

QUNIS

6. QUNIS DAY 31. AUGUST 2018
OVB MEDIENFORUM, ROSENHEIM

DIGITAL

POWER

LIVE

Agenda



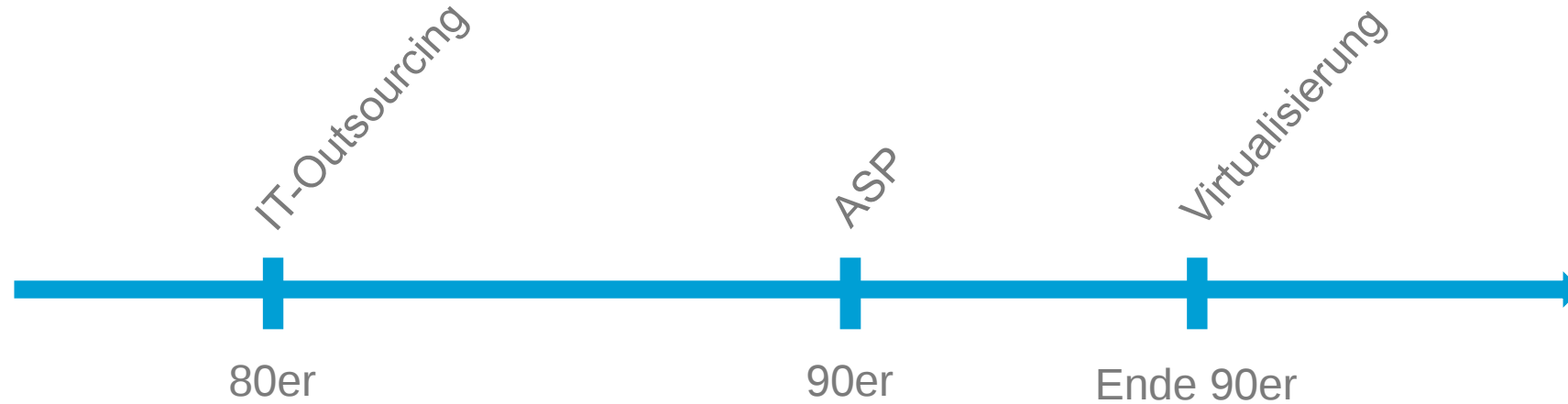
DIGITAL
POWER
LIVE

10.00	DIGITAL POWER LIVE Begrüßung durch Hermann Hebben und Steffen Vierkorn	
10.30	Cleveres Datenmanagement im Zeitalter von Digitalisierungsinitiativen: Agile Transformation von klassischen BI-Systemen zu Data-Lake-Konzepten! Steffen Vierkorn	
11.30	TRACK I BI und Data Warehouse Excellence	TRACK II Die analytische Evolution – es geht weiter!
11.30	Smart die Cloud nutzen – Ausgewählte Cloud-Services für den Einsatz in BI-Szenarien. Andy Löwen	AI als Enabler für digitale Geschäftsmodelle – Internet of Things zum Anfassen! Nico Damm
12.00	MITTAGSPAUSE	
13.00	QUNIS Automation Engine – Standardisierung und Automatisierung in der Data-Warehouse-Entwicklung. Raphael Henneke	Die Evolution im Controlling – Vom BI Power User zum Data-Scientisten. Dr. Franziska Deutschmann
13.30	Agiles Projektmanagement – Live-Präsentation der QUNIS-Methodik und -Werkzeuge. Ralph Gattinger	Artificial Intelligence zum Anfassen – Live-Präsentation ausgewählter Anwendungsfälle! Ilona Tag
14.00	BI, Big Data und AI brauchen Regeln – Data Governance leicht und verständlich. Annefried Simoneit	Analytics in der Cloud – Zum Mitmachen und selbst Erleben! Nico Damm
14.30	KAFEEPAUSE	
15.00	Herausforderung Digitalisierung – Wie Viessmann die Anforderungen an das Datenmanagement mit einem Data Lake managt. Marc Saure, Kundenvortrag Viessmann	Advanced Analytics in Tax – Praxisbeispiele von AI im Steuerbereich. Vanessa Just, Kundenvortrag WTS AI
15.30	Microsoft Power BI in Action – Live-Präsentation. Patrick Eisner	Ein Blick in das QUNIS Innovation Lab – Einblicke in die Research-Arbeit der QUNIS. Steffen Vierkorn
16.30	Digitalisierung, die uns bewegt! Zusammenfassung des Tages und Ausblick von Hermann Hebben und Steffen Vierkorn	

Ansprüche und Fragestellungen

- **Kernkompetenzen und Know-how im eigenen Team / Synergien**
 - Welches Know-how ist vorhanden und welches wird benötigt → Skill-Matrix
 - Aus- und Weiterbildung
 - Ressourcen
- **Betrieb**
 - Intern, Provider oder Managed Service
- **Verfügbarkeit (HA)**
- **Flexibilität und Skalierbarkeit**
 - Wie flexibel können Komponenten getauscht oder abgelöst werden
 - Wie skalierbar ist die Infrastruktur
- **Kosten**
- **Zugang für den Anwender von intern und extern**
- **Methodik und Regelwerk für die Entwicklung des Data Warehouse**
- **Hardware-Sizing**
 - Virtuell oder physikalisch
 - SMP oder MPP

