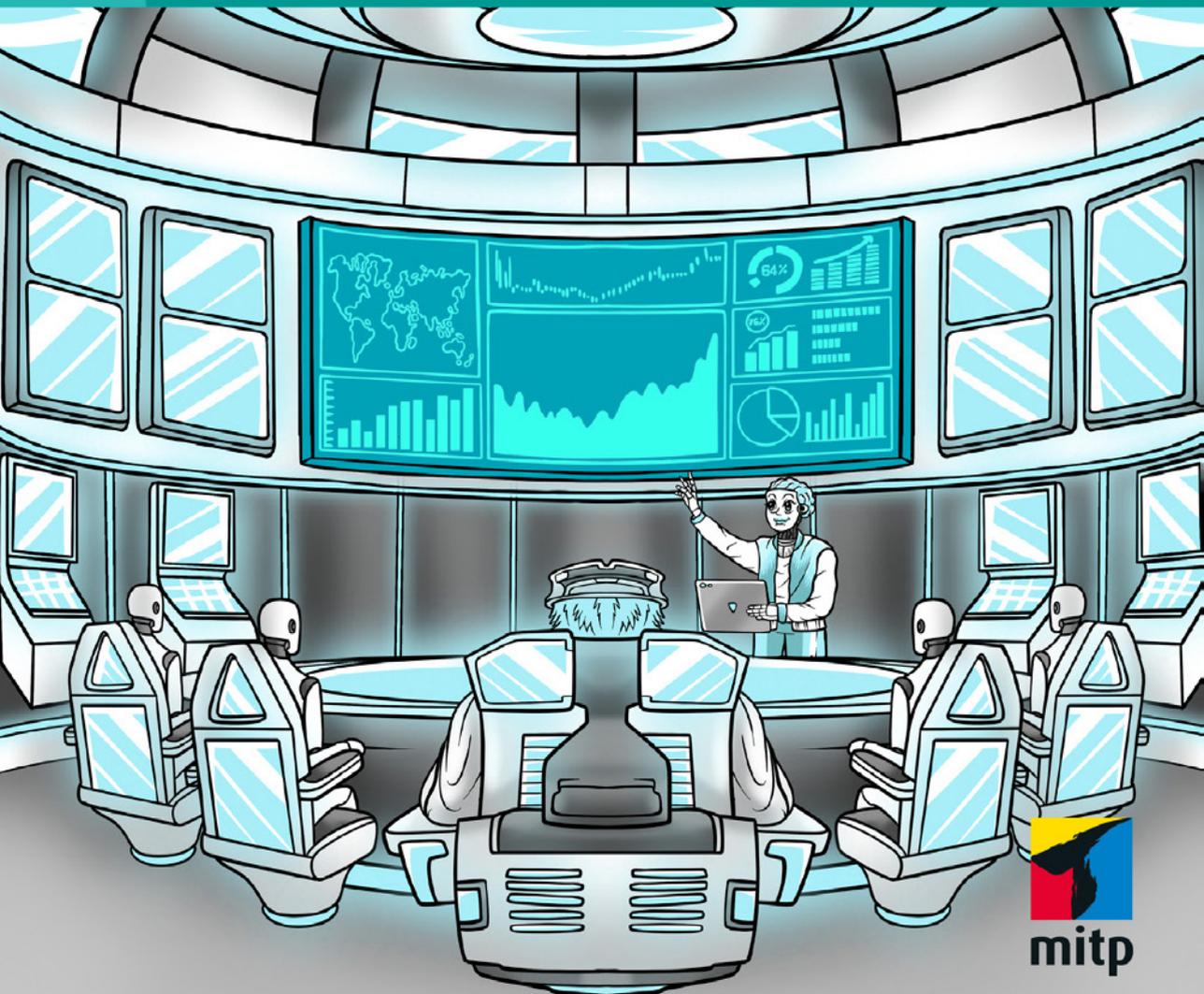


Manuel Hanik | Fabian Hanik

Microsoft Fabric

Der praktische Einstieg in die
All-in-One-Datenplattform



Hinweis des Verlages zum Urheberrecht und Digitalen Rechtemanagement (DRM)

Liebe Leserinnen und Leser,

dieses E-Book, einschließlich aller seiner Teile, ist urheberrechtlich geschützt. Mit dem Kauf räumen wir Ihnen das Recht ein, die Inhalte im Rahmen des geltenden Urheberrechts zu nutzen. Jede Verwertung außerhalb dieser Grenzen ist ohne unsere Zustimmung unzulässig und strafbar. Das gilt besonders für Vervielfältigungen, Übersetzungen sowie Einspeicherung und Verarbeitung in elektronischen Systemen.

Je nachdem wo Sie Ihr E-Book gekauft haben, kann dieser Shop das E-Book vor Missbrauch durch ein digitales Rechtemanagement schützen. Häufig erfolgt dies in Form eines nicht sichtbaren digitalen Wasserzeichens, das dann individuell pro Nutzer signiert ist. Angaben zu diesem DRM finden Sie auf den Seiten der jeweiligen Anbieter.

Beim Kauf des E-Books in unserem Verlagsshop ist Ihr E-Book DRM-frei.

Viele Grüße und viel Spaß beim Lesen

Ihr mitp-Verlagsteam



Manuel Hanik, Fabian Hanik

Microsoft Fabric

Der praktische Einstieg in die
All-in-One-Datenplattform



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de/opac.htm> abrufbar.

ISBN 978-3-7475-0954-8

1. Auflage 2025

www.mitp.de

E-Mail: mitp-verlag@lila-logistik.com

Telefon: +49 7953 / 7189 - 079

Telefax: +49 7953 / 7189 - 082

© 2025 mitp Verlags GmbH & Co. KG, Augustinusstr. 9a, DE 50226 Frechen

Dieses Werk, einschließlich aller seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt 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 benutzt werden dürften.

