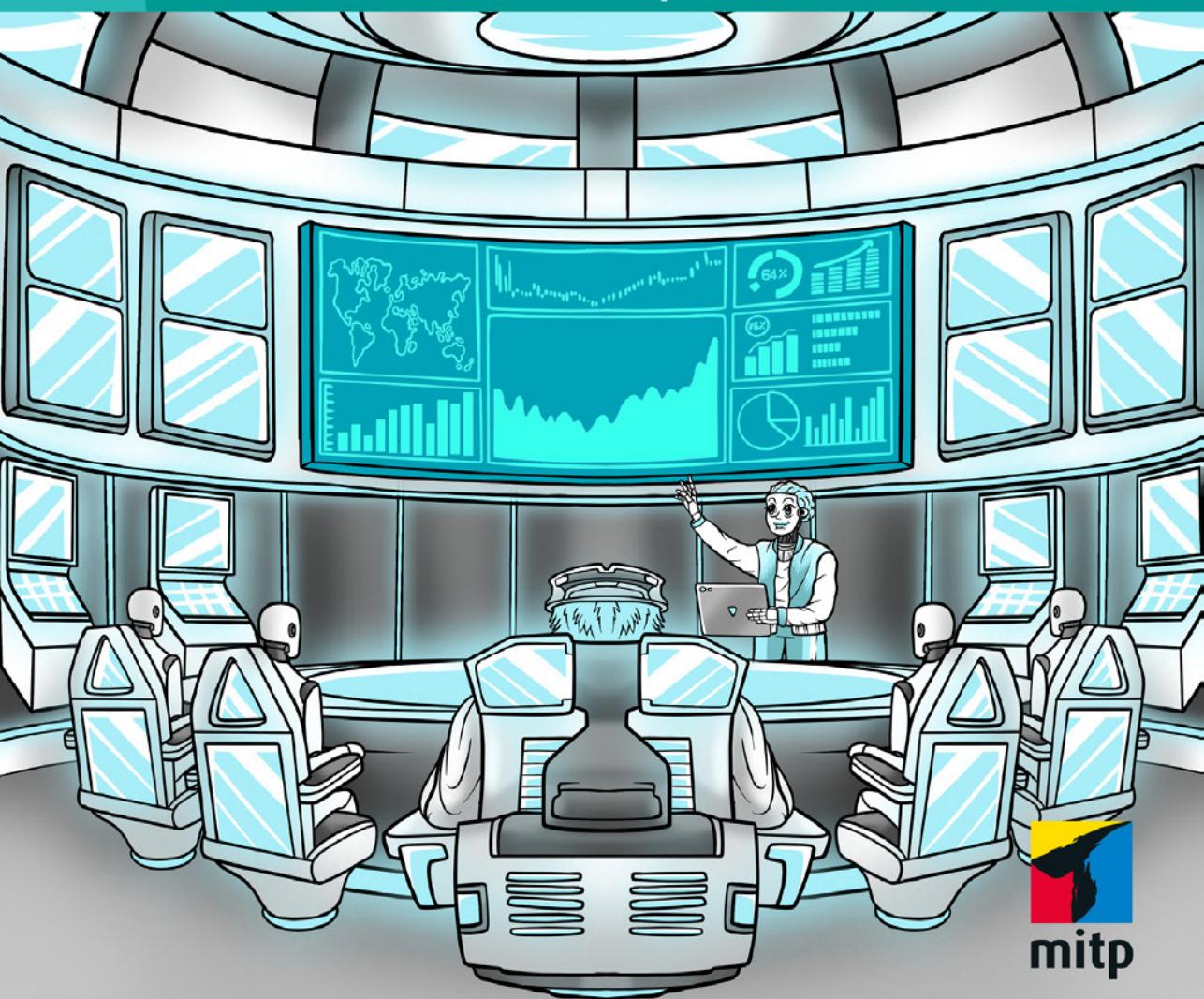


Manuel Hanik | Fabian Hanik

Microsoft Fabric

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



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

Inhaltsverzeichnis

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

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 Anwendungsbereiche.



