nutzenversprechen auf falschen Zahlen beruht? Was ist, wenn die Kosten dieser Migration falsch berechnet und wichtige Kostenfaktoren übersehen werden? Wie können die Führungskräfte eines Unternehmens sicher sein, dass die Teams, die einen Anwendungsfall umsetzen, Probleme transparent darstellen, die den Projekterfolg beeinträchtigen könnten? Was ist mit gescheiterten Anwendungsfällen? Wenn ein Unternehmen im Durchschnitt 100 000 USD pro fehlgeschlagenem Anwendungsfall verliert, muss es nach fünf fehlgeschlagenen Anwendungsfällen in einem erfolgreichen Anwendungsfall 500 000 USD gewinnen, um die Gewinnschwelle zu erreichen. Was ist, wenn sich alle Bemühungen nicht auszahlen? In diesem Kapitel möchten wir darlegen, warum es wichtig ist, den Übergang zu einer stärker datengesteuerten Vorgehensweise zu vollziehen, auch wenn dieser Übergang kostspielig sein kann.



Übungen

- In einigen Unternehmen erhalten Manager und Führungskräfte Bonuszahlungen für das Erreichen von Jahreszielen. Was könnte hier Ihrer Meinung nach ein Problem darstellen?
- Fallen Ihnen Maßnahmen ein, wie man die Datenkompetenz in einem Unternehmen verbessern kann?
- In diesem Kapitel behaupten wir, dass Tom, der Verkäufer, der nach seinem Bauchgefühl entscheidet, eine neue Kollegin namens Cherry bekommt, die nur auf Logik und Fakten vertraut. Angenommen, sie sind in ihrer täglichen Arbeit gleich gut qualifiziert, denken Sie, dass sie ihn schließlich übertreffen wird?

2.2 Kulturschock

Jedes Unternehmen hat eine Kultur, und es gibt immer Gründe, warum Unternehmen so sind, wie sie sind. In Kapitel 1 haben wir ein Unternehmen vorgestellt, das von der Persönlichkeit seines Gründers geprägt ist. Es hat strenge Hierarchien, und viele Mitarbeiter haben individuelle Strategien, um ihre täglichen Aufgaben zu erfüllen. Wir haben absichtlich einige Charaktere eingeführt, die manche als toxisch bezeichnen könnten. Die harte Realität ist, dass niemand erwarten kann, in einem Unternehmen zu arbeiten, in dem jeder ein Abbild des perfekten Mitarbeiters ist. Die Mischung von Persönlichkeiten mit Charaktereigenschaften in verschiedenen Rollen wird sich auf eine digitale Transformation auswirken, bei der die Wertschöpfung auf der Grundlage von Daten anders erfolgen soll. Manche Mitarbeiter lehnen den Wandel vielleicht absichtlich ab. Einige sind vielleicht von der Vorstellung geprägt, dass sie bald in Rente gehen werden, während andere Pessimisten sind, die daran zweifeln, dass

irgendetwas funktionieren könnte. Andere sehen vielleicht die Chance auf einen Karriereschub, wenn sie sich an einem solchen Vorhaben beteiligen, und beanspruchen für sich Führungsrollen, für die sie nicht geeignet sind. Diese Personen könnten in die Defensive geraten, wenn sie das Gefühl haben, dass andere mehr Wissen oder Erfahrung haben. Es gibt noch viele andere Gründe, warum sich Mitarbeiter seltsam verhalten können.



F+E-Abteilungen

Eine Möglichkeit, Konflikte in der Organisation zu lösen, besteht darin, Abteilungen zu schaffen, die außerhalb des Tagesgeschäfts tätig sind und deren Aufgabe es ist, nach strategischen Lösungen für die Gestaltung der Zukunft des Unternehmens zu suchen. Es scheint auf der Hand zu liegen: Während die einen erforschen, was in der Zukunft getan werden kann, können die anderen ungestört weiterarbeiten, um in der Gegenwart Werte zu schaffen.

Ein mögliches Risiko besteht darin, dass sich diese F+E-Abteilungen zu sehr vom eigentlichen Geschäft abkoppeln. Ihre Ideen könnten dann nicht mit der geschäftlichen Realität übereinstimmen. Stellen Sie sich Teams von talentierten technischen Experten vor, die Datenplattformen bauen, die niemand wirklich braucht, weil sie kaum mit den Nutzern zusammentreffen und deren wirkliche Probleme kennenlernen.

Ein weiteres Risiko besteht darin, dass die operativen Teams das Gefühl haben, dass die Forschungs- und Entwicklungsabteilung die ganze interessante Arbeit bekommt, während sie sich mit komplexen Kunden- oder Nutzeranfragen auf veralteten Datenplattformen herumschlagen müssen. Die Manager der operativen Teams könnten auch strategische Aufgaben zuweisen, um frustrierten Mitarbeitern eine neue Perspektive und neue Karriereziele zu bieten. Letztendlich könnten operative und F+E-Teams an strategischen Lösungen arbeiten, die nicht miteinander harmonieren, was zu Reibungen führen könnte.

Außerdem sind Mitarbeiter mit einzigartigen Persönlichkeiten in Unternehmen durch ihre Vergangenheit geprägt. Stellen Sie sich ein Unternehmen vor, das gehackt wurde und einen Teil seines Rufs und Geldes verloren hat. Dieses Unternehmen würde Sicherheitsfragen besonders ernst nehmen. Auch das Durchschnittsalter könnte einen Einfluss haben. Ein Unternehmen mit mehr jungen Mitarbeitern könnte anders handeln als ein Unternehmen mit im Durchschnitt älteren Mitarbeitern. Internationale Unternehmen sind vielen Kulturen ausgesetzt, was mehr Möglichkeiten und Missverständnisse zwischen Kollegen bedeuten könnte als bei Unternehmen, die nur in einem Land tätig sind.



Furcht vor Veränderung als Hemmnis

Das Bildungswesen ist ein hervorragendes Beispiel dafür, wie die Angst vor Veränderungen die Einführung von KI-gesteuerten Prozessen verhindern kann.

Wir sehen jetzt, was generative KI leisten kann, und es ist nicht schwer, sich vorzustellen, wozu die nächsten Iterationen von Frameworks wie GPT in der Lage sein könnten. KI-Algorithmen können die Leistungen der Schüler perfekt analysieren und umfassendes Feedback geben. Sie werden keine Fehler übersehen, weil sie müde sind oder von privaten Problemen betroffen sind, und kein Schüler wird sich jemals darüber beschweren können, dass er eine schlechtere Note bekommen hat, weil die KI ihn als Person nicht mochte.

Generative KI wird irgendwann in der Lage sein, detaillierteres Feedback zu geben als jeder Lehrer. KI könnte genau das Wissen herausfinden, das den Schülern für eine gute Note fehlt. Sie könnten sogar individuelle Tests für Schüler erstellen, die bestimmte Details noch nicht verstanden haben: Einige Sprachschüler könnten speziell auf ihre Grammatik getestet werden, andere auf ihren Wortschatz; bei einigen Mathematikschülern könnte man feststellen, dass sie zwar Formeln verstehen, aber nicht in der Lage sind, komplexe Probleme zu lösen, und sie könnten entsprechend gecoacht und getestet werden. Anhand von Daten lässt sich sogar feststellen, wann Schüler am besten lernen und welche Unterrichtsmethoden am effektivsten sind. Für jeden Schüler kann ein individuelles, KI-optimiertes Curriculum erstellt werden.

In einer vollständig digital transformierten und KI-gesteuerten Schule können Lehrer KI als Werkzeug sehen, das ihnen zum Erfolg verhilft, und sie können sich eher als Mentoren für Schüler sehen, um sie dabei zu unterstützen, das Beste aus der Arbeit mit KI herauszuholen. Oder sie sehen KI als Bedrohung, als ein unmenschliches System, das sie erniedrigt und ihnen ihre Freiheit nimmt. Am Ende könnten sie eine Gegenrevolution starten mit dem Slogan: „Hey, KI, leave us teachers alone“. Unabhängig davon, ob ihre Rebellion gerechtfertigt ist oder nicht, würden sie eine Firewall gegen KI fordern. Und da zum Bildungssystem auch Schüler und Eltern gehören, könnten auch sie eine Reihe positiver und negativer Ansichten zum KI-gestützten Lernen beitragen.

Wenn man über die Möglichkeit von vollständig KI-gesteuerten Organisationen nachdenkt, kann man sich leicht vorstellen, wovor die Menschen Angst haben könnten. Die Mitarbeiter fürchten sich vielleicht nicht unbedingt davor, gegen den Terminator kämpfen zu müssen, aber sie könnten dennoch Angst haben, ihre Autorität und Souveränität zu verlieren. Arbeitsplätze haben mit Status zu tun, und die Menschen sind stolz auf sie. Wenn Operationen automatisiert werden, wird das Prestige der „Halbgötter in Weiß“ sinken. Wenn Computer Wege finden, alles supereffizient zu machen, verlieren die Menschen ihre Spontaneität und die Freiheit, gelegentlich ineffizient zu sein.

Die harte Realität ist jedoch, dass keine Gesellschaft die künstliche Intelligenz ignorieren kann, es sei denn, sie will ausdrücklich ohne Technologie leben, wie es die Amish tun. Viele Länder sind mit dem demografischen Wandel konfrontiert, und die Unternehmen finden nicht genügend qualifizierte Mitarbeiter, um diejenigen zu ersetzen, die in den Ruhestand gehen. Wenn nicht mit Automatisierung durch KI, wie sollen wir dieses Problem sonst lösen?

Die KI wird jedoch den Status quo stören. Einige werden diese Veränderungen begrüßen, während andere sie fürchten werden. Zu Beginn dieses Abschnitts über den Kulturschock haben wir ein Unternehmen mit vielen Einzelpersonen vorgestellt, die alle ihre eigenen Prioritäten und Ziele haben. Einige sehen vielleicht die Chance, durch eine digitale Transformation mehr Einfluss zu gewinnen, andere fürchten sich davor und versuchen vielleicht sogar, sie zu sabotieren. Die kleinen politischen Spielchen, die manche Mitarbeiter spielen, könnten sich als die größten Herausforderungen erweisen, denen sich manche Unternehmen jemals stellen müssen.

In seinem Buch *The Geek Way* schlägt Andrew McAfee vor, dass der Erfolg in einem innovativen Umfeld eine bestimmte Denkweise erfordert.²⁾ Diese Denkweise muss seiner Meinung nach die folgenden Eigenschaften umfassen:

1. **Leidenschaft und fundiertes Wissen:** Geeks verfügen oft über fundierte Kenntnisse in bestimmten Bereichen, was in einem beruflichen Umfeld, in dem Fachwissen und detailliertes Verständnis gefragt sind, von großem Vorteil sein kann.
2. **Kreativität und Problemlösungsfähigkeit:** Die Fähigkeit, über den Tellerrand hinauszuschauen, wie sie in der Geek-Kultur in Form von Spielen, Fantasy und Science-Fiction häufig anzutreffen ist, ist für innovatives Denken am Arbeitsplatz von großem Wert.
3. **Technologische Versiertheit:** In Anbetracht des digitalen Charakters der modernen Welt betont das Buch den bedeutenden Vorteil, technisch versiert zu sein – eine Eigenschaft, die unter Geeks weit verbreitet ist.
4. **Gemeinschaft und Zusammenarbeit:** McAfee erörtert auch, wie sich die gemeinschaftlichen Aspekte der Geek-Kultur wie z. B. der Austausch von Wissen und die Zusammenarbeit an Projekten, gut in teambasierte Arbeitsumgebungen übertragen lassen.

² <https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle>



Übungen

- Braucht ein Unternehmen einen Kulturwandel, um datengesteuert zu werden? Fallen Ihnen weitere Ideen ein, wie man die Datenkompetenz in einem Unternehmen verbessern kann?
- Wenn ein Kulturwandel erforderlich *ist*, würde ein Unternehmen diesen Wandel auch dann vollziehen, wenn es zu viele „Dinosaurier“ und zu wenige „Geeks“ hätte?
- Halten Sie sich eher für einen Geek oder für einen Dinosaurier? Und selbst wenn Sie sich für einen Geek halten, wie können Sie sicherstellen, dass Sie keine Dinosaurier-Tendenzen in sich verbergen?
- Es gibt Leute, die behaupten, dass in jeder großen Organisation ein gewisses Maß an „organisatorischer Arroganz“ herrscht und dass die Qualität der Arbeit des Einzelnen von dem Maß an Arroganz abhängt, das vorhanden ist. Was halten Sie davon, wenn man bedenkt, dass organisatorische Arroganz ein kreatives Umfeld zerstören und den Aufbau einer Datenkultur behindern kann?
- In diesem Kapitel haben wir vorgeschlagen, dass GPT mit menschlichen Lehrern konkurrieren könnte. Gehen Sie zu OpenAI, wählen Sie ein benutzerdefiniertes GPT-Modul und denken Sie an ein Gebiet, von dem Sie glauben, dass Sie nicht viel darüber wissen. Versuchen Sie nun, in kurzer Zeit so viel wie möglich über dieses Gebiet zu lernen, indem Sie GPT bitten, Sie zu unterrichten. Was denken Sie? Haben Sie schneller Fortschritte gemacht als mit einem menschlichen Lehrer?
- Wenn Sie diese Frage mit „Ja“ beantworten, welche Auswirkungen könnte es Ihrer Meinung nach auf die Gesellschaft haben, wenn Bildungseinrichtungen auf KI-gestütztes Lernen umstellen? Und wären die traditionellen Lehrpläne weiterhin erforderlich? Könnten Länder, die frühzeitig auf KI-basiertes Lernen umstellen, einen wirtschaftlichen Vorteil erlangen?
- Wenn die KI so ausgereift ist, dass sie Schüler besser unterrichten kann als Lehrer, wie notwendig wird es dann noch sein, Schüler zu unterrichten? KI könnte auch in den Berufen, in denen diese Schüler nach ihrem Abschluss arbeiten würden, bessere Leistungen erbringen

2.3 Ideenfindung

In der Regel führen erfolgreiche Anwendungsfälle von Data Science zu mindestens einem der folgenden drei Ergebnisse:

- Senkung der Kosten,
- schnellere Entscheidungen
- oder das Eindringen in neue Märkte.

Wir haben gesehen, dass persönliche Meinungen und Gefühle nicht ausreichen, um über die Angemessenheit eines Data-Science-Falls zu entscheiden. Dennoch müssen wir uns zunächst die richtigen Fragen darüber stellen, was wir tun wollen.

In diesem Abschnitt zeigen wir Ihnen, wie Sie Ideation-Workshops organisieren, um die richtigen Ideen zu finden, und wie Sie diese nach Relevanz gruppieren und priorisieren können. Und schließlich zeigen wir Ihnen, wie Sie sie testen können.

Alles beginnt mit einem strategischen Ziel, das einem Data-Science-Team die Richtung vorgibt. Auf höchster Ebene sind die Ziele in verschiedenen Organisationen ähnlich: Mehr Gewinn machen, umweltfreundlicher sein, die Mitarbeiter zufriedener machen usw. Wenn man ins Detail geht, wird die Strategie präziser. Hier sind einige Beispiele:

- **Telekommunikation:** „Wir wollen den Umsatz steigern, indem wir die Kundenabwanderung reduzieren.“
- **Fertigung:** „Aufgrund des demografischen Wandels müssen wir automatisieren. Wir wollen in unseren Fabriken mehr produzieren, mit weniger Menschen.“
- **Automotive:** „Wir wollen die Fähigkeit zum autonomen Fahren der Stufe fünf erreichen. Wir wollen alle Hindernisse beseitigen, um unser Ziel innerhalb der nächsten zwei Jahre zu erreichen.“
- **Finanzen:** „Wir wollen der beste Investment-Broker sein und unseren Kunden einen besseren Einblick in Wertpapiere und Anleihen geben als unsere Konkurrenten.“
- **Jedes Unternehmen:** „Wir wollen die CO₂-Emissionen um 50 % reduzieren.“

Das Verständnis des Kontextes eines Ziels – zum Beispiel des geschäftlichen Umfelds – ist für die Herangehensweise an eine Lösung von wesentlicher Bedeutung.

Wenn ein Unternehmen seine Ziele kennt, kann es Workshops organisieren und Fachleute, Data Scientists und andere Beteiligte einladen. Das Ziel ist ein Brainstorming. Es ist sinnvoll, diesen Workshop außerhalb des üblichen Geschäftsumfelds abzuhalten. Die Teams sollten ein Umfeld erleben, das ihnen hilft, kreativ zu sein und Lösungen zu finden, um die Ziele des Unternehmens zu erreichen.

Mit den Zielen vor Augen können die Teams als Einstieg die aktuellen Prozesse und Arbeitsabläufe untersuchen. Dann können sie damit beginnen, Hypothesen aufzustellen, die jedoch hauptsächlich auf Intuition und persönlichen Überzeugungen beruhen können. Solche Annahmen wären etwa:

- „Ich glaube, dass die Kosten die Hauptsorge unserer Kunden sind. Wenn wir unseren Kunden garantieren, dass wir unsere Dienstleistungskosten nicht erhöhen, werden sie uns treu bleiben.“
- „Ich sehe eine Menge Ausschuss in unserer Produktionskette. Wenn wir einen Weg finden, die Anzahl der Ausschüsse zu reduzieren, können wir Geld sparen.“

- „Unsere Hauptverwaltung ist uralte. Sie muss eine Quelle der Energieverschwendung sein. Wir sollten in der Lage sein, eine Menge zu tun, um ihre Energieeffizienz zu verbessern und unseren Ruf als verantwortungsbewusstes Unternehmen zu stärken.“
- „Wir haben einen komplizierten Verkaufsprozess und werden mehr verkaufen, wenn wir ihn vereinfachen.“

Diese Hypothesen sind immer noch persönliche Annahmen und höchst subjektiv. Der nächste Schritt besteht darin, die Daten zu sammeln und nach Fakten zu suchen, die die unsere Hypothesen beweisen oder widerlegen.



Organisieren eines Workshops

Ideation-Workshops können dazu beitragen, Klarheit in die Datenstrategie eines Unternehmens zu bringen. Ihre Organisation erfordert oft eine gute Vorbereitung, um den Erfolg zu gewährleisten. Es ist sinnvoll, einen Ort zu wählen, an dem sich die Teilnehmer entspannen können und der nicht im Büro liegt. Die Veranstaltungen müssen an das Publikum angepasst werden. Teambuilding-Spiele können dazu beitragen, dass sich die Menschen öffnen, aber manche Teams sind nicht für Spiele geschaffen.

Wichtig ist auch, dass genügend Material für die Workshops zur Verfügung steht, z. B. Karton oder Buntstifte. Viele Hotels stellen Räume für Seminare zur Verfügung, die in der Regel gut ausgestattet sind.

