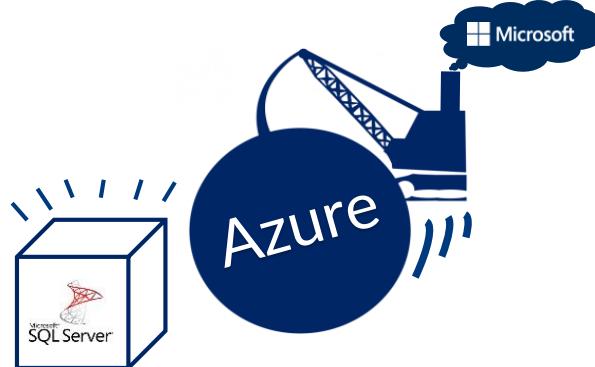




Azure Cloud kills SQL BI



PASS Bayern bei Microsoft München, 20.4.2017



Stefan Kirner

Teamleiter Business Intelligence Solutions
bei der inovex GmbH

- › Mehr als 15 Jahre Erfahrung mit dem Microsoft Business Intelligence Toolset
- › Microsoft Certified Systems Expert (MCSE) for BI
- › Microsoft Certified Systems Associate (MCSA) für Cloud Platform
- › Microsoft P-TSP Data Platform
- › Leitung SQL PASS e.V. Community Karlsruhe
- › Sprecher bei Konferenzen und User Groups zu BI- und Cloud-Themen
- › Mail: stefan.kirner@inovex.de
- › Twitter: @KirnerKa



inovex ist ein IT-Projekthaus
mit dem Schwerpunkt „Digitale Transformation“:

Product Ownership · Datenprodukte
Web · Apps · Smart Devices · BI
Big Data · Data Science · Search
Replatforming · Cloud · DevOps
Data Center Automation & Hosting
Trainings · Coachings



inovex gibt es in Karlsruhe · Pforzheim ·
Stuttgart · München · Köln · Hamburg

Und natürlich unter www.inovex.de

Wir nutzen Technologien,
um unsere Kunden glücklich zu machen.
Und uns selbst.

Agenda

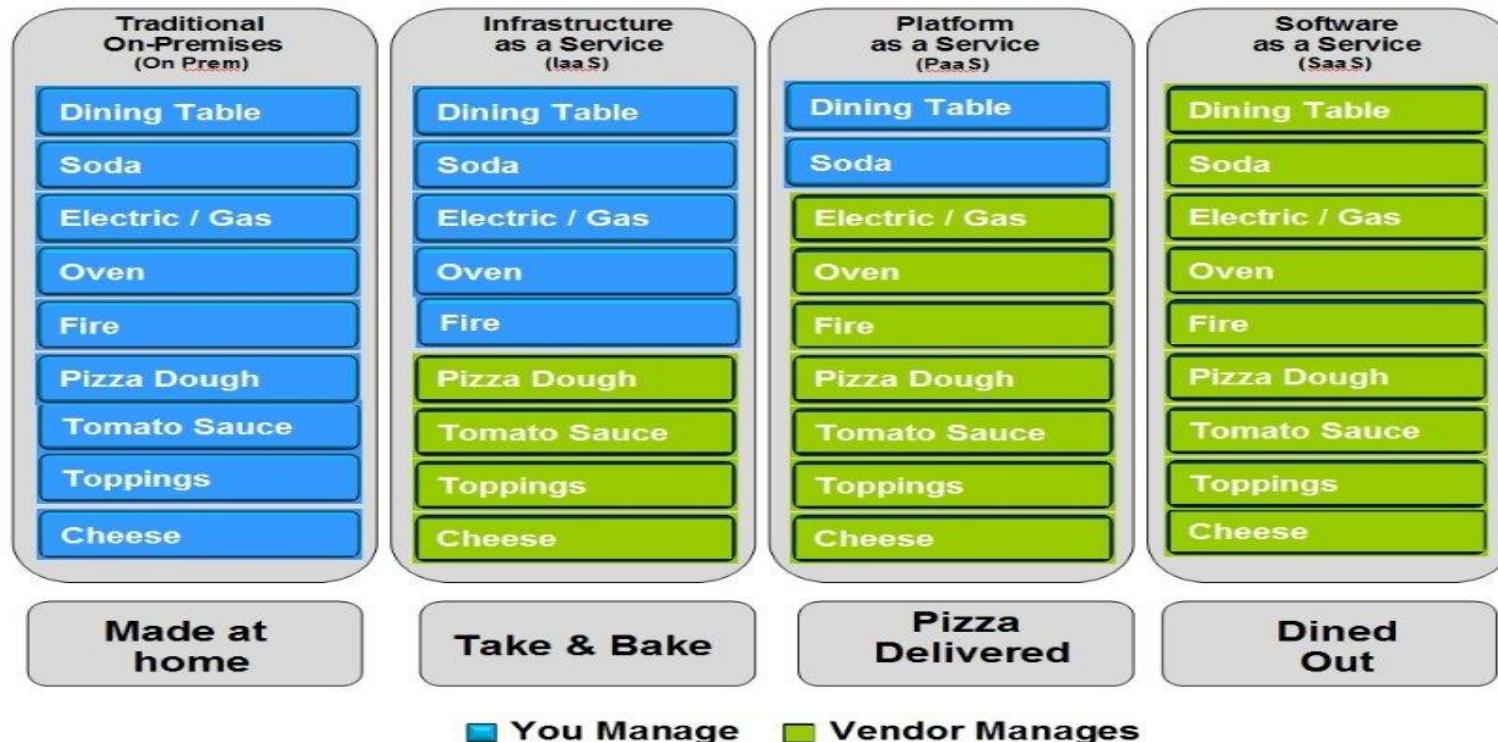
1. Grundbegriffe und Einführung
2. Data Warehouse relational
3. Data Management (ETL)
4. Semantisches Modell
5. Berichte und Dashboards
6. Scheduling
7. Advanced Analytics
8. Hybride Architekturen
9. Fazit

Sorry...out of Scope:

- Hadoop Stack auf Azure,
außer bei Advanced
Analytics
- Streaming

Grundbegriffe – as a Service einfach erklärt...

Pizza as a Service



Plattformvorteile Cloud BI

- › Skalierbarkeit
- › Technologievielfalt
- › Agiles Setup
- › Entlastung IT
- › Schnelle Innovationszyklen
- › Optimierung von Betriebskosten
- › Globale Szenarien
- › Erhöhte Ausfallsicherheit
- › Flexibilität für Testumgebungen



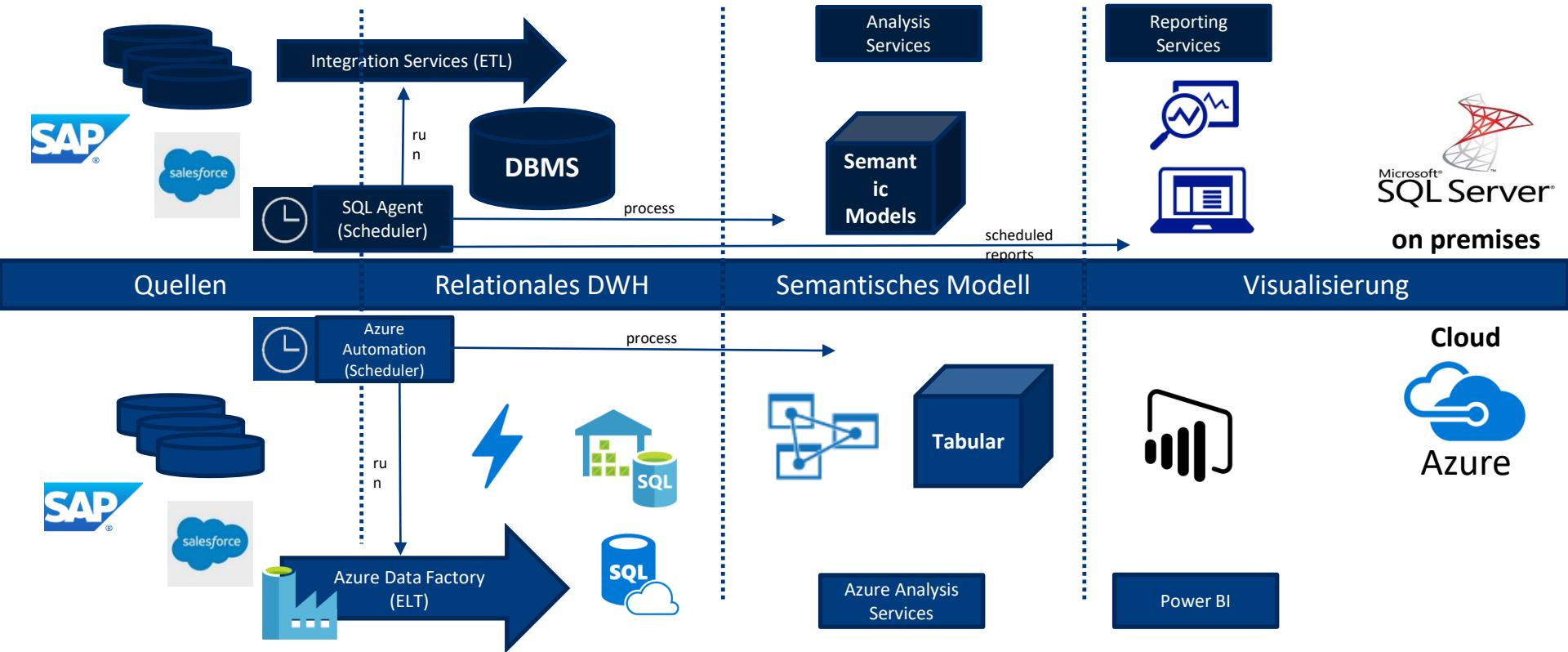
<http://allonzoinc.wordpress.com>

Plattform-“Herausforderungen“ Cloud BI

- › Viele Datentöpfe liegen on-premises
- › Datenschutz
- › Geringerer Reifegrad relativ junger Technologien
- › Intransparenz & erschwerte Fehlersuche durch Abstraktion
- › Installierten Versionen der PaaS Dienste hängen in hoch-agilen Bereichen wie OSS hinterher
- › Abgabe von Kontrolle und Abhängigkeit von Cloud Provider
- › Evergreening führt zu Verwirrung der Anwender (und der Techniker!)