In diesem Kapitel:

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

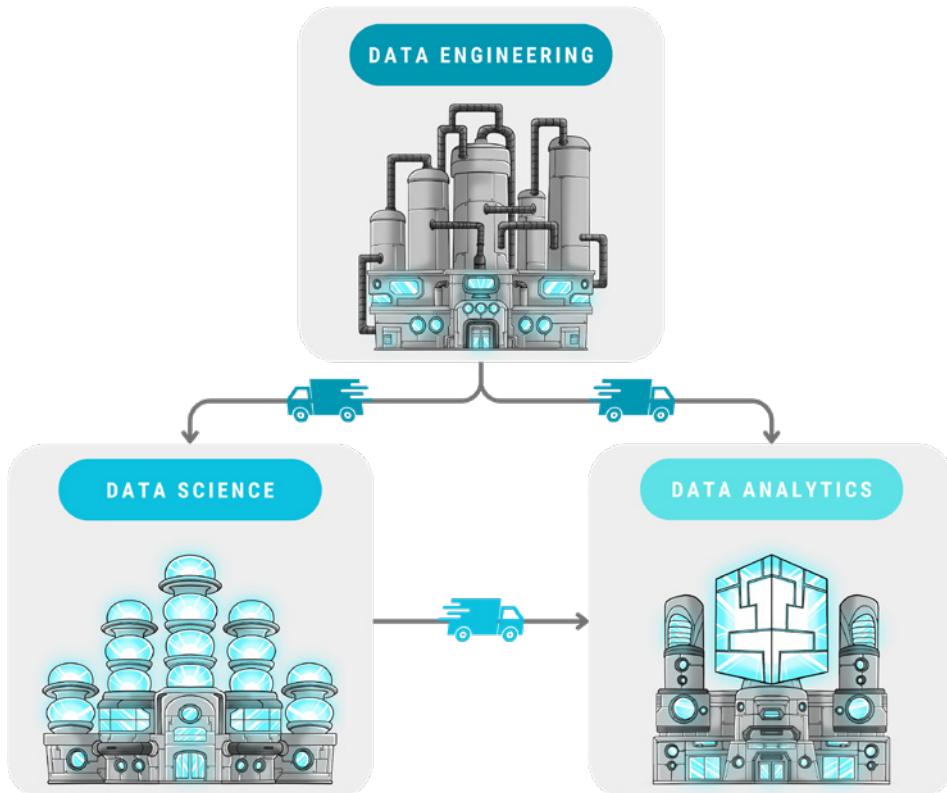
1.1 Traditionelle Dateninfrastruktur

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

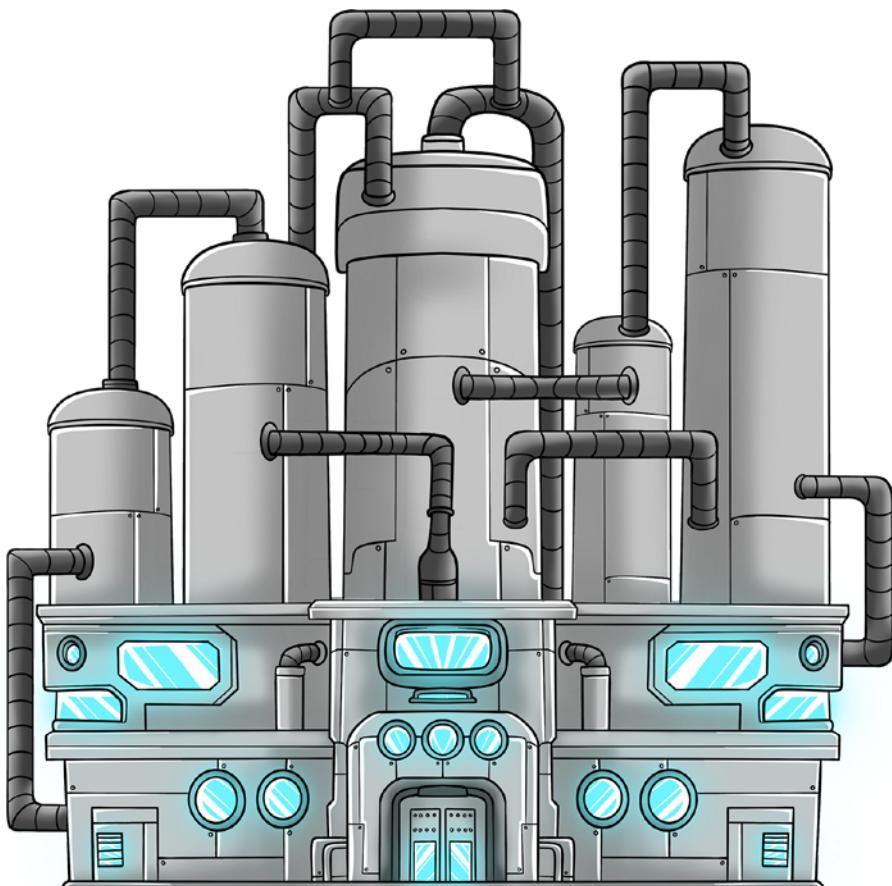
Prozessen. Zur Erstellung von Datenprodukten ist eine enge Zusammenarbeit zwischen den Fabriken erforderlich, da Daten zwischen den einzelnen Fabriken transportiert und ausgetauscht werden.