Der Dalai Lama wird mit den Worten zitiert: „Wenn du redest, wiederholst du nur, was du schon weißt. Aber wenn du zuhörst, kannst du etwas Neues lernen.“ Manche Menschen haben den Ruf, dass sie nur reden, um sich selbst reden zu hören. Aber wenn dieselben Leute, die in der Vergangenheit viel geredet haben, auch weiterhin reden, warum sollte sich dann etwas ändern? In solchen Konstellationen kann es hilfreich sein, mit einem externen Moderator zu arbeiten



Übungen

- Wie kann man in einem Workshop sicherstellen, dass Introvertierte mehr und Extravertierte weniger reden?
- Hand aufs Herz! Sind Sie normalerweise derjenige, der am meisten redet? Wie können Sie sich selbst stoppen?
- Was wäre Ihrer Meinung nach das perfekte Umfeld für einen Workshop, um neue Strategien für ein Unternehmen zu entwerfen?
- Sind Sie der Meinung, dass eingefleischte Pessimisten von visionären Workshops ausgeschlossen werden sollten, da ihre negativen Ansichten jeden kreativen Prozess zunichtemachen könnten? Oder sind die Pessimisten die einzige Kraft, die zwischen echten Lösungen und den Luftschlössern steht, die eingefleischte Optimisten aufbauen, wenn sie keiner bremst?

2.4 Entwurfsprozessmodelle

Modelle für Entwurfsprozesse sind Rahmenwerke, die Teams dabei helfen sollen, ihre Ziele durch die Anwendung bewährter Verfahren zu erreichen. Sie können für viele Produkttypen verwendet werden, nicht nur für Data-Science-Anwendungsfälle und KI-Produkte.

Alle Designprozessmodelle haben eines gemeinsam: Man muss zunächst verstehen, *was* man lösen muss, und dann mögliche Antworten auf das *Wie* finden. Schließlich wählen Sie die beste Lösung aus. Sie können die Prozesse pragmatisch an die Bedürfnisse Ihres Unternehmens anpassen. Wie bei jedem kreativen Prozess hängt der Erfolg bei der Erstellung eines Datenprozessmodells wesentlich davon ab, dass man die *richtigen* Leute an den *richtigen* Ort bringt und die *richtige* Atmosphäre schafft, um an den *richtigen* Problemen zu arbeiten.

Es gibt verschiedene Modelle für den Entwurfsprozess, jedes mit Vor- und Nachteilen für unterschiedliche Situationen. Wir werden nun zwei der gängigsten Modelle untersuchen: Design Thinking und das Double Diamond Framework. Anschließend erörtern wir, wie solche Workshops durchgeführt werden können.

2.4.1 Design Thinking

Design Thinking ist ein nichtlinearer, iterativer Prozess, den Teams nutzen, um die Nutzer zu verstehen, Annahmen in Frage zu stellen, Probleme neu zu definieren und innovative Lösungen zu entwickeln, die dann als Prototypen gebaut und getestet werden.³ Er umfasst fünf Phasen:

1. **Einfühlungsvermögen:** Verstehen Sie die Bedürfnisse des Benutzers.
2. **Definieren:** Formulieren Sie die Bedürfnisse des Nutzers klar und deutlich.
3. **Ideen entwickeln:** Brainstorming über mögliche Lösungen.
4. **Prototyp:** Erstellen Sie eine Musterlösung.
5. **Testen:** Probieren Sie die Lösung aus und verfeinern Sie sie anhand des Feedbacks.

Dies führt uns zu einer weiteren Hypothese. Wenn wir eine solche Methode brauchen, könnte eines unserer Probleme darin bestehen, dass wir glauben zu wissen, was unsere Nutzer oder Kunden wollen. Ein Zyniker könnte uns jetzt als „Möchtegern-Gedankenleser“ bezeichnen. Wenn wir dieses Verhalten im Kontext der Psychologie nachschlagen, könnten wir die Konsensverzerrung erforschen.

³ <https://www.interaction-design.org/literature/topics/design-thinking>

Es ist sehr verlockend zu sagen: „Ich habe die Arbeit unserer Kollegen gesehen, und gelegentlich habe ich in einer Pause mit ihnen geplaudert. Ich weiß, was sie brauchen, um besser zu arbeiten.“ Eine solche Haltung führt oft dazu, dass die Bedürfnisse der Benutzer nicht wirklich erfasst werden. Die Anwendung der Design-Thinking-Methode zwingt uns jedoch dazu, unsere Kunden ernst zu nehmen und uns selbst herauszufordern, einen systemischen Ansatz zu verwenden, um die Anforderungen der Nutzer zu verstehen.

Design Thinking wird oft als weniger strukturiert angesehen als das Double-Diamond-Modell, das wir als Nächstes kennenlernen werden. Es gibt einem Team mehr Freiheit und ist stärker auf die Nutzer ausgerichtet. Aber ist aus unternehmerischer Sicht die Konzentration auf den Kunden nicht das ultimative Erfolgsrezept?

2.4.2 Double Diamond

Der britische Rat für Design (British Design Council) kündigte dieses Modell des Designprozesses im Jahr 2005 an,⁴ in Anlehnung an das 1996 von dem ungarisch-amerikanischen Linguisten Béla H. Bánáthy vorgeschlagene Divergenz-Konvergenz-Modell.⁵

Die beiden Rauten stehen für das Erforschen eines Themas auf breiterer oder tieferer Ebene (divergentes Denken) und das anschließende **Ergreifen** gezielter Maßnahmen (konvergentes Denken). Leser, die an positiver Psychologie interessiert sind, könnten Parallelen zu Barbara Fredricksons *Broaden and Build*-Theorie erkennen.⁶

Die **Double-Diamond-Methode** legt nahe, dass der Entwurfsprozess vier Phasen umfassen sollte:

1. **Entdecken:** Verstehen Sie das Problem, anstatt es nur zu vermuten. Dazu müssen Sie mit den Menschen sprechen, die von den Problemen, die Sie lösen wollen, betroffen sind, und Zeit mit ihnen verbringen.
2. **Definieren:** Die in der Entdeckungsphase gewonnenen Erkenntnisse können dazu beitragen, die Herausforderung anders zu definieren.
3. **Entwickeln:** Geben Sie verschiedene Antworten auf das klar definierte Problem, lassen Sie sich von anderen inspirieren und gestalten Sie gemeinsam mit verschiedenen Personen.

⁴ https://www.designcouncil.org.uk/fileadmin/uploads/dc/Documents/ElevenLessons_Design_Council%2520%25282%2529.pdf

⁵ <https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond>

⁶ <https://positivepsychology.com/broaden-build-theory/>

4. **Umsetzen:** Dazu gehört es, verschiedene Lösungen in kleinem Maßstab zu testen, diejenigen zu verwerfen, die nicht funktionieren, und diejenigen zu verbessern, die funktionieren.

Wie das Design Thinking zwingt auch die Double-Diamond-Methode die Teams dazu, gemeinsam herauszufinden, was die Benutzer wirklich brauchen. In vielen Fällen stellt sich heraus, dass die Besserwisser, die Systeme ausschließlich auf der Grundlage ihrer eigenen Autorität, ihres Verstandes und ihrer Erfahrung entwerfen wollen, viele Details nicht kennen, über die sich normale Nutzer Gedanken machen.



Übungen

- Stellen Sie einem GenAI-Bot eine konkrete Frage zur Lösung der Probleme Ihres Unternehmens, z. B. „Wie kann KI dazu beitragen, die Kundenabwanderung bei einem Telekommunikationsanbieter zu verringern?“ oder „Glauben Sie, dass KI Menschen bei Designprozessmodellen ersetzen kann?“ Inwieweit glauben Sie, dass die Empfehlung von GenAI die dringendsten Probleme Ihres Unternehmens lösen würde?
- In vielen Rahmenwerken für Entwurfsprozesse ist auch die Rolle des Kritikers vorgesehen. Diese Person hat die Aufgabe, Fehler und Probleme aufzudecken. Glauben Sie, dass es wichtig ist, Menschen in Workshops einzubeziehen, die diese Rolle übernehmen können? Ist es gut, sie frühzeitig in den kreativen Prozess einzubinden?
- Schlagen Sie die „Disney-Methode“ und die „Six Thinking Hats“ nach. Glauben Sie, dass es für Workshops hilfreich sein kann, die Rollen der Teilnehmer zu wechseln und Übungen zu integrieren, bei denen die Teilnehmer in den verschiedenen Workshop-Phasen bewusst ihren Blickwinkel wechseln?

2.4.3 Durchführung von Workshops

2.4.3.1 Von der Strategie ...

Nehmen wir an, ein Unternehmen entscheidet sich für eine nicht-evidenzbasierte Strategie und ermutigt seine Mitarbeiter, nach ihrem Bauchgefühl zu handeln. Vielleicht ist ein solches „emotionsgesteuertes Unternehmen“ ein lustiger Arbeitsplatz. Jeder kann sich wie ein Held fühlen, der die fantastischsten Produkte entwickelt. Emotionsgesteuerte Unternehmen könnten mit Slogans wie diesem aufwarten:

- Wir sind ein führendes KI-Unternehmen, das die Welt verändern wird.
- Wir nutzen Daten für eine bessere Zukunft.
- Wir sind die Daten-Rockstars, die jedes Projekt rocken.

Wenn Teams dann aufgefordert werden, einem Kunden eine Value Proposition zu unterbreiten, wird der Mangel an evidenzbasiertem Denken deutlich, denn viele Slo-

gans dienen nur der Angeberei und zeigen nicht, wie ein Unternehmen ein bestimmtes Kundenproblem lösen will. Aussagen wie „Wir sind das führende KI-Unternehmen“ können alles Mögliche bedeuten. Ein Ausweg könnte sein, Technologien oder den Heldenstatus bestimmter Mitarbeiter als universelle Lösung für alle Probleme anzupreisen. Doch in der Regel haben diese Ansätze keinen langen Atem. Je tiefer ein Unternehmen in einen Lösungsraum eindringt, desto mehr muss es begründen, warum Alternative A und nicht Alternative B in Szenario C und warum es in Szenario D genau andersherum sein könnte. Um diese spezifischen Fragen beantworten zu können, müssen die Mitarbeiter aufhören, sich damit zu brüsten, wie toll sie sind, und anfangen, die richtigen Fragen zu stellen und sie mit bewährten strukturierten Methoden zur Lösung komplexer Probleme zu validieren.

Erfolgreiche Vorstände formulieren jedoch häufig SMART-Ziele (Specific, Measurable, Achievable, Relevant, Time-bound) für ihre Unternehmen, ähnlich wie John F. Kennedys Ziel, zum Mond zu fliegen.⁷

Hier sind einige Beispiele:

- Wir wollen im Jahr 2028 eine Million Produkte pro Jahr verkaufen.
- Aufgrund des demografischen Wandels bereiten wir unser Unternehmen darauf vor, bis 2030 mit 20 % weniger Mitarbeitern arbeiten zu können.
- Wir wollen bis 2035 CO₂-neutral sein.

In diesem Kapitel wird argumentiert, dass evidenzbasiertes Denken die Grundlage einer jeden AI-Strategie ist und Teil der Kultur eines Unternehmens sein kann. Die Qualität der strategischen Ziele kann einen Hinweis darauf geben, wie reif ein Unternehmen in Bezug auf evidenzbasiertes Arbeiten ist. Nehmen wir an, ein Unternehmen entscheidet sich für das erste Ziel in dieser Liste. Dies gibt den Abteilungen Raum, sich an den strategischen Zielen zu orientieren. Die Vertriebsabteilungen müssen sich vielleicht neue Märkte ausdenken, um mehr Produkte zu verkaufen. Die Produktdesigner müssen sich überlegen, wie sie ihre Produkte populärer machen können, und so weiter. Wenn das Unternehmen sein Ziel verfehlt, kann es auch untersuchen, was anders gemacht werden muss, um ein ähnliches Ziel zu erreichen, vielleicht später.

Fast alle sind sich einig, dass Ziele ohne einen Plan zu ihrer Umsetzung nur Wunschen sind. Daher werden in zielorientierten Unternehmen die strategischen Ziele in taktische Ziele für bestimmte Geschäftsbereiche aufgeteilt. Die Manager erstellen dann individuelle Ziele für ihre Teammitglieder, um ihnen zu helfen, zu den umfassenderen strategischen Zielen beizutragen.

Automatisierungsinitiativen können auf Abteilungen abzielen, in denen ein Mehrwert für die Kunden geschaffen wird. Durch die Standardisierung der Prozesse verbessert sich auch die Gesamtqualität. Die Mitarbeiter werden entlastet und können

⁷ <https://www.rice.edu/jfk-speech>

sich stärker auf die Innovation künftiger Produkte konzentrieren. Die Personalabteilung könnte sich spezifische Ziele setzen, um ein integrativer Arbeitgeber zu sein und die Vielfalt zu erhöhen, da dies alle Ziele unterstützt, als modernes Unternehmen wahrgenommen zu werden. Zu den Verkaufszielen könnte die Durchdringung neuer Märkte gehören, und die Produktionsabteilungen könnten die Produktqualität verbessern wollen.

Das entscheidende Element in der KI-Ära ist, dass KI oder Data Science für das Erreichen fast aller Unternehmensziele entscheidend sein kann. Das macht KI für Unternehmen mächtig und schafft Marktrisiken, wenn sie ignoriert wird. Dies gilt auch für nicht so offensichtliche Szenarien wie z. B. HR-Ziele. Mithilfe von Daten erfährt jede Personalabteilung, wie ein Unternehmen auf dem Markt wahrgenommen wird, und kann mögliche Problempunkte besser erkennen. In manchen Ländern ist es allerdings verboten, die Stimmung der Mitarbeiter zu messen. Bedenken Sie jedoch den Wert für Unternehmen, die Wert auf Integration legen, wenn Algorithmen Hassreden oder Diskriminierung im Verhalten der Mitarbeiter erkennen können.

Die Frage ist nicht, ob KI und Daten helfen können, sondern ob das Unternehmen bereit ist, KI und Daten zu nutzen. Wie bereits erwähnt, erfordert die Umstellung auf eine stärkere Daten- und KI-Orientierung manchmal, dass die Mitarbeiter ihre Komfortzone verlassen. Einige Unternehmen verknüpfen daher Ziele mit Bonuszahlungen. Geld kann ein solider Anreiz sein, um Mitarbeiter zu ermutigen, unbekanntes Terrain zu erkunden, was einigen von ihnen zunächst unangenehm erscheinen mag. Es braucht jedoch mehr als nur Bonuszahlungen; starke Führungsqualitäten sind ebenfalls erforderlich, um sicherzustellen, dass jeder zu den gemeinsamen Zielen beiträgt.

Je klarer die strategischen Ziele eines Unternehmens definiert sind und je besser das Team auf die Erreichung dieser Ziele ausgerichtet ist, desto erfolgreicher werden KI- und Data-Science-Workshops sein. Dies unterstreicht erneut, wie wichtig die Unterstützung durch das Top-Management ist und dass es besser sein kann, eine umfassende Strategie zu entwickeln, bevor man mit Workshops für KI- und Data-Science-Anwendungsfälle beginnt.



SWOT-Analysen

Eine Möglichkeit, die Strategie des Unternehmens zu untersuchen, ist die regelmäßige Durchführung einer sogenannten „SWOT-Analyse“, bei der die eigenen Stärken und Schwächen sowie die externen Chancen und Risiken untersucht werden. Eine ausführliche Erklärung, wie man eine detaillierte SWOT-Analyse durchführt, würde den Rahmen eines Buches über Data Science und KI sprengen; es genügt zu sagen, dass die Ergebnisse einer solchen Analyse für alle Datenprojekte von Nutzen sein können.

Sofern diese Ergebnisse unvoreingenommen sind (die SWOT-Analyse wurde z. B. von externen Beratern durchgeführt), können sie dazu beitragen, Veränderungsprozesse zu erleichtern. Stellen Sie sich zum Beispiel vor, dass Datenprojekte in der Vergangenheit durch IT-Prozesse gebremst wurden und die SWOT-Analyse zu dem Schluss kommt, dass die IT-Prozesse des Unternehmens starrer sind als die anderer Unternehmen desselben Sektors. Eine solche Feststellung gibt dem Datenteam mehr Möglichkeiten, Hindernisse in den IT-Prozessen zu beseitigen, um die Projektgeschwindigkeit zu erhöhen.

Beachten Sie, dass SWOT nicht der einzige Analyserahmen ist, der auf diese Weise nützlich sein kann. Es gibt viele andere wie z. B. die Bewertung der Datenreife, die sich speziell auf Stärken und Schwächen im Umgang mit Daten konzentriert.

Man beachte auch, dass einige der Meinung sind, SWOT sollte durch TOWS ersetzt werden sollte, was eine bessere Reihenfolge der Ausführung darstellt: Unternehmen sollten zuerst die Bedrohungen und Chancen auf dem Markt untersuchen, um ihre eigenen Stärken und Schwächen am besten kontextualisieren zu können. Andernfalls kann es zu „abstrakten Nabelschaudiskussionen“ kommen, wie es Michael D. Watkins ausdrückt, ein wichtiger Verfechter von TOWS.⁸⁾

2.4.3.2 ... zur Ausführung

Unternehmen haben oft funktionsübergreifende Teams, um strategische Ziele zu erreichen. Letztendlich sind die Data Scientists zwar einem Data Science Manager unterstellt, werden aber in der Regel auf Projekte verteilt, bei denen sie mit Mitgliedern anderer Geschäftsbereiche zusammenarbeiten. Angenommen, ein Ziel besteht darin, den Produktionsprozess zu automatisieren, um den Bedarf an 20 % der Mitarbeiter zu senken, dann wird ein funktionsübergreifendes Team aus Fabrikarbeitern, Data Scientists und anderen Beteiligten gebildet.

Teams treffen sich in der Regel in Workshops, um Probleme mithilfe von Entwurfsprozessen zu lösen, die auf Modellen basieren, wie sie in diesem Kapitel vorgestellt wurden. Eine Zutat für den Erfolg ist, dass alle Interessengruppen in den Workshops vertreten sind. Wenn ein Unternehmen den Schadenprozess in einem Data-Science-Prozess für Versicherungen optimieren möchte, müssen in diesem Workshop auch Kundenvertreter einbezogen werden, die in ihrem Tagesgeschäft Fälle bearbeiten. Wenn ein Unternehmen die Arbeit in einer Fabrik automatisieren möchte, um die Zahl der Mitarbeiter zu verringern, benötigen die Data Scientists ein detailliertes Verständnis der Fabrikprozesse im Allgemeinen sowie der für diese Fabrik spezifischen Prozesse. Nachdem sie sich dieses Wissen angeeignet haben, können sie eine erste Hypothese darüber aufstellen, wie die Fabrik mit weniger Mitarbeitern dieselbe oder sogar eine höhere Leistung erbringen könnte. Auch hier müssen Data Scientists mit

⁸ <https://hbr.org/2007/03/from-swot-to-tows-answering-a-readers-strategy-question>

Fachleuten zusammenarbeiten. Die Beiträge derjenigen, die täglich mit den Prozessen arbeiten, sind von unschätzbarem Wert; ohne sie beruhen die Schlussfolgerungen des Workshops lediglich auf Spekulationen. Aus diesem Grund ist es wichtig, den häufigen Fehler zu vermeiden, dass Fachexperten von einem Manager vertreten werden, der keine detaillierte Erfahrung mit dem Fachgebiet hat.