Lektorat: Janina Vervost

Sprachkorrektorat: Christine Hoffmeister

Covergestaltung: Christian Kalkert

Coverbild & Illustrationen: Manuel Hanik, Fabian Hanik

Satz: Petra Kleinwegen

Inhaltsverzeichnis

	Vorwort von Gabi Münster	7
	Vorwort von Goran Cvetanovski	9
	Einleitung	11
E.1	Für wen ist dieses Buch gedacht?	12
E.2	Aufbau des Buchs	12
E.3	Konventionen	13
E.4	Downloads	13
	Über die Autoren	15
1	Grundlagen	17
1.1	Traditionelle Dateninfrastruktur	17
1.2	Fabric als All-in-One-Plattform	21
1.3	Die Komponenten von Fabric	23
2	Praxisprojekt	31
2.1	Anforderungen und Ziele	31
2.2	Daten	32
2.3	Datenteam	34
2.4	Projektplan	37
3	Erste Schritte	41
3.1	Fabric-Konto einrichten	42
3.2	Arbeitsbereich einrichten	54
4	Datenexploration	59
4.1	CSV-Datei untersuchen	61
4.2	Excel-Datei untersuchen	65
4.3	XML-Datei untersuchen	69
4.4	JSON-Datei untersuchen	72

5	Datenspeicherung	77
5.1	Datenarchitektur	78
5.2	Delta Lake	82
5.3	Medaillon-Architektur	83
5.4	Lakehouse	86
5.5	Warehouse	89
5.6	Übersicht der Fabric-Elemente	92
6	Data Engineering	93
6.1	Landing-Zone	94
6.2	Bronze-Schicht	105
6.3	Silber-Schicht	121
6.4	Gold-Schicht	135
6.5	Orchestrierung	149
6.6	Übersicht der Fabric-Elemente	159
7	Data Science	161
7.1	Anwendungsfall des Datenprojekt-Beispiels	163
7.2	Workflow für Data Science	163
7.3	ML-Modell trainieren	165
7.4	ML-Modell bereitstellen	177
7.5	ML-Modell anwenden	179
7.6	Übersicht der Fabric-Elemente	182
8	Data Analytics	183
8.1	Daten mit SQL analysieren	185
8.2	Daten mit Power BI visualisieren	191
8.3	Power-BI-Elemente mit einer App bereitstellen	207
8.4	Übersicht der Fabric-Elemente	212
9	Monitoring	213
9.1	Überwachungshub nutzen	215
9.2	Warnungen für Berichte einrichten	215
9.3	Übersicht der Fabric-Elemente	218
	Zusammenfassung und Schlusswort	219
	Stichwortverzeichnis	227



Vorwort von Gabi Münster

Immer mehr Unternehmen erkennen heutzutage den Wert von Daten als eine bedeutsame und oft vernachlässigte Ressource für den Erfolg und die Optimierung ihrer Geschäftsprozesse. Datengetriebene Prozesse und Entscheidungen sollen als Unterstützung zum menschlichen Faktor und den wichtigen Erfahrungen der Mitarbeiter dienen, um Risiken und negative Einflussfaktoren zu reduzieren.

Wie viele andere Anbieter bietet auch Microsoft bereits seit Jahrzehnten Produkte und Services zur Datenhaltung, Datenanalyse und Datenvisualisierung an. In den letzten Jahren hat sich im Bereich der Anforderungen, der technischen Möglichkeiten, aber auch im Bereich der Zielgruppen vieles geändert. Dem müssen auch die Angebote Rechnung tragen.

Eine Vielzahl unterschiedlicher Produkte, die aufwendig in eine komplexe Landschaft integriert werden müssen und verschiedenste Kenntnisse und Fertigkeiten voraussetzen, sind nicht zeitgemäß für die Mehrheit der Projekte und Szenarien.

Mit Microsoft Fabric setzt Microsoft genau an diesem Punkt an. Die Vision ist die Bereitstellung einer Plattform, die ohne komplexe Integration, aufwendige Wartung und basierend auf bestehenden Standards und Kenntnissen Lösungsmöglichkeiten für einen Großteil an Anwendungsfällen bietet.

Nicht wegzudenken in der heutigen Zeit, und deswegen auch ein relevanter Aspekt der Microsoft Fabric Strategie, ist die Integration von generativer KI in die Gesamtheit der Plattform. Sei es zur Unterstützung von professionellen Entwicklern und Entwicklerinnen bei der Vereinfachung von Aufgaben, zum Erlernen von neuen Fähig- und Fertigkeiten oder zur Navigation und Beantwortung von Fragen: Jedem Fabric-Nutzer steht immer ein Copilot zur Seite.

Betrachten wir zunächst die Zielgruppen, die in Microsoft Fabric ein geeignetes Angebot finden können. Ein Strategiewechsel bedeutet natürlich nicht, dass die bisherigen Kernanwendungen von Data&Analytics-Diensten weniger bedeutsam für die Gestaltung des Produkts sind. Data und BI Professionals, Entwickler und Entwicklerinnen sowie Administratoren und Administratorinnen sind nach wie vor die Kernzielgruppe, die für den Betrieb der Dienste und die Entwicklung und Bereitstellung von zentralen Lösungen in den Unternehmen verantwortlich ist. Demzufolge ist es relevant, dass Microsoft Fabric die Voraussetzungen für professionelle Entwicklung generischer, flexibler und performanter Lösungen bereitstellt.

Mit Power BI als ihrem ersten erfolgreichen Software-as-a-Service-Angebot in diesem Bereich hat Microsoft in der Vergangenheit bereits den Grundstein für ein stabiles Angebot im Bereich Self-service Analytics gelegt. Die zunehmende Affinität und Sicherheit der Citizen Developer begründet die Notwendigkeit, auch die restlichen Bereiche der Datenanalyse-Wertschöpfungskette für interessierte Benutzer und Benutzerinnen nutzbar zu machen. Die Bereitstellung von Low-Code-/No-Code-Lösungen, eine intuitiv nutzbare Integration von relevanten Kernprozessen wie Quellcodeverwaltung und Bereitstellungsprozessen und die Vereinfachung der Gestaltung und Realisierung von Gesamtprozessen sind hierfür von essenzieller Bedeutung.

Genauso wichtig sind aber selbstverständlich die Konsumierenden der zu erstellenden Datenprodukte. Sei es in Form von flexiblen und interaktiven Berichten, operativen Berichten zur pixelgetreuen Darstellung oder über die Ad-hoc-Analysemöglichkeiten in Excel, die auch in der heutigen Zeit aus den Fachabteilungen nicht wegzudenken sind.

Aber neben diesen offensichtlichen Zielgruppen gibt es auch weitere Anwenderbereiche, die über die erweiterten Angebote in Fabric erschlossen werden sollen. Hier sind unter anderem Datenanalysten und Data Scientists zu nennen.

Damit stellt sich auch die Frage nach den Anwendungsfällen, die in Fabric realisierbar sind. Neben den traditionellen Warehouse- und Reporting-Szenarien sowie modernen Lakehouse-Konzepten bietet Fabric auch Lösungen für Streaming und Echtzeitszenarien. Real-Time Intelligence basiert auf einer etablierten Technologie, aber gerade die nahtlose Integration in die Plattform eröffnet neue Anwendungsfälle, nicht nur im Bereichen wie Internet of Things und Sensordatenintegration, sondern auch in Echtzeit- und Textanalysen.