Vorreiter zur Cloud



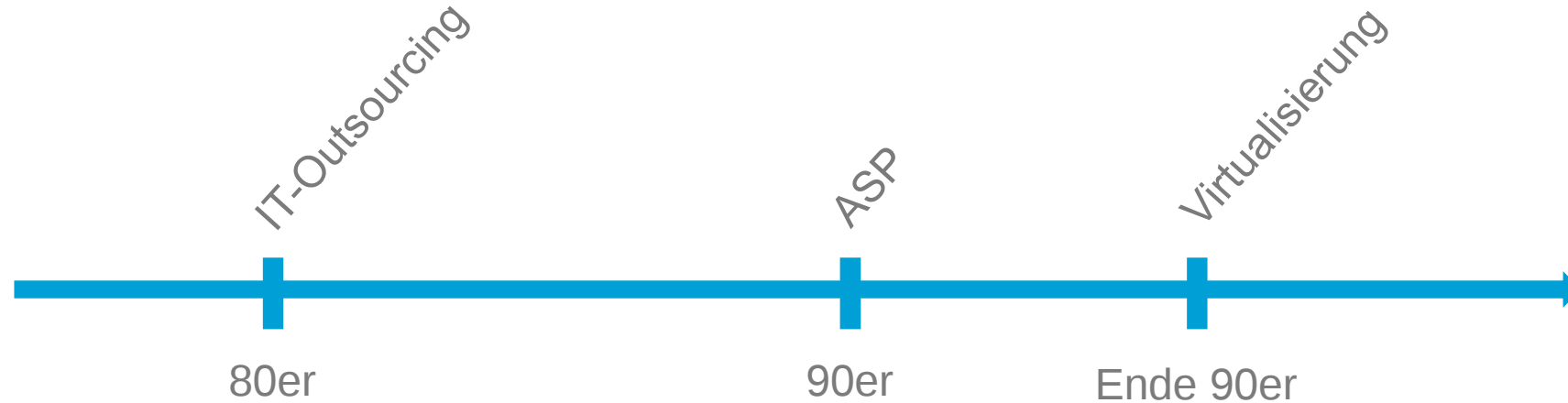
Vorteile

- Kostensenkung
- Konzentration auf Kernkompetenzen
- Steigerung der Qualität
- Zeitersparnis / freie Ressourcen
- Zugriff auf Experten Know-How

Nachteile

- Abhängigkeit
- erhöhte Kommunikation
- Kontrollabgabe
- mangelnde Fortbildung der eigene Mitarbeiter
- Einblicke in sensible Daten

Vorreiter zur Cloud



Vorteile

- Weniger physikalische Systeme: Kosten- und Energieeinsparung
- Geringerer Platzbedarf
- Schnellere Server-Provisionierung
- Lebenszeit alter Applikationen verlängern
- VM-Management (Clone, Snapshot, Live-Migration)
- Erhöhte Verfügbarkeit für alle VM

Nachteile

- benötigt leistungsstarke Server
- Lizensierung
- Programme, die nicht auf virtuellen Maschinen laufen
- höhere Latenzzeiten bei unausgeglichenen Ressourcen
- Ausfall aller VM wenn Host nicht verfügbar ist
- Hardware-Unabhängigkeit

Chance und Risiko Cloud

Cloud ist die Bereitstellung von Infrastruktur (Rechenleistung, Speicherkapazität, Software) als Dienstleistung über das Internet. Die verfügbaren Dienstleistungen umfassen neben Infrastruktur auch Plattformen und Software.

Servicemodelle:

- Infrastructure **as a S**ervice (z.B. Virtuelle Maschine)
- **P**latform **as a S**ervice (z.B. Datenbank)
- **S**oftware **as a S**ervice (z.B. Email-Programm)