Überblick Architektur für Vergleich

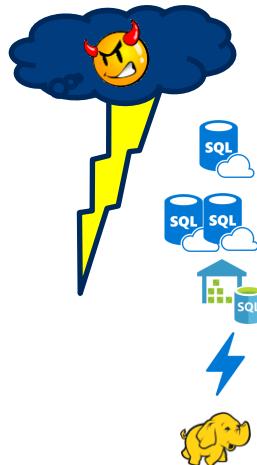


Relationales DWH



On Premises

SQL Server relationales
Datenbanksystem
(Analytical Platform System)



Azure Cloud

SQL Azure - SQL DB PaaS

SQL Azure Elastic – Skalierende DBs

Azure SQL Datawarehouse

Azure Data Lake

(Azure HD Insight)

Data Warehouse SMP vs. MPP

› Symmetric multi-processing (SMP)

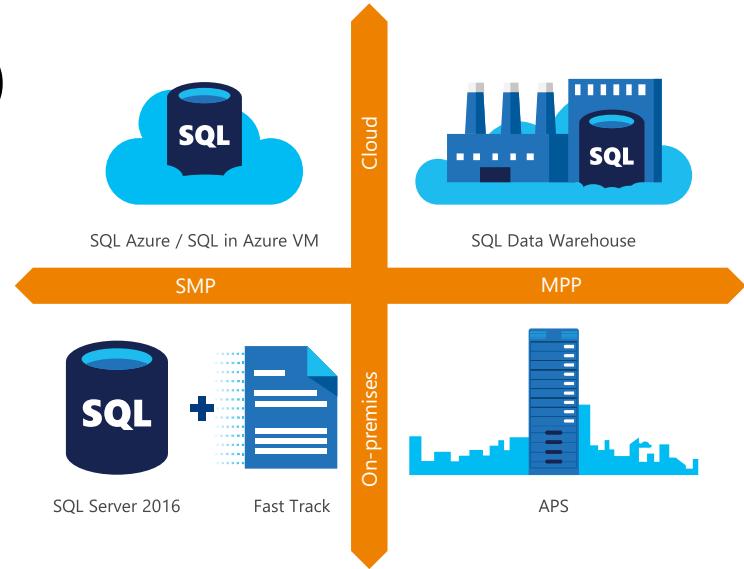
On-premises: SQL Server 2016 or SQL Server Fast Track Data Warehouse

Cloud: SQL Azure / SQL Server in Azure VM

› Massively parallel processing (MPP)

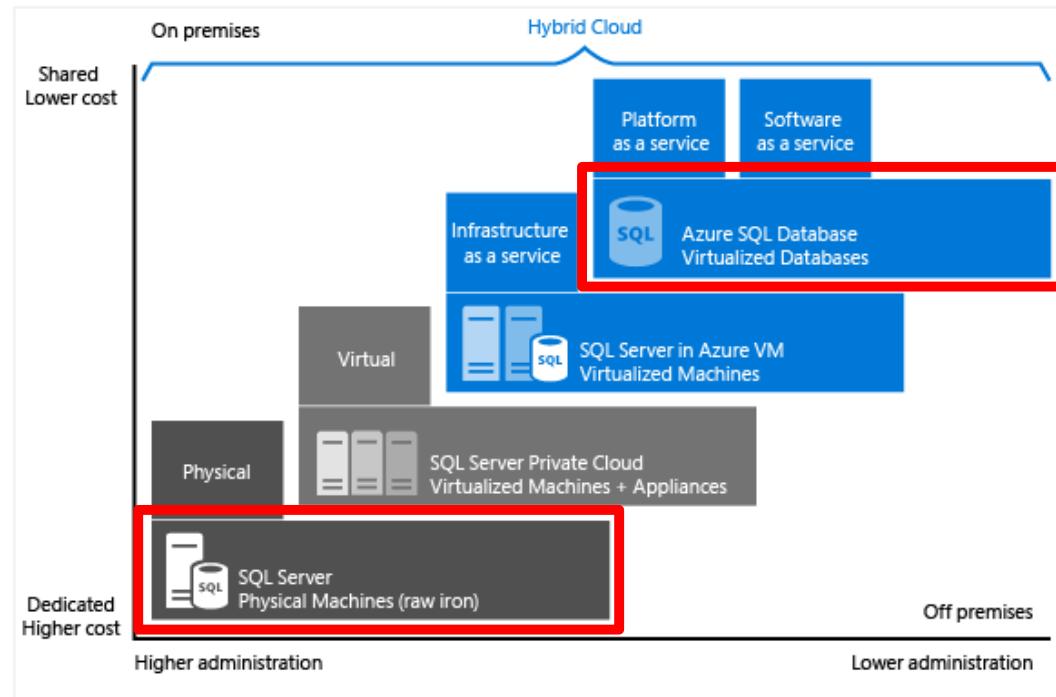
On-premise: Analytics Platform System (APS)

Cloud: Azure SQL Data Warehouse



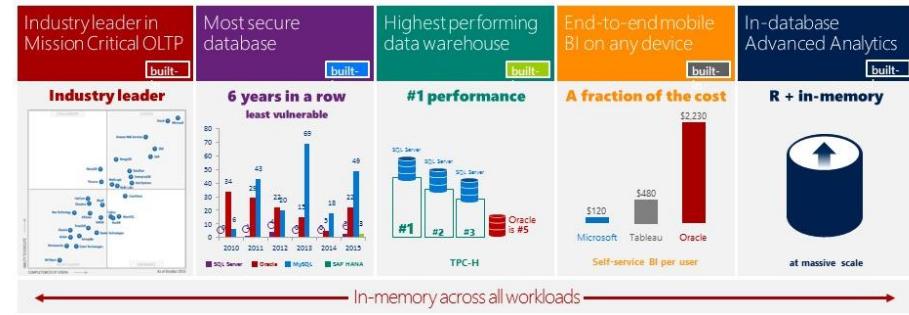
Data Warehouse on premises to managed cloud

- › Geringere Kosten in Cloud durch Synergieeffekte
- › Weniger administrativer Aufwand in der Cloud
- › Zur Vereinfachung betrachten wir hier SQL Server physical on-prem und Azure SQL Databases als cloud service (PaaS)



SQL Server relational on premises

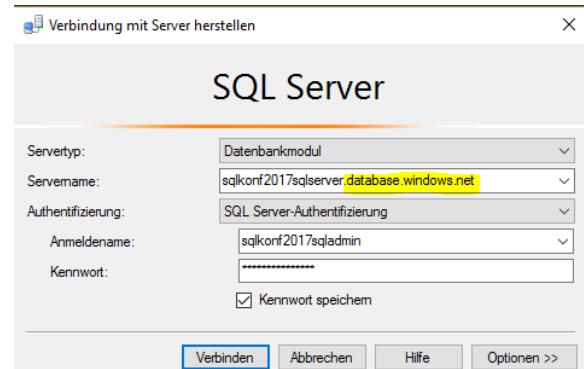
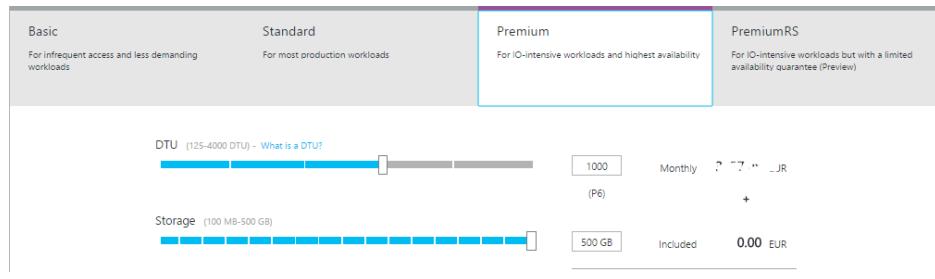
- › seit ca. 20 Jahren am Markt und hohe Verbreitung
- › Spitzenpositionen in einigen Gardner Quadranten
- › Editionen Express bis Enterprise
- › Feature complete für DWH, aber nicht ab Werk dafür optimiert (Ausnahme: Fast-Track)
- › Polybase Unterstützung für unstrukturierte Daten
- › Max: 524 PB Storage, 512 Cores 24 TB RAM
- › Kein Scale out



SQL Azure



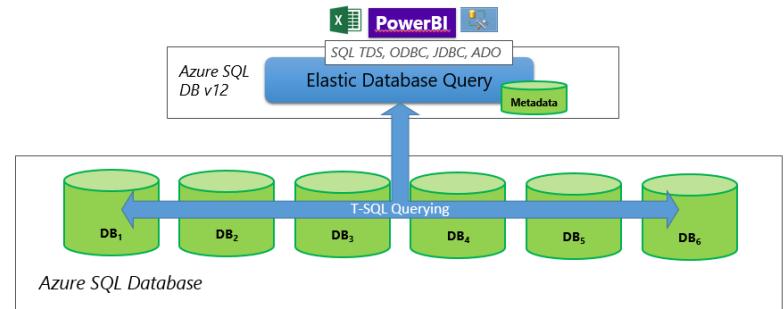
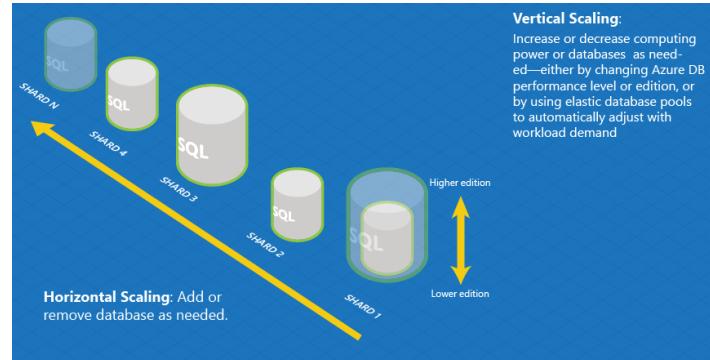
- › SQL Datenbank als Cloud Service
- › seit ca. 6 Jahren am Markt
- › PaaS mit den Tiers Basic, Standard und Premium – Unterschiede in Performance und Features
- › Max: 4 TB Storage und 4000 DTUs im Premium Tier mgl., 30000 Conc. Sessions
- › Feature complete für DWH, aber nicht ab Werk dafür optimiert, hier fehlt Polybase, Columnstore ab P1
- › Scale up/down in Sekunden
- › Bedienung über gewohntes Tooling (SSMS, Data Tools Projekte...)