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.



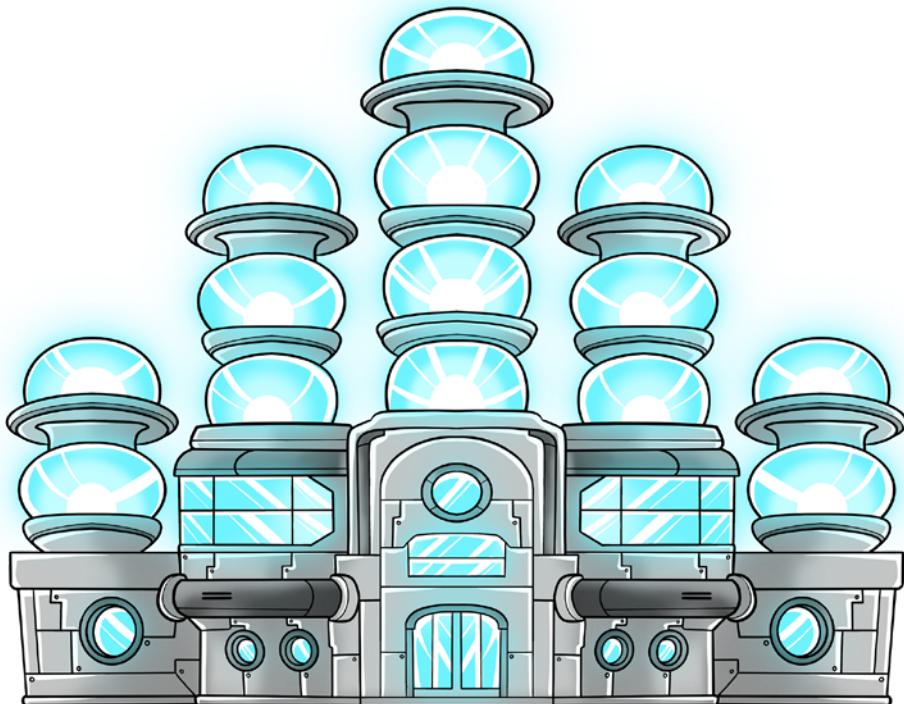
Data Engineering ist ein Bereich, der sich mit der Sammlung, Speicherung und Verarbeitung von Daten beschäftigt. Dabei wird die Datenqualität mittels Datenintegrationen, -bereinigungen und -transformationen sichergestellt, sodass die Daten für nachgelagerte Prozesse nutzbar gemacht werden. Die Fachleute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



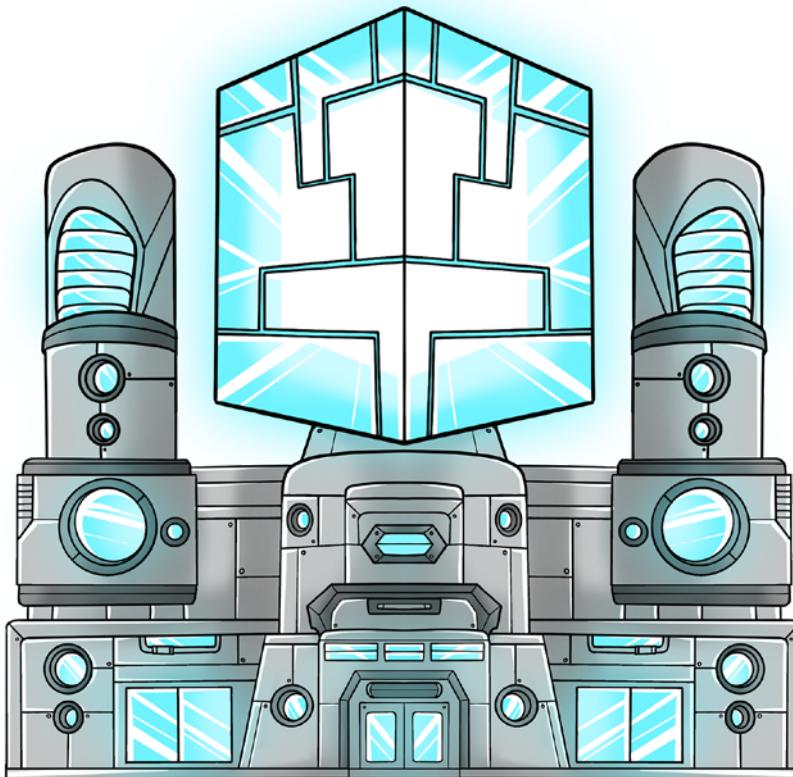
Data Science ist ein Bereich, der sich mit der Anwendung statistischer Methoden zur Erkennung von Mustern und Zusammenhängen in den Daten befasst. Dabei werden Machine-Learning-Modelle entwickelt und trainiert, um datenbasierte Prognosen für die Zukunft zu treffen. Die Fachleute in diesem Bereich werden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