Vorteile

zeitnahe Skalierung

gesteigerte Agilität der IT

geringere Administration

spart Kosten & reduziert Investitionen

Spielraum für Innovationen

bleibt kompatibel

in der Regel höhere Sicherheitsstandards

Nachteile

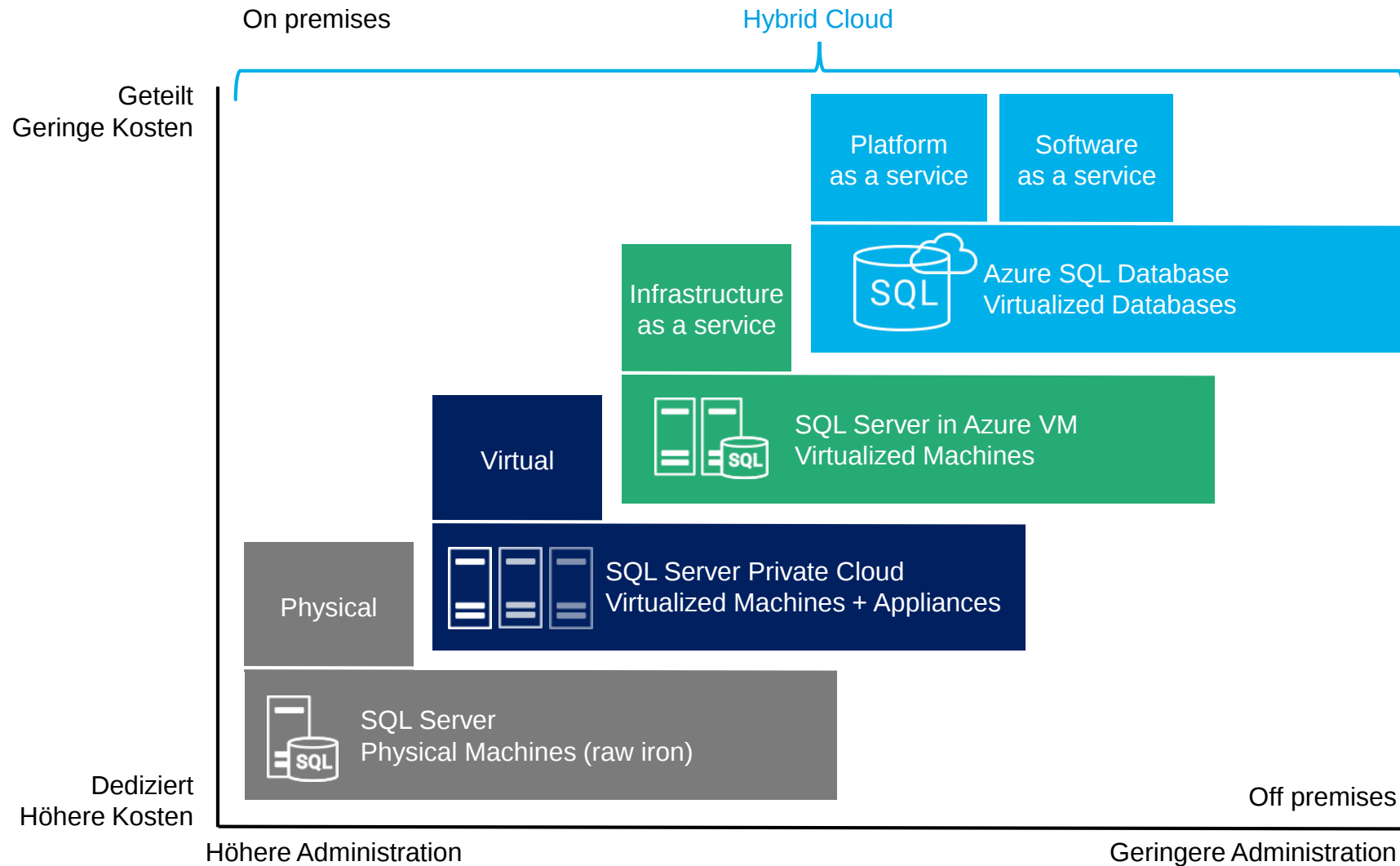
Abhängigkeit

Datenschutz

mangelnde Fortbildung der eigenen Mitarbeiter

schnelle Internetanbindung

Infrastructure as a Service (IaaS) vs. Platform as a Service (PaaS)



Welche Komponenten benötigt eine BI-Lösung

Welche Komponenten können für eine BI-Lösung relevant sein?

Überblick

Hardware



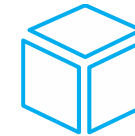
Relationale Datenbanken



Multidimensionale Datenbanken
(Cubes)



TABULAR



MULTIDIMENSIONAL

ETL-Werkzeug



Frontend



Benutzer- und
Rechteverwaltung



Welche Komponenten sind für eine BI-Lösung in Azure verfügbar

Hardware



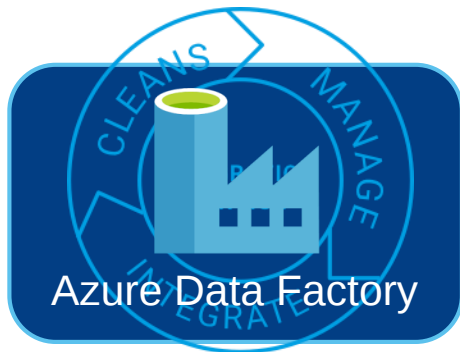
Relationale Datenbanken



Multidimensionale Datenbanken
(Cubes)