Data Warehouse SMP Elastic Option



- › Horizontales Skalieren - “Sharding” Verteilung von Abfragen auf identische DBs
- › > 4 TB Speicher Workloads
- › Auch hier: Vertikales Skalieren – Erhöhen der Leistung einzelner Datenbanken
- › Abfragen über Elastic Database Query gehen über alle Knoten
- › Klingt richtig cool, aber vieles muss selbst gemacht werden, z.B. Replikation der Shards



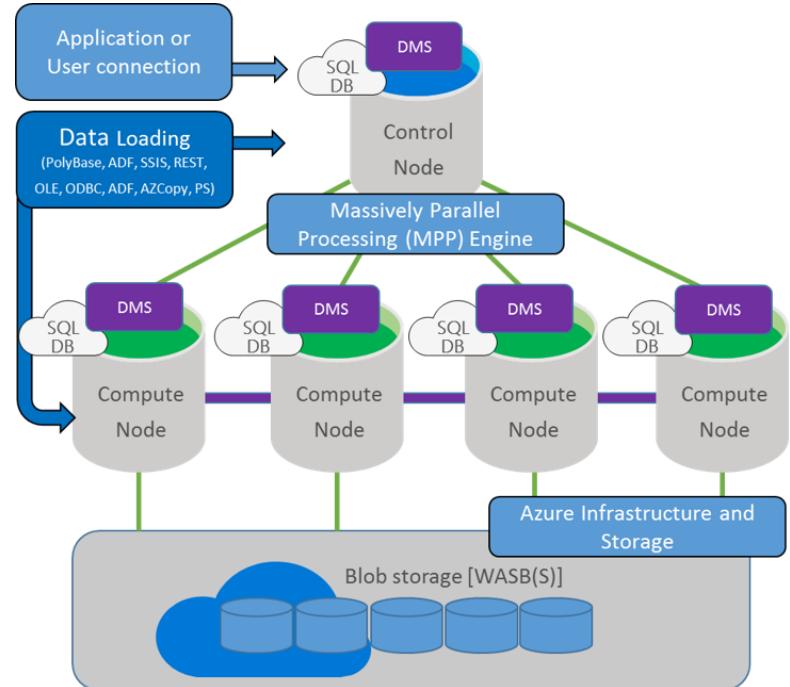
<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview>

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-scale-introduction>

Azure SQL Data Warehouse (MPP)

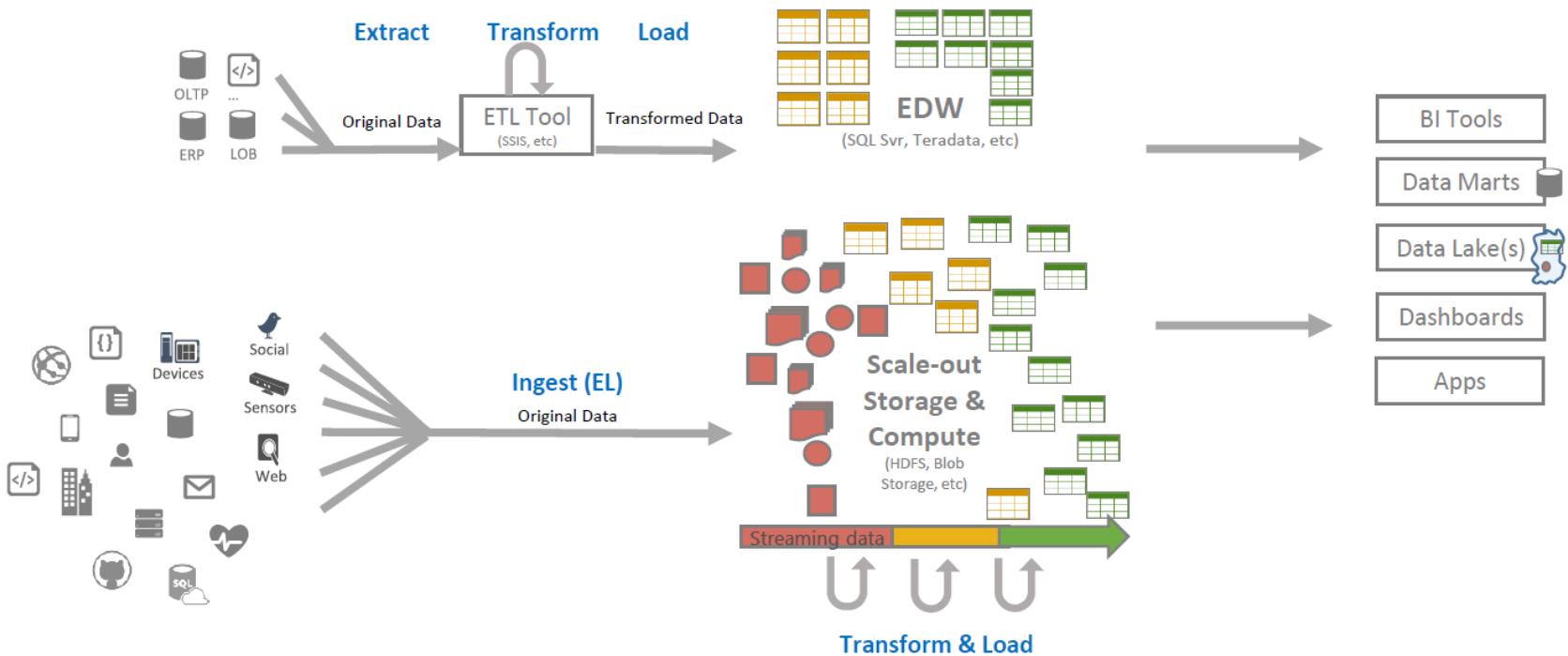


- › Für DWH optimierte MPP Verarbeitung in parallel aufgesetzten SQL Server Instanzen, zentral durch Kontrollknoten gesteuert
- › Integriert in Azure Infrastruktur
- › Skalieren zeitgesteuert und per slider bar möglich, unbegrenzte Ressourcen verfügbar
- › T-SQL Dialekt mit angepasstem Funktionsumfang
- › Andere Ladepattern für Data Loads (ELT)
- › Automatische Verteilung der Daten nach Konfiguration über Kontrollknoten
- › Polybase Unterstützung für unstrukturierte Daten
- › Kann pausiert werden
- › Fehlende Features: row level security, indexed views, constraints
- › Max: Datenbankvolumen 240 TB, Concurrrent Connections: 1024, Concurrent Queries: 32



<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is>

Data Lake - Neue Ladepattern



Reza Rad : Azure Data Factory vs. SSIS, PASS Summit 2015

Azure Data Lake



- › Verteilter Speicher- und Analysedienst semi-strukturierter Daten
- › 2 Komponenten:
 - ⚡ › ADL Store – hochoptimierte Writes/Reads der Daten mit Abstraktion der Storage Accounts
 - ⚡ › ADL Analytics – Abfrageengine, auch parallel über andere Stores wie Azure SQL usw.
- › Dynamische Skalierung
- › Verteilte Hardware und Parallelisierung gänzlich abstrahiert
- › U-SQL – TSQL und C# oder Hive, Pig
- › Visual Studio Unterstützung
- › Textdaten ablegen und Abfragen ohne Import in ein Schema
- › Auch Einbindung von Cognitive Services
- › Abfragen können weitere Cloud Stores enthalten
- › Demnächst auch in Europa verfügbar!

```
@t = EXTRACT date string  
      , time string  
      , author string  
      , tweet string  
  FROM "/Input/MyTwitterHistory.csv"  
  USING Extractors.Csv();  
  
@res = SELECT author AS author  
        , COUNT(*) AS tweetcount  
    FROM @t  
   GROUP BY author;  
  
OUTPUT @res TO "/Output/MyTwitterAnalysis.csv"  
ORDER BY tweetcount DESC  
USING Outputters.Csv();
```



	Preparing	Queued	Running	Finalizing
40 seconds				
Job Result: Succeeded				
Duration: 6.9 minutes				
Submit Time: 9/8/2015 8:59:08 AM				
Start Time: 9/8/2015 9:00:32 AM				
End Time: 9/8/2015 9:06:04 AM				
Compilation: 40 seconds				
Queued: 2 seconds				
Running: 5.5 minutes				
Account: mwinkle				
Author: mwinkle@microsoft.com				
Runtime: kobo_ynext_4204497_e0215c				
Priority: 0				
Root Process Id: 93ff9106-4145-4549-9936-b02eb1c05c6e				
Application Id: 1441686056117_0530				
DLAU Count: 35				
Bytes Left: 0				
Bytes Read: 50,610,927,788				
Bytes Written: 34,616,688,131				
Vertices: 74				

Fazit Data Relationales DWH

SQL Server 2016 on premises



DWH Nutzung möglich und erprobt mit vielen Features.
Geeignet für große Datenmengen, aber SMP.

SQL Azure



Geeignet für Data Marts, Fachbereichs DWH mit überschaubaren Datenmengen

SQL Data Warehouse



MPP Enterprise DWH Lösung für sehr große Datenmengen, Einschränkungen bei Anzahl Benutzer und Sessions

Elastic Database



SQL Azure Scale out, für Reporting für große Benutzerzahlen bei großen Datenmengen

Azure Data Lake



Für sehr große Menge auch semi-strukturierter Daten aus versch. Töpfen, eher explorative Analyse als Dashboards

Data Management



On Premises

SQL Server Integration Services



Azure Cloud

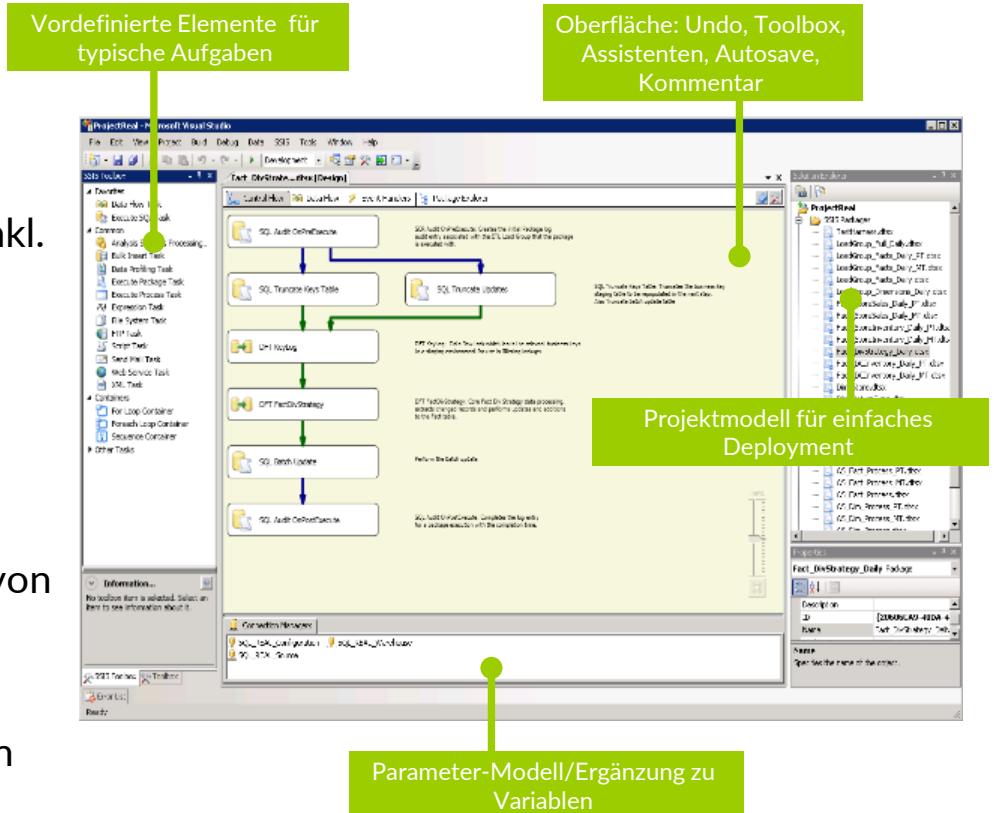


Azure Data Factory

Ergänzungen & Alternativen kurz
erläutert

ETL on prem - SSIS

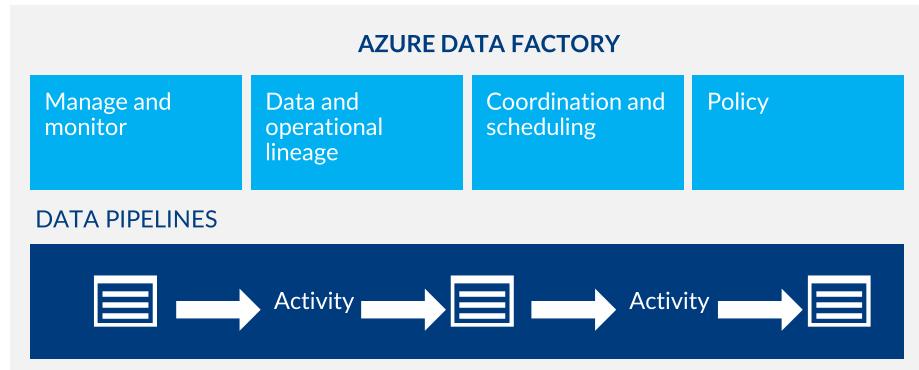
- › Komfortable Entwicklungsumgebung inkl. Debugger
- › Zugriff auf viele vordefinierte Datenquellen
- › Datenbank-Management- und weitere Spezial-Aufgaben
- › Erweiterbar durch offene .NET-Schnittstellen
- › Wartbarkeit und Wiederverwendung von ETL-Paketen
- › Überwachung und Monitoring mit eingebautem SSIS-Dashboard
- › Sicherstellung der unternehmensweiten Datenkonsistenz