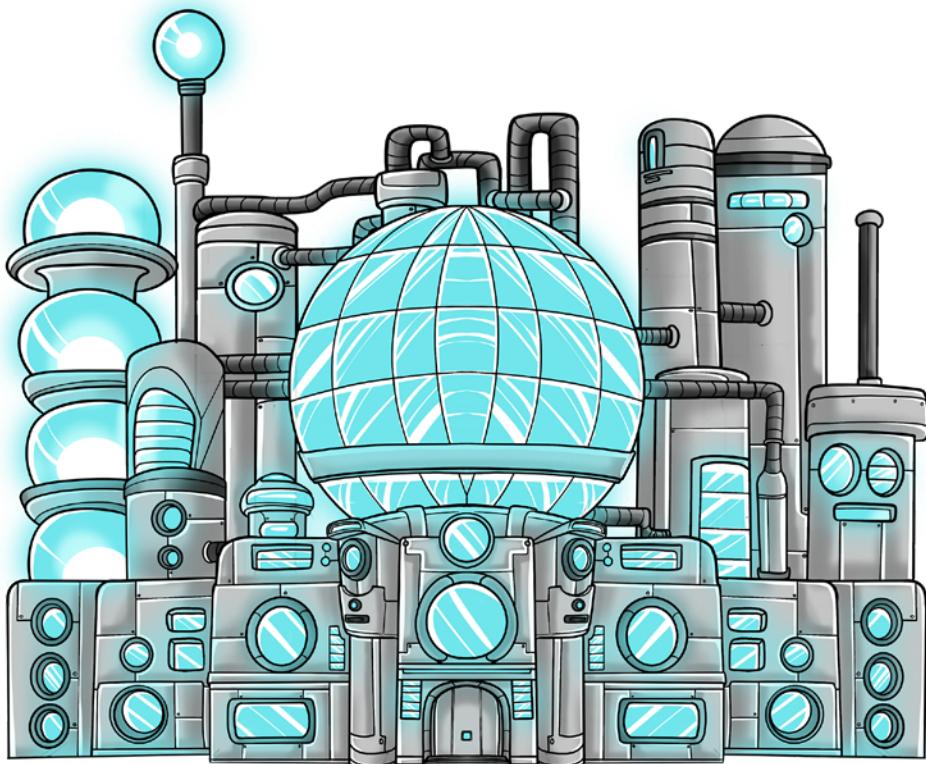
Data Analytics ist ein Bereich, der sich auf die Analyse und Visualisierung von Daten fokussiert. Dabei werden interaktive Visualisierungen, aussagekräftige Berichte und Dashboards erstellt, um wertvolle Erkenntnisse aus den Daten zu gewinnen und fundierte, datenbasierte Entscheidungen zu treffen. Die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