ETL-Werkzeug



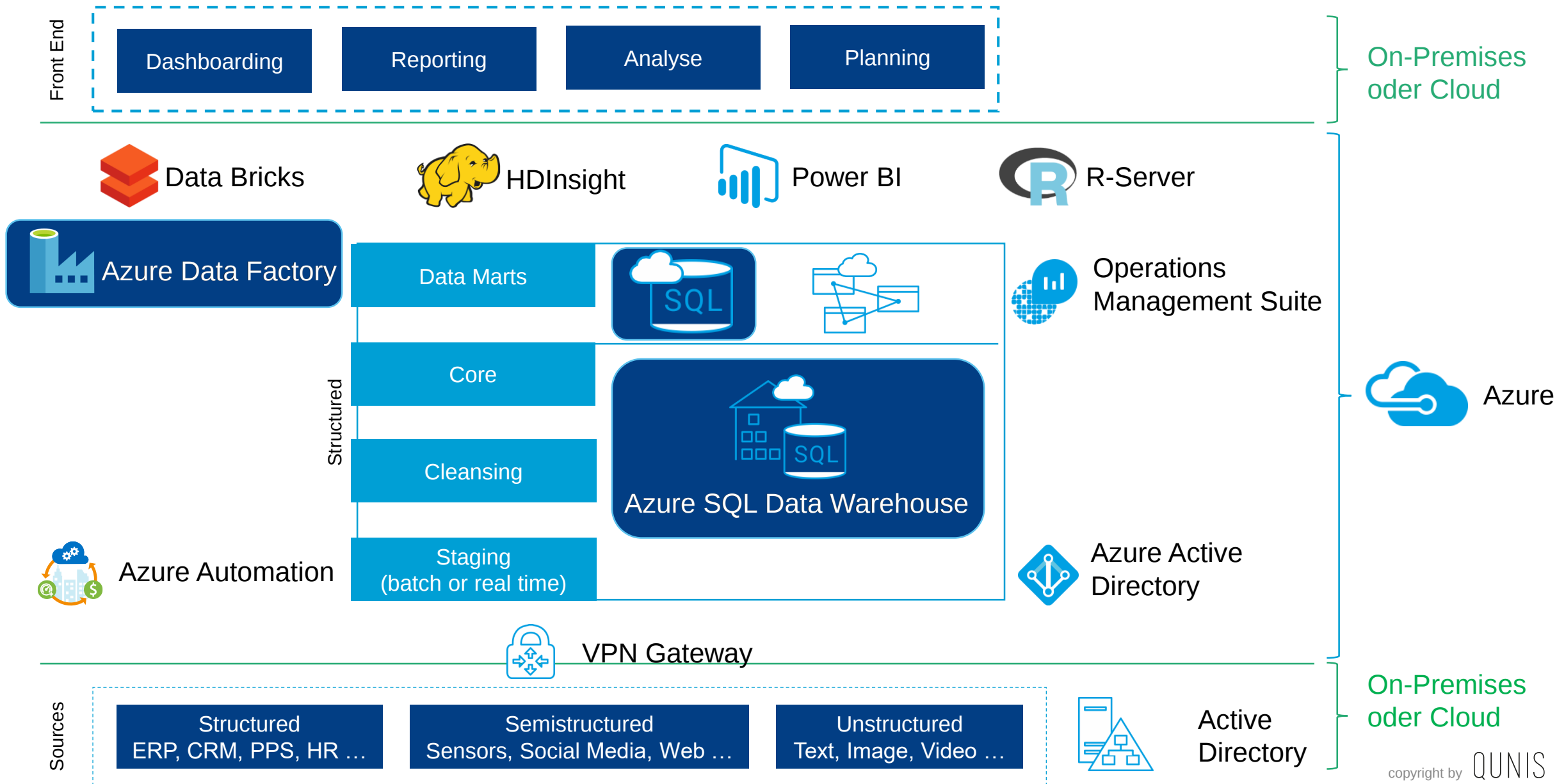
Frontend



Benutzer- und
Rechteverwaltung



Data-Lake Beispielarchitektur



Ausgewählte Azure Komponenten

Azure SQL Database

Azure SQL Database



- Verfügbar seit 2010
- Database-as-a-Service (DBaaS)
- Fällt damit in die Kategorie PaaS
- Vergleichsprodukt: Amazon's RDS, Google Cloud SQL oder Cloud Spanner
- Nutzt statt SQL-Server die Version 12 der Azure SQL Database
- Optimizer versteht T-SQL, ein Migration von SQL-Server auf Azure SQL Database problemlos möglich
- Basiert auf standardisierter Hardware und Software, die Microsoft gehört und von Microsoft gehostet und verwaltet wird.
- Azure SQL Database wird Kunden nicht mit einer Lizenz, sondern als Dienst verkauft.

Ausgewählte Azure Komponenten

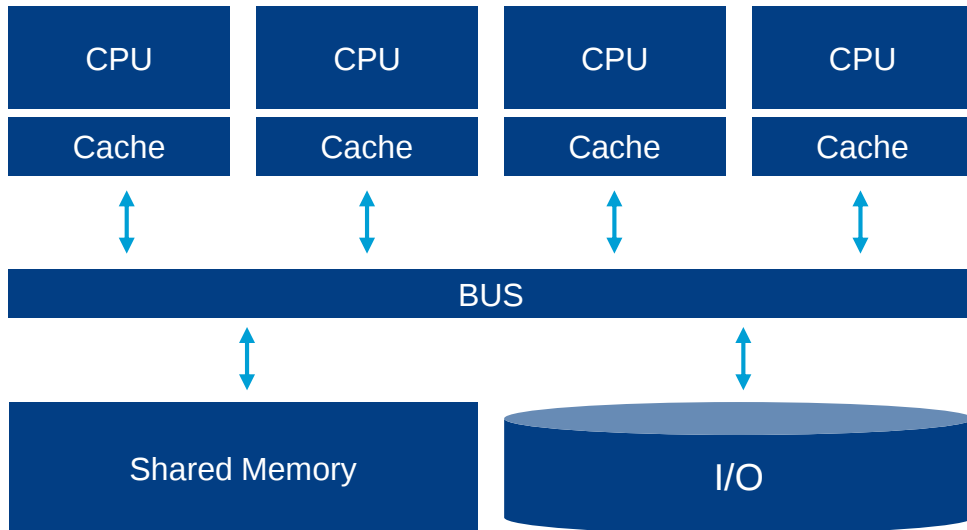
Azure SQL Data Warehouse

Datenbankarchitekturen: SMP vs. MPP

Kurzüberblick

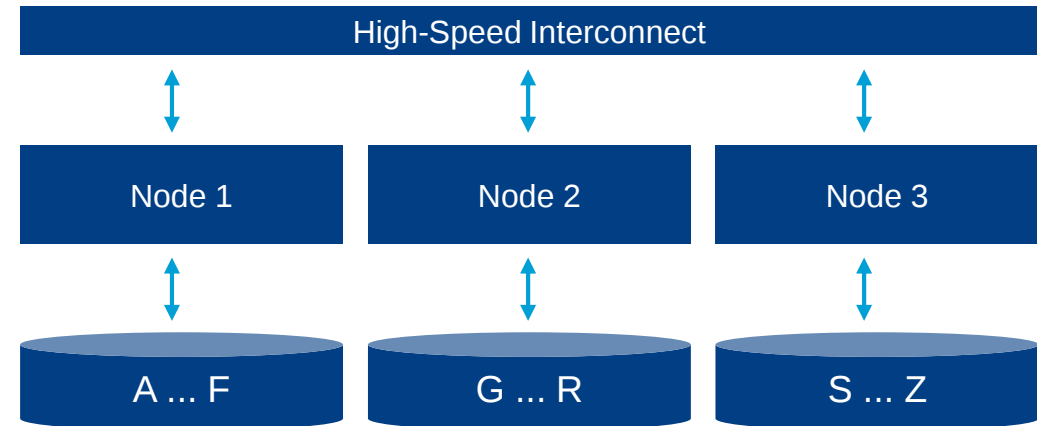
SMP (Symmetrisches Multiprozessorsystem)

- Ein Server, in dem sich alle CPU's den verfügbaren Hauptspeicher und Datenspeicher teilen müssen
- Eine SQL-Operation (SELECT, INSERT, ...) wird auf einem Server ausgeführt
- Der Flaschenhals in der Performance liegt meist beim Disk-IO und nicht der CPU



MPP (Massively Parallel Processing)