Je nach Erfahrung und Reifegrad der Daten gehen die Unternehmen unterschiedlich an die Umsetzung heran. Zu Beginn ihrer Datenreise könnten die Unternehmen damit beginnen, Ergebnisse messbar zu machen. Betrachten wir das Ziel der CO₂-Reduzierung: Bis zu welchem Detailgrad weiß ein Unternehmen bereits, wie viel CO₂ es produziert? Beruhen die Berichte auf analytischen Modellen, die den Ausstoß anhand der Größe der Branche schätzen? Gibt es weitere Möglichkeiten, die tatsächliche Produktion zu messen? Könnten verfeinerte Messverfahren eingesetzt werden, um deutlicher zu zeigen, welche genauen Prozesse in der Wertschöpfungskette CO₂ erzeugen? Viele Unternehmen hoffen bei diesen Workshops auf tiefhängende Früchte, und manchmal ist das auch möglich. Vor allem im Tagesgeschäft kann es einfach zu implementierende Änderungen geben, die sofortige Vorteile bringen. Dies gilt vor allem dank moderner Cloud-Infrastrukturen, die es Unternehmen ermöglichen, schnell Datenplattformen zur Untersuchung von Daten einzurichten. Aber in den meisten Fällen braucht es Zeit, um mithilfe von Data Science einen Mehrwert zu schaffen. So kann es beispielsweise einige Zeit dauern, bis alle erforderlichen Daten gesammelt sind, um mit den Untersuchungen zu beginnen. Die Dauer, die für den ersten sichtbaren Erfolg benötigt wird, kann für Unternehmen mit einem geringen Budget für Data Science entmutigend sein.

Manche Lösungen erfordern mehr Technik, manche weniger. Es gibt keine eierlegende Wollmilchsau, die alle Probleme löst. Einige Ansätze können auch zusätzliche Software und technischen Aufwand erfordern. Um Veränderungen in einer Fabrik zu erforschen, um beispielsweise die Zahl der benötigten Mitarbeiter oder die Menge des ausgestoßenen CO₂ zu verringern, können Digital Twins Data Scientists eine Grundlage für Experimente bieten. Die Digitalisierung einer ganzen Fabrik, um Änderungen zu simulieren, könnte jedoch eine Weile dauern, wenn ein Unternehmen noch am Anfang seiner Transformation zum datengetriebenen Unternehmen steht. Im Gegensatz dazu möchte zum Beispiel ein Versicherungsunternehmen erforschen, wie generative KI dazu beitragen kann, den Bedarf an Menschen bei der Bearbeitung von Versicherungsansprüchen zu verringern. In diesem Fall können die Data Scientists frühzeitig mit Experimenten beginnen.

Zusammenfassend lässt sich sagen, dass die Ansätze umso unterschiedlicher werden, je mehr wir uns mit den Details der Data-Science-Anwendungsfälle beschäftigen. Die Teams werden auf spezifische Probleme stoßen. Es kann eine beträchtliche Zeit dauern, bis man Zugang zu den benötigten Daten erhält. Bestehende Softwareanwendungen in einem Unternehmen müssen möglicherweise angepasst oder aktualisiert werden, um die Daten in einem Format bereitzustellen, das die Data Scientists benötigen.

In anderen Fällen müssen die Data Owners überzeugt werden, die Daten freizugeben. Gelegentlich verlangt die Abteilung für Informationssicherheit viele Details zu den vorgeschlagenen Data-Science-Nutzungsfällen, was die Geschwindigkeit der Erkundung verlangsamt. Um alle Beteiligten zur Zusammenarbeit zu bewegen, sind manchmal Überzeugungsarbeit und Verhandlungsgeschick erforderlich, insbesondere wenn die „Datenkultur“ noch nicht etabliert ist.

Unternehmen, die auf ihrem Weg weiter fortgeschritten sind, werden KI-Anwendungen schneller integrieren. Es wird für sie selbstverständlicher werden, maschinelles Lernen einzusetzen, um Ausreißer zu erkennen und darauf zu reagieren. Sie können verschiedene Algorithmen zur Vorhersage des Kundenverhaltens und zur Verbesserung der Vertriebsprozesse einsetzen. Dennoch vertrauen alle ausgereiften Datenunternehmen auf Fakten, die aus Daten abgeleitet werden, denn ohne evidenzbasierte Ansätze und kritisches Denken würde die Erforschung von Daten letztlich keinen Sinn machen.



Übungen

- Was bedeutet „Management Talk“, einschließlich Phrasen wie „in unserer Branche führend im Bereich KI sein“ oder „der beliebteste Arbeitgeber in der Region sein“? Stellen Sie sich vor, Ihre Aufgabe ist es, Data Science in einem Unternehmen einzuführen, und Sie werden mit vagen oder nicht vorhandenen Zielen konfrontiert: Wie würden Sie damit umgehen?
- Schlagen Sie „Data Maturity“ nach oder fragen Sie einen GenAI-Bot zu diesem Thema. Versuchen Sie, sich über verschiedene Reifegradmodelle zu informieren, die den Reifegrad einer Organisation definieren. Wo sehen Sie Ihre Organisation nach einem solchen Modell?
- Einige Mitarbeiter sehen vielleicht nicht, wie sie von einer gut definierten Datenstrategie profitieren können. Wie würden Sie sicherstellen, dass diese Mitarbeiter dennoch an den Workshops zur Strategieentwicklung teilnehmen?
- Wie würden Sie die Kultur in Ihrem Unternehmen beschreiben? Werden Entscheidungen von oben nach unten getroffen, vom Management bis zur Basis, oder gibt es in Ihrem Unternehmen flache Hierarchien? Fließen die Ideen für Datenprojekte tatsächlich von unten nach oben, angetrieben von daten- und technikbegeisterten Mitarbeitern?
- Wie würden Sie während eines Workshops damit umgehen, wenn einige Teilnehmer nicht so aufgeschlossen sind wie die anderen?
- Suchen Sie nach Fachleuten, die seit vielen Jahren in einem Ihnen unbekanntem Bereich arbeiten, und stellen Sie ihnen detaillierte Fragen zu ihrer Arbeit.
- Halten Sie es für sinnvoll, jemanden in ein Workshop-Team aufzunehmen, der nur moderiert und völlig ergebnisoffen ist? Was könnten die Vor- und Nachteile sein?

- Was würden Sie Unternehmen mit einem geringen Budget für Data Science empfehlen? Ist es gut, mit einem niedrigen Budget weiterzumachen? Wäre es besser, ganz auf Investitionen in Daten zu verzichten, da die Arbeit mit einem niedrigen Budget wahrscheinlich überhaupt keine Ergebnisse bringen wird? Oder sollten sie risikofreudiger werden und versuchen, Investitionsgelder zu beschaffen, um ihre Datenumstellung zu beschleunigen?
- Wenn Sie für eine Organisation arbeiten, kennen Sie Ihre Unternehmensziele? Erfüllen sie die SMART-Kriterien?
- Stimmen Sie den hervorgehobenen Aussagen im Text zu: „Das entscheidende Element in der KI-Ära ist, dass KI oder Data Science entscheidend sein kann, um die Ziele fast jeder Organisation zu erreichen“ und „Die Frage ist nicht, ob KI und Daten helfen können, sondern ob das Unternehmen bereit ist, KI und Daten zu nutzen“?
- Zu Beginn des Kapitels haben wir erwähnt, dass Skeptiker die Umstellung auf Daten als kostspielig und mit dem Risiko, dass viele Anwendungsfälle scheitern, betrachten könnten. Selbst diejenigen, die dafür plädieren, dass ein Unternehmen datengesteuert werden soll, können dies nicht leugnen. Gleichzeitig wird in diesem Kapitel angedeutet, dass Unternehmen, die KI ignorieren, wahrscheinlich aufhören werden zu existieren. Was ist Ihre Meinung dazu? Kann man den „KI-Skeptikern“ antworten, dass das Ignorieren von KI das größere Risiko darstellt?

2.4.3.3 Taktische Details

Es ist auch sinnvoll, sich mit Problemen zu befassen, die während der Strategieentwicklungsworkshops oder bei der Umsetzung einer Strategie auftreten können.

Einige Unternehmen wissen nicht, wo sie anfangen sollen

Externe Berater sind vielleicht gut darin, ihren Kunden zu sagen, was ähnliche Unternehmen in der Vergangenheit getan haben, um erfolgreich zu sein. Die Kunden könnten entgegen, dass das, was bei einem Unternehmen funktioniert, nicht unbedingt auch bei einem anderen funktionieren muss. Einige Unternehmen könnten in eine Sackgasse geraten, weil sie nicht wissen, wie sie anfangen sollen.

Eine Möglichkeit, eine solche Sackgasse zu durchbrechen, ist eine SWOT-Analyse. Die Mitarbeiter können Ideen dazu entwickeln, was ein Unternehmen gut kann. Wenn ein Unternehmen in der Vergangenheit nicht erfolgreich gewesen wäre, würde es nicht mehr existieren. Was war es also, das das Unternehmen erfolgreich gemacht hat? Worin sind sich alle Kunden oder Mitarbeiter einig, dass dieses Unternehmen besser ist als andere?

Bei der Untersuchung der Stärken werden häufig auch Schwächen festgestellt, die das Unternehmen möglicherweise verbessern muss. Mit diesen Schwächen im Hinterkopf ist es möglich, über Chancen und Gefahren nachzudenken. Wenn ein Unterneh-

men Dinge besser macht als die Konkurrenz und die Kunden bereit sind, für seine Dienstleistungen oder Produkte zu bezahlen, hat es ausgezeichnete Chancen, im Geschäft zu bleiben.

Verschiedene Menschen können sich auf unterschiedliche Weise ausdrücken, so als ob sie verschiedene Sprachen sprechen würden.

Ziel jeder Data-Science-Initiative ist es, dass Mitarbeiter aus verschiedenen Bereichen zusammenarbeiten, um etwas zu erreichen. Dies führt zu Situationen, in denen Mitarbeiter in der Regel in völlig unterschiedlichen Abteilungen arbeiten und für einen Workshop gemeinsam in einem Besprechungsraum sitzen.

Stellen Sie sich ein Unternehmen vor, das versucht, die Produktionsprozesse in seinen Fabriken zu automatisieren. In sehr komplexen Sätzen könnten promovierte Data-Science-Nerds Hypothesen darüber aufstellen, wie ihre Datenprojekte die Verfahren des Unternehmens für immer verändern könnten. Sie könnten genau beschreiben, wie sie Daten extrahieren möchten und welche Algorithmen am besten geeignet sind, um die Chancen auf ein akzeptables Konfidenzintervall zu maximieren. Währenddessen wird von den Fabrikarbeitern, die neben ihnen sitzen, erwartet, dass sie lediglich Fragen darüber beantworten, wie sie die Maschinen benutzen, mit denen sie arbeiten. Und sobald sie gefragt werden, beginnen sie mit einer Reihe von technischen Beschreibungen, die nur ihre Kollegen in der Fabrik verstehen können.

Solche Situationen lassen sich nur durch exzellente Führung lösen, da diejenigen, die solche Workshops leiten, wissen müssen, wie man die Veranstaltung so organisiert, dass die Arbeiter am Ende genug Zeit finden, ihre Arbeitsrealität zu erklären, und die Data Scientists überlegen können, wie sie die Situation verbessern können.

Workshops können emotionale Achterbahnfahrten mit vielen Höhen und Tiefen sein

Manchmal werden Workshops von extremen Positionen dominiert, von „wir verändern die Welt mit Daten“ bis „egal was wir tun, es wird in einer Katastrophe enden“. Extreme Positionen sind üblich, und Konflikte zwischen extremen Positionen sind häufig. Der eingefleischte Pessimismus der einen Seite könnte diejenigen frustrieren, die darauf bedacht sind, die Innovationszyklen des Unternehmens endlich zu beschleunigen, um ein führendes datengesteuertes Unternehmen in ihrem Bereich zu werden.

Eine Möglichkeit, dieses Problem zu lösen, besteht darin, die Perspektiven beider Seiten zu verstehen. Versetzen Sie sich in die Lage eines Betriebsleiters, dessen Aufgabe es ist, ein Unternehmen am Laufen zu halten, und der Anrufe von verärgerten Benutzern und Managern erhält, wenn Anwendungen ausfallen. Im schlimmsten Fall müssen sie Wochenend- oder Nachtschichten einlegen, um die Dinge wieder zum Laufen zu bringen. Für sie kann zu viel Veränderung als Bedrohung empfunden werden.

Parallel dazu können Sie sich vorstellen, dass Sie ein Unternehmer sind, der mit der Innovation des Unternehmens beauftragt ist. Vielleicht sind Sie sich einiger strategischer Bewertungen Ihres Unternehmens bewusst und wissen, dass Ihr Unternehmen Gefahr läuft, wegen der langsamen Durchführung von Innovationen hinter die Konkurrenz zurückzufallen. Für Sie riskiert jeder, der die Innovation verlangsamt, den langfristigen Erfolg des Unternehmens.

Die einzige Möglichkeit, Pattsituationen zwischen extremen Positionen aufzubrechen, besteht darin, dass man versucht, die Position der anderen Seite zu verstehen. Ohne diejenigen, die auf Innovation drängen, gäbe es keinen Fortschritt, denn schließlich gibt es in der Geschichte viele Beispiele, in denen visionäre Denker die Welt gegen alle Kritik verändert haben. Manchmal ist visionäres Denken jedoch wie der Bau von Luftschlossern, und Pessimisten können uns dabei helfen, Ideen zu erkennen, die Bestand haben können, und jene davon zu unterscheiden, deren Verfolgung eine Verschwendung von Zeit und Geld wäre.

Viele Ansätze schlagen vor, verschiedene flexible Standpunkte einzunehmen.^{9),10),11)} Die Idee ist, jeden Standpunkt als eine flexible Rolle zu sehen. Als Pessimist ist es die Aufgabe einer Person, Gründe zu finden, warum eine Idee scheitern könnte. Derselben Person kann jedoch eine neue Rolle zugewiesen werden. Nachdem sie den Advokaten des Teufels gespielt hat, kann sie die Rolle wechseln und versuchen, Argumente zu finden, warum die Lösung erfolgreich sein wird.

Es mag einige Zeit dauern, sich an den Gedanken zu gewöhnen, dass ein Standpunkt nicht feststeht und durch einen Rollentausch verändert werden kann, aber es kann helfen, extreme Positionen aufzubrechen.

2.5 Kurz und bündig



Bias

Bias (Voreingenommenheit) ist ein natürliches Phänomen, und es ist menschlich, voreingenommen zu sein. Es ist jedoch möglich, den eigenen Bias zu überwinden. Wenn wir voreingenommen sind, können wir nicht objektiv sein. Um bei Datenprojekten erfolgreich zu sein, braucht man eine Kultur, die sich auf Fakten konzentriert. Um sich auf Fakten zu konzentrieren, braucht man kritisches Denken. Kritisches Denken erfordert ein Bewusstsein für Voreingenommenheit.

⁹ <https://positivepsychology.com/broaden-build-theory/>

¹⁰ https://en.wikipedia.org/wiki/Disney_method

¹¹ https://en.wikipedia.org/wiki/Six_Thinking_Hats

SWOT- und Datenreifegrad-Analysen

Eine Analyse der Datenreife eines Unternehmens und seiner Stärken, Schwächen, Chancen und Bedrohungen kann helfen, die eigene Strategie zu definieren. Dies wiederum führt zu Data-Science-Projekten, die sich mit den wichtigsten Herausforderungen des Unternehmens befassen.

Datenkompetenz

Bei der Datenkompetenz geht es darum, die Zahlen zu verstehen, die wir betrachten. In dem Moment, in dem wir verstehen, welche Zahlen wichtig sind, können wir mit ihnen arbeiten.

Entwurfsprozess

Es gibt viele Prozesse, die bei der Definition von Datenprodukten helfen können. Double Diamond und Design Thinking sind nur Beispiele, die wir im Detail untersucht haben.

3

Cloud-Dienste

Stefan Papp

„Vollkommenheit ist nicht erreicht, wenn es nichts mehr hinzuzufügen gibt, sondern wenn es nichts mehr wegzulassen gibt.“

Antoine de Saint-Exupéry



Fragen, die in diesem Kapitel beantwortet werden:

- Welche Systemumgebungen sind für Data-Science-Projekte erforderlich?
- Warum sind Cloud-Plattformen ideal für experimentelle Data-Science-Projekte?
- Welche Bedeutung haben GPUs und andere Hardwarekomponenten für Data-Science-Projekte?
- Wie können Sie Plattformen dynamisch aufbauen, indem Sie Infrastructure as Code verwenden, und sie mit Versionsverwaltungswerkzeugen verwalten?
- Was unterscheidet eine Microservice-Architektur von einer monolithischen Architektur?
- Wie können Linux-Systeme effizient für Data Science genutzt werden?
- Was sind die jeweiligen Merkmale, Stärken und Schwächen der verfügbaren „Everything-as-a-Service“-Modelle (XaaS)?
- Welche Cloud-Dienste können Datenexperten dabei helfen, Cloud-native Lösungen zu entwickeln?

3.1 Einführung

Dieses Buch beschreibt, wie **künstliche Intelligenz, Machine Learning** und **Deep Learning** unser Leben beeinflussen werden. Künstliche Intelligenz ist in letzter Zeit zu einem beliebten Thema in den Medien geworden, und das aus gutem Grund. Die KI-Revolution ist heute in aller Munde, und Sam Altman ist fast allen, die sich für dieses Thema interessieren, ein Begriff, ein bekannter Name geworden. Sein Unternehmen, OpenAI, wird oft als entscheidender Wegbereiter der generativen KI dargestellt. Unternehmen wie Microsoft oder Google werden als diejenigen angesehen, die die KI schließlich der Bevölkerung zugänglich machen werden.

Während die Unternehmen, die sich mit den softwarebezogenen Aspekten der KI befassen, weltweit Beachtung finden, sind die Unternehmen, die die Hardware herstellen, auf der die KI läuft, und die somit ebenfalls für viele der jüngsten Erfolge der KI verantwortlich sind, die unbesungenen Helden. Zumindest solange, bis wir den Erfolg solcher Unternehmen wie Nvidia an der Börse betrachten. Aber während Nvidia seine Millionen mit der Produktion von Grafikprozessoren verdient hat, gehen die Hardwareanforderungen für KI und Data Science weit darüber hinaus. Bei daten-gesteuerten Ansätzen untersuchen Datenexperten enorme Datenmengen, manchmal bis zu Petabytes. Um diese Daten zu hosten und zu verarbeiten, ist eine beträchtliche Menge an Hardware-Ressourcen erforderlich.

In diesem Kapitel erörtern wir, wie wir Cloud-Infrastrukturen für unsere Datenprojekte nutzen können. Wir werden uns mit **Infrastructure-as-a-Service (IaaS)** Lösungen beschäftigen, die uns mehr Freiheit bei der Installation der von uns gewünschten Anwendungen und Tools geben. Wir werden uns auch ansehen, wie wir über eine **Platform as a Service (PaaS)** vorausgewählte Pakete eines Cloud-Anbieters nutzen können.