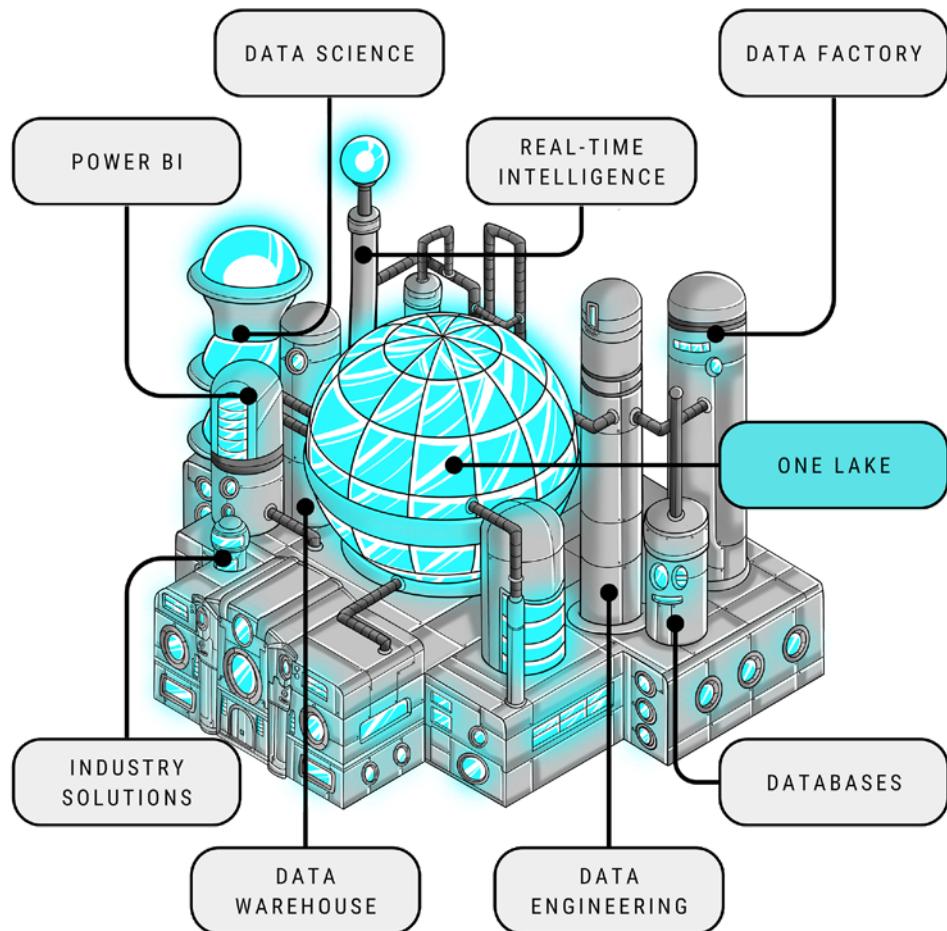
In Microsoft Fabric können Datenexperten – von Data Engineers über Data Analysts bis hin zu Data Scientists – auf einer gemeinsamen Plattform zusammenarbeiten. Dabei gibt es für jede Rolle einen speziellen Bereich, in dem die benötigten Dienste gebündelt zur Verfügung gestellt werden. Durch eine nahtlose 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.

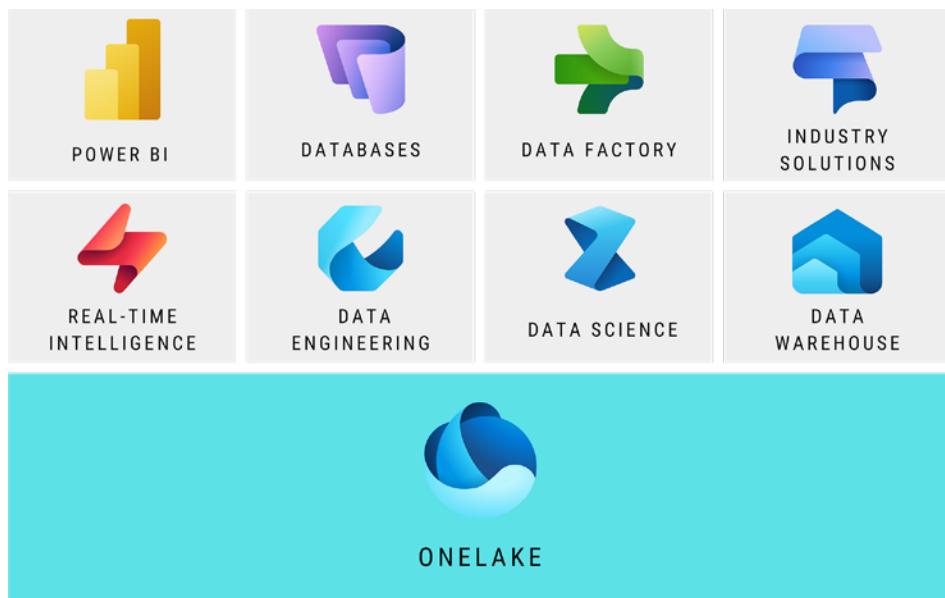
DATENFABRIK

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.