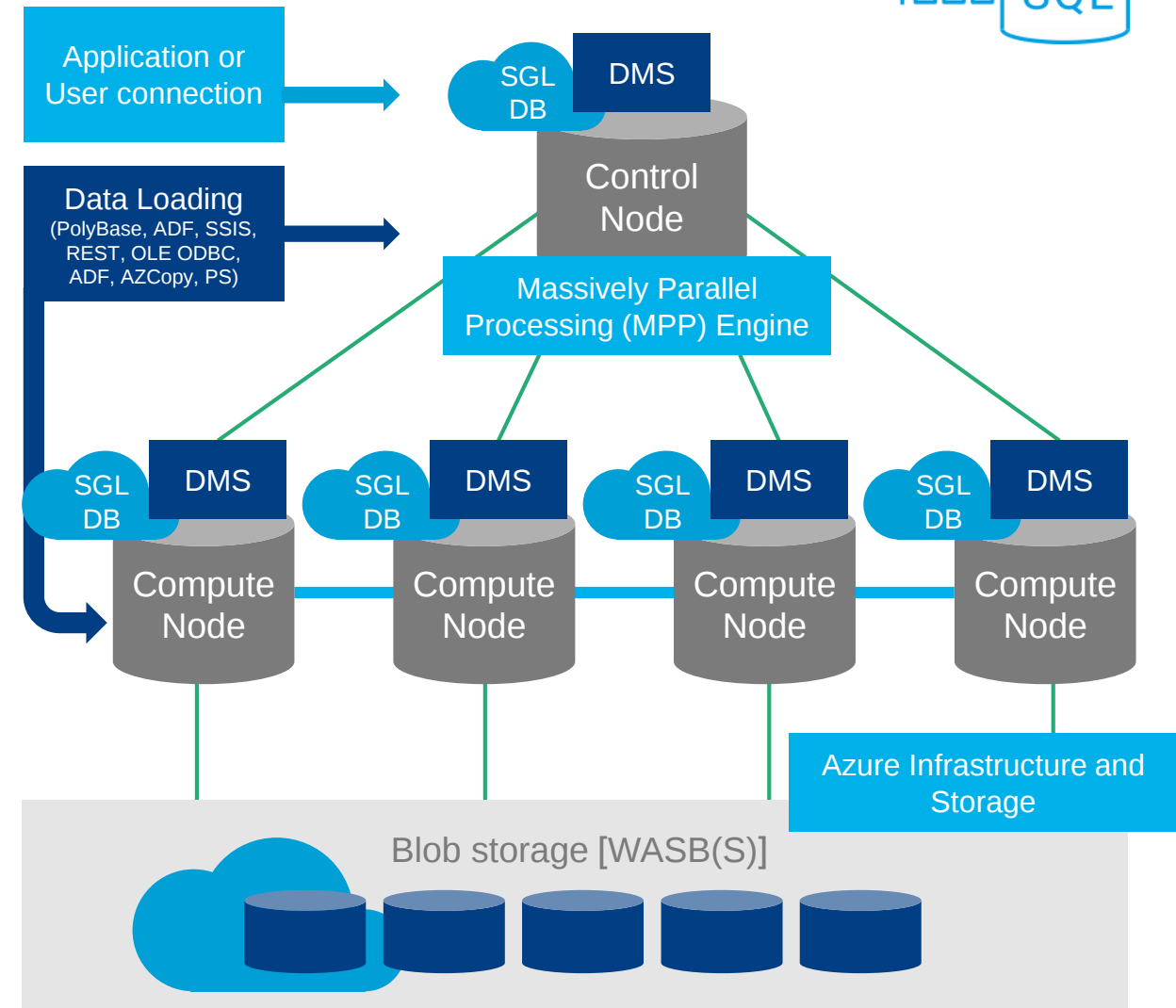
- Eine MPP Architektur besteht aus mehreren Servern (Compute-Nodes), die virtuell zu einem zusammengefasst werden
- Ein Server wie bei SMP entspricht in MPP einem Node
- Eine SQL-Operation (SELECT, INSERT, ...) wird über mehrere Nodes verteilt und später wieder zusammengeführt



Azure SQL Data Warehouse



- Verfügbar ab ca. Juli 2016
- Data Warehouse as a Service (DWaaS)
- Fällt damit in die Kategorie PaaS
- Vergleichsprodukt: Amazon's Redshift, Google Big Query
- Basiert auf der gleichen Grundtechnologie wie die Analytics Plattform System (APS / PDW)
- Nutzt statt SQL-Server die Version 12 der Azure SQL Database
- Compute ist nun vom Storage getrennt, so dass diese unabhängig skaliert werden können