Die Integration von Data Science ermöglicht den oben genannten Benutzergruppen das Erkunden und Realisieren von Experimenten und Machine-Learning-Projekten ohne die Notwendigkeit von Vorkenntnissen.

Ich möchte mich bei Fabian und Manuel für dieses Buch bedanken, das neben den bereits genannten Zielgruppen auch Entscheidungsträgern auf der Suche nach zielführenden Informationen, Fachanwendern mit dem Wunsch, ihren technischen Horizont zu erweitern, oder Studierenden zum Zweck der Auseinandersetzung mit State-of-the-art-Lösungen im Bereich der ganzheitlichen Datenaufbereitung und -analyse eine anschauliche und umfangliche Lektüre bereitstellt. Gerade der Ansatz, am Beispiel eines Projekts den gesamten Ablauf von Beginn bis Ende zu erläutern, ist optimal geeignet, um für jeden Interessenten die richtigen Informationen gezielt und effizient zur Verfügung zu stellen.

Gabi Münster
Senior Program Manager
Fabric Customer Advisory Team



Vorwort von Goran Cvetanovski

In meinen Jahren als Gründer von *Hyperight* und Chefredakteur der *Data Innovation Summit* hatte ich das Privileg, mit einigen der klügsten Köpfe im Bereich Datenmanagement, Analytics und KI in Kontakt zu kommen. Durch diese Erfahrungen wurde eines klar: Während KI die Schlagzeilen beherrscht und die Vorstellungskraft vieler beflügelt, ist es das Fundament eines robusten Datenmanagements, das diese transformative Technologie tatsächlich so mächtig macht. Ohne eine solide Basis fehlt selbst den fortschrittlichsten KI-Systemen die zuverlässigen, sicheren und hochwertige Daten, die sie benötigen, um sich zu entwickeln.

Dieses Buch – ein praxisnaher Leitfaden zu Microsoft Fabric – könnte zu keinem besseren Zeitpunkt erscheinen. In einer Welt, die mit exponentiellem Datenwachstum, komplexen Anforderungen an Governance und Compliance, steigenden Cloud-Kosten und einem nicht endenden Bedarf an verwendbaren Erkenntnissen kämpft, benötigen Unternehmen dringend Lösungen, die sowohl skalierbar als auch praxisnah sind. Microsoft Fabric könnte ein entscheidendes Puzzleteil sein – eine umfassende Plattform, die es Organisationen ermöglicht, Daten End-to-End zu verwalten und gleichzeitig messbaren geschäftlichen Nutzen zu erzielen.

Was dieses Buch besonders macht, ist sein Fokus auf Praxisnähe und Klarheit. Die Autoren führen die Leser durch alle Phasen eines Datenprojekts – von den Grundlagen von Datenplattformen über die Implementierung von Machine-Learning-Modellen und die Erstellung verwendbarer Analysen bis hin zur Sicherstellung eines stabilen Betriebs. Dies ist keine rein theoretische Abhandlung, sondern ein praxisorientierter Leitfaden zur Bewältigung realer Herausforderungen. Mit Best Practices, Tipps und konkreten Handlungsanweisungen ausgestattet, ermöglicht es dieses Buch, Datenprojekte erfolgreich umzusetzen – sowohl aus technischer als auch aus geschäftlicher Perspektive.

Beim Lesen dieses Buchs werden Sie Schritt für Schritt durch die Implementierung von Microsoft Fabric geführt – von den Grundlagen von Datenplattformen und den Kernkomponenten bis hin zur Entwicklung und Optimierung umfassender Datenprojekte. Beginnend mit den grundlegenden Konzepten stellt das Buch ein Beispielprojekt mit einem realen Anwendungsfall vor und leitet die Leser durch die Einrichtung der Umgebung, die Datenexploration und die Speicherung in Lakehouses und Data Warehouses. Es behandelt den Aufbau robuster Datenpipelines mit PySpark und SQL, das Trainieren von Machine-Learning-Modellen

zur Gewinnung vorausschauender Erkenntnisse und die Erstellung interaktiver Power-BI-Dashboards für aussagekräftige Analysen. Abschließend wird die Bedeutung operativer Stabilität durch effektives Monitoring hervorgehoben, um Workflows effizient und leistungsoptimiert zu halten. Ein unverzichtbares Werk für alle, die Microsoft Fabric meistern wollen.

In einer Zeit, in der KI sich von einem Schlagwort zu einer allgegenwärtigen Technologie entwickelt, die in Software, Produkten und Strategien eingebettet ist, muss der Fokus auf dem Fundament bleiben: den Daten. Die Vision einer Zukunft autonomer Systeme – in der Systeme intelligent auf menschliche Bedürfnisse reagieren und sich anpassen – hängt wesentlich von unserer Fähigkeit ab, Daten effektiv zu managen. Dieses Buch liefert die Werkzeuge, um diese Zukunft zu gestalten.

Wenn Sie dieses Buch in die Hand nehmen, lernen Sie nicht nur eine neue Technologie kennen, Sie legen auch das Fundament für den Erfolg im Zeitalter von KI. Ich danke den Autoren für die Erstellung eines so umfassenden und praxisorientierten Leitfadens und ermutige jede Leserin und jeden Leser, sich mit Neugier, Entschlossenheit und Engagement für hohe Datenqualität darauf einzulassen. Die Reise, die vor Ihnen liegt, ist eine spannende. Möge dieses Buch Ihr verlässlicher Begleiter sein.

Herzliche Grüße
Goran Cvetanovski

*Gründer von Hyperight
Chefredakteur von Data Innovation Summit*



Einleitung

Die Bedeutung von Daten hat in den letzten Jahren massiv zugenommen. Zusammen mit der rasanten Entwicklung von künstlicher Intelligenz entfalten sie ihr volles Potenzial. Dadurch sind Daten in der heutigen Zeit der entscheidende Treiber für Effizienz, Innovation und Wettbewerbsvorteile. Unternehmen, die Daten effizient nutzen und ihr Potenzial entfalten, können ihre Geschäftsprozesse optimieren, Kosten senken und die Produktivität nachhaltig steigern.