Azure Data Factory



- › Versteht sich als Orchestrierungs-, nicht als ETL Tool. Copy als einzige Transformation
- › Konnektivität zu vielen Datenquellen, auch on-prem
- › Unterstützt Auflösung komplexer zeitlicher Abhängigkeiten verschiedener Quelldaten
- › Integriertes Scheduling über Zeitscheiben
- › Optimiert für cloud born data: production policies für Retry, Concurrency und late data handling
- › Managed Service
- › Nutzt für Processing T-SQL, U-SQL, HIVE, Pig und was die Compute Knoten sonst noch bieten



Data Factory im Portal

- › Werkzeuge für das Erstellen der Factory
- › Zustandsüberwachung
- › Manuelles Starten von Vorgängen
- › Auslastung und Diagnosen
- › Copy Data Assistent für einfache Tasks

COPY DATA

Source data store

Specify the source data store for the copy task. You can use an existing data store linked service (or) specify a new data store.

[FROM EXISTING LINKED SERVICES](#) [CONNECT TO A DATA STORE](#)

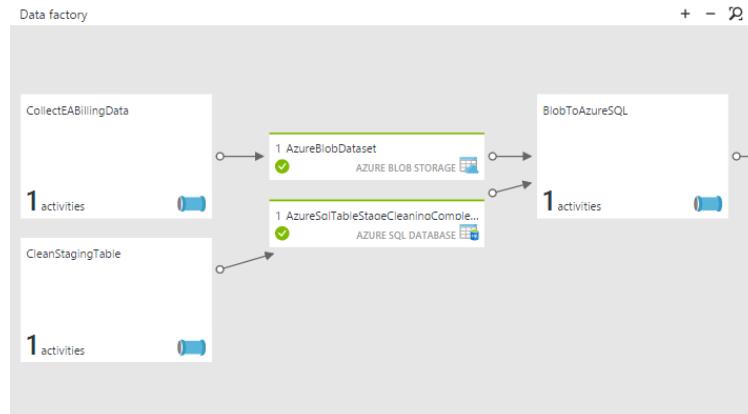
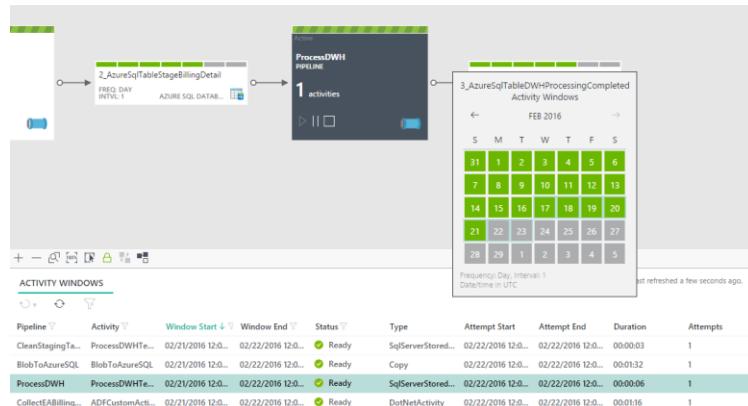
 Azure Blob Storage

 Azure SQL Database

 Azure SQL Data Warehouse

 On-premises SQL Server

 Azure Data Lake



Data Factory in Visual Studio

- › Templates erlauben schnelles Erstellen einfacher Anwendungsfälle
- › Deployment aus Visual Studio oder Factory Dashboard
- › JSON-Dokumente ohne Toolunterstützung füllen und verknüpfen
- › Design by Copy and Paste
- › Integration in Application Lifecycle Management durch Code Versioning

```
{
  "name": "ADFTutorialPipeline",
  "properties": {
    "description": "Copy data from a blob to Azure SQL table",
    "activities": [
      {
        "type": "Copy",
        "typeProperties": {
          "source": {
            "type": "BlobSource"
          },
          "sink": {
            "type": "SqlSink",
            "writeBatchSize": 10000,
            "writeBatchTimeout": "60.00:00:00"
          }
        },
        "inputs": [
          {
            "name": "MasterDataAzureBlobLocation"
          }
        ],
        "outputs": [
          {
            "name": "EmpSQLTable"
          }
        ],
        "policy": {
          "timeout": "01:00:00",
          "concurrency": 1,
          "executionPriorityOrder": "NewestFirst",
          "style": "StartOfInterval"
        },
        "scheduler": {
          "frequency": "Hour",
          "interval": 1
        },
        "name": "CopyFromBlobToSQL",
        "description": "Push Regional Effectiveness Campaign data to Azure SQL database"
      }
    ],
    "start": "2015-07-12T00:00:00Z",
    "end": "2015-07-13T00:00:00Z",
    "isPaused": false
  }
}
```

Azure Data Factory vs. SSIS

	SSIS	Data Factory
Einsatzziel	Traditionelles ETL-Tool	Orchestrierungsdienst
Skalierbarkeit	Für DWH ausreichend	Für Big Data gedacht
„Reichweite“	Firmennetzwerk	Global in Azure
Art der Software	On Prem Software	Managed Service
Produktzyklen	Lang: Jahre	Kurz: Wochen
Design-GUI	Ja	Naja.., Assistent ist ganz ok
Transformationen	Viele	Copy
Bezahlmodell	Lizenz & Advanced Features (EE)	Nutzungsbasiert
Error Handling	Alerts, Logging, Handling	Alerts & Logging
Scheduling	SQL Agent, Workflows für Logik	Etwas abstrakte Steuerlogik, berücksichtigt Abhängigkeiten über Data Slices

Alternativen/Ergänzungen noch kurz erläutert

1. M in Power BI – für self-service BI und die Power BI Datenmodelle gut geeignet, für große Datenmengen nicht wegen fehlender Persistenz
2. Polybase – Integration von Azure Blob Store oder HDFS in SQL DWH, Transformation durch **Create Table As Selects**
3. U-SQL bei ADLA oder T-SQL bei Azure SQL/SQL DWH in Stored Procedures können auch direkt gescheduled und ausgeführt werden z.B. per Azure Automation
4. R für Datentransformation, verfügbar in Power BI und HD Insight
5. HDInsight in unterschiedlichen Varianten (Spark, Hadoop..) mit versch. Methoden (Hive, MapReduce, SparkSQL..)
6. Integration Services über eine VM in Azure mit den Azure Komponenten ist auch eine mögliche Variante (IaaS)

Fazit Data Management

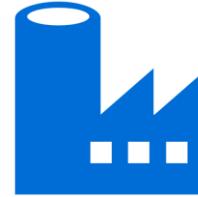
Integration Services



Microsoft®
SQL Server®

Erprobtes Tooling für ETL
Prozesse im Unternehmen,
steile Lernkurve durch drag
& drop experience

Azure Data Factory



Ausrichtung auf Big Data,
Transformationen auf den
Data Stores (ELT), eher
schwierig zu bedienen

Semantisches Modell

Performance, Navigation & KPIs

On Premises



SQL Server Analysis Services

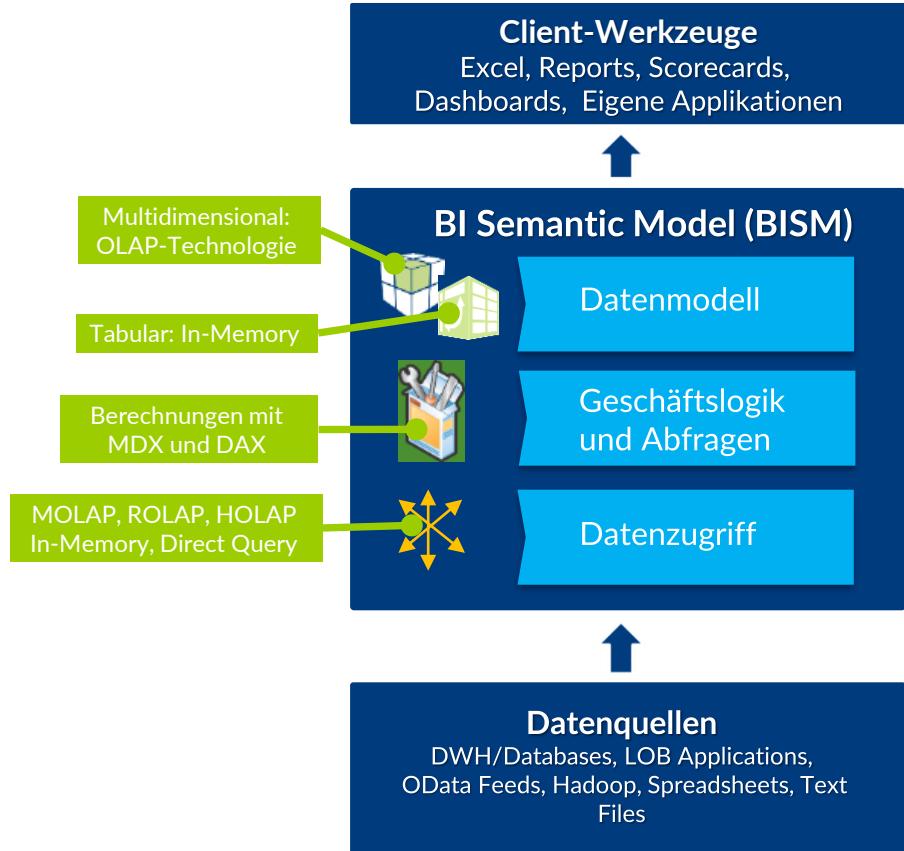


Azure Cloud

Azure Analysis Services

Analysis Services on premises

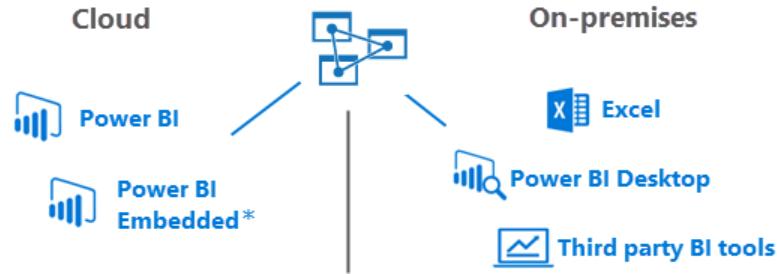
- › Erstellung von Analysemodellen für die effektive Datenanalyse
- › Kennzahlen, Dimensionen, Hierarchien, KPI's, Perspektiven, ...
- › Abbildung der „Geschäftslogik“ durch berechnete Kennzahlen
- › Analysemodelle für Vertrieb, Marketing, Controlling, Planung
- › Zugriff via Excel Pivot, Power View, Reporting, Third Party ...
- › Rollenbasierte Steuerung der Zugriffsrechte
- › Mehrsprachen-Unterstützung und Währungsumrechnung
- › Hohe Benutzerakzeptanz durch kurze Antwortzeiten
- › Advanced Analytics mit Data Mining Modellen
- › Integration in Active Directory