- Azure SQL Data Warehouse kann mit APS kombiniert werden, um z.B. sensible Daten on-premises zu halten während nicht sensible Daten in der Cloud liegen



Ausgewählte Azure Komponenten

Azure Data Factory

Azure Data Factory (ADF)

Serverloser hybrider Datenintegrationsdienst (ETL) in Azure



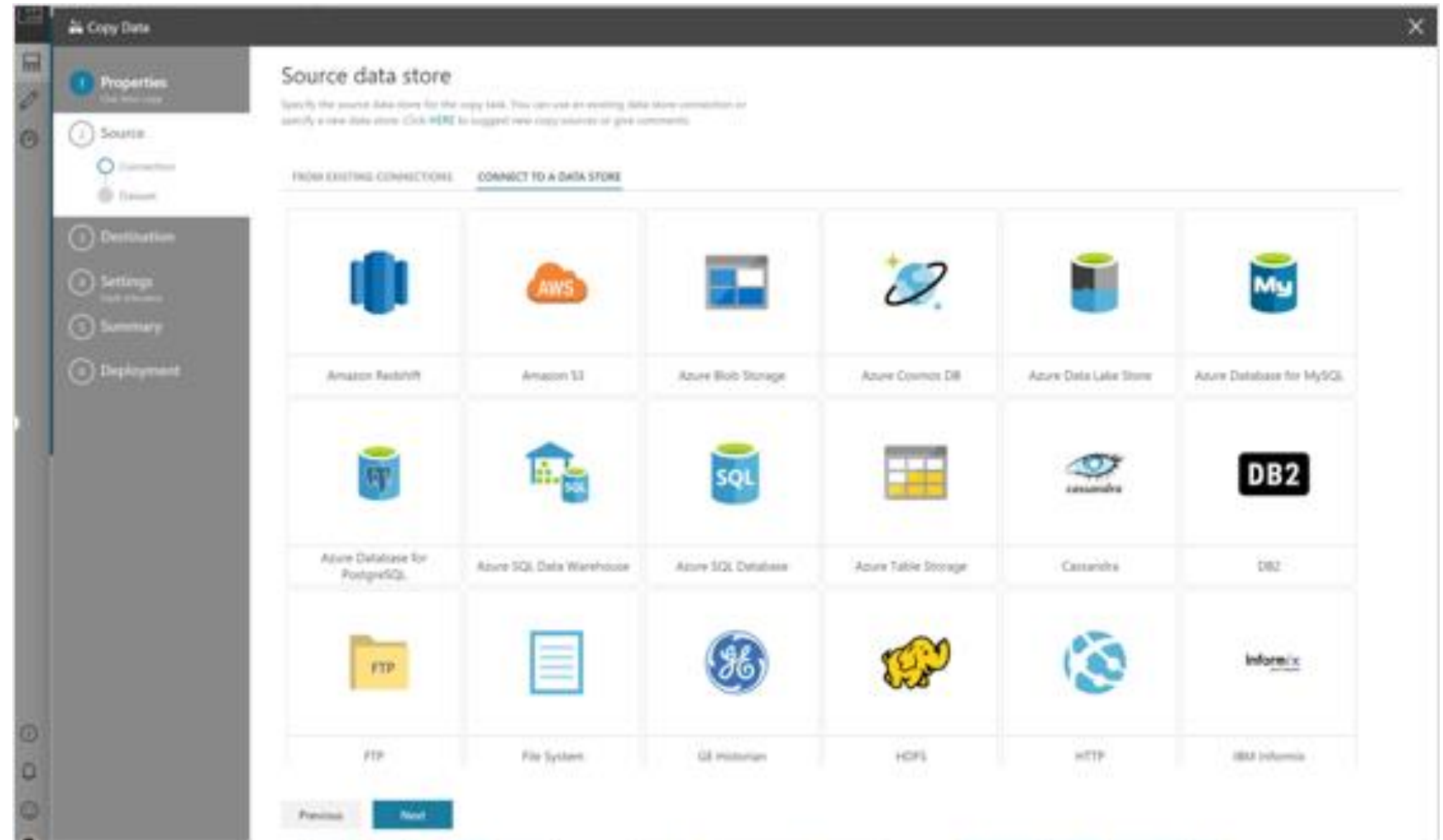
- Graphische Benutzeroberfläche in Azure, um Datenpipelines per Drag & Drop zu erstellen oder zu verwalten
- Bereits mehr als 70 verfügbare Datenquellenkonnektoren z.B. Azure Data Services, AWS S3 und Redshift, Google BigQuery, SAP HANA, Oracle, DB2, MongoDB, etc.
- Bereitstellung von SSIS-Paketen in Azure
- Zugriff auf Datenquellen in der Cloud oder On-Premises
- Skalierbar
- Einbindung von benutzerdefiniertem Code in verschiedenen Programmiersprachen (Python, .Net, ARM)