Aus diesem Grund haben sich zahlreiche neue Technologien entwickelt, mit denen Daten gespeichert, verarbeitet und analysiert werden können. Gleichzeitig sind weitere Technologien entstanden, die für die Entwicklung von Dashboards und KI-Modellen genutzt werden. Unternehmen stehen dadurch vor der großen Herausforderung, für den gesamten End-to-End-Prozess – von der Datenaufbereitung bis hin zu finalen Dashboards und trainierten KI-Modellen – auf eine Vielzahl unterschiedlicher Technologien zurückgreifen zu müssen. Infolgedessen sind oftmals mehrere verschiedene Lizenzen notwendig, die zu hohen Kosten führen. Zusätzlich müssen die einzelnen Technologien noch in der Lage sein, miteinander zu kommunizieren. So ist es sehr herausfordernd, einen effizienten End-to-End-Datenprozess zu realisieren.

Microsoft hat dieses Problem erkannt und mit Microsoft Fabric eine Lösung entwickelt, die es erlaubt, alle relevanten Datenprozesse auf einer einzigen Plattform abzubilden. Microsoft Fabric vereint neue und bestehende Technologien aus dem Microsoft-Ökosystem und ermöglicht es, diese mit nur einer Lizenz in einer einheitlichen Umgebung zu nutzen. Durch die enge Verzahnung dieser Technologien wird eine nahtlose Kommunikation zwischen den Komponenten ermöglicht. Zudem bietet die benutzerfreundliche Oberfläche von Microsoft Fabric eine intuitive Möglichkeit, komplexe Datenprozesse zu erstellen und zu verwalten.

Dieses Buch ermöglicht einen praxisorientierten Einstieg in Microsoft Fabric. Anhand eines fiktiven Beispiels wird ein komplettes End-to-End-Datenprojekt in Microsoft Fabric umgesetzt, wobei alle relevanten Schritte – von der Datenanbindung über Datentransformationen bis hin zum finalen Dashboard – detailliert erklärt werden. Es vermittelt sowohl grundlegende Konzepte und Best Practices für die Umsetzung von Datenprojekten als auch spezifisches Wissen zur Nutzung der einzelnen Komponenten von Microsoft Fabric.

Sie haben die Möglichkeit, aktiv mitzumachen und alle Schritte selbst umzusetzen. Um das Lernen noch anschaulicher zu gestalten, werden die Erklärungen durch eine fiktive Geschichte über eine futuristische Datenfabrik begleitet, die Microsoft Fabric symbolisiert, und durch zahlreiche visuelle Elemente ergänzt. Sie werden auf eine exklusive Besichtigung mitgenommen, bei der Sie die verschiedenen Bereiche und Prozesse der Datenfabrik hautnah erleben und Kapitel für Kapitel tiefer in die Datenfabrik – und somit auch in Microsoft Fabric – eintauchen. Das Buch ist damit mehr als nur ein theoretischer Leitfaden – es ist eine interaktive Reise in die Welt von Microsoft Fabric.

E.1 Für wen ist dieses Buch gedacht?

Dieses Buch eignet sich sowohl für Einsteiger als auch für Experten im Datenbereich. Ein grundlegendes Verständnis von Daten sowie erste Erfahrungen mit Python und SQL sind an einigen Stellen von Vorteil. Es richtet sich an:

- **Interessierte**, die Microsoft Fabric näher kennenlernen möchten und einen praxisnahen Einstieg suchen.
- **Nutzer des Microsoft-Ökosystems**, die das Potenzial von Microsoft Fabric als All-in-One-Datenplattform entdecken möchten.
- **Datenexperten** wie Data Engineers, Data Scientists und Data Analysts, die sich in Microsoft Fabric einarbeiten möchten.
- **Studierende und angehende Datenexperten**, die lernen möchten, wie End-to-End-Datenprojekte in der Praxis effizient realisiert werden.
- **Entscheidungssträger**, die eine effiziente Datenstrategie für ihr Unternehmen entwickeln und umsetzen möchten.

E.2 Aufbau des Buchs

Dieses Buch gleicht einem Rundgang durch die Datenfabrik Microsoft Fabric, bei der Sie Schritt für Schritt alle Bereiche und Prozesse der All-in-One-Datenplattform kennenlernen und ein komplettes End-to-End-Datenprojekt nachvollziehen:

- Kapitel 1 beleuchtet Microsoft Fabric als All-in-One-Lösung und liefert einen Überblick über die Komponenten und deren Anwendungsgebiete.
- Kapitel 2 stellt ein fiktives Datenprojekt vor, das im Laufe des Buchs End-to-End mit Microsoft Fabric umgesetzt wird.
- Kapitel 3 zeigt, wie ein Microsoft-Fabric-Konto erstellt und die Arbeitsumgebung eingerichtet wird.
- Kapitel 4 beleuchtet die Rohdaten des Datenprojekt-Beispiels und untersucht deren Datenqualität.

- Kapitel 5 stellt Konzepte und Architekturen zur effizienten Datenspeicherung vor und erläutert deren Umsetzung in Microsoft Fabric.
- Kapitel 6 beschäftigt sich mit der Implementierung effizienter Datenpipelines in Microsoft Fabric zur Integration, Bereinigung und Transformation von Daten.
- Kapitel 7 zeigt, wie Machine Learning-Modelle in Microsoft Fabric entwickelt werden.
- Kapitel 8 erläutert die Datenanalyse- und -visualisierung in Microsoft Fabric.
- Kapitel 9 beschäftigt sich mit der Überwachung von Daten und Prozessen in Microsoft Fabric.
- Das Buch schließt mit einer Zusammenfassung über die gelernten Inhalte.

E.3 Konventionen

In diesem Buch werden folgende Formatierungen verwendet:

- *Fachbegriffe*
- **Wichtige Begriffe**
- SCHALTFLÄCHEN und BUTTONS
- Namen von Dateien, Ordnern und Fabric-Elementen

Nachfolgend werden die Bedeutungen der verschiedenen Kästen erläutert.

DATENFABRIK

Dieser Kasten erzählt die **symbolische Geschichte** über die Datenfabrik.

HINWEIS

Dieser Kasten vermittelt wichtige **Hinweise**.

TIPP

Dieser Kasten beinhaltet wichtige **Tipps**.

E.4 Downloads

Auf der folgenden Seite finden Sie Begleitmaterialien zu dem Datenprojekt-Beispiel:

<https://www.deeplearningnerds.com/sample-data-university-dataset/>

Über die Autoren

Fabian Hanik und Manuel Hanik sind Gründer von **Deep Learning Nerds**. Sie betreiben unter diesem Namen eine der größten Instagram-Seiten zu Themen aus den Bereichen Data Engineering, Data Science und Data Analytics. Darüber hinaus stellen sie auf ihrer Lernplattform regelmäßig Tutorials zu diesen Themen bereit.