3.2 Cloud-Essentials

In ihrer einfachsten Definition bedeuten Cloud-Dienste die Nutzung des Computers eines anderen. Anstatt Rechenzentren und Serverfarmen zu bauen, können IT-Unternehmen die Anschaffung von Hardware vermeiden und für das bezahlen, was sie nutzen.



Warum „On Premise“ schwierig ist

In Kapitel 1 haben wir ein IT-Unternehmen mit übermütigen IT-Mitarbeitern vorgestellt. Sie schlagen eine „No-Cloud-Strategie“ vor, weil sie glauben, dass sie die Datenressourcen des Unternehmens besser verwalten können als andere Cloud-Anbieter. Leider sind in vielen Fällen wie diesem die Lehren aus einer solchen

Strategie bitter. Um ein Rechenzentrum in Eigenregie zu betreiben, muss ein Unternehmen viele Experten mit unterschiedlichen Fähigkeiten beschäftigen, um einen 24/7-Service anbieten zu können:

- Systemarchitekten entwerfen die Lösung auf der Grundlage der Anforderungen.
- Operations Engineers ersetzen defekte Hardware.
- Netzwerktechniker bauen das Netzwerk auf, einschließlich Router und Verkabelung.
- Betriebssystemexperten installieren und konfigurieren Betriebssysteme.
- Facility Manager kümmern sich um Systeme wie Klimaanlage und Brandschutz.
- Das Sicherheitspersonal sichert den Zugang zum Rechenzentrum gegen unbefugtes Eindringen.

Data-Science-Projekte können eine Menge Hardware-Ressourcen benötigen, und Fehler bei der Dimensionierung – egal ob CPU, GPU, RAM, Festplatten oder Netzwerk – können teuer werden, da sie zu Engpässen führen können.

Darüber hinaus erfordern einige Data-Science-Projekte teure GPUs. Sobald diese gekauft sind, kann die nächste Generation besserer GPUs verfügbar sein. Da eine Cloud ständig mit der neuesten Hardware aufgerüstet wird, gerät ein Nutzer von Cloud-Diensten nie ins Hintertreffen, wohingegen es nicht garantiert ist, dass On-Premises-Installationen regelmäßig aufgerüstet werden.

Betriebssysteme und ihre Treiber werden regelmäßig aktualisiert. Linux-Betriebssysteme beispielsweise sind mit moderneren Kernen – der Kernfunktionalität jedes Betriebssystems – und Treibern oft für moderne Hardware optimiert und bieten schnellere Durchsätze bei Datenübertragungen und eine bessere Leistung durch spezielle Algorithmen. Die Hersteller von Betriebssystemen versuchen, ihre Software so lange wie möglich abwärtskompatibel zu halten. Das bedeutet, dass Software, die auf älteren Versionen eines Betriebssystems lief, auch auf neueren Versionen ausgeführt werden kann. Das Ziel der Softwarehersteller ist es, so viele Kunden wie möglich zu bedienen. Wenn die Programmierer die Software mit den modernsten Funktionen der neuesten Kernel und Treiber optimiert haben, kann es sein, dass viele Kunden diese Software nicht ausführen können, weil die Richtlinien die Installation der neuesten Versionen eines Betriebssystems verbieten. Dies ist wahrscheinlich, da die IT-Politik vieler Unternehmen eher konservativ ist und die Unternehmen ihre Server erst spät auf neuere Versionen aktualisieren, da sie die Anzahl der parallel gepflegten Betriebssystemversionen begrenzen wollen.

Cloud-Anbieter können die neuesten Treiber und Betriebssystem-Kernelversionen für ihre nativen Cloud-Dienste verwenden. Sie sind nicht aus Kompatibilitätsgründen an alte Betriebssystemkomponenten gebunden. Dadurch können sie die neuesten Funktionen nutzen und den Nutzern ein insgesamt besseres Erlebnis bieten.

Viele Experten sehen die Cloud nicht nur als eine neue Technologie, sondern vielmehr als ein alternatives Geschäftsmodell für die Beschaffung und Verwaltung von Hardware selbst.¹⁾ Es muss nicht unbedingt ein bestimmter Dienst sein wie eine Cloud-basierte Datenbank oder ein Cloud-basierter Dateispeicher, der neue Kunden anzieht. Ein Cloud-Anbieter gewinnt wahrscheinlich einen neuen Kunden, wenn ein Unternehmen beginnt, an den Vorteil des Mietens von IT-Ressourcen gegenüber deren Besitz zu glauben.

Statistiken zeigen, dass die meisten Unternehmen die Cloud nutzen oder irgendwann nutzen werden.²⁾ Auch wenn es für einige Unternehmen immer noch besser sein mag, betriebliche Kernsysteme in lokalen Umgebungen abzuschirmen, bevorzugen die meisten Unternehmen bereits die Cloud für analytische Workloads. Darüber hinaus ist die Cloud auch ein perfektes Experimentierlabor für Data Scientists, da sie die benötigten Ressourcen dynamisch erzeugen und stilllegen können.

Jeder Cloud-Anbieter definiert den Begriff „Cloud“ etwas anders.^{3),4),5)} Einige Wertversprechen richten sich auch an unterschiedliche Zielgruppen. Buchhalter beispielsweise hören gerne, dass sie ihre Fixkosten senken können, da sie nicht viel Geld für den Kauf einer neuen Plattform ausgeben wollen. Ein Pay-as-you-go-Modell macht die Sache für sie einfacher. Technische Teams werden den Vorteil sehen, dass sie neue Lösungen schnell einsetzen und bei Bedarf skalieren können.

Bevor wir uns näher mit den funktionalen Diensten eines Cloud-Anbieters befassen, ist es wichtig, die Anzahl der Rechenzentren der Cloud-Anbieter weltweit als einen entscheidenden Faktor für die Leistungsfähigkeit eines Cloud-Anbieters hervorzuheben. Mit den Big 3 – Amazon, Google und Microsoft – können Kunden ihre Dienste global skalieren.^{6),7),8)} Vor allem für kleinere Cloud-Anbieter ist es schwer, mit ihnen zu konkurrieren, da sie nicht über die Ressourcen verfügen, um weltweit Rechenzentren aufzubauen.

Dieses Wertversprechen ist allgemein gehalten und kann auch auf andere Cloud-Anbieter angewendet werden, die möglicherweise ihre eigenen Varianten davon haben.

¹ <https://medium.com/@storjproject/there-is-no-cloud-it-s-just-someone-else-s-computer-6ecc37cdcfe5>

² <https://www.cloudzero.com/blog/cloud-computing-statistics/>

³ <https://docs.aws.amazon.com/whitepapers/latest/aws-overview/six-advantages-of-cloud-computing.html>

⁴ https://microsoft.firstdistribution.com/wp-content/uploads/2021/08/Microsoft-Azure-Value-Proposition_2-1.pdf

⁵ <https://cloud.google.com/why-google-cloud>

⁶ https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

⁷ <https://azure.microsoft.com/en-us/explore/global-infrastructure/geographies/#geographies>

⁸ <https://cloud.google.com/about/locations>

3.2.1 XaaS

Das Modell „**Everything-as-a-Service**“ (**XaaS**) skizziert verschiedene Szenarien und definiert, wer – der Kunde oder der Cloud-Anbieter – für was verantwortlich ist. Die vier bekanntesten XaaS-Modelle sind:

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)
- Function-as-a-Service (FaaS)

IDENTIFY WHAT YOU ARE ALREADY DOING IN THE CLOUD



Gartner.

Bild 3.1 Hosting- und Cloud-Modelle⁹⁾

Traditionelle On-Premises-IT bedeutet, dass Sie alles selbst machen, einschließlich des Baus Ihres Rechenzentrums mit Serverräumen, Kühlung, Racks und unterbrechungsfreier Stromversorgung (USV). Dazu gehören auch Themen im Zusammenhang mit dem Gebäudemanagement, z. B. Feueralarm und Zugangskontrolle zu den Serverräumen.

Colocation ermöglicht es den Kunden, das Facility Management auszulagern. Die Kunden mieten einen Serverraum, bauen ihre Hardware auf und sind für alle Themen im Zusammenhang mit der Computerhardware verantwortlich. Sie müssen

⁹⁾ Copyright: Gartner, Inc. (<https://www.zdnet.com/article/xaas-why-everything-is-now-a-service/>)

nach wie vor spezialisierte Fachleute für alles von Netzwerken bis zu Betriebssystemen beschäftigen, und die Mitarbeiter benötigen weiterhin Zugang zu den Einrichtungen, um beschädigte Hardware zu ersetzen.

Hosting ist der erste Schritt, wenn ein Kunde keine Hardware mehr warten muss. Allerdings muss der Kunde immer noch Software auf der nackten Hardware installieren, und seine Mitarbeiter müssen mit dem Hardware-Layout des Rechenzentrums vertraut sein.

Infrastructure as a Service bedeutet die Anmietung von virtualisierten Hardware-Ressourcen mit einem installierten Basisbetriebssystem. Die Mitarbeiter werden einen Serverraum der bereitgestellten IaaS-Dienste nie von innen sehen und müssen sich nicht darum kümmern, wie ein Rechenzentrum organisiert ist. Dennoch müssen sie das Betriebssystem verwalten, einschließlich der Systemaktualisierungen.

Platform as a Service entlastet einen Kunden von der Verwaltung der zugrunde liegenden Software. Die Kunden verwenden vorkonfigurierte Software, die sie entweder über die grafische Benutzeroberfläche oder über die Befehlszeile konfigurieren. Denken Sie an Datenbanken oder Dateisysteme, bei denen der Anbieter die darunter liegende Software verwaltet.

Software as a Service bedeutet einfach die Nutzung von Software. Während ein Kunde bei PaaS noch an Speicherdienste wie Amazon S3 oder Azure Blob Storage denkt, nutzt er bei SaaS Dienste wie Dropbox. Es besteht keine Notwendigkeit, Details zu konfigurieren, die darüber hinausgehen, wie viel Speicherplatz der Kunde zu zahlen bereit ist.

Function as a Service spiegelt den Trend zum serverlosen Computing wider. Ein Client definiert kleine Teile des Quellcodes, und Ereignisse können diese Funktionalitäten auslösen. Stellen Sie sich einen kleinen Dienst vor, der einmal täglich den Aktienmarkt abfragt und dem Client eine E-Mail schickt, wenn eine Aktie einen bestimmten Wert erreicht.

3.2.2 Cloud-Anbieter

Der Gartner-Quadrant vom Oktober 2023 spiegelt den Zustand des Marktes für Cloud-Anbieter wider, wie er von diesem weltweit angesehenen Beratungsunternehmen gesehen wird. Diese Analyse der weltweit anerkannten Strategie unterstreicht die Dominanz der Big 3, die sich auch in dieser Nasdaq-Studie widerspiegelt.¹⁰

¹⁰ <https://www.nasdaq.com/articles/the-big-3-in-cloud-computing>



Bild 3.2 Magischer Quadrant für Cloud-Infrastruktur als Service (weltweit), 2023¹¹⁾

Herausforderer wie Alibaba Cloud oder IBM könnten in den nächsten Jahren aufholen. Geopolitische Spannungen schaffen auch spezifische Märkte für Cloud-Anbieter in bestimmten Ländern. So könnten Cloud-Anbieter mit Sitz in China in einigen Ländern mit größeren Herausforderungen konfrontiert sein, während sie in anderen Ländern weniger Schwierigkeiten haben als ihre US-amerikanischen Konkurrenten.¹²⁾ Jeder Cloud-Anbieter bietet ähnliche Dienste an, die sich zwar im Detail unterscheiden, aber dennoch einen ähnlichen Anwendungsfall lösen. Tabelle 3.1 stellt eine minimale Teilmenge dieses schnell wachsenden Ökosystems dar.

¹¹ Copyright: Gartner, Inc. (<https://aws.amazon.com/blogs/aws/read-the-2023-gartner-magic-quadrant-for-strategic-cloud-platform-services/>)

¹² https://www.andrew.cmu.edu/user/hongs/files/Alibaba_ICS2022.pdf

Tabelle 3.1 Cloud-Anbieter und eine ihrer Referenzlösungen für verschiedene Cloud-Anforderungen

Dienst	AWS	Azure	Google Cloud
Lagerung	Amazon S3	Azure Blob-Speicher	Google Cloud Storage (GCS)
Berechnen Sie	Amazon EC2	Azure-VM	Google Compute Engine (GCE)
Entwicklung	AWS Code Commit	Azure DevOps	Google App Engine (GAE)
Veraltetes Kubernetes	Amazon Elastic Kubernetes Service (EKS)	Azure Kubernetes-Dienst (AKS)	Google Kubernetes Engine (GKE)
Datenbank (NoSQL)	Amazon DynamoDB	Azure Cosmos DB	Firebase
Data Warehouse	Amazon Redshift	Azure SQL	BigQuery
Serverless Computing	AWS Lambda	Azure-Funktionen	Cloud-Functions (GCF)

Cloud-Anbieter können sich in ihren spezialisierten, neuen Diensten unterscheiden. Eine Faustregel besagt jedoch, dass jeder seit Jahren etablierte Standarddienst, der den Kunden zur Verfügung gestellt wird, genügend Funktionen bietet, um die Anforderungen der normalen Cloud-Kunden zu erfüllen.

3.2.3 Native Cloud-Dienste

Speicherdienste

Jeder Cloud-Anbieter bietet einen File-Hosting-Dienst an, um seinen Kunden unbegrenzten Datenspeicher und ergänzende Dienste für den Zugriff auf diesen Speicher über verschiedene Kanäle zur Verfügung zu stellen. Diese Dienste werden oft als „Objektspeicher“ bezeichnet, um sie von Dateisystemen zu unterscheiden. Objektspeicherdienste bieten häufig zusätzliche Funktionen wie z. B. das Hosten einer Webseite zusätzlich zu diesem Dateispeicher. Außerdem sind Objekte in einem Objekt unveränderlich; Sie können nichts in einem Objektspeicher bearbeiten, sondern nur überschreiben. Dateisystemdienste hingegen haben oft höhere Anforderungen an den Datendurchsatz beim Lesen oder Schreiben von Daten als Datei-Hosting-Dienste. Das Hauptkriterium für beide ist es, die Datenverfügbarkeit hoch und die Kosten niedrig zu halten.

Cloud-Anbieter mischen Speicherlösungen mit anderen Bausteinen. Bei Datenbanklösungen kann der Cloud-Anbieter beispielsweise seine Speicherlösung zur physischen Speicherung der Daten verwenden, während die entsprechende Datenbanklösung auch eine Datenbank-Engine enthält.

Ein Qualitätsindikator für Cloud-Speicherdienste ist die Haltbarkeit, die angibt, wie gesund und widerstandsfähig Ihre Daten sind, wenn sie in der Cloud gespeichert sind. Die großen 3 versprechen eine Haltbarkeit von 11 Neunen oder mehr. Dies entspricht einem durchschnittlich erwarteten jährlichen Verlust von 0,000000001 % der Objekte. Wenn Sie beispielsweise 10 000 000 Objekte speichern, können Sie im Durchschnitt alle 10 000 Jahre mit dem Verlust eines einzigen Objekts rechnen.

Compute

Der häufigste Anwendungsfall ist die Instanziierung virtueller Maschinen (VMs), auf denen Sie nach Belieben Software installieren können.

Stellen Sie sich einen Benutzer vor, der eine spezielle Software für die Datenanalyse von einem Anbieter einsetzen möchte. Er sucht nach einer Hardwarekonfiguration, die seinen Anforderungen entspricht. Möglicherweise findet er spezielle Vorlagen für Data Science, z. B. GPU-optimierte Instanzen. Der Benutzer kann auch ein Betriebssystem für die VM auswählen.

Der größte Unterschied zwischen Ihrem lokalen Rechenzentrum, das ein mittelgroßes Unternehmen betreiben kann, und den großen Cloud-Anbietern ist die verfügbare Konfiguration. Die großen Cloud-Anbieter bieten den Nutzern die neuesten und leistungsfähigsten GPUs in Serverfarmen, die sich viele kleinere Wettbewerber nicht leisten können. Wenn die großen Cloud-Anbieter neue Rechenzentren in einem neuen Land bauen, können sie auf die Erfahrung dessen zurückgreifen, was in vielen Ländern bereits seit Jahren funktioniert. Sie optimieren jedes winzige Detail, wie die einzelnen Komponenten miteinander kommunizieren und wie die Server über das Netzwerk miteinander kommunizieren.

Datenbanken

Unternehmen können Open-Source- und fast jede proprietäre Datenbank in der Cloud hosten. Jeder der drei großen Cloud-Anbieter bietet seine Datenbankdienste auch in der Cloud an. Das Produkt von Microsoft ist Azure SQL, eine Cloud-basierte Version von **MS SQL Server**. **CosmosDB** ist eine ergänzende NoSQL-Datenbank. Amazon hat mehrere eigene Systeme, wobei **DynamoDB** und **Redshift** die bekanntesten Vertreter sind. DynamoDB punktet mit NoSQL-Fähigkeiten, während Redshift ein Data Warehouse ist. Die Lösungen von Google sind **BigTable** und **BigQuery**. BigTable ist die Technologie, aus der auch ein berühmter Vertreter einer Datenbank für Hadoop hervorgegangen ist: **HBase**.

Die Datenbanksysteme der Cloud unterstützen das **polyglotte Persistenz-Paradigma**.¹³⁾ Spezifische Anforderungen erfordern spezifische Lösungen. Infolgedessen sind einzelne Speicherlösungen darauf ausgelegt, einen bestimmten Anwendungsfall zu lösen.

¹³ <https://martinfowler.com/bliki/PolyglotPersistence.html>



Polyglotte Lagerung und Skifahren

Der polyglotte Speicher kann mit dem Skisport verglichen werden, der mehrere Disziplinen umfasst. Auf dem Höhepunkt seiner Karriere dominierte Hermann Meier, ein pensionierter Skirennfahrer, den Weltcup in Abfahrt, Riesenslalom und Super-G.

Man hätte ihn auch bitten können, Slalom – eine vierte Disziplin – zu fahren. Als erfahrener Skirennfahrer hätte er ein Rennen beendet, aber es ist zweifelhaft, dass er eine Leistung hätte erreichen können, die mit der eines Skirennfahrers vergleichbar ist, der sich auf Slalom konzentriert.

Ein ähnliches Prinzip gilt auch für Datenbanken. Einige Anwendungsfälle erfordern NoSQL-Datenbanken, andere wiederum funktionieren besser mit relationalen Datenbanken. Zu wissen, welche Datenbank zu verwenden ist, kann nützlich sein.