Azure Data Factory (ADF)

Datenintegration mit Azure Data Factory



Schritt 1: Zugreifen auf Daten und Erfassen von Daten mit integrierten Konnektoren



Azure Data Factory (ADF)

Datenintegration mit Azure Data Factory



Schritt 2: Erstellen skalierbarer Datenflüsse mit einer codefreien Benutzeroberfläche oder Schreiben von eigenem Code

The screenshot shows the 'Monitor Pipeline Runs' page for a pipeline named 'testdfv2datafactory'. The interface includes a 'Refresh' button, a date range selector set to 'Last 30 Days' (from 11/21/2017 11:53 AM to 12/02/2017 11:53 AM), and a time zone dropdown set to 'UTC-08:00 Los Angeles'. Below these are tabs for 'All', 'Succeeded', 'In Progress', 'Failed', and 'Cancelled'. The main table lists pipeline runs with columns: Pipeline Name, Actions, Run Start, Duration, Triggered By, Status, Parameters, Error, and RunID. The table shows a mix of successful and failed runs, with the most recent run failing.

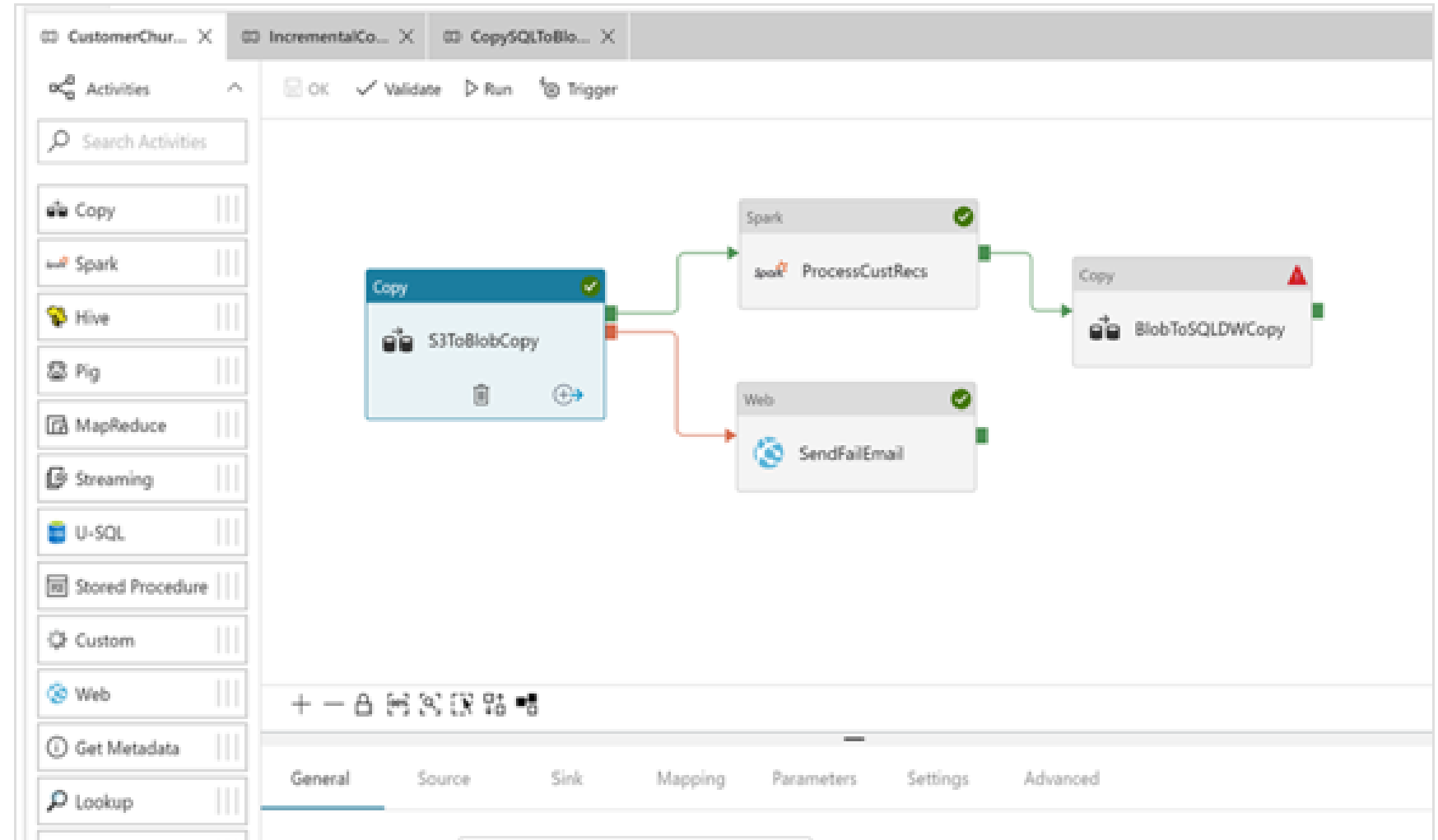
Pipeline Name	Actions	Run Start	Duration	Triggered By	Status	Parameters	Error	RunID
CopySQLToBlobPipeline	[Icon]	12/15/2017, 4:17:19 PM	00:00:29	Manual trigger	Failed			ee78628b-35e0-4258-9576-57b0c47a0624
CopyPipeline-gtr	[Icon]	12/15/2017, 2:52:16 PM	00:00:22	Manual trigger	Succeeded			34b71005-7a03-4e38-a8de-64b4b0b9128
CopyPipeline-1ca	[Icon]	12/15/2017, 2:46:37 PM	00:00:22	Manual trigger	Succeeded			4ebb2672-6cad-4e50-a98e-21a881406136
CopyPipeline-bfu	[Icon]	12/15/2017, 7:57:22 AM	00:00:24	Manual trigger	Succeeded			5b9037bc-84e1-4719-91db-45e630d9104
CopyPipeline-12j	[Icon]	12/14/2017, 11:23:39 AM	00:00:30	Manual trigger	Succeeded			ae188a83-4b10-4c7d-948b-518c28190043
CopyBlobToBlob	[Icon]	12/14/2017, 11:07:08 AM	00:00:26	Manual trigger	Succeeded			3f28402e-630b-490d-8992-896217c80487
CopySQLToBlobPipeline	[Icon]	12/14/2017, 9:36:38 AM	00:00:47	Manual trigger	Failed			afa946d0-4eb2-4d12-8666-6e4296d36c57
CopyBlobToBlob	[Icon]	12/13/2017, 10:52:51 PM	00:00:23	Manual trigger	Succeeded			855477f9-8112-4822-4225-38ae89916714
CopyBlobToBlob	[Icon]	12/13/2017, 10:52:51 PM	00:00:24	Manual trigger	Succeeded			a7f99cd5-0005-40d1-a74b-7534461111a8
CopyBlobToBlob	[Icon]	12/13/2017, 10:52:49 PM	00:00:23	Manual trigger	Succeeded			6e8623e5-d0d8-406d-w004-3x2d87727e35
CopySQLToSQLDW	[Icon]	12/12/2017, 12:44:39 PM	00:00:31	Manual trigger	Succeeded			2bf9034-d0b6-4955-ac34-455788998823
CopyPipeline-81a	[Icon]	12/12/2017, 7:22:06 AM	00:00:21	Manual trigger	Succeeded			9ee54c09-28f9-473c-952f-12d2c831920d
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:14:57 AM	00:00:30	ScheduleTrigger	Failed			5d774a43-c5f9-4c57-b8a6-8ca435a02e51
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:13:54 AM	00:00:30	ScheduleTrigger	Failed			ad50631f-3999-488c-a0d8-c10b-18c771e
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:12:54 AM	00:00:29	ScheduleTrigger	Failed			191ae308-8a06-48cd-96da-7566a712d853
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:11:57 AM	00:00:29	ScheduleTrigger	Failed			30d6e7f9-8fcd-44c5-8a23-4c88a710351d
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:10:54 AM	00:00:29	ScheduleTrigger	Failed			25e2ee75-dfeb-43cd-b572-ba6d8cd2350d
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:09:54 AM	00:00:44	ScheduleTrigger	Failed			26380d63-5a05-42f9-b0c3-bd01024b5268
CopySQLToBlobPipeline	[Icon]	12/08/2017, 11:08:54 AM	00:01:16	ScheduleTrigger	Failed			3c5ec111-32cd-47d6-a027-64e1a98ab496

Azure Data Factory (ADF)