Beide haben an der HFT Stuttgart ihr Bachelor- und Masterstudium in Mathematik absolviert. Im Studium sind sie zum ersten Mal mit Datenthemen in Berührung gekommen und haben sich im Rahmen ihrer Master-Thesis intensiv mit dem Bereich Deep Learning beschäftigt. Während ihres Auslandssemesters an der *University of Aberdeen* konnten sie ihr Wissen rund um Machine Learning vertiefen und wertvolle Erfahrungen in interkultureller Zusammenarbeit sammeln.

Manuel Hanik hat mehrere Jahre Berufserfahrung in den Bereichen Data Science und Data Engineering gesammelt und insbesondere in der Finanz- und Automobilbranche verschiedene Datenprojekte umgesetzt.



Fabian Hanik hat mehrere Jahre Berufserfahrung in den Bereichen Business Intelligence und Data Analytics und insbesondere in der Finanzbranche und im E-Commerce verschiedene Datenprojekte umgesetzt.



Schon während ihres Mathematik-Studiums war es das Ziel der beiden, komplexe Themen aus der Datenwelt anschaulich und verständlich aufzubereiten. Diese Herangehensweise ist auch die Grundlage für dieses Buch. Sie erklären Microsoft Fabric auf spielerische und unterhaltsame Weise – unterstützt durch visuelle Elemente und symbolische Eselsbrücken, die Ihnen helfen, die Themen besser zu verstehen und zu verinnerlichen.

Grundlagen

Als Einstieg werden in diesem Kapitel die Grundlagen und Funktionen von **Microsoft Fabric** erläutert. Zunächst werden traditionelle Dateninfrastrukturen vorgestellt und deren Herausforderungen beleuchtet, bevor der innovative Ansatz von Microsoft Fabric als All-in-One-Lösung aufgezeigt wird. Darüber hinaus erhalten Sie einen Überblick über die verschiedenen Komponenten von Microsoft Fabric und deren Anwendungsgebiete.



In diesem Kapitel:

- Ansatz traditioneller Dateninfrastrukturen
- Themengebiete im Datenumfeld
- Microsoft Fabric als All-in-One-Datenplattform
- Architektur und Komponenten von Microsoft Fabric



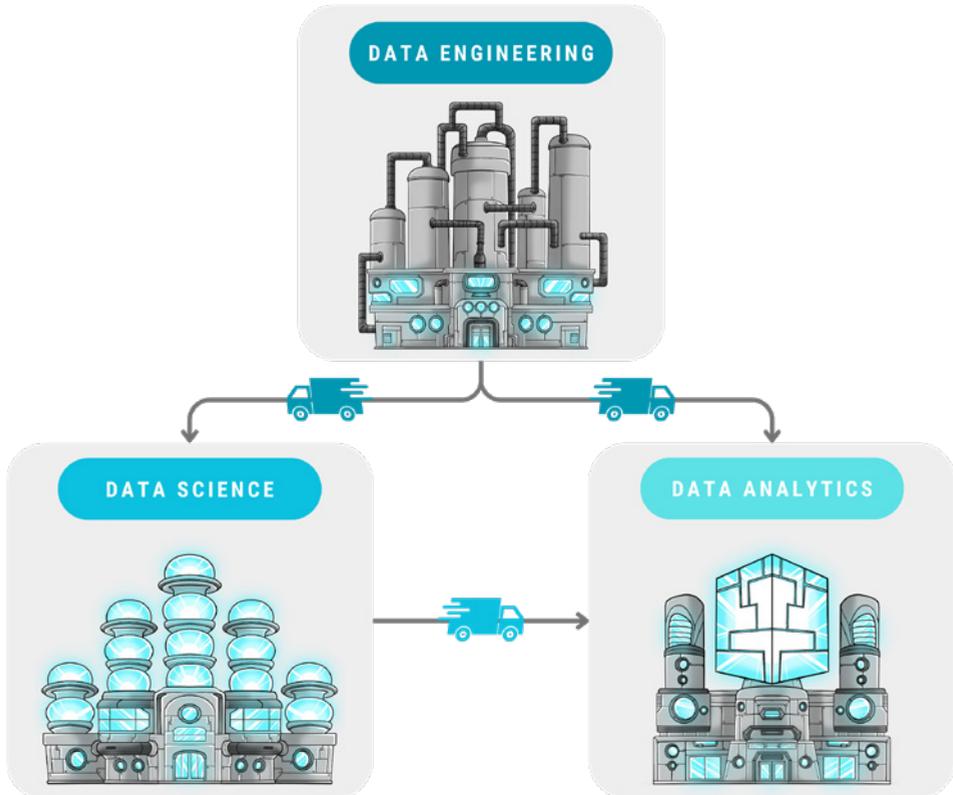
1.1 Traditionelle Dateninfrastruktur

Lassen Sie uns zunächst einen Blick auf traditionelle Dateninfrastrukturen werfen und wie diese in der Praxis bisher typischerweise eingesetzt werden. Dazu werden im Folgenden die wesentlichen Themengebiete im Datenumfeld erläutert: **Data Engineering**, **Data Science** und **Data Analytics**.

DATENFABRIK

In der symbolischen Geschichte, die Sie beim Lesen dieses Buchs begleitet, werden Datenprozesse in **Fabriken** durchgeführt. Dabei existieren unterschiedliche Fabriken, die sich in ihren Spezialisierungen unterscheiden, beispielsweise für die Themenbereiche Data Engineering, Data Science und Data Analytics. Diese spezialisierten Fabriken sind jeweils auf die spezifischen Anforderungen ihres Themengebiets ausgelegt und unterscheiden sich in ihrer Architektur und ihren