Das Herzstück einer Datenbank ist ihre Architektur. Vor-Ort-Lösungen werben oft mit einer **Shared-Nothing-Architektur**, um die vorhandene Infrastruktur optimal zu nutzen. Auf der anderen Seite bauen viele Cloud-Anbieter ihre Cloud-nativen Datenbanklösungen auf einer **Multi-Cluster-Architektur mit gemeinsamen Daten auf**. Das Herzstück der zweiten Architektur ist eine Schicht aus superschnellen SSD-basierten Festplatten, die Daten mehrfach replizieren. Die Knoten, auf denen Abfrageoptimierer einen Ausführungsplan erstellen, befinden sich schließlich in einer separaten Schicht.

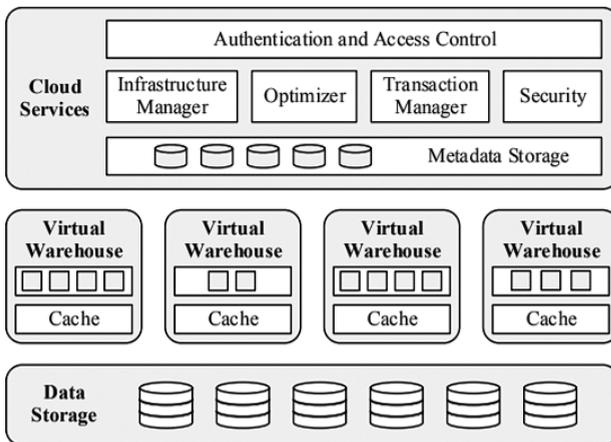


Bild 3.3

Architektur mit mehreren Clustern und gemeinsamen Daten (<https://dl.acm.org/doi/pdf/10.1145/2882903.2903741>)

Entwicklung

Google bietet viele Dienste rund um **Kubernetes**, den De-facto-Standard für die Container-Orchestrierung. Da Kubernetes sehr populär geworden ist, kann die Nutzung der Cloud-Plattform des Entwicklers für einige Unternehmen ein Entscheidungskriterium sein.

Microsoft hat eine lange Tradition in der Entwicklung von Softwareentwicklungswerkzeugen und bietet eine gute Integration von Cloud-Diensten in seine Produkte für Entwickler wie z. B. **Visual Studio Code**.

Directories

Microsoft verwendet **Active Directory**, was dazu beiträgt, Kunden zu gewinnen, die bereits an Windows-basierte Umgebungen in ihren lokalen Einrichtungen gewöhnt sind. Dies kann ein entscheidendes Entscheidungskriterium sein, wenn ein Kunde bereits ein Microsoft-Kunde ist.

AWS Identity and Access Management (IAM) ist ein Verzeichnisdienst, der Benutzer und Ressourcen für Organisationen verwaltet, die Amazons Cloud-Services abonnieren. Sie können IAM verwenden, um Benutzer zu erstellen und zu gruppieren und ihren Zugriff auf praktisch jeden AWS-Service zu kontrollieren, wie z. B. **EC2** oder **S3**.



Übungen

- Lassen Sie sich zertifizieren. In Vorbereitung auf eine Cloud-Anbieterzertifizierung lernen Sie alles, was Sie für die professionelle Arbeit mit der Cloud brauchen. Es gibt auch spezielle Zertifizierungen für maschinelles Lernen und andere analytikbezogene Fähigkeiten.
- Studieren Sie Richtlinien, Sicherheitsgruppen und Firewall-Einstellungen im Detail. Sie werden es sich danken, wenn Sie unter Zeitdruck stehen und ein Zugriffsproblem lösen müssen, während Sie kurz vor einer Demo stehen.
- Experimentieren Sie mit einem Kostenrechner, um ein Gefühl für die zu erwartenden Kosten einer Plattform in der Cloud zu bekommen.¹⁴⁾ Würden Sie einem Unternehmen empfehlen, seine On-Premises-Architektur per Lift and Shift in die Cloud zu verlagern?
- Erkunden Sie verschiedene Cloud-Anbieter und finden Sie heraus, wo sie ihre Rechenzentren haben. Welche Auswirkungen hat es Ihrer Meinung nach, wenn ein Cloud-Anbieter der erste ist, der ein Rechenzentrum in einem bestimmten Land baut?

3.2.4 Cloud-native Paradigmen

Zwei wichtige Architekturbauusteine für Cloud-native Lösungen sind **Microservices** und **serverlose Architekturen**.

¹⁴ <https://calculator.aws/#/>

Microservices

Microservices kapseln Funktionalität, um eine einzige Aufgabe zu erfüllen. Einige Entwickler kennen diese Idee bereits aus dem **Single Responsibility Principle**. In einem Datenprojekt könnte eine solche Funktion darin bestehen, Daten aus dem Data Warehouse zu extrahieren und die Informationen als Datei in der Cloud zu speichern. Um Microservice-Architekturen zu erstellen, müssen Entwickler atomare Funktionen in einem Container verpacken, der die Funktionen kapselt, und ein Mini-Betriebssystem zur Ausführung dieser Funktionen.

Kubernetes ist ein Orchestrierungstool für Container. Es steuert die automatische Instanziierung von Containern und den Ersatz von defekten Containern. Die Kubernetes-Läufe sind auf mehrere Knoten verteilt und verwenden zwei Kategorien von Diensten. Die **Steuerungsebene** verwaltet Prozesse und Interaktionen zwischen Prozessen. Neue Anfragen werden beispielsweise über eine Rest-API empfangen, und die Kontrollebene koordiniert die entsprechenden Aktionen. **Kubernetes-Worker** – die zweite Kategorie – führen dann die Funktionalität aus. Ein Pod ist eine Einheit, die einen oder mehrere Container beherbergt.

Das Tolle an Kubernetes ist, dass niemand die Software auf den Host-Servern aktualisieren muss, sobald die Techniker eine neue Microservice-Version bereitgestellt haben. Kubernetes fungiert auch als sichere Sandbox für die Ausführung, falls ein Container bösartigen Code einkapselt.

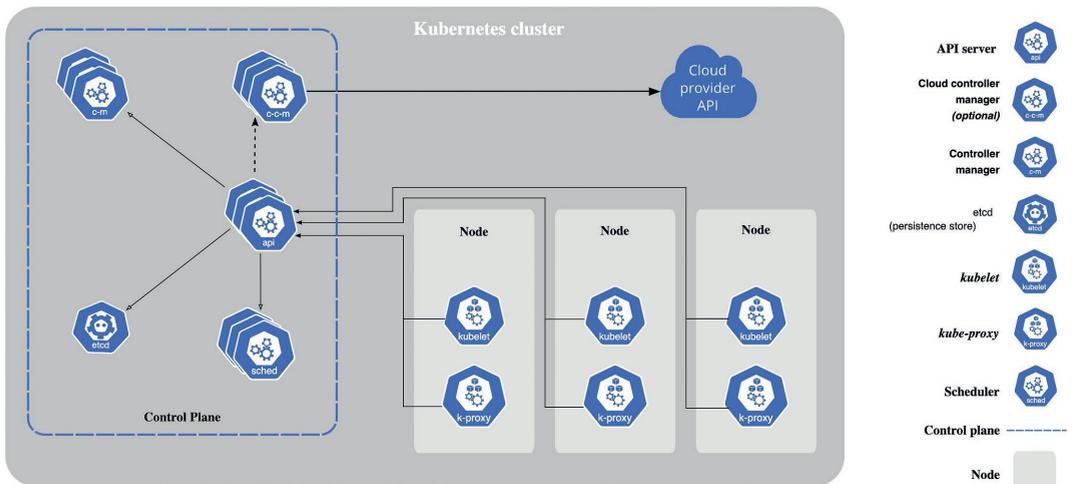


Bild 3.4 Kubernetes in Aktion¹⁵⁾

¹⁵⁾ Copyright: Kubernetes, <https://kubernetes.io/docs/concepts/overview/components/>

Serverless Computing

Beim **serverless Computing** bietet der Cloud-Anbieter einen Dienst an, mit dem Nutzer Funktionen ausführen können, ohne sich um die Verwaltung von Servern kümmern zu müssen. Die Kunden zahlen nur für die Rechenzeit; Ausfallzeiten fallen nicht an.

Serverloses Computing ist für Anwendungsfälle konzipiert, die **Elastizität** erfordern, in denen die Anzahl der Ereignisse, die eine Verarbeitung auslösen, also stark schwankt. Ob massiv oder winzig, die Anfragen müssen rechtzeitig bedient werden. Stellen Sie sich einen Dienst vor, der am Ende eines jeden Monats eine monatliche Statistik erstellt. Ein ständig laufender physischer Server, der diese Aufgabe nur einmal im Monat ausführt, würde Ressourcen verschwenden.

Alle drei wichtigen Anbieter haben eigene serverlose Systeme: **AWS Lambda**, **Azure Functions** und **Google Cloud Functions**.



Übungen

- Verwenden Sie Infrastructure as Code, um einen Kubernetes-Cluster zu erstellen.
- Versuchen Sie Alpine Linux anstelle von Ubuntu Linux als Betriebssystem für Ihre Container.
- Erfahren Sie, wie Pods in Kubernetes miteinander kommunizieren.
- Verstehen, was ein Data Mesh ist und wie man es konfiguriert.
- Legen Sie sensible Daten in Hashicorp Vault ab und lernen Sie, wie man einen Secrets Manager mit Kubernetes verwendet.
- Vertiefen Sie Ihr Wissen über Kubernetes und lassen Sie sich bei Bedarf zertifizieren.
- Vertiefen Sie sich in Lambda und andere serverlose Anwendungen. Es gibt auch ein Framework für serverlose Architekturen namens „serverless“.¹⁶⁾

3.3 Infrastructure as a Service

Das Einrichten einer virtuellen Maschine in der Cloud ist ein Referenzbeispiel für Infrastruktur als Service. Als Benutzer wählen Sie die Hardware für einen Server aus und installieren Anwendungen. Eine ideale Einrichtung muss die folgenden Anforderungen erfüllen:

- Robustes Design und Redundanz müssen die geringste Wahrscheinlichkeit eines Datenverlusts aufweisen (**Haltbarkeit**).

¹⁶ <https://www.serverless.com/>

- Die Sicherheitsmechanismen müssen die Daten vor unberechtigtem Zugriff schützen (**physische Sicherheit**).
- Die Datenplattform muss die geltenden Datenschutzrichtlinien einhalten (**Datenschutz**).
- Die Datenplattform muss die Ergebnisse in einer angemessenen Zeit liefern (**Leistung**).
- Die Nutzer müssen jederzeit auf die Plattform und ihre Daten zugreifen können (**Verfügbarkeit**).

Wir müssen kurz Hardware, Linux und Infrastructure as Code erklären, um mit Infrastructure as a Service arbeiten zu können.

3.3.1 Hardware

Um Aufträge in der Cloud auszuführen, müssen Datenexperten die Hardwarekonfiguration der Dienste festlegen, auf denen sie ihre Aufträge ausführen. Auch wenn Cloud-Abonnenten die Hardware nicht kaufen und Cloud-Dienste jederzeit kündigen können, können Konfigurationsfehler dennoch kostspielig sein.

In vielen Fällen muss das Team, während es den Entscheidungsträgern potenzielle Datennutzungsfälle vorstellt, auch die OPEX-Kosten für die Aufrechterhaltung der operativen Dienste in der Endphase des Anwendungsfalls abschätzen. Wenn die Hardware-Anforderungen falsch sind, kann sich das auf diese Schätzung auswirken.



Reservierte und Spot-Instanzen

Die Kosten können durch reservierte und Spot-Instanzen optimiert werden. Kurz gesagt, mit einer reservierten Instanz verpflichten Sie sich zur Nutzung der Rechenleistung und kaufen die Kapazität im Voraus zu einem niedrigeren Preis. Wenn Sie also wissen, dass Sie eine kontinuierliche Datenverarbeitungslast haben werden, können Sie die Kosten senken.

Spot-Instanzen beziehen sich auf spezielle Angebote, bei denen Sie Rechenleistung zu einem Spotpreis erhalten. Sobald der Preis jedoch steigt und der Spotpreis nicht mehr eingehalten wird, verlieren Sie den Zugang zu den Berechnungen.

IaaS und PaaS hängen von der Hardwarekonfiguration ab. Lassen Sie uns untersuchen, was wir in der Cloud konfigurieren können.

Prozessoren (CPU/GPU)

Da Geräte Berechnungen mit Daten durchführen, werden Prozessoren häufig in ihre gängigsten Formen unterteilt: **Zentraleinheiten** (CPUs) und **Grafikverarbeitungseinheiten** (GPUs).

Viele Ingenieure vergleichen CPUs mit einem menschlichen Gehirn, das viele **verschiedene Operationen durchführt**. Der Schlüsselindikator für die Stärke eines Prozessors ist die **Million Instructions per Second (MIPS)**, die angibt, wie viele Aufgaben ein Prozessor pro Sekunde parallel bearbeiten kann. Von 0,002 MIPS für die ersten Prozessoren im Jahr 1951 werden Spitzenwerte von 2 356 230 MIPS für moderne CPUs angegeben.¹⁷⁾ Eine schwache CPU-Konfiguration kann sich beispielsweise auf den Datentransfer zwischen Knoten auswirken, wenn diese CPU-gebunden sind.

Das ursprüngliche Ziel der GPUs war es, die CPU von der hohen Rechenlast zu befreien, die für 3D-Animationen erforderlich ist. GPUs sind für die Ausführung vieler Matrixoperationen optimiert, was sie perfekt für neuronale Netze macht. GPU-Prozessoren von Herstellern wie **Nvidia** oder **Habana Labs** (einer Intel-Tochter), sind auf zahlreichen Cloud-basierten Data-Science-Plattformen installiert.

Netzwerk

Die Nutzer von Cloud-Systemen müssen sich zwar keine Gedanken über die Verkabelung machen, aber sie müssen dennoch **virtuelle Netzwerke** in der Cloud konfigurieren, um den Zugriff auf Netzwerksegmente zu kontrollieren. Ein zentraler Aspekt ist die Verschlüsselung und Nutzung privater Kanäle über **VPC**.

Ein Aspekt, der auch Datenexperten betrifft, sind Netzwerkkonfigurationen für Firewalls. In vielen Fällen werden bei der Bereitstellung neuer Plattformen Ports geschlossen, und Datenexperten müssen ein Mindestmaß an Wissen darüber haben, wie sie dies angehen können.

Speicher

Typische Big-Data-Cluster haben 64 GB bis 512 GB **RAM** pro Knoten, je nach Anwendungsfall. Mit diesen Größen können enorme Datenmengen in den Speicher geladen werden, und die Speichergröße pro Knoten ist auch ein entscheidender Faktor dafür, wie viele Nutzer parallel mit dem Cluster arbeiten können.

E/A

Festplatten galten lange Zeit als der „Flaschenhals“ in Computersystemen.¹⁸⁾ Wir stießen schnell an mechanische Grenzen, und jahrelang war die E/A-Innovation langsamer als bei anderen Hardwarekomponenten. Mit dem Aufkommen von SSDs boten sich Computersystemen neue Möglichkeiten zur Leistungssteigerung. Cloud-Anbieter bieten zum Beispiel unter¹⁹⁾ **bereitgestellte IOPS** Konfigurationen mit bis zu 256 000 IOPS an. Sie sind deutlich schneller als die Standardkonfiguration (bis zu 16 000 IOPS)

¹⁷⁾ https://handwiki.org/wiki/Instructions_per_second

¹⁸⁾ https://en.wikipedia.org/wiki/I/O_bound

¹⁹⁾ <https://aws.amazon.com/ebs/volume-types/>

und daher ausdrücklich für Datenbanken mit geringen Anforderungen an die Antwortzeit geeignet.



Übungen

- Fragen Sie ChatGPT, welche Grafikprozessoren für Data-Science-Anwendungen verfügbar sind und wie sie sich auf die Modellerstellung auswirken können. Sie können Details zu bestimmten Produkten erfahren und fragen, wie sie sich unterscheiden.
- Stellen Sie sich vor, Sie bereiten Daten für die analytische Modellierung vor. Sie schreiben Spark-Jobs, die einen Cluster verwenden. Die Leistung ist langsamer als erwartet. Ihr Chef fragt Sie nach Möglichkeiten, die Leistung zu steigern. Wo fangen Sie an?
- Wählen Sie eine Netzwerkzertifizierung Ihres bevorzugten Cloud-Anbieters oder lesen Sie die für die Zertifizierung empfohlenen Bücher, um alles über Protokollstapel zu erfahren.
- Lesen Sie das PolarFS-Papier²⁰ und erfahren Sie, wie sich PolarFS von verteilten Dateisystemen wie dem **Hadoop Distributed File System** (HDFS) unterscheidet.

3.3.2 Verteilte Systeme

Zwei Prinzipien, die ursprünglich aus der agilen Softwareentwicklung stammen, sind für den Aufbau von Data-Science-Plattformen unerlässlich:

- **KISS** (Keep it simple, stupid) empfiehlt, Systeme so einfach wie möglich zu halten. Je weniger komplex eine Einheit ist, desto einfacher ist es, ein System zu entwickeln, das sie skaliert.
- **YAGNI** (You ain't gonna need it) empfiehlt, Systeme einfach zu halten, indem ausdrücklich nicht gebaut wird, was nicht zur Ausführung der Kernfunktion einer Einheit benötigt wird.

In Kapitel 1 haben wir Cesar vorgestellt, einen ehrgeizigen Datenexperten, der eine Menge Technologie vorschlägt, um etwas Modernes aufzubauen. Wenn eine Technologie einen Zweck nicht erfüllt, verstößt sie gegen beide Prinzipien.

Nehmen wir an, wir haben unsere Dienste aufgebaut, die die erforderlichen Funktionen auf einfachste Weise bereitstellen. Plötzlich steigt die Zahl der Anfragen, und die erste Einrichtung kann sie nicht mehr bewältigen. Was können wir tun?

- **Scale-up** oder **vertikale Skalierung** verbessert die Leistung des Systems, indem eine Hardwarekonfiguration durch eine leistungsfähigere ersetzt wird. Doch

²⁰ <http://www.vldb.org/pvldb/vol11/p1849-cao.pdf>

selbst die beste Hardware wird ein Problem nicht lösen, das über eine bestimmte Last hinausgeht.

- **Scale-out** oder **horizontale Skalierung** bedeutet, dass die Last auf mehr Hardware verteilt wird. Experten sprechen auch von verteilter Verarbeitung oder Parallelisierung.

Schauen wir uns an, wie dies in einem praktischen Beispiel aussehen kann.

3.3.2.1 Von-Neumann-Architektur

Eine **Von-Neumann-Architektur**²¹ ist die Grundlage für einen modernen Computer, wie in Bild 3.5 dargestellt, und besteht aus einem System, das eine CPU, Speicher und eine E/A-Komponente umfasst, die über einen Bus miteinander verbunden sind. Ein Cluster von Computern (gemeinhin als **Knoten** bezeichnet) in einem Serverraum besteht daher in der Regel aus mehreren, über ein Netzwerk verbundenen Von-Neumann-Systemen.