Azure Analysis Services



- › Momentan nur Tabular (1200) unterstützt.
Keine multidimensionalen Modelle
- › MDX and DAX query support
- › VertiPaq and DirectQuery Mode supported
- › Ideal for hybrid and cloud solutions (on premises data gateway)
- › Powerful bi-directional relationship support
- › KPIs, translations, perspectives, row level security
- › Familiar developer tools: SQL Server Data Tools (SSDT) & SQL Server Management Studio (SSMS)
- › Bis zu 100 Gbyte Arbeitsspeicher, verschiedene Tiers wählbar
- › Pausieren und elastisch skalieren
- › AAD Auth



<https://docs.microsoft.com/en-us/azure/analysis-services/analysis-services-overview>

Fazit Semantisches Modell

Analysis Services



Microsoft®
SQL Server®

Erprobtes Tooling für
OLAP Cubes und
„erwachsene“ Power BI
Modelle (Tabular)

Azure Analysis Services



Managed Tabular Modelle
für Cloud und on premises
Datenquellen. OLAP leider
noch nicht unterstützt.

Berichte und Dashboards

Visualisierung der Kennzahlen und Daten

On Premises



SQL Server Reporting Services

Paginated / Mobile



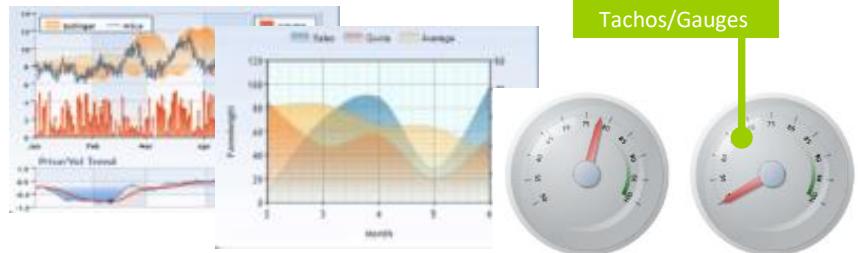
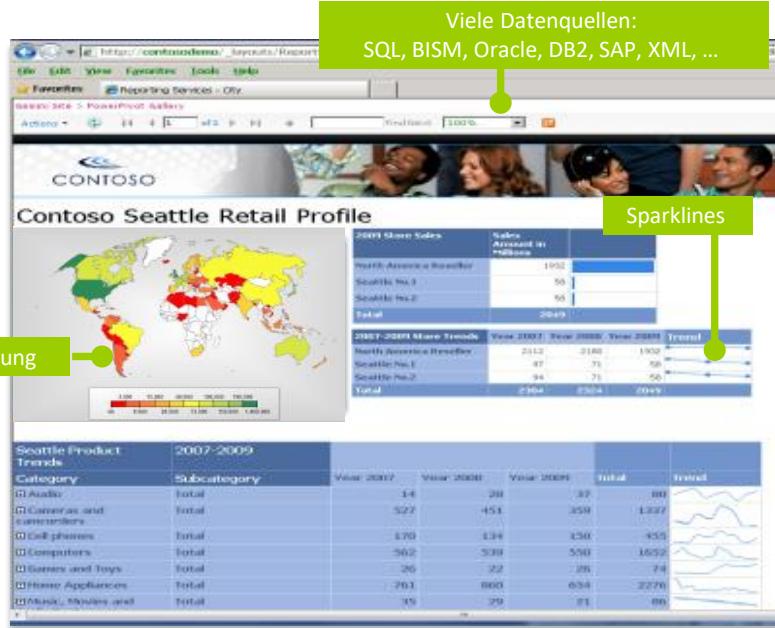
Azure/Productivity Cloud



Power BI

Reporting Services (paginated)

- › Webbasierte Standard- und Ad-Hoc-Berichte mit beliebigen Datenquellen
- › Ausgabeformate: PDF, Excel, Word, TIFF, XML, ...
- › Optimal in Verbindung mit SharePoint
- › Report Builder für die Fachabteilung
- › „pixel perfect“-Ausdrucke von Geschäftsberichten
- › Automatische Berichtsverteilung über Abos und Benachrichtigung
- › Reports als Quelle für Self-Service-BI/Power Pivot
- › Eingebaute Wiederverwendbarkeit von Grafiken, Abfragen ...



Reporting Services (mobile)

- › Ehemals Datazen
- › Rapidly publish BI
- › Perfect experience on any device
- › Mobile BI designer
- › Integrated publishing system
- › Support native Apps for all platforms
- › Enable advanced data access
- › Real-time queries
- › Output caching and offline support
- › Personalized results
- › Integrate with existing infrastructure





Power BI

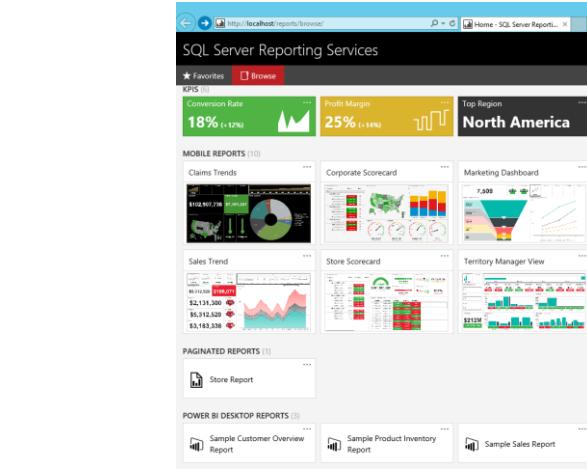
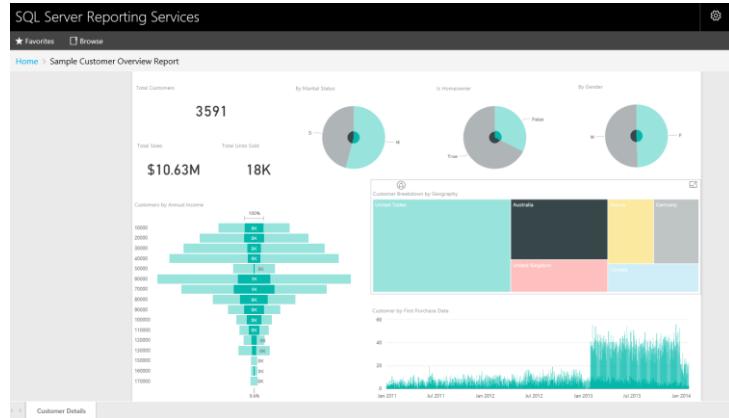
- › Explorative Analyse von Geschäftsdaten
- › Kombination verschiedener Datentöpfe
- › Interaktive Visualisierung
- › Teilen mit Kollegen
- › Datenspeicherung In-Memory Datenbank, Direct Access oder Realtime-Push/Streaming Sets
- › Support von nativen Apps aller mobile platformen
- › Power BI Desktop oder Excel für Entwicklung
- › Power BI Embedded
- › Rest APIs
- › Email Zustellung derzeit im Preview
- › Im Portal: Dashboards, Q&A, Quick Insights
- › Aber: Zum Drucken eher ungeeignet



Reporting Services

(Preview Power BI Integration)

- › Interaktive explorative Analyse
- › Geplant für SQL Server vNext
- › PBI Reports im SSRS Portal
- › Auf der Roadmap: Query Integration
- › Nur Reports, derzeit keine Roadmap für weitere Funktionen von powerbi.com
- › Keine: Dashboards, Q&A, Streaming/Push Datasets, Quick Insights usw.



Fazit Berichte und Dashboards

Reporting Services



Microsoft®
SQL Server®

Berichte zum Ausdrucken und
Dashboards für mobile
Devices, komplett wenn auch
Power BI Reports integriert

Power BI



Dashboards und explorative
Analyse gut unterstützt, für
klassisches Berichtswesen
eher ungeeignet

Scheduler

Starten von Verarbeitungsschritten

On Premises



SQL Server Agent



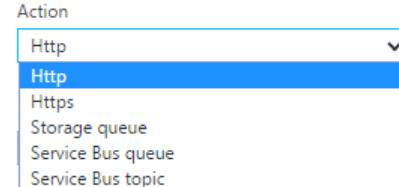
Azure/Productivity Cloud

Azure Data Factory

Azure Automation

(Azure Schedule)

Kann zu wenig



SQL Server Agent

- › Aufrufen von SSIS Paketen, Cube Processings und anderen
- › Workflow-Unterstützung (bei Fehler gehe zu...)
- › Gute Integration der BI Tools (drag und drop zur Auswahl)
- › Benachrichtigungen an Operatoren
- › Volle Integration in DBMS, alles auch per TSQL Script ausführbar

The screenshot shows the SQL Server Management Studio (SSMS) interface. On the left, the Object Explorer pane displays the database structure of 'INOVEXSQL2016' (SQL Server 13.0.2186.6 - INOVEXSQL2016\sqladmin). In the center, the 'SQL Server Agent' node under 'Management' is selected, and a context menu is open, with 'New Job...' highlighted. To the right, the 'Properties' dialog for the 'SQL Server Integration Services Package' is displayed, showing the 'Step name:' field set to 'test' and the 'Type:' dropdown set to 'Transact-SQL script (T-SQL)'. Below this, a list of job types is shown, with 'SQL Server Integration Services Package' selected. At the bottom of the dialog, there are 'OK', 'Cancel', and 'Apply' buttons.

Object Explorer

Connect ▾

INOVEXSQL2016 (SQL Server 13.0.2186.6 - INOVEXSQL2016\sqladmin)

- Databases
- Security
- Server Objects
- Replication
- PolyBase
- AlwaysOn High Availability
- Management
- Integration Services Catalogs
- SQL Server Agent

New Job...

Manage Schedules

Manage Job Categories

View History

Filter

Start PowerShell

Reports

Refresh

Step name:

test

Type:

Transact-SQL script (T-SQL)

Operating system (CmdExec)

PowerShell

Replication Distributor

Replication Merge

Replication Queue Reader

Replication Snapshot

Replication Transaction-Log Reader

SQL Server Analysis Services Command

SQL Server Analysis Services Query

SQL Server Integration Services Package

Transact-SQL script (T-SQL)

Open

Select a page

All Jobs

Status

Last Refresh: 11/30/2012 4:20:48 PM

Next Refresh: Manual

Filter: None

Connection

Server: KMUGHAL\SQL2012

Agent Job Activity:

Name	Enabled	Status	Last Run	Next Run	Category
Backup	Enabled	Idle	Succes...	11/30/...	Data...
sysjob	Enabled	Stop Job	Idle	never	Uncat...