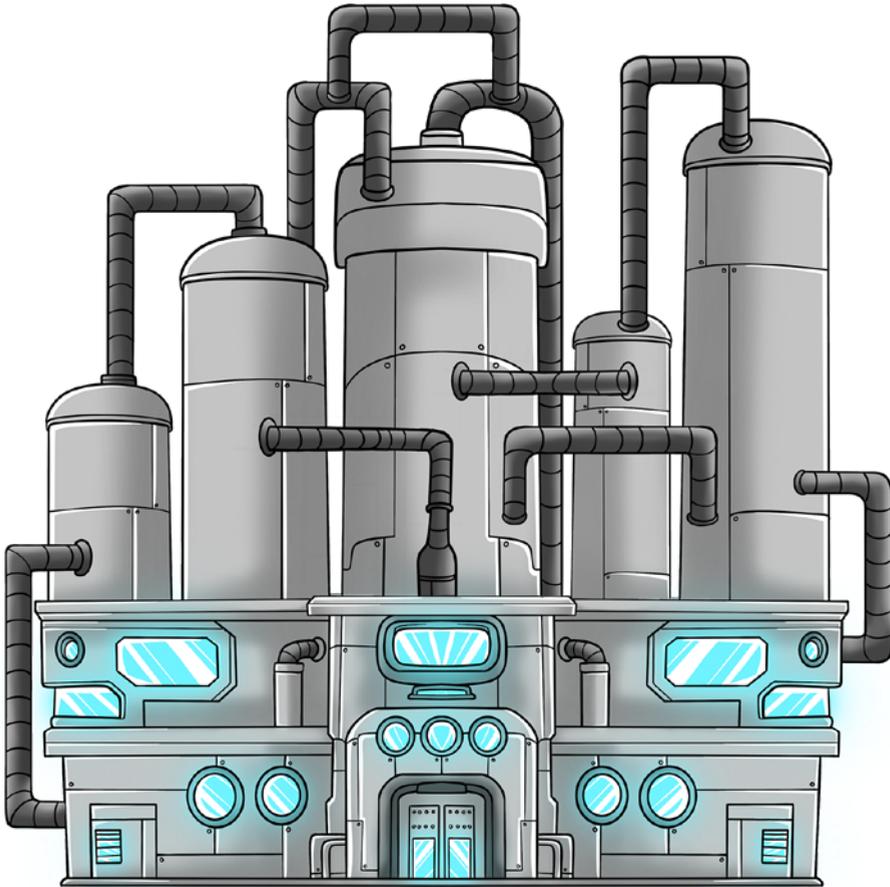
Prozessen. Zur Erstellung von Datenprodukten ist eine enge Zusammenarbeit zwischen den Fabriken erforderlich, da Daten zwischen den einzelnen Fabriken transportiert und ausgetauscht werden.



1.1.1 Data Engineering

DATENFABRIK

Die »Data Engineering«-Fabrik besteht aus speziellen Bereichen, in denen Rohdaten gesammelt, bereinigt und transformiert werden. Die verschiedenen Gebäude der Fabrik werden zur Datenspeicherung und -verarbeitung genutzt und sind durch Pipelines miteinander verbunden. Zunächst werden Daten in ihrer Rohform von Kunden an die Fabrik geliefert. Anschließend werden diese Daten gesammelt, zusammengeführt und bereinigt. Am Ende der Prozesskette liegen die Daten in einem nutzbaren und bereinigten Zustand vor und werden an die »Data Science«-Fabrik und die »Data Analytics«-Fabrik geliefert.



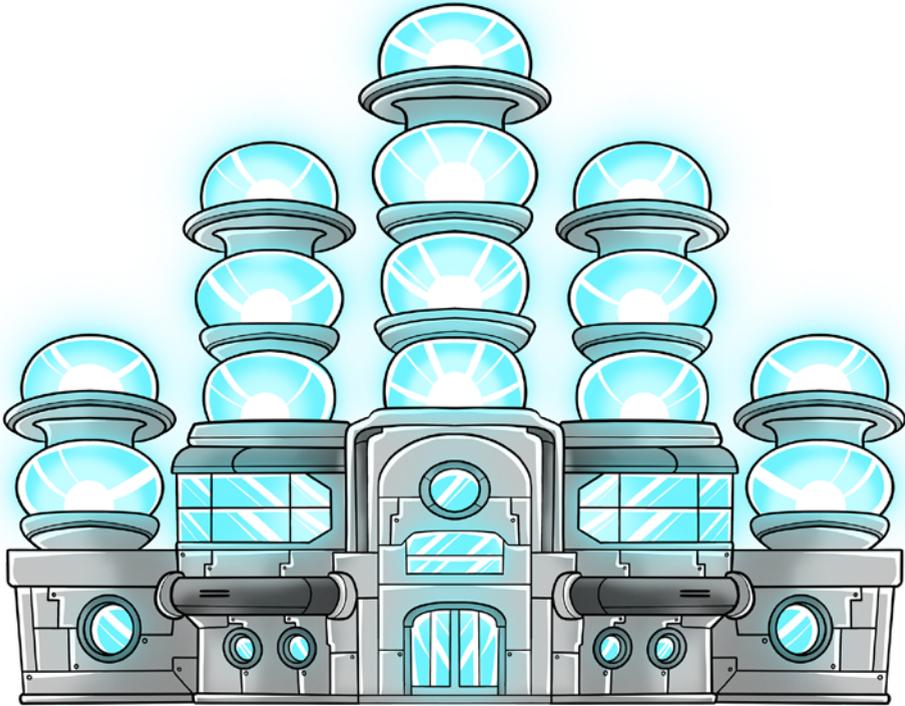
Data Engineering ist ein Bereich, der sich mit der Sammlung, Speicherung und Verarbeitung von Daten beschäftigt. Dabei wird die Datenqualität mittels Datenintegrationen, -bereinigungen und -transformationen sichergestellt, sodass die Daten für nachgelagerte Prozesse nutzbar gemacht werden. Die Fachleute in diesem Bereich werden als *Data Engineers* bezeichnet.

1.1.2 Data Science

DATENFABRIK

Die »**Data Science**«-Fabrik umfasst spezielle Bereiche, in denen KI-Experimente mit Daten durchgeführt werden. Sie besteht aus mehreren leuchtenden und hochmodernen Türmen, in denen die KI-Experimente stattfinden. Die Fabrik kann nur Daten verarbeiten, die bereits in der notwendigen Datenqualität vorliegen. Die Daten werden daher in einem bereinigten Zustand von der »Data Engineering«-Fabrik geliefert. Anschließend werden für verschiedenste Anwen-

dungsfälle Machine-Learning-Modelle entwickelt. Die entwickelten KI-Modelle werden je nach Anforderung entweder direkt an den Kunden übermittelt oder zur Weiterverarbeitung an die »Data Analytics«-Fabrik geliefert.