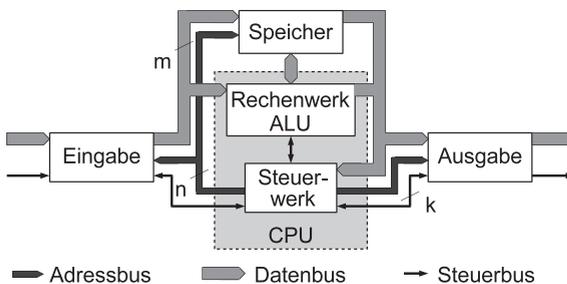


Bild 3.5

Von-Neumann-Architektur²²

Stellen Sie sich vor, wir haben einen Knoten, der einen Auftrag ausführt. Mit zunehmender Arbeitslast beschweren sich die Benutzer über lange Wartezeiten. Eine Skalierung würde bedeuten, diesen Knoten durch einen leistungsfähigeren Knoten zu ersetzen. Aber was wäre, wenn wir uns entschließen würden, zwei Knoten anstelle eines Knotens mit mehr Hardwareressourcen zur Skalierung zu verwenden?

Es gibt nun zwei Szenarien.

- **Lastausgleich:** In diesem Szenario gibt es eine zweite Komponente, einen Load Balancer. Der Load Balancer verteilt die Anfragen, anstatt sie zu verarbeiten. Wir können bei Bedarf weitere Knoten hinzufügen. In diesem Szenario sind alle Arbeitsknoten völlig unabhängig. Wenn ein Knoten ausfällt, werden die Aufträge von den anderen Knoten weiterverarbeitet.
- **Verteiltes Rechnen:** Wir erwarten von der Software, dass sie die Algorithmen zur Verarbeitung auf mehrere Knotenpunkte verteilt.

²¹ <https://www.computerscience.gcse.guru/theory/von-neumann-architecture>

²² Copyright: Kapooh (https://en.wikipedia.org/wiki/File:Von_Neumann_Architecture.svg)

Das letztere Szenario ist komplexer. Sie benötigen Funktionen, die die Datenverarbeitung auf mehreren Rechnern koordinieren. Dazu gehört auch die Synchronisierung von Zwischenergebnissen und Zuständen zwischen diesen Verarbeitungskomponenten. Die Software muss auch auf Störungen reagieren, z. B. auf einen Knotenausfall oder eine verzögerte Synchronisierung. Schauen wir uns an, wie dies erreicht werden kann.

3.3.2.2 MapReduce

Bei der lokalen Verarbeitung von Daten auf einem einzelnen Computer lädt ein Prozess die Daten in den Speicher. Dann wird ein Algorithmus auf die geladenen Daten angewandt, um sie zu verändern oder ihre Ergebnisse zu berechnen. Diese Ergebnisse können auf dem Bildschirm angezeigt oder als Datei gespeichert werden. Wenn die zu ladende Datenmenge größer ist als der verfügbare Arbeitsspeicher, kann es zu Verzögerungen kommen, da die Datenverarbeitungsmaschine während der Verarbeitung Zwischenergebnisse auf die lokale Festplatte auslagern muss.

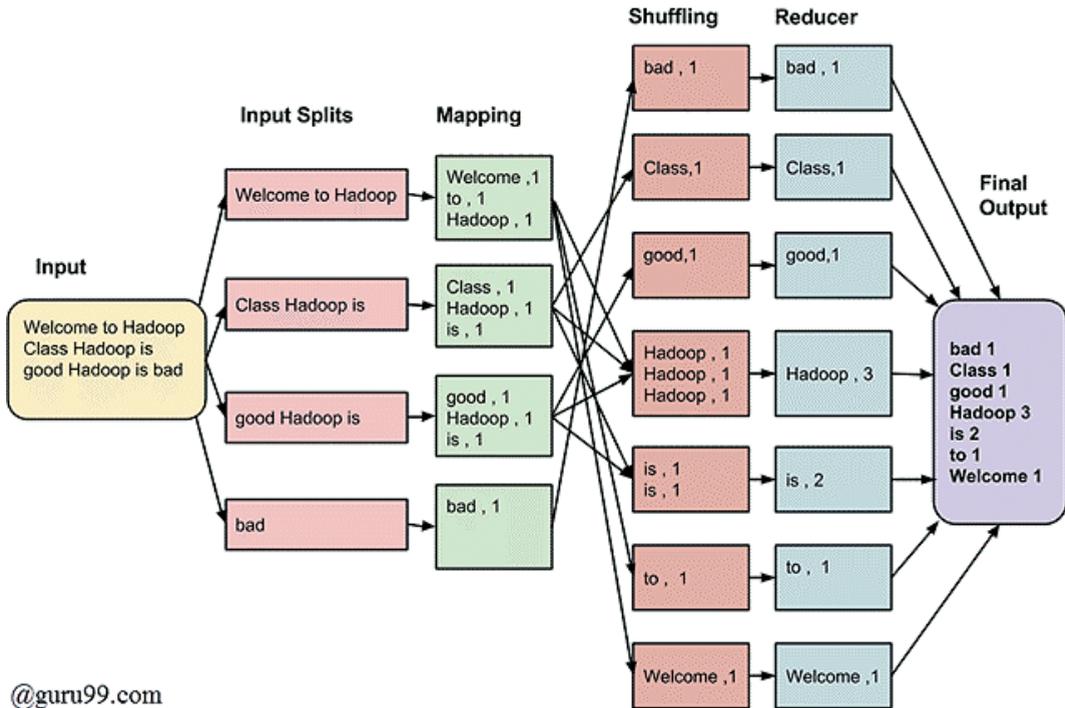
Bild 3.6 zeigt den zusätzlichen Overhead anhand einer vereinfachten Darstellung der verteilten Verarbeitung. Wir müssen uns vorstellen, dass wir alle Daten auf verschiedenen Knoten in einem verteilten System speichern. Dieser Mechanismus bedeutet, dass jeder Knoten so viel wie möglich verarbeiten muss, bis ein Rahmenwerk reduzierte Daten auf einigen wenigen Knoten sammeln kann. In einer lokalen Umgebung sind all diese Vorverarbeitungen zunächst auf separaten Knoten und die spätere Zusammenführung unnötig.

Die Algorithmen werden zunächst auf den Daten ausgeführt, die auf den sogenannten „Datenknoten“ gespeichert sind. Dann laden Prozesse, die auf den Datenknoten laufen, analog zur lokalen Verarbeitung die Daten in den Arbeitsspeicher, wenden die Algorithmen an und speichern die Ergebnisse zwischen. Schließlich werden die Zwischenergebnisse auf den einzelnen Knoten in einem zweiten Schritt zentral zusammengeführt. Die Details dieses Prozesses, genannt **MapReduce**, werden auf der Seite Hadoop beschrieben, einem Rahmenwerk, das MapReduce verwendet.²³⁾

In praktischen Anwendungen abstrahieren Schichten diese Art der Verarbeitung und bieten eine einheitliche API. Ein solches Framework ist Apache Spark, das in Kapitel 4.4.2 näher erläutert wird.

Jeder Ingenieur sollte sich bewusst sein, dass die lokale Verarbeitung einen geringeren Overhead hat. Es ist immer schneller, wenn die Daten klein genug sind, um lokal verarbeitet zu werden.

²³ <https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>



@guru99.com

Bild 3.6 Verarbeitung mit MapReduce²⁴⁾



Übungen

- Lesen Sie mehr über Paradigmen wie Moore's Law oder Amdahl's Law. Ein Standardwerk zu diesem Thema ist das Buch *Designing Data-Intensive Applications* von Martin Kleppmann.²⁵⁾
- Recherchieren Sie im Internet und versuchen Sie zu verstehen, was Begriffe wie „PolarDB“, „Paxos“, „Round Robin“, „CAP Theorem“, „BASE“ und „Eventual Consistency“ bedeuten.
- Lernen Sie, den Unterschied zwischen Shared-Nothing-Architekturen und Multi-Cluster-Shared-Data-Architekturen zu erklären.
- Lernen Sie, wie Indizes in Datenbanken funktionieren und wie verschiedene Plattformen den Zugriff durch Baumdatenstrukturen optimieren. Versuchen Sie auch, Partitionierungsstrategien zu verstehen und wie man mit schiefen Daten arbeitet.

²⁴⁾ <https://www.guru99.com/introduction-to-mapreduce.html>

²⁵⁾ <https://www.amazon.com/Designing-Data-Intensive-Applications-Reliable-Maintainable/dp/1449373321>

3.3.3 Linux Essentials für Datenexperten

Obwohl viele Dienste im Zusammenhang mit der Datenanalyse über Cloud-basierte PaaS- und SaaS-Lösungen in Anspruch genommen werden, gibt es immer noch viele Anwendungsfälle, in denen Datenexperten Linux-basierte VMs verwenden müssen. Sie können in der Cloud gehostet werden oder in manchen Fällen müssen Daten aus lokalen Lösungen extrahiert werden, und der Datenexperte erhält lediglich eine Verbindung zur Konsole.

Auch wenn der mit der Einrichtung von Systemdatenplattformen beauftragte DevOps- oder Operations-Ingenieur oft nicht mit dem Datenteam identisch ist, hilft es, die Sprache des Teams zu sprechen, um die gewünschte Einrichtung durchzuführen. Außerdem gehen die Data Engineers gelegentlich zur Konsole, um Verbindungen mit Standard-Linux-Tools wie *netstat* zu testen.



Linux in der Unternehmenspraxis

Die Unternehmensrichtlinien für Betriebssysteme schreiben oft eine konservative Betriebsstrategie vor. Viele Linux-Server, auf denen Unternehmensanwendungen laufen, verwenden einen älteren Linux-Kernel. Das Wissen über Ubuntu, CentOS/RedHat oder SuSE scheint ausreichend.

Viele Plattformen verwenden Technologien, die alte Betriebssystemversionen unterstützen sollen, sodass die verwendete Software auch alte Kernel verwendet. Daher können exotischere Distributionen, die neuere Linux-Kernel verwenden, die Leistung steigern. Ein Beispiel für ein solches System ist Unum.²⁶⁾ Die Entwickler dieser Software verwenden die neuesten Treiber der neuesten Linux-Kernel-Version, um Leistungsbenchmarks gegen ihre Konkurrenten zu gewinnen.

Einige argumentieren, dass mit der Zeit immer weniger Menschen mit der Kommandozeile arbeiten werden und mehr Prozesse durch Tools abstrahiert werden. Generative KI könnte jedoch die Kommandozeile wiederbeleben. ChatGPT gibt genaue Antworten darauf, welche Linux-Befehle verwendet werden können, um das zu erreichen, was in der Eingabeaufforderung gefragt wurde. Es ist eine größere Herausforderung zu zeigen, wie man es in einer grafischen Benutzeroberfläche macht.

Terminal-Umgebung

Wer intensiv mit Betriebssystemen arbeitet, muss mit der Kommandozeile vertraut sein. Daher ist es hilfreich, eine beispielhafte Terminalumgebung zu wählen. Wer mit einem Mac arbeitet, kann zum Beispiel das Programm **iTerm2** verwenden, das mehr Funktionen bietet als der Standard-Terminal emulator. Windows-Benutzer werden oft **PuTTY** verwenden.

²⁶ <https://www.unum.cloud/>

Docker ist wiederum der Standard für Container und kann als gekapselte Umgebung innerhalb eines Betriebssystems verstanden werden. Nach der Installation wird mit dem folgenden Befehl ein Container auf der Grundlage des `devenv`-Images instanziiert:

```
docker run -it stefanpapp/devenv /bin/zsh
```

Programmierer, die mit Docker nicht vertraut sind, können sich die Beziehung zwischen Images und Containern wie folgt merken: Ein Image ist wie eine Klasse, und ein Container ist wie ein Objekt. Deshalb verwenden wir gerne den Begriff „Instanziierung“; was er bedeutet, ist für jeden Programmierer sofort ersichtlich.



Bild 3.7 Ein Docker-Container aus dem `devenv`-Image

Sobald Sie den Befehl ausgeführt haben, haben Sie eine virtualisierte Umgebung auf diesem Betriebssystem eingerichtet, die ein Mini-Linux kapselt. Windows-Benutzer können auch einige Linux-Befehle unten über diesen Docker-Container ausprobieren.

Shell

Wenn Sie mit der Konsole arbeiten, müssen Sie sich viele Befehle merken. Eine einfache Lösung ist die Suche nach zuvor verwendeten Befehlen in der Befehlszeile mit **Strg+R**. Wenn Sie zum Beispiel den Befehl `docker run` ausgeführt haben, können Sie mit dem Tastaturkürzel **Strg+R** eine Suche aktivieren. Wenn Sie `run` eingeben, werden Sie zum letzten Befehl mit `run` geführt. Sie können die Tastenkombination **Strg+R** auch wiederholen, um zu früheren Verwendungen von `run` zu springen. Sie können diese Suche über `hstr`²⁷⁾ erweitern, sodass Sie Ihren Befehlsverlauf interaktiv durchsuchen können.

Es lohnt sich, einen Blick auf die verfügbaren Tastaturbefehle zu werfen, um zu vermeiden, dass der Cursor ständig hin und her bewegt wird. Beachten Sie, dass einige Shells wie `zsh`²⁸⁾, ein Bash-Ersatz im Detail variieren können.

Paketmanager

Sie sollten Software mit einem Paketmanager installieren. Paketmanager rufen Repositories ab, die für das jeweilige Betriebssystem nutzbare Softwarepakete enthalten. Sie können diese Repositories über das Internet aktualisieren.

²⁷⁾ <https://github.com/dvorka/hstr>

²⁸⁾ <https://www.zsh.org/>

Für jedes Betriebssystem und die verschiedenen Linux-Distributionen gibt es verschiedene Standard-Paketmanager. Hier ist eine kleine Auswahl:

Tabelle 3.2 Paketmanager für Betriebssysteme

OS	Paketmanager
Alpine	apk (https://wiki.alpinelinux.org/wiki/Alpine_Linux_package_management)
Ubuntu	apt (https://linux.die.net/man/8/apt-get)
CentOS	yum (https://www.centos.org/docs/5/html/yum)
Mac OS X	brew (https://brew.sh)

Überwachung und Benchmarking

Datenexperten müssen häufig die Ressourcennutzung überprüfen. Es gibt zahlreiche kommerzielle Lösungen, aber dieser Abschnitt stellt einige Open-Source-Alternativen vor.

Wenn Sie die Ressourcenauslastung wissen wollen, können Sie dies mit `htop` tun. Viele Datenexperten verwenden diesen Befehl regelmäßig, um Engpässe in einer Datenpipeline zu entdecken.

```

1. htop
x htop  #1 x zsh  #2

 1  [|||||||||||||] 50.7% Tasks: 470, 2503 thr; 3 running
 2  [|||||||||] 28.9% Load average: 2.81 3.16 3.69
 3  [|||||||||] 52.3% Uptime: 3 days, 21:50:24
 4  [|||||||||] 28.3%
Mem [|||||||||||||] 10.3G/16.0G
Swp [|||||||||] 4.05G/5.00G

PID USER PRI NI VIRT RES S CPU% MEM% TIME+ Command
81665 stefanpap 17 0 5956M 197M ? 104. 1.2 4h12:31 /Applications/Google Ch
80938 stefanpap 17 0 5361M 288M ? 7.5 1.8 31:22.33 /Applications/Google Ch
81940 stefanpap 24 0 4558M 43304 W 6.5 0.3 1:00.49 /Applications/iTerm.app
13327 stefanpap 17 0 5153M 165M ? 4.5 1.0 0:10.49 /Applications/Google Ch
 538 stefanpap 17 0 4803M 235M ? 3.4 1.4 8:56.68 /Applications/Fantastic
15044 stefanpap 17 0 4380M 10480 ? 2.0 0.1 0:46.88 /Applications/Lightshot
79547 stefanpap 17 0 6411M 132M ? 1.6 0.8 16:29.16 /Applications/Skype for
 579 stefanpap 24 0 6514M 960 ? 1.0 0.0 6h59:17 com.docker.hyperkit -A
76707 stefanpap 17 0 7178M 444M ? 0.6 2.7 51:06.92 /Applications/Microsoft
82109 stefanpap 17 0 7805M 195M ? 0.4 1.2 1:42.09 /Applications/Google Ch
99186 stefanpap 17 0 5224M 142M ? 0.4 0.9 1:12.29 /Applications/Google Ch
81366 stefanpap 17 0 5372M 145M ? 0.3 0.9 1:27.42 /Applications/Google Ch
12602 stefanpap 17 0 5149M 178M ? 0.3 1.1 0:31.96 /Applications/Google Ch
 302 stefanpap 17 0 5720M 98M ? 0.3 0.6 2:37.77 /usr/sbin/distnoted age
81360 stefanpap 17 0 5366M 148M ? 0.3 0.9 1:08.63 /Applications/Google Ch
F1Help F2Setup F3Search F4Filter F5Tree F6SortBy F7Nice F8Kill F9Quit F10Quit

```

Bild 3.8 Der Befehl `htop` ist ein Werkzeug zur Überwachung von Ressourcen.

Der Befehl `du` liefert Informationen über den Speicherplatz auf der Festplatte, der möglicherweise knapp wird. Der Parameter `-h` gibt an, dass die Ausgabe für Menschen lesbar sein soll. Ein weiterer Befehl, den Sie ausprobieren sollten, ist `df -hu`. ChatGPT kann weitere Einblicke in die Systemüberwachung geben.



Linux-Profi-Tipp: Arbeiten mit mehreren Terminals

Um mit mehreren Fenstern in einem Terminal zu arbeiten, gibt es das Tool **tmux**. Es hilft, mehrere Terminals parallel in der Kommandozeile zu verwalten. Außerdem bietet es zahlreiche weitere wertvolle Funktionalitäten.²⁹⁾

Netzwerk

Mit dem Befehl `curl ifconfig.me` lässt sich die öffentliche IP-Adresse eines Hosts ermitteln. Dies kann bei der Behebung von Verbindungsproblemen zwischen Servern hilfreich sein. Sie sollten in der Lage sein, sich mithilfe eines SSH-Schlüssels mit verschiedenen Servern zu verbinden, ohne ein Passwort eingeben zu müssen. ChatGPT bietet ausführliche Anleitungen dazu.



Container-Betriebssysteme: Je kleiner, desto besser

Container sind gekapselte Subsysteme, die in eine Host-Umgebung eingebettet sind und auf einem konfigurierbaren Gastbetriebssystem laufen. Im Sinne der Microservices-Idee, auf die wir später in diesem Kapitel noch näher eingehen werden, sollte ein Container genau eine Aufgabe perfekt lösen.

Eine Linux-Distribution für Server oder Desktops wie Ubuntu oder CentOS enthält unnötigen Overhead, der in Containern nicht benötigt wird, wie grafische Benutzeroberflächen oder viele Treiber.

Daher haben die Entwickler schlanke Distributionen für Gastbetriebssysteme für Container mit minimaler Funktionalität entwickelt. Alpine, CoreOS, RancherOS und andere unterscheiden sich in bestimmten Details.³⁰⁾ Wer viel mit Containern arbeitet, wird sich in dieses Thema vertiefen müssen.