Refresh Filter ... Help

New Job...

Start Job at Step...

Stop Job

Enable Job

Disable Job

Refresh job

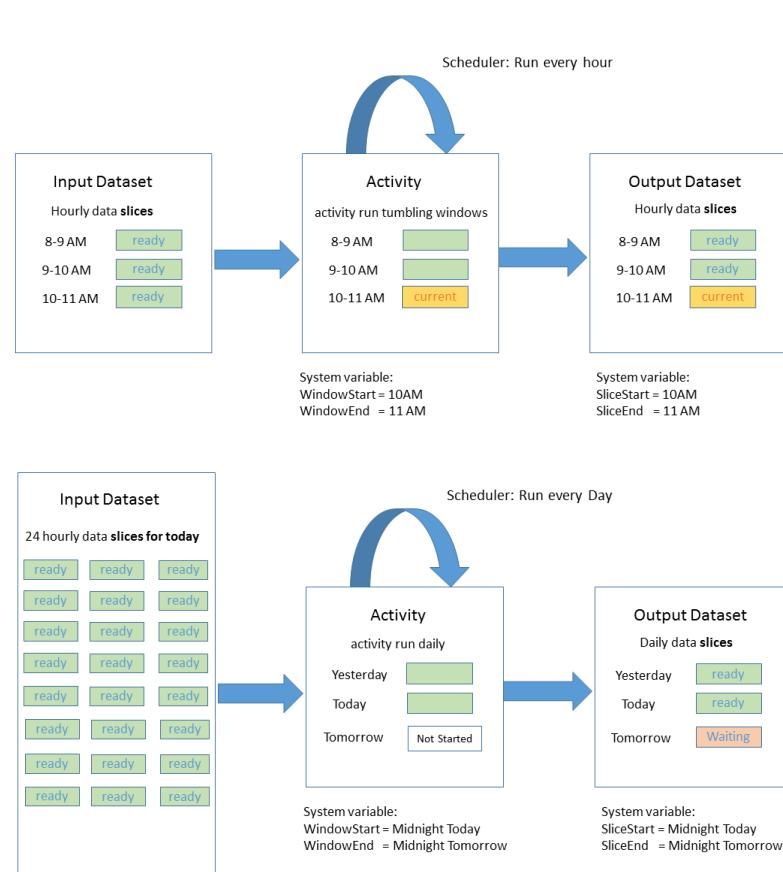
Delete job

View history

Properties

Azure Data Factory

- › Data Factory hat eigenes Konzept mit Time Slices
- › Ausgefeilt und deckt vieles ab, z.B. Check alle 24 h Slices für Vortag vorhanden vor Aggregation
- › eher Use Case abhängig brauchbar
- › Relativ kompliziert einzustellen
- › Kann nur Data Factory Pipelines/Activities triggern
- › z.B. Kein Processing von SSAS Cubes





Azure Automation

- › Runbooks für das verwalten, schedulun und definieren von Jobs
- › Runbooks enthalten command script JSON Documente mit PowerShell Kommandos
- › Gallery enthält cmdlets Sammlungen, auch zu ADF und Analysis Services
- › Vorgegebene oder eigene PS Runbooks hochladen, ggf. editieren und schedulen

AzureRM.AnalysisServices

[View Source](#) [Import](#)

Microsoft Azure PowerShell - Analysis Services

Created by: azure-sdk
Tags: Azure ResourceManager ARM AzureAS PSModule
Dependencies: AzureRM.Profile (= 2.5.0)
[View Source Project](#)

Learn more
[View in PowerShell Gallery](#)
[Documentation](#)
[Licensing information](#)

Content

TYPE	NAME
Cmdlet	Resume-AzureRmAnalysisServicesServer
Cmdlet	Suspend-AzureRmAnalysisServicesServer
Cmdlet	Get-AzureRmAnalysisServicesServer
Cmdlet	Remove-AzureRmAnalysisServicesServer
Cmdlet	Set-AzureRmAnalysisServicesServer
Cmdlet	Test-AzureRmAnalysisServicesServer
Cmdlet	New-AzureRmAnalysisServicesServer

New Schedule

Name: every day 6

Description:

Starts: 2017-03-15 6:00:00 AM

Timezone: Germany - Central European Time

Recurrence: Once Recurring

Recur every: 1 Day

Set expiration: Yes No

Browse Gallery

Search: datafactory

AzureRM.DataFactories

Microsoft Azure PowerShell - DataFactories service cmdlets for Azure Resource Manager

Tags: Azure ResourceManager ARM DataFactory PSModule

Guter Blogeintrag von Jorg Klein dazu mit Beispiel: <http://bit.ly/2gi5fss>

Beispiel: Ausführen einer Data Factory Pipeline

Check in Test pane Feedback

```
1 # Variables; modify
2 $rgn = "sqlconf2017" #Resource Group Name
3 $acn = "infstefankirneraad" #Automation Credential Name -has to be an AAD Account, no Microsoft Account
4 $dfn = "sqlkonf2017adf" #Data Factory Name
5 $pln = "sqlkonf2017adfp1" #Pipeline Name
6 $dsn = "OutputDataset-uwo" #DataSet Name (output dataset of pipeline that needs to be produced)
7
8 # To test from PowerShell client, uncomment the 2 rows below and provide subscription ID
9 #Login-AzureRmAccount
10 #Set-AzureRMContext -SubscriptionId "f1019f96-8da6-418a-93a8-95685e05c2e7"
11
12 # Authenticate
13 # To test from PowerShell client, comment out the 2 rows below
14 $AzureCred = Get-AutomationPSCredential -Name $acn
15 Add-AzureRmAccount -Credential $AzureCred | Out-Null
16
17 # Get data factory object
18 $df=Get-AzureRmDataFactory -ResourceGroupName $rgn -Name $dfn
19 If($df) {
20     Write-Output "Connected to data factory $dfn in resource group $rgn."
21 }
22
23 # Create start/end DateTime (yesterday)
24 $sdt = [System.DateTime]::Today.AddDays(-1) #Yesterday 12:00:00 AM
25 $edt = [System.DateTime]::Today.AddSeconds(-1) #Yesterday 11:59:59 PM
26
27 # Update active period to yesterday
28 $apr=Set-AzureRmDataFactoryPipelineActivePeriod -DataFactory $df -PipelineName $pln -StartTime $sdt -
29 If($apr) {
30     Write-Output "Pipeline $pln of data factory $dfn updated with StartDateTime $sdt and EndDateTime :
31 }
32
```

Completed

Pipeline sqlkonf2017adfp1 of data factory sqlkonf2017adf updated with StartDateTime 02/10/2017 00:00:00 and EndDateTime 02/10/2017 23:59:59.

Pipeline sqlkonf2017adfp1 resumed.

Slice status is:

State

InProgress
Wait 15 seconds

Slice status is:

State

InProgress
Wait 15 seconds

Slice status is:

State

InProgress
Wait 15 seconds

Slice status is:

State

InProgress
Wait 15 seconds
Slice finished.
Pipeline sqlkonf2017adfp1 paused.

Fazit Scheduling

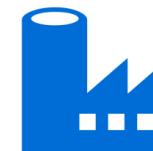
SQL Agent



Microsoft®
SQL Server®

Einfach zu bedienendes Tool mit hohem Integrationsgrad in SQL BI Stack.

Azure Data Factory



Integrierter Scheduler gut für komplexe Abhängigkeiten, aber eingeschränkte Funktionen

Azure Automation



Ausführen von Powershell Runbooks bietet viele Möglichkeiten, ist aber komplexer als SQL Agent Jobs

Advanced Analytics

How about data science?

On Premises



SQL Server Analysis Services

SQL Server R Integration



Azure/Productivity Cloud

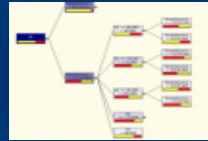
 **Azure Machine Learning**

 **R Server on HDInsight**

SQL Server Analysis Services

- › Idee
 - › Nutzung statistischer Verfahren und Algorithmen aus der künstlichen Intelligenz zur Auffindung von Mustern in Datenbeständen
- › Merkmale
 - › Einige Algorithmen enthalten und parametrisierbar, aber ziemlich veraltet
 - › Visual Studio & Excel für Erstellung, Training, Visualisierung und Validierung
 - › Daten aus allen gängigen Datenbanken nutzbar
 - › Integration in den gesamten BI-Stack und DMX-Query-Sprache
- › Nutzen & Vorteile
 - › Geringe Einstiegskosten, da in SQL Server Lizenz enthalten
 - › Einfacher Zugang durch Excel-Add-Ins (Table Analysis & Data Mining)
 - › Nutzung von Data Mining in Excel auch ohne fundiertes Know-How

Data Mining- und Vorhersage-Modelle mit SQL Server und Excel



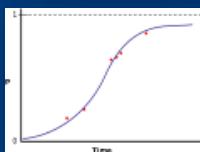
Entscheidungsbäume



Clustering



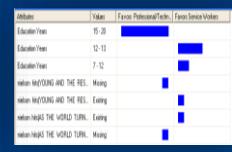
Zeitreihenanalysen



Regression



Sequence Clustering



Naïve Bayes



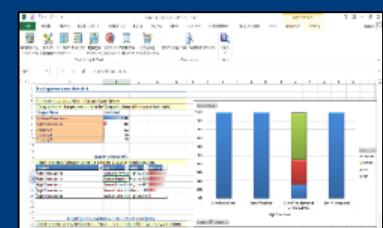
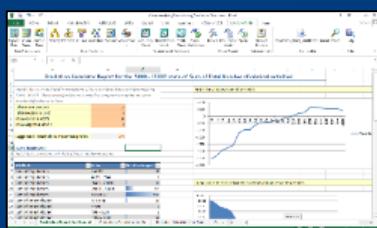
Neuronale Netze