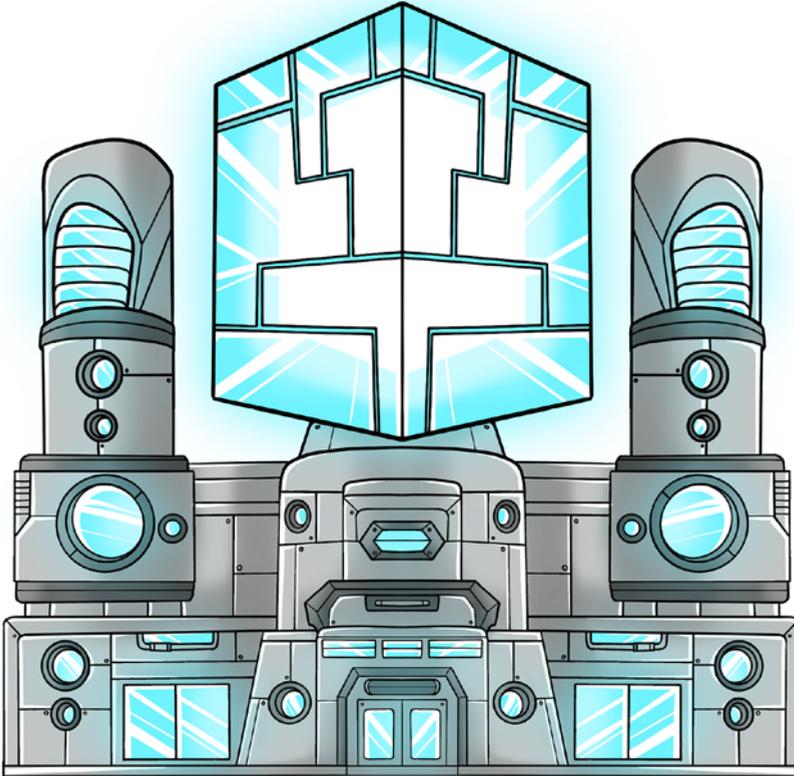
Data Science ist ein Bereich, der sich mit der Anwendung statistischer Methoden zur Erkennung von Mustern und Zusammenhängen in den Daten befasst. Dabei werden Machine-Learning-Modelle entwickelt und trainiert, um datenbasierte Prognosen für die Zukunft zu treffen. Die Fachleute in diesem Bereich werden als *Data Scientists* bezeichnet.

1.1.3 Data Analytics

DATENFABRIK

Die »Data Analytics«-Fabrik besteht aus speziellen Bereichen, in denen Daten modelliert und mit interaktiven Dashboards veranschaulicht werden. Sie besteht aus einem großen Würfel und mehreren Türmen. Die Datenmodellierung erfolgt in dem Würfel und die Datenvisualisierung in den Türmen. Die Fabrik kann dabei nur Daten verarbeiten, die bereits in der notwendigen Datenqualität

vorliegen. Zunächst werden die Daten von der »Data Engineering«-Fabrik und der »Data Science«-Fabrik geliefert. Anschließend werden die Daten modelliert, analysiert und in interaktiven Dashboards visualisiert. Die entwickelten Dashboards werden schließlich an den Kunden übermittelt.



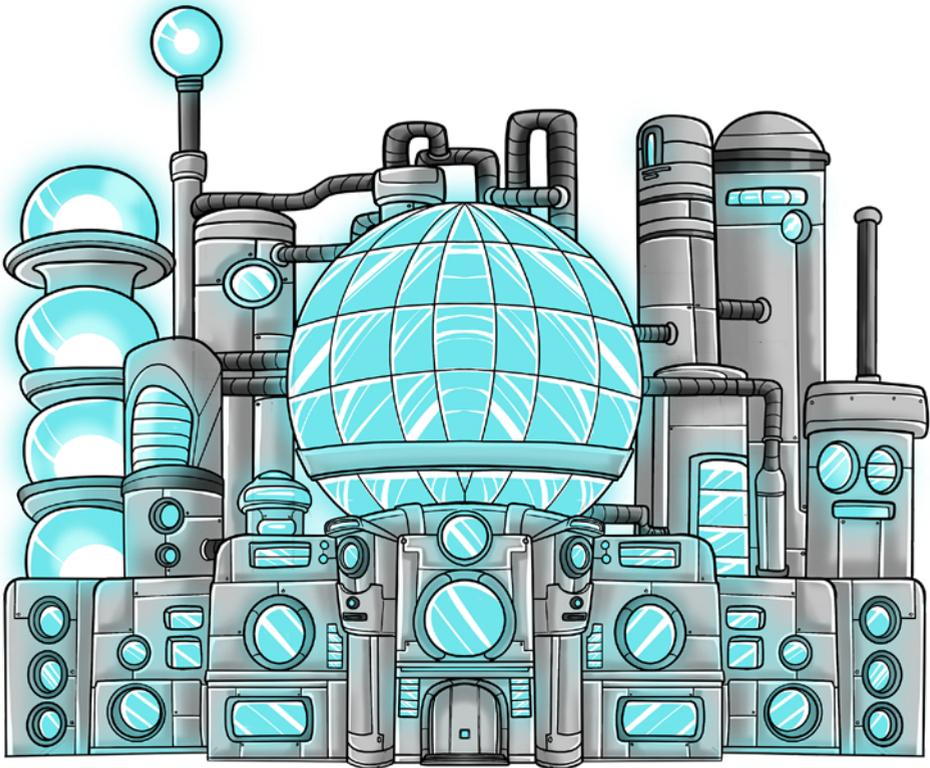
Data Analytics ist ein Bereich, der sich auf die Analyse und Visualisierung von Daten fokussiert. Dabei werden interaktive Visualisierungen, aussagekräftige Berichte und Dashboards erstellt, um wertvolle Erkenntnisse aus den Daten zu gewinnen und fundierte, datenbasierte Entscheidungen zu treffen. Die Fachleute in diesem Bereich werden als *Data Analysts* bezeichnet.

1.2 Fabric als All-in-One-Plattform

Nun haben Sie die zentralen Themenbereiche im Datenumfeld kennengelernt. Im Folgenden wird erläutert, wie Fabric als **All-in-One-Datenplattform** diese Bereiche kombiniert und dadurch eine umfassende Lösung bietet, die sich deutlich von traditionellen Dateninfrastrukturen abhebt.

DATENFABRIK

Es wurde eine neuartige Fabrik konstruiert, um die Prozesse der »Data Engineering«-Fabrik, der »Data Science«-Fabrik und der »Data Analytics«-Fabrik zu vereinen. Diese **Datenfabrik** ist ein großer und innovativer Komplex, in dem Datenprodukte vollumfänglich produziert werden und die komplette Wertschöpfungskette abgebildet wird. Dank der fortschrittlichen Architektur ist es daher nicht mehr erforderlich, mehrere separate Fabriken zu betreiben. Die verschiedenen Prozesse sind innerhalb der Fabrik auf unterschiedliche Gebäude verteilt. So gibt es für die Themenbereiche Data Engineering, Data Science und Data Analytics eigene Gebäude mit hochmodernen Maschinen. Zudem umfasst die Fabrik auch Gebäude, in denen Daten auf neue und innovative Weise gespeichert und verarbeitet werden. Die Rohdaten werden zunächst von den Kunden angeliefert und durchlaufen anschließend die erforderlichen Prozesse der Bereiche Data Engineering, Data Science und Data Analytics. Am Ende dieses ganzheitlichen Prozesses werden die fertigen Datenprodukte an den Kunden übermittelt.