Im Folgenden werden die einzelnen Fabric-Komponenten sowie deren Elemente näher beleuchtet. Dabei werden ausschließlich diejenigen Komponenten und Elemente erläutert, die für das Datenprojekt-Beispiel von Relevanz sind. Dies ermöglicht eine fokussierte Betrachtung der Dienste und Funktionalitäten, die für die Umsetzung des Datenprojekt-Beispiels benötigt werden.

HINWEIS

Fabric umfasst insgesamt neun Komponenten. Die Komponenten Databases und Industry Solutions werden jedoch im Rahmen dieses Buchs nicht behandelt, da sie für die Umsetzung des Datenprojekt-Beispiels nicht benötigt werden. Die Komponente Databases ermöglicht die Erstellung und Verwaltung von SQL-Datenbanken innerhalb von Fabric. Industry Solutions bietet branchenspezifische vorkonfigurierte Analyse- und Datenlösungen für verschiedene Industriezweige.

1.3.1 OneLake



Das Herzstück von Fabric bildet der *OneLake*. Er dient als zentraler Speicherort für alle Arten von Daten, egal ob strukturiert oder unstrukturiert. Alle Daten werden mit der Technologie *Delta Lake* gespeichert, die im weiteren Verlauf des Buchs näher erläutert wird. Aufgrund des einheitlichen Speicherformats wird eine konsistente Datenspeicherung für die gesamte Plattform gewährleistet. Darauf aufbauend lassen sich mit den weiteren Komponenten Datenintegrationen, Datentransformationen, Machine-Learning-Modelle, Berichte oder Dashboards implementieren.

1.3.2 Data Warehouse



Data Warehouse ist eine Lösung zum Erstellen und Verwalten von Data Warehouses. Darunter fällt auch das Laden, Transformieren und Abfragen von Daten mittels *Transact SQL*, kurz *T-SQL*. Dabei handelt es sich um eine von Microsoft entwickelte Erweiterung von SQL, die zusätzliche Funktionen zur Abfrage und Verarbeitung von Daten bietet.

Symbol	Name	Beschreibung
	Warehouse	In einem <i>Warehouse</i> können strukturierte Daten gespeichert, verwaltet und abgefragt werden.

1.3.3 Data Engineering



Data Engineering ist eine Lösung zum Speichern, Integrieren, Transformieren und Laden von Daten. So lässt sich die Dateninfrastruktur aufbauen und die Datenverarbeitung implementieren.

Symbol	Name	Beschreibung
	Lakehouse	In einem <i>Lakehouse</i> können strukturierte und unstrukturierte Daten gespeichert, verwaltet und ausgewertet werden.

Symbol	Name	Beschreibung
	Notizbuch	In einem <i>Notizbuch</i> kann Python-Code interaktiv geschrieben und ausgeführt werden, beispielsweise um Daten zu bereinigen und zu transformieren.
	Datenpipeline	Eine <i>Datenpipeline</i> verkettet mehrere Aktivitäten zu einem Workflow. Mithilfe von Zeitplänen und Triggern lassen sich Pipelines automatisch ausführen.

1.3.4 Data Factory



Data Factory basiert auf dem Dienst Azure Data Factory und ermöglicht das Integrieren, Extrahieren, Transformieren und Laden von Daten aus unterschiedlichen Datenquellen. Über eine Benutzeroberfläche lassen sich Pipelines erstellen, um Workflows zu orchestrieren und automatisch auszuführen.

Symbol	Name	Beschreibung
	Dataflow Gen2	Ein <i>Dataflow Gen2</i> bietet eine interaktive Benutzeroberfläche zur Erstellung von Datenoperationen. Mit Dataflows können beispielsweise Datenbereinigungen und Datentransformationen durchgeführt werden.
	Datenpipeline	Eine <i>Datenpipeline</i> verkettet mehrere Aktivitäten zu einem Workflow. Mithilfe von Zeitplänen und Triggern lassen sich Pipelines automatisch ausführen.

HINWEIS

Sicher ist Ihnen aufgefallen, dass das Datenpipeline-Element sowohl in **Data Factory** als auch in **Data Engineering** vorhanden ist. Der Grund dafür ist, dass dieses Element thematisch beiden Bereichen zugeordnet werden kann, jedoch mit unterschiedlichen Schwerpunkten. In Fabric kann grundsätzlich ein und dasselbe Element in mehreren Komponenten Anwendung finden. Unter anderem gilt dies auch für das Notizbuch-Element.

1.3.5 Data Science



Data Science ist eine Lösung zum Erstellen, Trainieren, Testen und Bereitstellen von Machine-Learning-Modellen.