Assoziation

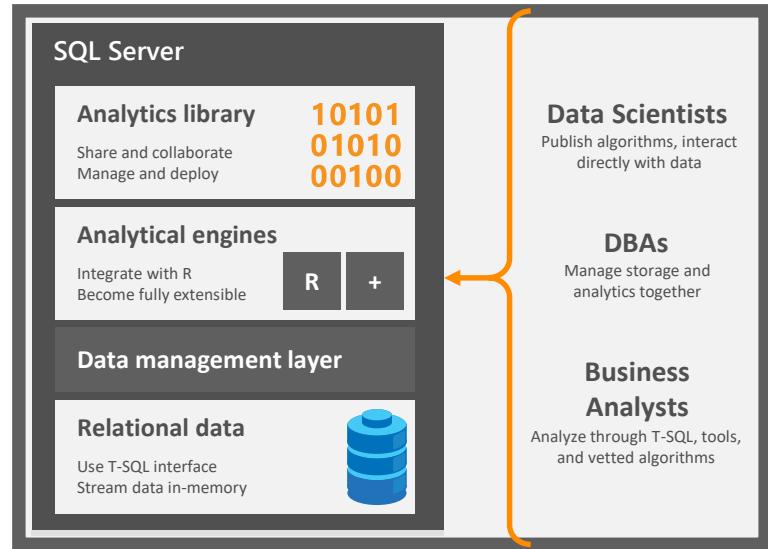


Third-Party



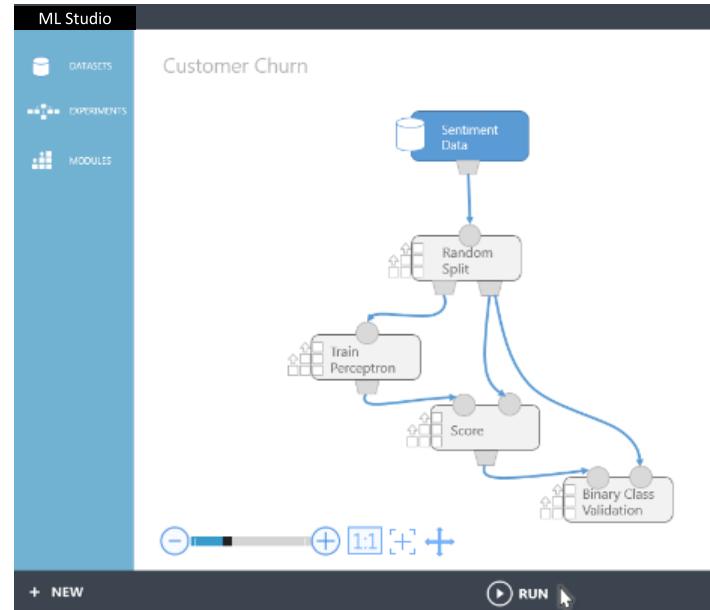
SQL Server 2016 R integration scenario

- › Exploration
 - › Use Revolution R Enterprise (RRE) from R integrated development environment (IDE) to analyze large data sets and build predictive and embedded models with compute on SQL Server machine (SQL Server compute context)
- › Operationalization
 - › Developer can operationalize R script/model over SQL Server data by using T-SQL constructs
 - › DBA can manage resources, plus secure and govern R runtime execution in SQL Server



Azure Machine Learning

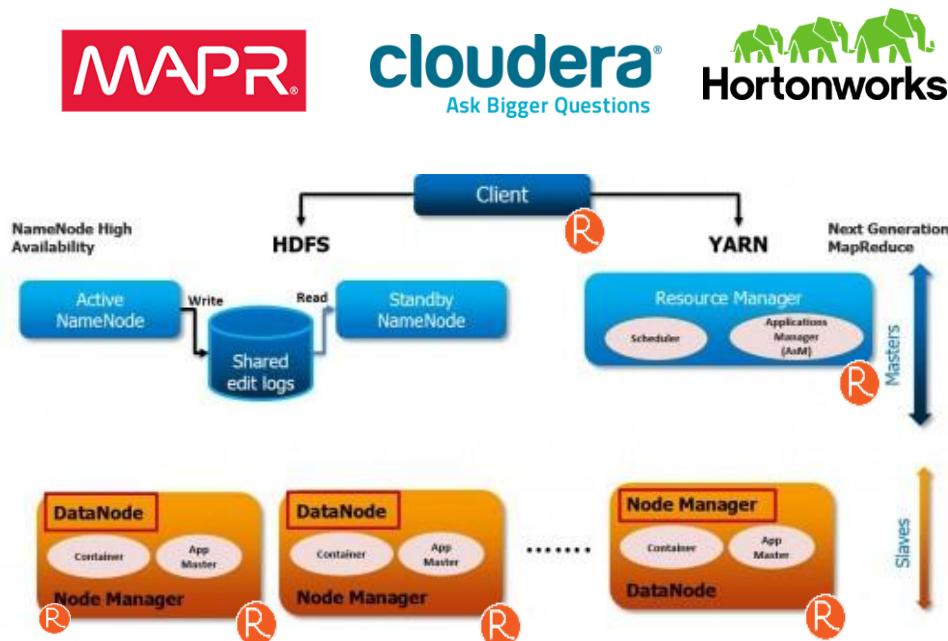
- › Cloud service with 3 components
 - › Environment to build predictive models
 - › Run Time Environment for models with SLA
 - › Gallery / Templates for predictive models
- › Targets Data Scientists & Business Analysts
 - › Data Scientists build models
 - › Business Analysts consume the models
- › Key Differentiators
 - › Web based authoring and easy deployment
 - › Collaboration and Isolation
 - › Intuitive UI, Integration of „R“ and Phyton



R on Hadoop

(also on HDInsight as PaaS)

- › Uses Hadoop nodes for R computations
- › Eliminate data movement latency on very large data
- › Remove data duplication
- › Faster model development
- › No MapReduce R coding
- › Develop better models using all the data



R on HDInsight

Cluster configuration

* Cluster type ⓘ R Server * Operating system Linux * Version R Server 9.0 (HDI 3.5)

* Cluster tier ⓘ STANDARD PREMIUM

R Server : Terabyte-scale, enterprise grade R analytics with transparent parallelization on top of Spark and Hadoop.

Configuration Options:

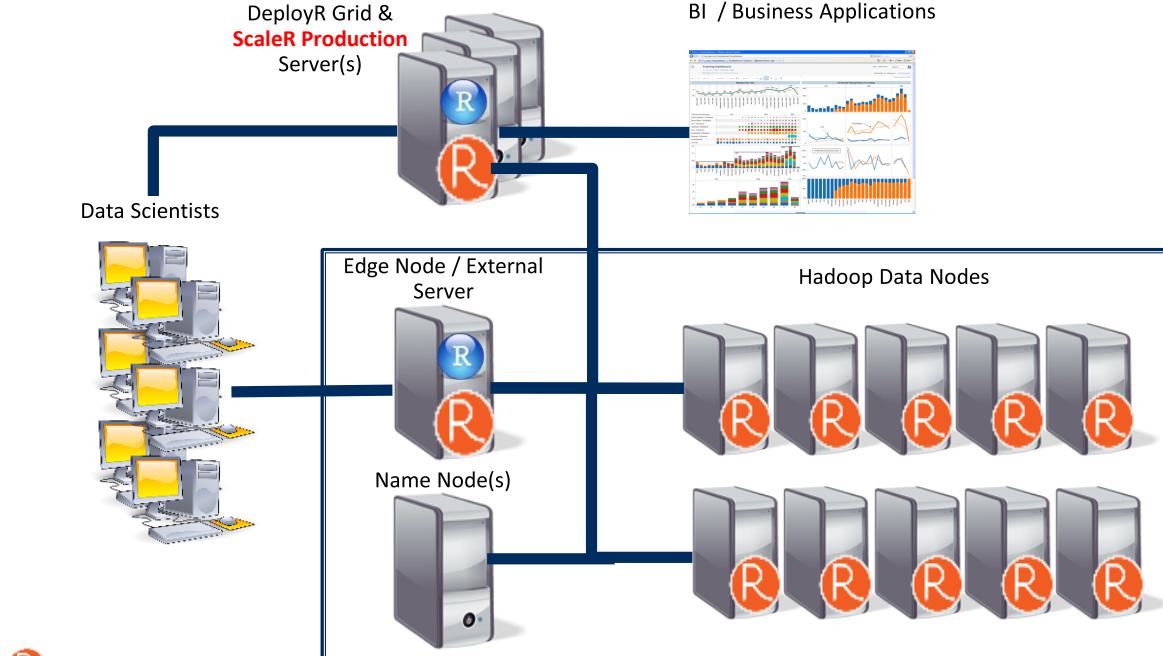
- R Server 9.0 on Spark 2.0 with Java 8
- R Server 8.0 on Spark 1.6 with Java 7

Adds \$0.06746 per Core-Hour.

Features

* denotes preview feature

Available	Not available
<input checked="" type="checkbox"/> R Studio community edition for R Server	+ Apache Ranger* (PREMIUM) ⓘ
+ Secure shell (SSH) access	+ Domain joining* (PREMIUM) ⓘ
+ HDInsight applications	+ Remote Desktop access ⓘ
+ Custom virtual network	
+ Custom Hive metastore	
+ Custom Oozie metastore	



Fazit Advanced Analytics

SQL Server Analysis Services



Data Mining Modul mit guter Integration in den MS BI Stack aber veraltete Algorithmen

SQL Server R Integration



R Script Ausführung direkt in der DB ist effizient und kann gut operationalisiert werden.

Azure Machine Learning



Modernes Tooling für den Data Scientisten, Webservices für Zugriff von außen, Erweiterbarkeit über R und Python

R on HDInsight

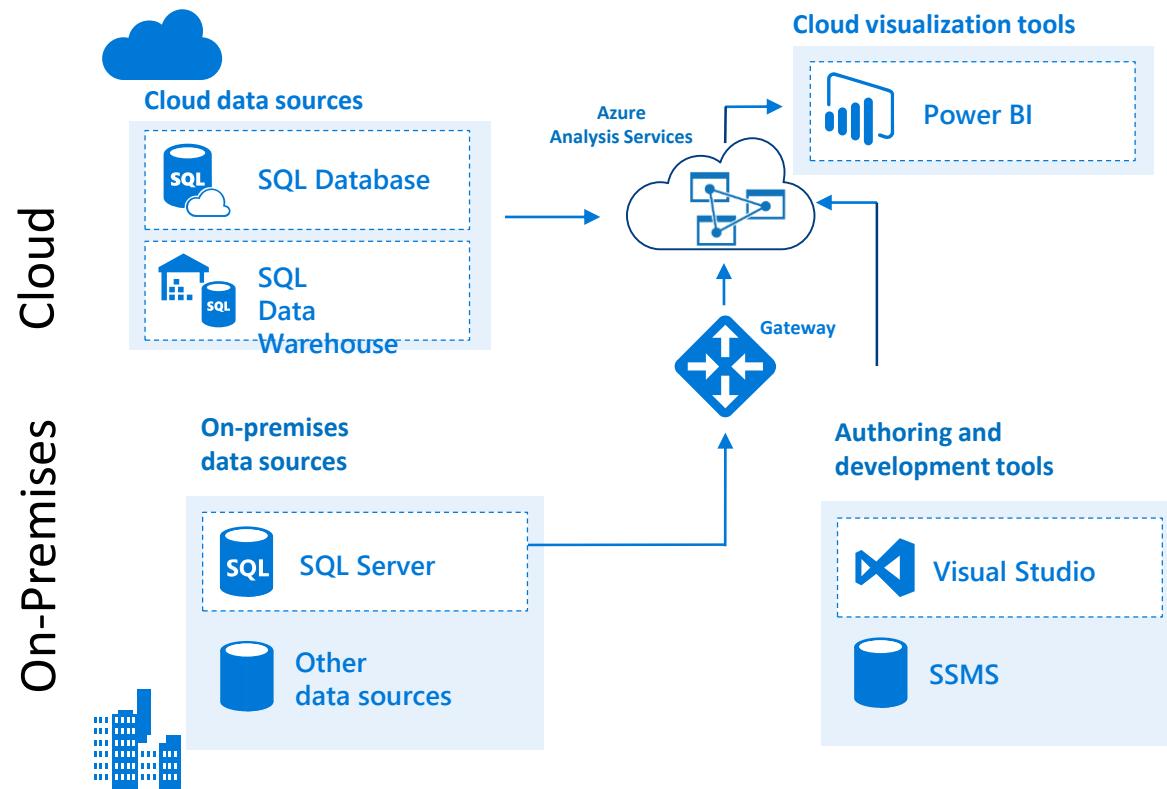


Lösung für R Skalierung über HDInsight Spark Knoten, geeignet für sehr große Datenmengen