Eine weitere gängige Praxis ist es, die Leistung von zwei Servern mit Tools wie **iperf3** zu testen. Auch hier kann ChatGPT helfen. Nachfolgend finden Sie einige Netzwerkbefehle.

²⁹⁾ <https://github.com/rothgar/awesome-tmux>

³⁰⁾ <https://computingforgeeks.com/minimal-container-operating-systems-for-kubernetes/>

Tabelle 3.3 Weitere notwendige Netzwerkbefehle im Überblick

Befehl	Bedeutung
nmap	<i>nmap</i> ist ein <i>Port-Scanner</i> für Netzwerke und Hosts. <i>nmap</i> sendet rohe IP-Pakete an alle Knoten und deren Ports. Ziel ist es, herauszufinden, welche Hosts in einem Netzwerk verfügbar sind, welche Dienste angeboten werden und ggf. welches Betriebssystem installiert ist. Sie können mit <i>nmap</i> auch Informationen über Paketfilter und Firewalls erhalten.
lsuf	Der Befehl <i>lsuf</i> (<i>list open files</i>) verknüpft offene Ports mit Diensten. Mit <i>grep</i> können Sie die Ausgabe filtern.
netstat	<i>netstat</i> zeigt den Inhalt einzelner netzbezogener Datenstrukturen an. Sie können die Ausgabe konfigurieren. Der Befehl <i>man</i> listet alle Optionen auf, die ein Benutzer in <i>netstat</i> einstellen kann.

Wenn Sie diese Befehle ausprobieren möchten, können Sie Folgendes eingeben:

```
nmap localhost
nmap google.com
```

```

1. exit (zsh)
nmap localhost
Starting Nmap 7.70 ( https://nmap.org ) at 2018-07-12 09:33 +04
Nmap scan report for localhost (127.0.0.1)
Host is up (0.0010s latency).
Other addresses for localhost (not scanned): ::1
Not shown: 498 closed ports, 498 filtered ports
PORT      STATE SERVICE
22/tcp    open  ssh
631/tcp   open ipp
3689/tcp  open rendezvous
8080/tcp  open http-proxy
Nmap done: 1 IP address (1 host up) scanned in 3.04 seconds

2. zsh
nmap google.com
Starting Nmap 7.70 ( https://nmap.org ) at 2018-07-12 09:35 +04
Nmap scan report for google.com (172.217.169.142)
Host is up (0.077s latency).
rDNS record for 172.217.169.142: sof02s32-in-f14.1e100.net
Not shown: 998 filtered ports
PORT      STATE SERVICE
80/tcp    open  http
443/tcp   open  https
Nmap done: 1 IP address (1 host up) scanned in 7.64 seconds

```

Bild 3.9 Beispiel nmap mit localhost und google.com

Wie erwartet, zeigte der Befehl, dass google.com zwei offene Ports hat: HTTP und HTTPS. Der Versuch, sich dort über SSH an Port 22 anzumelden, wäre sinnlos. Anders verhält es sich mit localhost, das Port 22 für SSH-Verbindungen geöffnet hat.

Verwenden Sie den folgenden Befehl, um festzustellen, welche Dienste auf dem System existieren, die den Port 8080 abhören. Verwenden Sie *grep*, um die Ausgabe zu filtern:

```
cat /etc/services | grep 8080
```

```

2. zsh
cat /etc/services | grep 8080
http-alt      8080/udp    # HTTP Alternate (see port 80)
http-alt      8080/tcp    # HTTP Alternate (see port 80)

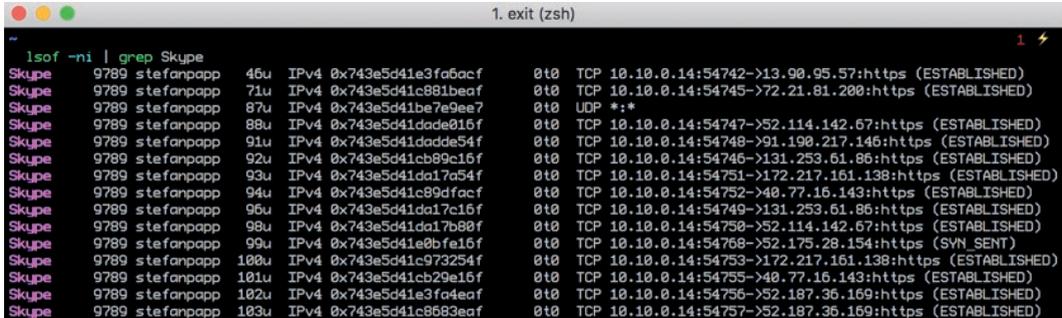
```

Bild 3.10

Mit *grep* als Parameter werden alle Dienste an Port 8080 gefiltert.

Die folgende Ausgabe zeigt, welche Verbindungen Skype verwendet, und Sie können sehen, dass einige Verbindungen offen sind:

```
lsof -ni | grep Skype
```



```

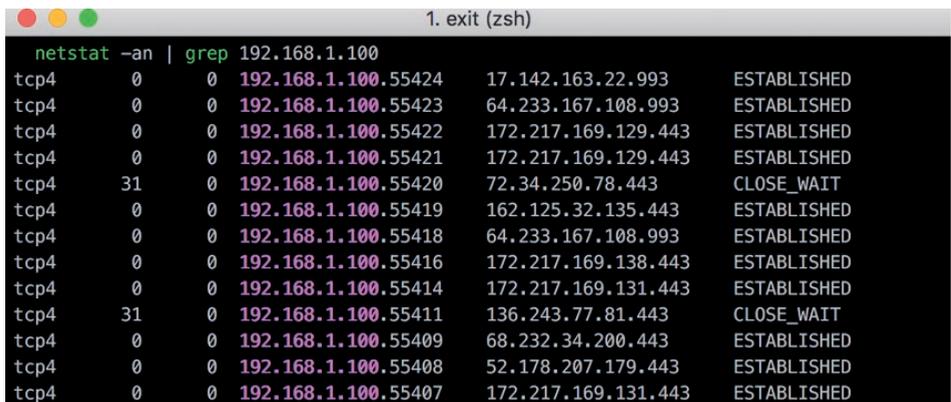
1. exit (zsh)
lsof -ni | grep Skype
Skype  9789 stefanpapp 46u IPv4 0x743e5d41e3fa6acf 0t0 TCP 10.10.0.14:54742->13.90.95.57:https (ESTABLISHED)
Skype  9789 stefanpapp 71u IPv4 0x743e5d41c881beaf 0t0 TCP 10.10.0.14:54745->72.21.81.200:https (ESTABLISHED)
Skype  9789 stefanpapp 87u IPv4 0x743e5d41be7e9ee7 0t0 UDP *: *
Skype  9789 stefanpapp 88u IPv4 0x743e5d41dada015f 0t0 TCP 10.10.0.14:54747->52.114.142.67:https (ESTABLISHED)
Skype  9789 stefanpapp 91u IPv4 0x743e5d41dada554f 0t0 TCP 10.10.0.14:54748->91.190.217.146:https (ESTABLISHED)
Skype  9789 stefanpapp 92u IPv4 0x743e5d41cb99e15f 0t0 TCP 10.10.0.14:54746->131.253.61.86:https (ESTABLISHED)
Skype  9789 stefanpapp 93u IPv4 0x743e5d41da17a54f 0t0 TCP 10.10.0.14:54751->172.217.161.138:https (ESTABLISHED)
Skype  9789 stefanpapp 94u IPv4 0x743e5d41c89dfacf 0t0 TCP 10.10.0.14:54752->40.77.16.143:https (ESTABLISHED)
Skype  9789 stefanpapp 96u IPv4 0x743e5d41da17c16f 0t0 TCP 10.10.0.14:54749->131.253.61.86:https (ESTABLISHED)
Skype  9789 stefanpapp 98u IPv4 0x743e5d41da17b80f 0t0 TCP 10.10.0.14:54750->52.114.142.67:https (ESTABLISHED)
Skype  9789 stefanpapp 99u IPv4 0x743e5d41eb9fe16f 0t0 TCP 10.10.0.14:54768->52.175.28.154:https (SVN_SENT)
Skype  9789 stefanpapp 100u IPv4 0x743e5d41c973254f 0t0 TCP 10.10.0.14:54753->172.217.161.138:https (ESTABLISHED)
Skype  9789 stefanpapp 101u IPv4 0x743e5d41cb29e16f 0t0 TCP 10.10.0.14:54755->40.77.16.143:https (ESTABLISHED)
Skype  9789 stefanpapp 102u IPv4 0x743e5d41e3fa4eaf 0t0 TCP 10.10.0.14:54756->52.187.36.169:https (ESTABLISHED)
Skype  9789 stefanpapp 103u IPv4 0x743e5d41c8683eaf 0t0 TCP 10.10.0.14:54757->52.187.36.169:https (ESTABLISHED)

```

Bild 3.11 „lsof“-Beispiel mit Anzeige aller Verbindungen zu Skype

Schließlich wenden wir `netstat` an, um die Adresse `192.168.1.100` von einer anderen Seite aus zu betrachten:

```
netstat -an | grep 192.168.1.100
```



```

1. exit (zsh)
netstat -an | grep 192.168.1.100
tcp4  0  0  192.168.1.100,55424  17.142.163.22.993  ESTABLISHED
tcp4  0  0  192.168.1.100,55423  64.233.167.108.993  ESTABLISHED
tcp4  0  0  192.168.1.100,55422  172.217.169.129.443  ESTABLISHED
tcp4  0  0  192.168.1.100,55421  172.217.169.129.443  ESTABLISHED
tcp4  31  0  192.168.1.100,55420  72.34.250.78.443  CLOSE_WAIT
tcp4  0  0  192.168.1.100,55419  162.125.32.135.443  ESTABLISHED
tcp4  0  0  192.168.1.100,55418  64.233.167.108.993  ESTABLISHED
tcp4  0  0  192.168.1.100,55416  172.217.169.138.443  ESTABLISHED
tcp4  0  0  192.168.1.100,55414  172.217.169.131.443  ESTABLISHED
tcp4  31  0  192.168.1.100,55411  136.243.77.81.443  CLOSE_WAIT
tcp4  0  0  192.168.1.100,55409  68.232.34.200.443  ESTABLISHED
tcp4  0  0  192.168.1.100,55408  52.178.207.179.443  ESTABLISHED
tcp4  0  0  192.168.1.100,55407  172.217.169.131.443  ESTABLISHED

```

Bild 3.12 Beispiel: `netstat` zeigt offene Verbindungen.

Prozess-Management

Auch Prozesse müssen verwaltet werden. Normalerweise würden Sie Hintergrundprozesse beenden oder ihre Priorität ändern.

Sie können den Befehl `ps` verwenden, um die aktuell laufenden Prozesse zu finden. Mit `grep` können Sie die Ausgabe nach bestimmten Begriffen filtern.

```
ps -ef | grep Skype
```

Auf diese Weise erhalten Sie auch die Prozess-ID. Sie können Prozesse mit `kill` beenden.

```
kill <pid>
```

Verwaltung von Rechten

Um sicherzustellen, dass die Datenzugriffsrechte korrekt sind, sollten Sie mit Befehlen wie `chmod` und `chown` vertraut sein, mit denen sich Zugriffsrechte und Dateibesitzer ändern lassen.

Es ist auch hilfreich, Befehle wie `useradd`, `passwd` und `gpasswd` zu kennen, um Benutzer anzulegen und Passwörter in Testumgebungen zu generieren.

Vim

Ein Texteditor wie **Vim** ist für jeden Ingenieur unerlässlich, der mit Open-Source-Plattformen arbeitet. Normalerweise verbindet man sich mit Servern über SSH und arbeitet an der Konsole.

Was diesen Editor einzigartig macht, ist die große Anzahl von Tastenkombinationen. Es erfordert eine lange Lernkurve und ständiges Editieren, aber jemand, der sich an die Arbeit mit Vim gewöhnt und sich die wichtigsten Tastenkombinationen merkt, wird schneller Ergebnisse erzielen als mit einem GUI-basierten Editor. Auf die Details von Vim einzugehen, würde den Rahmen sprengen. Es gibt eine Menge exzellenter Tutorials im Internet. Sie sollten sich mit dem Editor `nano` vertraut machen, wenn Vim nicht installiert und die Installation von Software blockiert ist.

Nachfolgend finden Sie eine Zusammenfassung einiger Aktivitäten für diejenigen, die diesen Bereich weiter erforschen möchten.



Übungen

- Lernen Sie, wie man eine Linux-Distribution benutzt. Wenn Sie eine Standarddistribution wie Debian oder Ubuntu beherrschen, können Sie sich an eine komplexere Distribution wie Arch Linux wagen.
- Enthusiasten sollten lernen, `grep` und `logwatch` zu verwenden, um Protokolldateien schneller zu analysieren.
- Informieren Sie sich über Lösungen wie `Cilium` und verstehen Sie, wie sie funktionieren.
- Werden Sie ein Meister im Umgang mit Befehlszeilenabkürzungen wie `Strg+A` oder `!!`.
- Absolvieren Sie eine Linux-Zertifizierung Ihrer Wahl.
- Machen Sie sich mit Werkzeugen wie Vim, `tmux` und `zsh` vertraut. Wenn Sie alle Abkürzungen beherrschen, werden Sie in Rekordzeit arbeiten. Besuchen Sie auch Websites wie `awesome Vim`, `awesome tmux` und `awesome zsh`.

- Verstehen, wie man Merge-Konflikte in Git auflöst.
- Beschäftigen Sie sich mit Wireshark. Viele Netzwerkexperten empfehlen das Tool, um den Datenverkehr effizient zu filtern.
- Um die Sicherheit zum Schutz von Datenplattformen zu vertiefen, sollten Sie sich mit ethischem Hacking und dem Schutz von Systemen vor Angriffen beschäftigen. Distributionen wie Kali Linux sind ein guter Anfang.
- Sie können Zertifizierungen wie CompTIA+ für Netzwerke erwerben, die fast keine Frage im Bereich Netzwerke unbeantwortet lassen.
- Abonnieren Sie einen Linux-Podcast.³¹⁾
- Gehen Sie über Linux hinaus und erkunden Sie andere Plattformen wie z. B. FreeBSD.
- Verwenden Sie Paketmanager, um die Software zu installieren. Sie könnten zum Beispiel Anaconda unter Linux installieren.

3.3.4 Infrastructure as Code

Ein zentraler Grundsatz im Umfeld von Data Science ist die Reproduzierbarkeit. So ist ein Ergebnis ohne die Möglichkeit der Überprüfung durch Dritte aus wissenschaftlicher Sicht nicht sinnvoll.

Reproduzierbarkeit in Data-Science-Projekten bedeutet auch, dass eine unabhängige Partei in die Lage versetzt wird, Infrastrukturen zu schaffen, um Experimente auf Knopfdruck durchzuführen. **Konfigurationsdrift**, die manuelle Änderung bestehender Infrastrukturen, kann die Abfrageergebnisse beeinflussen.

Ein Systemadministrator kann Infrastructure as Code verwenden, um Infrastrukturen in Textdateien zu deklarieren. Automatisierungswerkzeuge interpretieren dann diese deklarative Sprache und automatisieren die Erstellung dieser Infrastruktur auf einer Zielpattform wie AWS oder Azure. In diesem Zusammenhang verstehen wir Infrastrukturen als Bausteine wie virtualisierte Server, Netzwerksegmente oder Speicher. Die Devise lautet: Beschreiben Sie Ihre Infrastruktur zunächst als Code und lassen Sie sie von Tools auf Knopfdruck erstellen. Diese Tools interpretieren nicht nur den Bauplan richtig, sondern fangen auch Fehler beim Aufbau der Infrastruktur ab und reagieren darauf.

Ein Merkmal von IaC-Tools, die oft als **Provisioning Tools** bezeichnet werden, ist die **Idempotenz**. Idempotenz bedeutet, dass jeder Aufruf von Funktionen zu identischen Ergebnissen führt. Indem die Infrastruktur als deklarativer Code dargestellt wird, werden Systemkonfigurationen versionierbar. Dadurch können sie in Versionsverwaltungs-Repositories wie Git gespeichert werden.

³¹ <https://linuxunplugged.com/>

Jeder der drei großen Public-Cloud-Anbieter (Amazon, Microsoft und Google) verfügt über ein proprietäres Produkt, das Infrastrukturen für seine Plattform als Code abbildet. Administratoren bevorzugen jedoch oft eine anbieterunabhängige Lösung. Infolgedessen hat sich **Terraform** von Hashicorp zum Standard entwickelt, da es AWS, GCP, Azure und sogar noch exotischere Anbieter als Zielplattformen unterstützt.

Stellen wir uns vor, wie die Programmierlogik aussehen würde, wenn wir den Aufbau der Infrastruktur selbst durchführen würden. Der folgende Code könnte zum Beispiel Teil einer Bereitstellung sein.

```
try:
    CreateNetwork(CIDR)
    for i in range(6):
        CreateServerinstance(i)
    print( "Es wurden 6 Instanzen erstellt!")
except(PlatformCreationError e):
    print("Fehler")
```

Man müsste diese wenigen Codezeilen erheblich erweitern, um zahlreiche Ausnahmefälle abzudecken. Dieser Auszug beschreibt nur einen winzigen Teil der möglichen Probleme beim Aufbau einer Infrastruktur:

- Der Cloud-Anbieter benötigt mehr Ressourcen für neue Infrastrukturen oder der Aufbau des Zielsystems würde die Kontingente überschreiten.
- Das Dienstkonto, das Sie zum Ausführen des Automatisierungstools und zum Erstellen von Infrastrukturen verwenden, verfügt nicht über die Rechte zum Erstellen einzelner Infrastrukturen auf der Zielplattform.
- Elemente, die ein Automatisierungswerkzeug erstellen möchte, sind bereits vorhanden.