In der Praxis tritt bei traditionellen Dateninfrastrukturen häufig das Problem auf, dass aufgrund der Vielzahl unterschiedlicher Technologien schnell eine hochgradig komplexe Tool-Landschaft entsteht. Dies führt dazu, dass viel Zeit und Ressourcen aufgebracht werden müssen, um die notwendige Infrastruktur aufzubauen, zu betreiben und zu warten. Zudem entstehen aufgrund unterschiedlicher Lizenzmodelle oft hohe Infrastrukturkosten. Die Zusammenarbeit zwischen den Datenexperten der unterschiedlichen Bereiche wird darüber hinaus deutlich erschwert, da die Integration und Verknüpfung der unterschiedlichen Technologien und Dienste oft mit erheblichen technischen Herausforderungen verbunden ist.

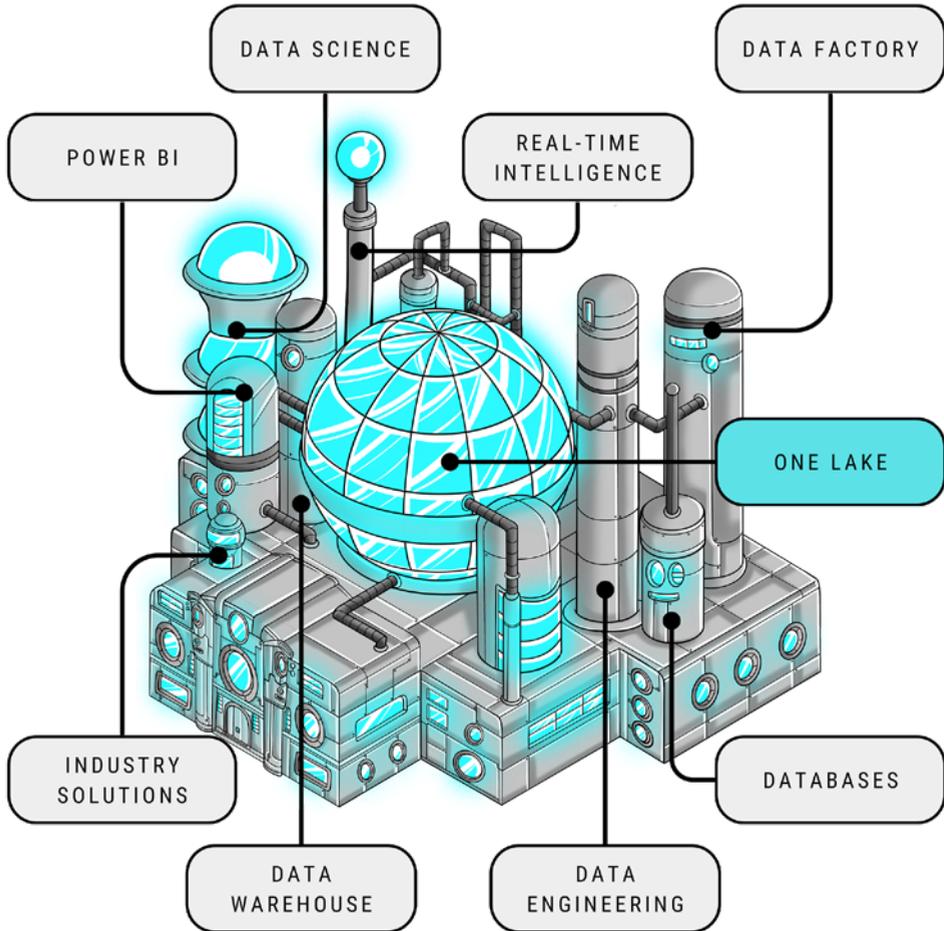
In Microsoft Fabric können Datenexperten – von Data Engineers über Data Analysts bis hin zu Data Scientists – auf einer gemeinsamen Plattform zusammenarbeiten. Dabei gibt es für jede Rolle einen speziellen Bereich, in dem die benötigten Dienste gebündelt zur Verfügung gestellt werden. Durch eine nahtlose Verknüpfung der Dienste sowie einen gemeinsamen Datenzugriff wird die Zusammenarbeit zwischen den verschiedenen Rollen beträchtlich vereinfacht und alle erforderlichen Aspekte eines Datenprojekts auf einer einzigen Plattform abgedeckt, was zu einer erheblichen Vereinfachung der Wertschöpfungskette führt.

1.3 Die Komponenten von Fabric

Im vorangegangenen Abschnitt haben Sie bereits die Vorteile von Fabric kennengelernt. Lassen Sie uns nun einen genaueren Blick auf die Architektur, den Aufbau sowie die verschiedenen **Komponenten** werfen.

DATENFABRIK

Die Datenfabrik haben Sie bereits von außen gesehen. Bevor Sie nun die Fabrik betreten, werfen Sie zunächst einen Blick auf den **Übersichtsplan**. Dieser veranschaulicht den Aufbau der Datenfabrik. Die Fabrik besteht aus mehreren Abteilungen, die auf unterschiedliche Gebäude verteilt sind. In jeder Abteilung wird ein bestimmter Prozessschritt durchgeführt. Sobald Sie die Fabrik betreten, erleben Sie hautnah, wie die einzelnen Abteilungen in der Praxis arbeiten.



Betrachten wir zunächst die Architektur und den Aufbau von Fabric. Zu den Fabric-Komponenten zählen sowohl bereits etablierte Microsoft-Dienste wie Power BI als auch neu entwickelte Dienste. All diese Dienste werden nun in einer einheitlichen Plattform zusammengeführt. Während einzelne Microsoft-Dienste bisher separat lizenziert werden mussten, können sie jetzt über eine gemeinsame Lizenz innerhalb der Plattform genutzt werden. Durch den Zusammenschluss dieser Dienste und ihrer nahtlosen Verknüpfung ermöglicht Fabric die Umsetzung von End-to-End-Datenprojekten auf einer einzigen Plattform. Jeder dieser Dienste deckt dabei einen spezifischen Prozessschritt innerhalb eines Datenprojekts ab.

Die nachfolgende Abbildung veranschaulicht alle Fabric-Komponenten.