Symbol	Name	Beschreibung
	ML-Modell	Ein <i>ML-Modell</i> enthält eine oder mehrere Versionen eines trainierten Machine-Learning-Modells.
	Experiment	Ein <i>Experiment</i> ist eine interaktive Benutzeroberfläche, mit der das Training von ML-Modellen protokolliert und nachverfolgt werden kann.
	Notizbuch	In einem <i>Notizbuch</i> kann Python-Code interaktiv geschrieben und ausgeführt werden, beispielsweise um Machine-Learning-Modelle zu trainieren.

1.3.6 Power BI



Power BI ist eine benutzerfreundliche Business-Intelligence-Lösung zur Modellierung, Visualisierung und Analyse von Daten. Mit interaktiven Berichten und Dashboards wird eine anschauliche Darstellung der Daten ermöglicht.

Symbol	Name	Beschreibung
	Semantikmodell	Ein <i>Semantikmodell</i> oder <i>semantisches Modell</i> stellt ein Datenmodell dar, in dem Berechnungen, Aggregationen und Tabellenbeziehungen abgebildet werden.
	Bericht	Ein <i>Bericht</i> ist eine Anordnung von Visualisierungen auf einer oder mehreren Seiten.

Symbol	Name	Beschreibung
	Dashboard	Ein <i>Dashboard</i> ist eine Sammlung von Visualisierungen, die aus unterschiedlichen Berichten stammen können.
	Scorecard	Eine <i>Scorecard</i> bietet die Möglichkeit, Metriken zu erstellen, um Ziele zu definieren, darzustellen und nachzuverfolgen.

1.3.7 Real-Time Intelligence



Real-Time Intelligence ist eine Lösung zur Speicherung und Analyse von Echtzeitdaten sowie zur Überwachung von Prozessen und Datenänderungen.

Symbol	Name	Beschreibung
	Aktivator	Der <i>Aktivator</i> dient zur Überwachung von Daten und Prozessen. Hierbei lassen sich Aktionen definieren, die automatisch durch Datenänderungen oder vordefinierte Bedingungen ausgelöst werden.

Stichwortverzeichnis

A

Abfrage 185
Abhängigkeit
 hinzufügen 156
Abnehmer 82
ACID-Transaktion 82
ADLS Gen2 100
Aggregation 195
Aktivator 29, 217
Ansicht 138
App 207
 erstellen 208
 veröffentlichen 211
Arbeitsbereich 54
 erstellen 54
Arbeitsbereich-Ordner 56, 92, 159, 182, 212, 218
Architektur 78, 164
Ausführung 163
 automatische 157
Authentifizierungsart 102
Azure 96, 100
 Anmeldung 100
 Kontoschlüssel 103

B

Bereinigung 122
Bericht 28, 191, 192, 196, 197, 200, 215
 Power BI 196
Bronze-Schicht 85, 105

C

Column-Level Security 79
Copy-Aktivität 117
CSV 33, 61, 107, 117

D
Dashboard 21, 29, 191, 192, 201, 203
 erstellen 201
Data Analyst 21, 36, 59, 183
Data Analytics 21, 39, 183
Data Architect 35, 77
Data Engineer 19, 36, 93
Data Engineering 19, 26, 38, 54, 86, 93, 122
Data Factory 27, 106
Dataflow 27, 106, 107, 115, 116
 CSV 107
 erstellen 107
 Excel 115
 JSON 116
 XML 116
Dataflow-Aktivität 151
DataFrame 126, 168
 reduzieren 169
Data Governance 80
Data Lake 79
Data Lakehouse 80
Data Manipulation Language 83
Data Science 20, 28, 38, 161, 165
Data-Science-Anwendungsfall 163
Data-Science-Workflow 163
Data Scientist 20, 36, 161
Data Warehouse 26, 79, 89
DATE 128, 130