Datenintegration mit Azure Data Factory

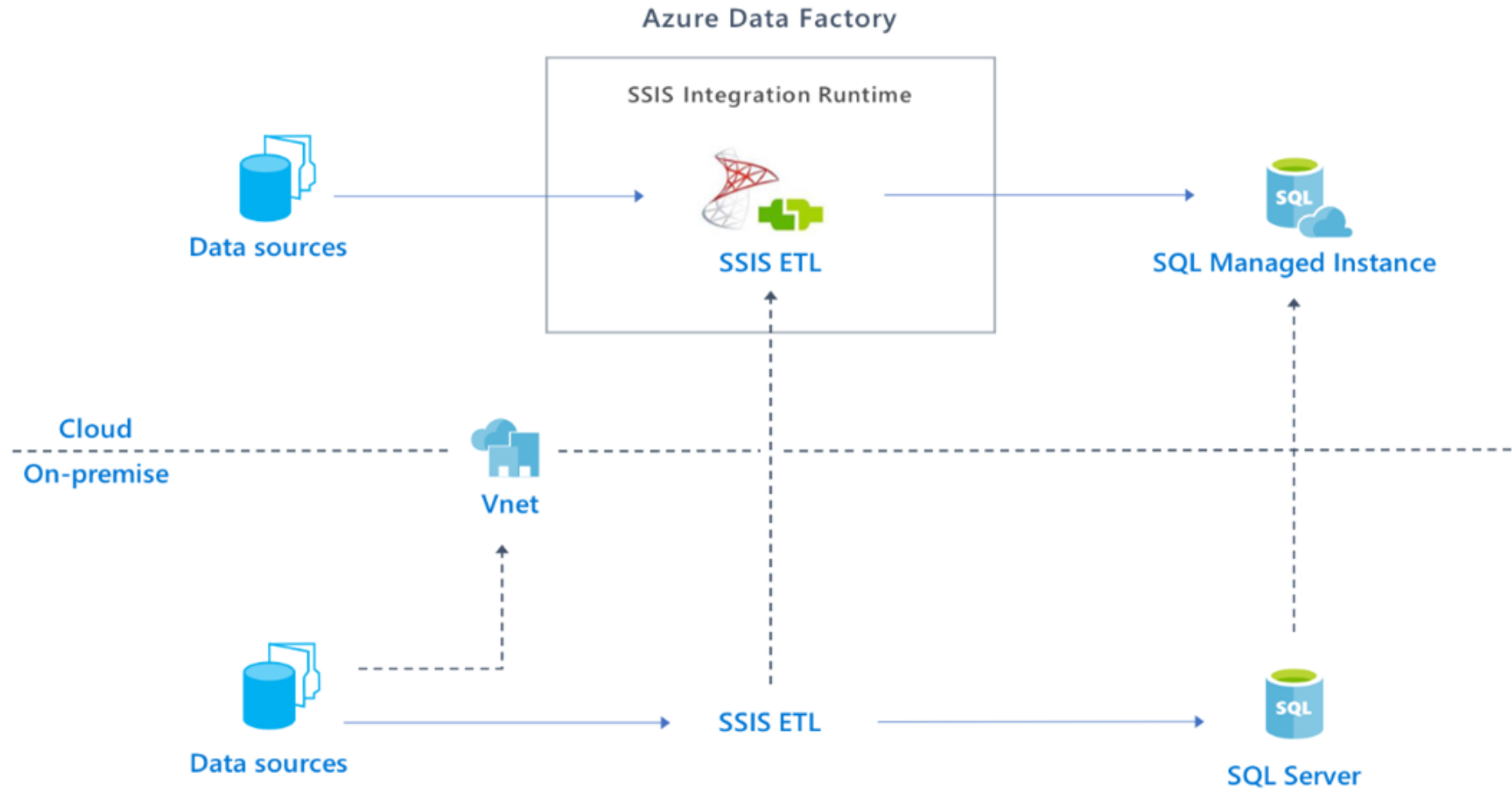


Schritt 3: Planen, Ausführen und Überwachen Ihrer Pipelines



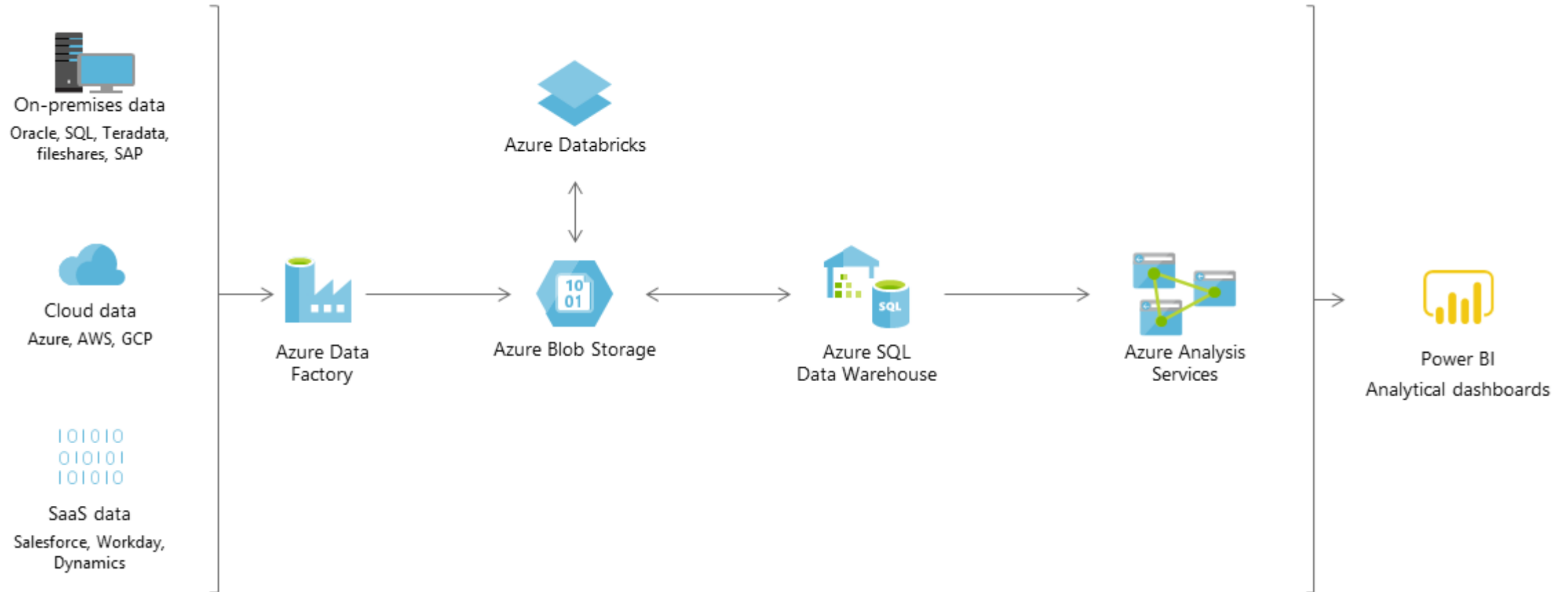
Azure Data Factory (ADF)

SSIS-Paketausführung in Azure



Azure Data Factory (ADF)

Orchestrierung mit Azure Data Factory





Andy Löwen

Senior Consultant, QUNIS GmbH

Phone: +49 176 114802 16

E-Mail: andy.loewen@qunis.de



THE
INFORMATION
COMPANY

WWW.QUNIS.DE