Außerdem müssen die Werkzeuge Variablen abfragen, die einzelne Konfigurationsparameter beschreiben. Im Folgenden finden Sie einen kleinen Auszug solcher Variablen:

- Die Ressourcen (CPU, RAM, Festplatten usw.) von Compute-Server-Instanzen
- Die CIDR-Bereiche der Netze
- Die Konfiguration der Verschlüsselung von Daten und Diensten
- Firewall-Regeln und Zugangskonfigurationen
- Lastverteiler und ihre Konfiguration

Um eine umfassende Lösung zu programmieren, müssen Sie viele Routinen zum Lesen und Auswerten der Parameter schreiben. Die Komplexität Ihres Codes würde mit der Anzahl der Elemente in Ihrer Infrastruktur exponentiell ansteigen. Irgendwann wäre er nur noch schwer zu lesen. Aus diesem Grund ist eine vorgefertigte Lösung

wie Terraform eine gute Wahl. Um Terraform zu installieren, folgen Sie den Anweisungen auf der Seite des Herstellers.³²⁾

Im ersten Beispiel laden wir ein Docker-Image herunter und verwenden die here-doc-Notation.

```
cat << EOF > main.tf
# Laden Sie das neueste Ghost-Image herunter.
Ressource "docker_image" "image_id" {
  name = "ghost:latest"
}
EOF
```

Wenn Sie Terraform 0.14 oder höher verwenden, müssen Sie auch die folgende Datei hinzufügen.

```
cat << EOF > versionen.tf
terraform {
  required_providers {
    docker = {
      Quelle = "Kreuzer/Docker"
    }
  }
  required_version = ">= 0.13"
}
EOF
```

Der nächste Schritt besteht darin, die definierte Erklärung als Grundlage für ein Rollout zu verwenden. Hierfür sind drei Befehle wichtig: `init`, `plan` und `apply`. Sie müssen im Verzeichnis der Datei `main.tf` ausgeführt werden.

```
Terraform init
```

Nach dem Aufruf des `init`-Befehls erstellt Terraform ein lokales Repository in einem Unterverzeichnis im Ausführungsverzeichnis. Terraform sucht in dem lokalen Ordner nach einer `main.tf`-Datei. Diese Datei erwartet die Infrastrukturkonfiguration, die definiert, was es bauen muss. Terraform sucht auch nach einer Deklaration einer Zielplattform wie AWS, Azure oder GCP in `*.tf`-Dateien. In unserem speziellen Fall wird Terraform feststellen, ob Docker benötigt wird. In unserem Beispiel wird Terraform ein Provider-Plug-in für Docker lokal in das Repository laden.

```
terraform plan
```

Dieser Befehl prüft, ob die Deklarationen in den `*.tf`-Dateien syntaktisch korrekt sind und Terraform sie ausführen kann. Terraform bereitet einen Ausführungsplan vor, muss aber noch Befehle gegen ein Zielsystem ausführen.

```
terraform apply
```

³² www.terraform.io

Sie können einen Rollout auslösen, indem Sie den Befehl `terraform apply` aufrufen. Terraform wird versuchen, alle benötigten Infrastrukturelemente zu erstellen und einen Fehler zurückgeben, wenn dies nicht möglich ist. Bild 3.13 zeigt einen erfolgreichen Rollout. Abschließend können Sie den Status mit `terraform show` überprüfen.

```

midgard 22 1 zsh
> terraform apply

An execution plan has been generated and is shown below.
Resource actions are indicated with the following symbols:
+ create

Terraform will perform the following actions:

# docker_image.image_id will be created
+ resource "docker_image" "image_id" {
+   id      = (known after apply)
+   latest = (known after apply)
+   name    = "ghost:latest"
}

Plan: 1 to add, 0 to change, 0 to destroy.

Do you want to perform these actions?
Terraform will perform the actions described above.
Only 'yes' will be accepted to approve.

Enter a value: yes

docker_image.image_id: Creating...
docker_image.image_id: Still creating... [10s elapsed]
docker_image.image_id: Still creating... [20s elapsed]
docker_image.image_id: Still creating... [30s elapsed]
docker_image.image_id: Still creating... [40s elapsed]
docker_image.image_id: Creation complete after 40s [id=sha256:f8916cf136c60c14b635e82fa45b7c6657521fe224ad6c8fbbc1826a3e72bccfghost:latest]

Apply complete! Resources: 1 added, 0 changed, 0 destroyed.

```

Bild 3.13 Erfolgreiche Ausführung von `terraform apply`

Sie können den folgenden Befehl verwenden, um zu überprüfen, ob die Bereitstellung korrekt erfolgt ist und ob wir ein Ghost-Image bereitgestellt haben:

```
docker image ls
```

Im Run-Verzeichnis finden Sie das Unterverzeichnis `.terraform`, in dem alle geladenen Provider-Plug-ins gespeichert sind. Ein Plug-in kapselt die Funktionalität zur Steuerung des Rollouts. Alternativ können Sie die Provider auch mit dem folgenden Befehl auflisten:

```
terraform providers
```

Die Blöcke in einer Terraform-Erklärung sind immer wie folgt aufgebaut:

```

<BLOCK TYPE> "<BLOCK LABEL>" "<BLOCK LABEL>" {
  # Block body
  <IDENTIFIER> = <EXPRESSION> # Argument
}

```

Nachdem wir bereits ein Docker-Image deklariert haben, deklarieren wir eine Cloud-Ressource des Provider-Plug-ins für AWS, wie unten gezeigt. Der Block in diesem Beispiel selbst ist minimalistisch. Nur eine Variable bestimmt, dass ein CIDR-Block deklariert wird. Es gibt viele andere Konfigurationsoptionen für dieses Element.³³⁾

```
Ressource "aws_vpc" "main" {  
  cidr_block = "10.0.0/16"  
}
```

Dieser Block deklariert `aws_vpc`, einen Platzhalter, der ein Netzwerk auf AWS darstellt. Neben den Ressourcentypen gibt es in Terraform noch weitere Elemente wie z. B. Variablen und Ausgaben. Auch hier können Sie die Details in der Terraform-Dokumentation finden.

Sie können den folgenden Befehl verwenden, um Installationen wieder zu löschen.

```
terraform destroy
```

Neben Terraform werden auch andere Tools wie Ansible, Chef oder Puppet häufig in Verbindung mit Terraform verwendet. Sie haben einen anderen Zweck und werden gewöhnlich als **Konfigurationsmanagement**-Werkzeuge bezeichnet. Während Terraform für die Erstellung der Infrastruktur verwendet wird, installieren diese Tools die Software automatisch. Wenn Sie also eine Kafka-Standardinstallation in der Cloud einrichten möchten, würden Sie zunächst virtuelle Maschinen bei einem Cloud-Provider mit Terraform einrichten und Kafka mit Ansible, Chef oder Puppet darauf installieren.



Übungen

- Lernen Sie die Module von Terraform kennen und erfahren Sie, wie Sie Infrastrukturen modularisieren können.
- Wenn Sie einen bevorzugten Cloud-Anbieter haben, erkunden Sie die Infrastructure-as-Code-Lösungen, die dieser für seine Plattformen anbietet, z. B. AWS CloudFormation oder Azure Resource Manager.
- Lernen Sie, wie Sie Software-Rollouts wie Apache Kafka oder Apache Webserver mit Ansible, Puppet und Co. automatisieren können.
- Nehmen Sie sich etwas Zeit, um darüber nachzudenken, wie Sie bereits bestehende Plattformen aktualisieren würden. Würden Sie es zum Beispiel vorziehen, bestehende Plattformen zu löschen und eine neue Version zu erstellen? Oder würden Sie stattdessen eine bestehende Plattform mit Änderungen an der Infrastruktur aktualisieren?

³³ <https://registry.terraform.io/providers/hashicorp/aws/latest/docs/resources/vpc>

3.4 Platform as a Service

Bei Infrastructure as a Service stellt der Cloud-Anbieter Komponenten wie virtuelle Maschinen oder Kubernetes-Cluster bereit, und der Kunde verwaltet, was darauf läuft. Bei Platform-as-a-Service werden ganze Plattformen bereitgestellt, und die Nutzer müssen sich nicht mehr um die Verwaltung der Infrastruktur kümmern. Stellen Sie sich in einem einfachen Beispiel einen Kunden vor, der eine Datenbank für seine maßgeschneiderte Lösung nutzen möchte. Bei einer IaaS-Lösung kann er eine virtuelle Maschine mieten und dort sein Datenbankmanagementsystem installieren; bei einer PaaS-Lösung stellt der Cloud-Anbieter das DBMS zur Verfügung; die Cloud-Administratoren des Kunden müssen sich nicht mehr um die Wartung der Systemsoftware kümmern, z. B. um System-Upgrades oder Backups.

Jeder Anbieter von Cloud-Plattformen bietet seinen Kunden zahlreiche Dienste in vielen verschiedenen Kategorien an. Künstliche Intelligenz und Data-Science-Dienste sind nur eine Kategorie von vielen anderen Diensttypen. Es ist fast unmöglich, alle KI- und Datendienste eines Anbieters in einem Buch zu beschreiben. Angesichts der raschen Entwicklung dieser Dienste wäre das Buch bei seiner Veröffentlichung bereits veraltet. Die Dienste können spezifisch sein und sich mit konkreten Problemen im Zusammenhang mit der Nutzung von KI befassen wie z. B. Cloud-Anbieter, die spezielle Dienste für Computer Vision, Betrugserkennung oder vorausschauende Wartung anbieten. Andere Dienste richten sich an Entwickler, um ihnen eine Plattform mit einer breiten Palette von Tools für die Entwicklung ihrer KI-Dienste zur Verfügung zu stellen.

In diesem Buch konzentrieren wir uns auf die drei großen Cloud-Anbieter: AWS, Azure und GCP. Es ist wichtig zu betonen, dass viele andere Anbieter ebenfalls geeignete Lösungen anbieten.

Cloud-Anbieter bieten in der Regel eigene und externe (oder Marktplatz-) Lösungen an. Bei der ersten Kategorie handelt es sich um Artefakte, die die Cloud-Anbieter selbst erstellen, und viele dieser Dienste sind einem breiten Spektrum von Datenexperten bekannt, wie AWS Redshift oder Azure SQL Server. In der zweiten Kategorie bieten Dritunternehmen ihren Kunden Lösungen über den Cloud-Anbieter an. Marktplätze enthalten auch IaaS-Lösungen wie z. B. spezifische virtuelle Images für bestimmte Zwecke, aber häufiger sind Plattformen. So stellen beispielsweise DMBS-Entwickler wie MongoDB Inc. ihr Produkt MongoDB als Marktplatzdienst in der Cloud für ihre Kunden bereit.

3.4.1 Cloud Native PaaS-Lösungen

Alle Anbieter bieten Cloud-native Suiten von PaaS-Lösungen für maschinelles Lernen, Datenanalyse und künstliche Intelligenz-Workloads an. In diesem Abschnitt fassen wir die gängigsten Lösungen zusammen. Es gibt jedoch Seiten für jeden Cloud-Anbieter, auf denen Sie tiefer einsteigen können.^{34),35),36)}

3.4.1.1 AWS

Amazon SageMaker

Amazon SageMaker ermöglicht Data Scientists das Erstellen, Trainieren und Bereitstellen von Modellen für maschinelles Lernen. Es erleichtert die Entwicklung hochwertiger Modelle, indem es jeden Schritt des Machine-Learning-Prozesses entlastet. Es bietet mehrere integrierte Algorithmen, One-Click-Training und Funktionen zur Modellabstimmung und unterstützt die Bereitstellung von Modellen in großem Umfang ohne Ausfallzeiten.

AWS Glue

AWS Glue ist ein ETL-Service, der Daten kostengünstig kategorisiert, bereinigt, anreichert und zuverlässig zwischen verschiedenen Datenspeichern verschiebt. Datenexperten verwenden ihn bei der Vorbereitung und Umwandlung von Daten für Analysen und maschinelles Lernen. AWS Glue bietet eine serverlose Umgebung, die automatisch skaliert wird, um die Verarbeitungsanforderungen ihrer Datenaufträge zu erfüllen.

Amazon Redshift

Amazon Redshift ist ein schnelles, skalierbares Data Warehouse, das Daten über Data Warehouses und Data Lakes hinweg analysieren kann. Redshift erzielt außergewöhnliche Leistungen durch maschinelles Lernen, massiv parallele Abfrageausführung und spaltenweise Speicherung auf Hochleistungsfestplatten. Es lässt sich in Data Lakes integrieren und ermöglicht Datenexperten die Durchführung einheitlicher Analysen über eine komplette Datenlandschaft hinweg.

Amazon Athena

Amazon Athena ist ein interaktiver Abfrageservice, der Daten in Amazon S3 mit SQL analysiert. Da es sich um eine serverlose Komponente handelt, müssen Datenexperten keine Infrastruktur für Athena verwalten und zahlen nur für die Abfragen, die sie

³⁴ <https://aws.amazon.com/solutions/ai-ml/>

³⁵ <https://azure.microsoft.com/en-us/products>

³⁶ <https://cloud.google.com/products>

ausführen. Dies macht es zu einem einfachen, aber leistungsstarken Tool für die Analyse großer Datensätze, ohne dass komplexe ETL-Aufträge zur Vorbereitung der Daten für die Analyse erforderlich sind.

Amazon Comprehend

Amazon Comprehend ist ein NLP-Service, der maschinelles Lernen nutzt, um Erkenntnisse und Beziehungen in Texten aufzudecken. Er bietet Funktionen zur Stimmungsanalyse, Entity-Erkennung, Themenmodellierung und Spracherkennung, um Textdaten effizient zu organisieren und zu kategorisieren.

Amazon EMR

Amazon EMR ist eine Plattform für die Ausführung und Skalierung von Big-Data-Arbeitslasten wie Apache Spark, Hive und Presto. Datenexperten können Aufträge auf verschiedenen Infrastrukturdiensten wie Amazon EC2 ausführen.

Amazon Forecast

Amazon Forecast nutzt maschinelles Lernen, um hochpräzise Prognosen zu erstellen. Es erkennt automatisch, wie sich Produktmerkmale, Feiertage und Wetter auf Ihre Geschäftsmetriken auswirken, um fundierte Entscheidungen zu treffen. Forecast erfordert keine Erfahrung mit maschinellem Lernen und kann zur Vorhersage von Verkäufen, Lagerbeständen und der betrieblichen Nachfrage verwendet werden.

AWS Lake Formation

AWS Lake Formation vereinfacht und automatisiert viele der manuellen Schritte, die zum Erstellen eines Data Lakes erforderlich sind, einschließlich des Sammelns, Bereinigens und Katalogisierens von Daten und der sicheren Bereitstellung dieser Daten für Analysen und maschinelles Lernen.

Amazon Rekognition

Amazon Rekognition hilft bei der Identifizierung von Objekten, Personen, Text, Szenen und Aktivitäten in Bildern und Videos und erkennt unangemessene Inhalte. Rekognition bietet auch hochpräzise Gesichtsanalyse- und Gesichtssuchfunktionen, die Kunden zum Erkennen, Analysieren und Vergleichen von Gesichtern für verschiedene Benutzerüberprüfungen, Personenzählungen und Anwendungsfälle der öffentlichen Sicherheit nutzen können.

3.4.1.2 Azure

Azure Machine Learning

Azure Machine Learning ist ein Cloud-Service, mit dem sich Modelle für maschinelles Lernen schnell erstellen, trainieren und bereitstellen lassen. Er bietet Tools für jede Phase des Lebenszyklus des maschinellen Lernens, einschließlich Datenvorbereitung, Modelltraining, Bereitstellung und Verwaltung. Mit Azure Machine Learning können Teams die Modellauswahl und -abstimmung automatisieren, ihre Modelle in der Produktion verwalten und überwachen und sogar Modelle vor Ort oder im Netzwerk bereitstellen.

Azure Synapse

Azure Synapse Analytics kombiniert Big Data und Data Warehousing, um grenzenlose Analysedienstfunktionen bereitzustellen. Es bietet eine einheitliche Erfahrung für das Aufnehmen, Aufbereiten, Verwalten und Bereitstellen von Daten für unmittelbare BI- und Machine Learning-Anforderungen. Synapse Analytics lässt sich nahtlos in Azure Machine Learning und Power BI integrieren, sodass Kunden große Datenmengen analysieren und Modelle für maschinelles Lernen direkt in ihrem Data Warehouse erstellen können. Azure Synapse Analytics kann auch Apache Spark integrieren.

Azure Databricks

Azure Databricks ist eine Apache Spark-native Analyseplattform, die für Microsoft Azure-Cloud-Dienste optimiert ist. Sie wurde gemeinsam mit den Gründern von Apache Spark entwickelt und lässt sich in Azure-Dienste integrieren, um eine Ein-Klick-Einrichtung, optimierte Arbeitsabläufe und einen interaktiven Arbeitsbereich zu bieten, der die Zusammenarbeit zwischen Data Scientists, Data Engineers und Businessanalysten ermöglicht. Azure Databricks unterstützt verschiedene Aufgaben der Datenanalyse und des maschinellen Lernens und ist damit ein leistungsstarkes Tool für die Verarbeitung und Analyse von Big Data.

Azure Data Factory

Azure Data Factory ist ein hybrider Datenintegrationsdienst, der es Teams ermöglicht, ihre ETL/ELT-Workflows zu erstellen, zu planen und zu orchestrieren. Er unterstützt die Verbindung mit einer Vielzahl von Datenquellen, die Umwandlung von Daten mithilfe von Azure Data Lake Analytics, Azure Databricks und Azure HDInsight sowie das Laden von Daten in verschiedene Azure-Datenspeicher für weitere Analysen oder maschinelles Lernen.

Azure Cognitive Services

Azure Cognitive Services bietet eine Sammlung von APIs, SDKs und Diensten, die Entwicklern zur Verfügung stehen, um ihre Anwendungen intelligenter, ansprechender und besser auffindbar zu machen. Sie umfassen eine Vielzahl von KI-Funktionen wie z. B. Computer Vision, Verarbeitung natürlicher Sprache, Sprache und Entscheidungsalgorithmen. Mit diesen Diensten können Kunden KI-Funktionen wie Bild- und Videoanalyse, Gesichtserkennung, Spracherkennung und -synthese, Knowledge Mining und Sprachverständnis problemlos in ihre Anwendungen integrieren.

Azure Bot Services

Azure Bot Services bietet ein umfassendes Framework für die Erstellung von KI-Konversationserlebnissen auf Unternehmensniveau. Es ermöglicht Entwicklern das Erstellen, Testen, Bereitstellen und Verwalten intelligenter Bots, die auf natürliche Weise mit Benutzern über Websites, Apps, Cortana, Microsoft Teams, Skype, Slack, Facebook Messenger und mehr interagieren können. Azure Bot Services nutzt die KI-Fähigkeiten von Microsoft, um Bots zu erstellen, die menschliche Bedürfnisse verstehen und interpretieren können, und erleichtert so die Implementierung von Lösungen, die Benutzer ansprechen und unterstützen.

Azure HDInsight

Azure HDInsight ist ein vollständig verwalteter Cloud-Service für Open-Source-Analysen. Er macht die Verarbeitung großer Datenmengen einfach, schnell und kostengünstig. Sie können beliebige Open-Source-Frameworks wie Hadoop, Spark, Hive, LLAP, Kafka, Storm und R verwenden. HDInsight lässt sich auch mit Azure Data Lake Storage, Azure Synapse Analytics und Azure Machine Learning integrieren, um End-to-end-Analyse-Pipelines zu erstellen.

3.4.1.3 Google

Google AI Platform

Google AI Platform ist eine umfassende Suite für ML-Entwickler und Data Scientists. Sie bietet einen verwalteten Dienst, der den gesamten ML-Workflow erleichtert, von der Dateneingabe und -vorbereitung bis hin zum Modelltraining, der Bewertung, dem Einsatz und der Vorhersage. Sie unterstützt bekannte ML-Frameworks, darunter TensorFlow, PyTorch und scikit-learn, und bietet eine benutzerfreundliche Oberfläche und Befehlszeilen-Tools.

BigQuery

BigQuery ist ein serverloses, hoch skalierbares, kosteneffizientes Multi-Cloud-Data-Warehouse, das für geschäftliche Agilität und die Durchführung von Analysen im großen Maßstab entwickelt wurde. Es bietet Echtzeiteinblicke mit SQL-ähnlichen Abfra-