- Dateiformat
 CSV 61
 Excel 65
 JSON 72
 XML 69
Datenarchitektur 78, 81, 94, 162, 184, 214
Datenbereinigung 125, 126, 128, 130, 132, 133
Datenexploration 38, 59
Datenfabrik 17, 22
Datenfluss 95
Dateninfrastruktur
 traditionelle 17
Datenintegrität 82
Datenpipeline 27, 106, 117
 erstellen 150
 planen 157
Datenprojekt
 Phasen 38
Datenprojekt-Beispiel 32
 Ordner 56
Datenqualitätsproblem 64
Datenquelle 95
 externe 96, 100
 hochgeladene 96
 interne 96
 Konfiguration 118
Datenspeicherung 38, 77, 81
Datenstruktur 91
Datentransformation 136, 139, 141, 144, 147
Datentyp 63, 68, 71, 75
 definieren 127
Datenversionierung 82
Datenziel
 konfigurieren 119
DAX 195
dbo 89
Delta Lake 26
Delta-Tabelle 82, 89
 Schema 89
DevOps Engineer 36, 213
Dimensionstabelle 137
Direct Lake 197
DML-Operation 83
Download
 Datenprojekt-Beispiel 34, 61
Duplikate 102
- E**
- Entscheidungsbaum 173
Excel 33, 65, 115
Experiment 28, 163, 165, 176
 erstellen 166
 setzen 174
- F**
- Fabric
 einrichten 41
Faktentabelle 136
Fehler 215
Filter 199
Formelsprache 112
- G**
- Gold-Schicht 85, 136
- I**
- ID 65
INNER JOIN 168
- J**
- JSON 33, 68, 72, 116, 128
 Array 73
 Objekt 73
- K**
- Klassifikationsproblem
 binäres 163
Konfiguration 52
Kontoschlüssel 103

L

- Label 163
- Lakehouse 26, 86
 - erstellen 86
 - Explorer 88
 - Ordner 88
 - verknüpfen mit Notizbuch 124
- Landing-Zone 84, 86, 94
 - Datenfluss 95
- LEFT JOIN 147

M

- Machine Learning 20, 163
- Measure 195, 216
 - visualisieren 198
- Medaillon-Architektur 81, 84
 - Implementierung 85
- Merkmal 163
 - kategoriales 170
 - kategoriales umwandeln 171
 - numerisches 170, 171
- Metadaten 85
- MLflow 165
- ML-Modell 28, 161, 163, 165, 177
 - erstellen 173
 - trainieren 165
- Modell
 - semantisches 28, 191, 193
 - semantisches erstellen 193
- Modell-Anwendung 179
- Modell-Training 165
- Monitoring 39, 82, 213

N

- Notizbuch 27, 28, 123
 - Abfrage 190
 - ausführen 127
 - erstellen 123
- Notizbuch-Aktivität 155
- Null-Wert 170

O

- One-Hot-Encoding 171
- OneLake 26
- Orchestrierung 150
- Ordner 92
 - erstellen 56

P

- Pandas DataFrame 169
- Parquet 82
- Pipeline-Aktivität 153
- Power BI 24, 28, 183, 191, 201
- Power Query M 112
- Prognose 161
- Projektleiter 35, 41
- Projektplan 37
- PySpark 125
 - leere Werte 126
- Python 61, 167

Q

- Quellen 81

R

- Real-Time Intelligence 29
- Rohdaten 32, 94, 96
- Rollback 82
- Row-Level Security 79

S

- Schema 89
 - importieren 120
- Schema Enforcement 82
- Schema Evolution 82
- Schema-on-Read-Prinzip 79
- Schema-on-Write-Prinzip 79
- Schlüssel-Wert-Paar 72
- Scikit-learn 169
 - Autologging 174
- Scikit-learn-Pipeline 170
- Scorecard 29, 191, 203
 - erstellen 204

Semantisches Modell
 Siehe Modell, semantisches
Setup 38, 41
Silber-Schicht 85, 122
Spaltenzuordnung 113
Spark-Connector 168
SQL-Abfrage 136, 138, 186
Standardisierung 171
Startseite 51
Status 215
 Anzeige 206
Sternschema 85, 136, 137

T

Tabelle
 dynamische 138
 virtuelle 138
Tabellenschema 82
Tabellenverknüpfung 147
Teamleiter 35
Testdatensatz 174
Testzugang 43, 49
 Aktivierung 49
Time Travel 82
Trainingsdatensatz 174
T-SQL 26, 138, 183, 187

U

Überwachungshub 215

V

Verknüpfung 100
 erstellen 102
Verwaltungsportal 52
Virtuelle Tabelle 138
Visualisierung 21, 191
Visuelle Abfrage 187
Vorhersagen 165, 179

W

Warehouse 26, 89
 Aufbau 91
 erstellen 89
 Explorer 90
 Ordner 91
 Zugriffsrechte 92
Warnung 215
 konfigurieren 216
 testen 217

X

XML 33, 69, 116

Z

Zeitplan 157
Zielvariable 163
Zugriffsberechtigung 54
Zugriffsrecht 92