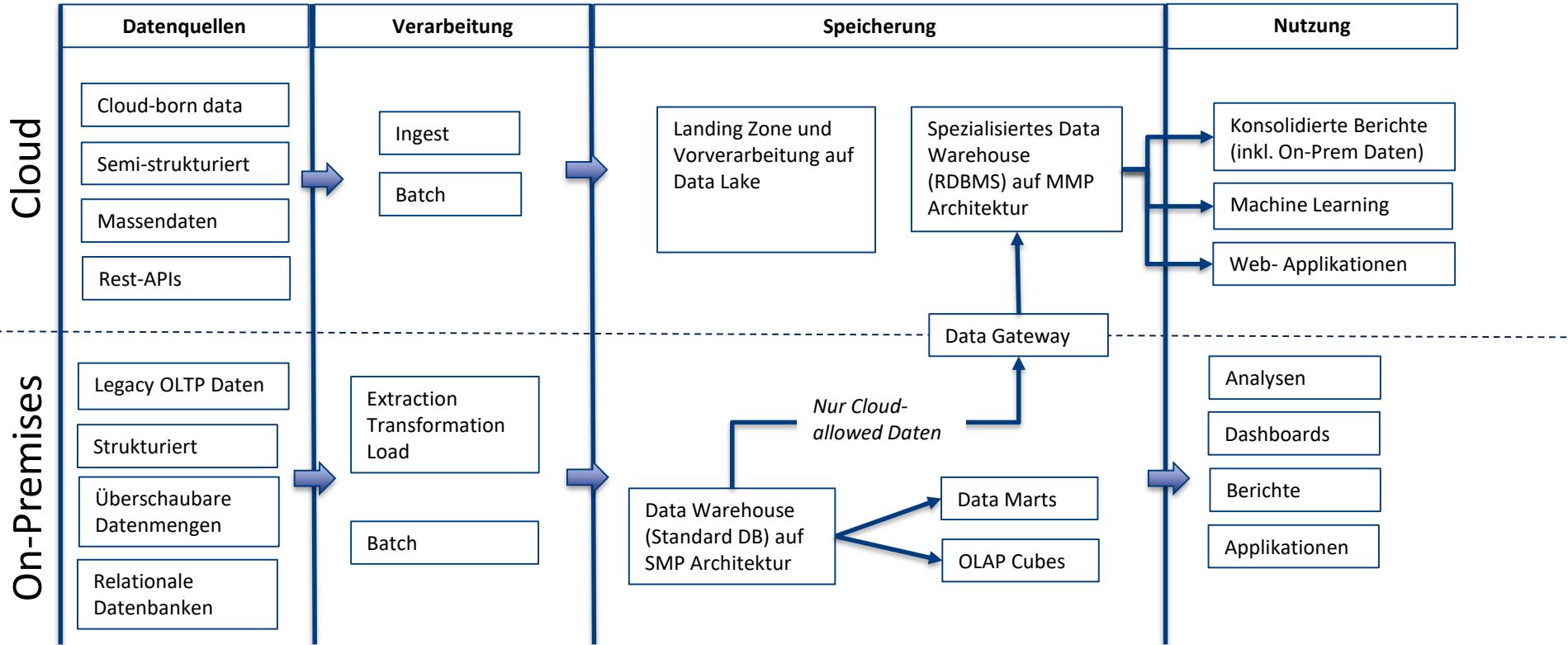
Hybride Szenarien

Azure Analysis Services als Data Hub

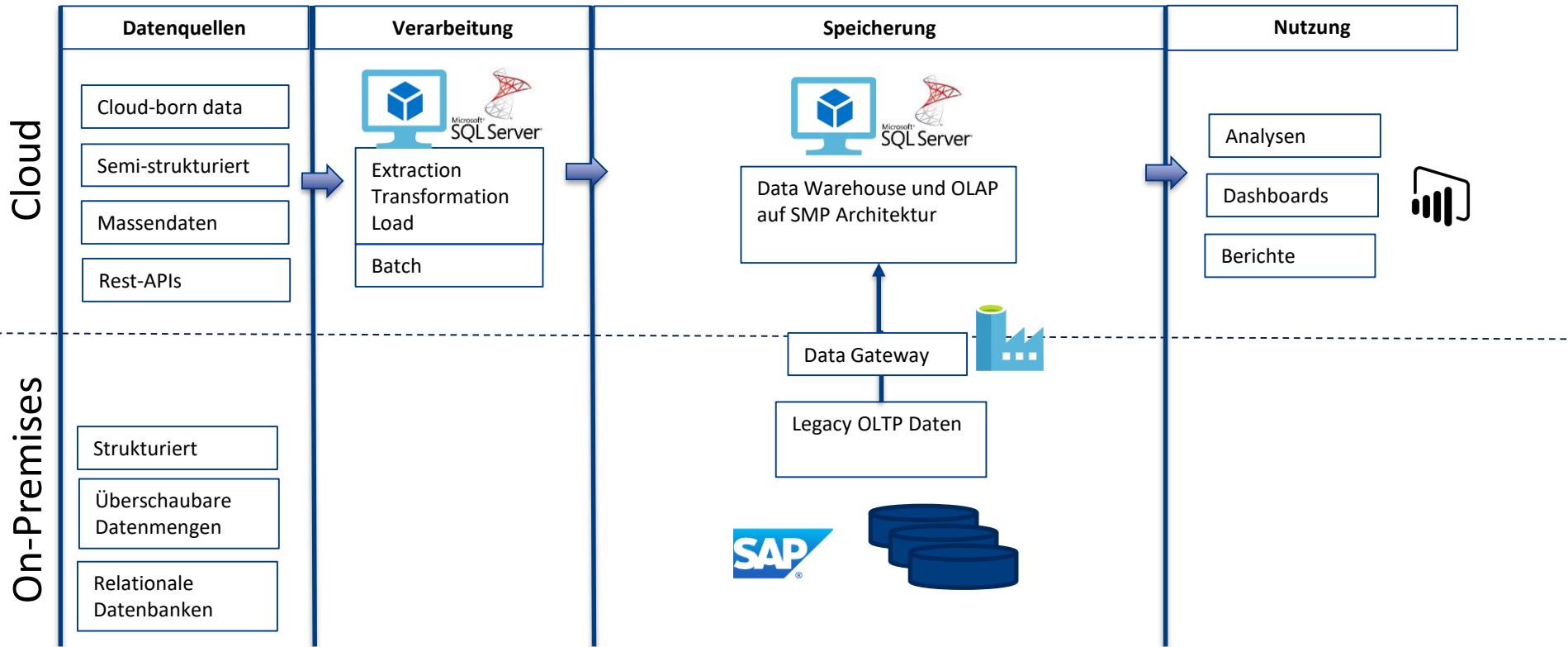
Analysis Services als
Data Hub für Cloud
und on-premises
Datenquellen und
semantische
vorgelagerte Schicht
vor Power BI



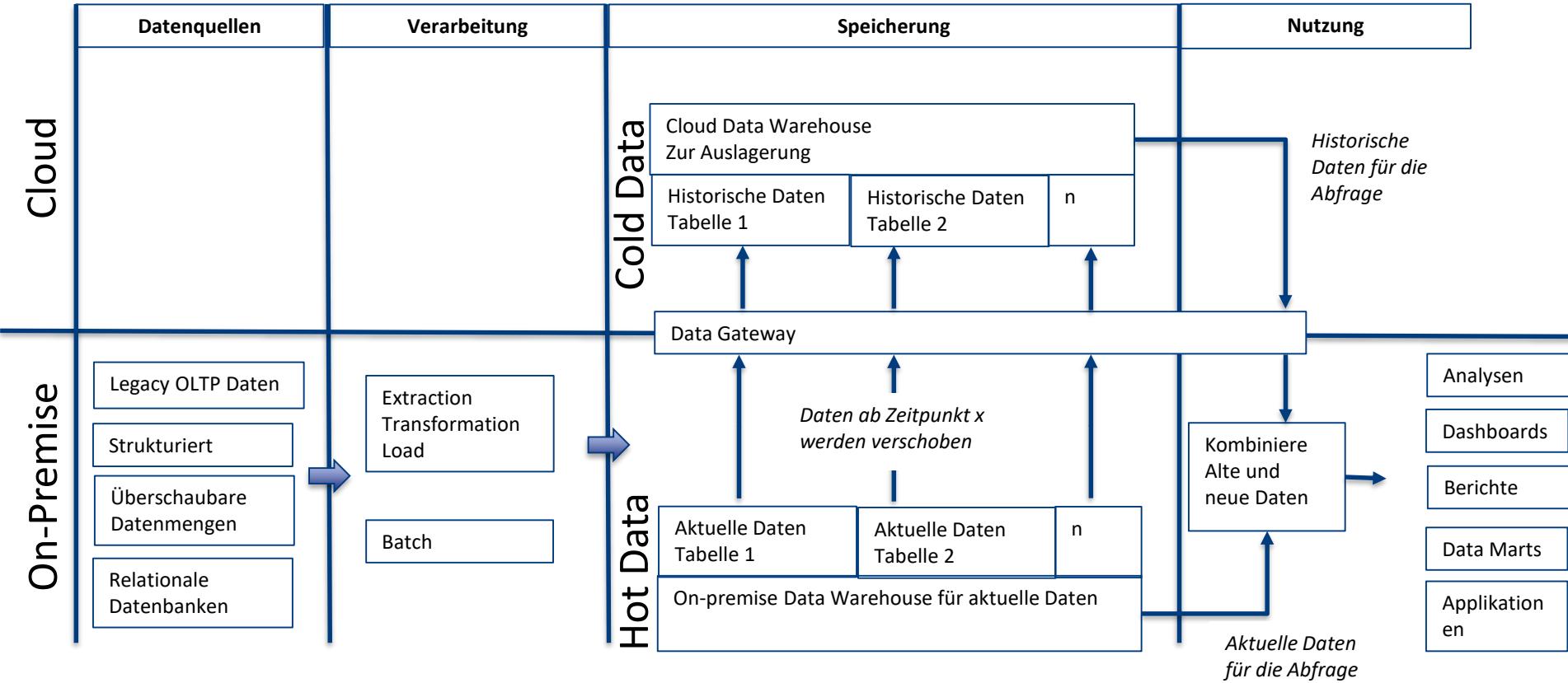
Skalierbarkeit in der Cloud



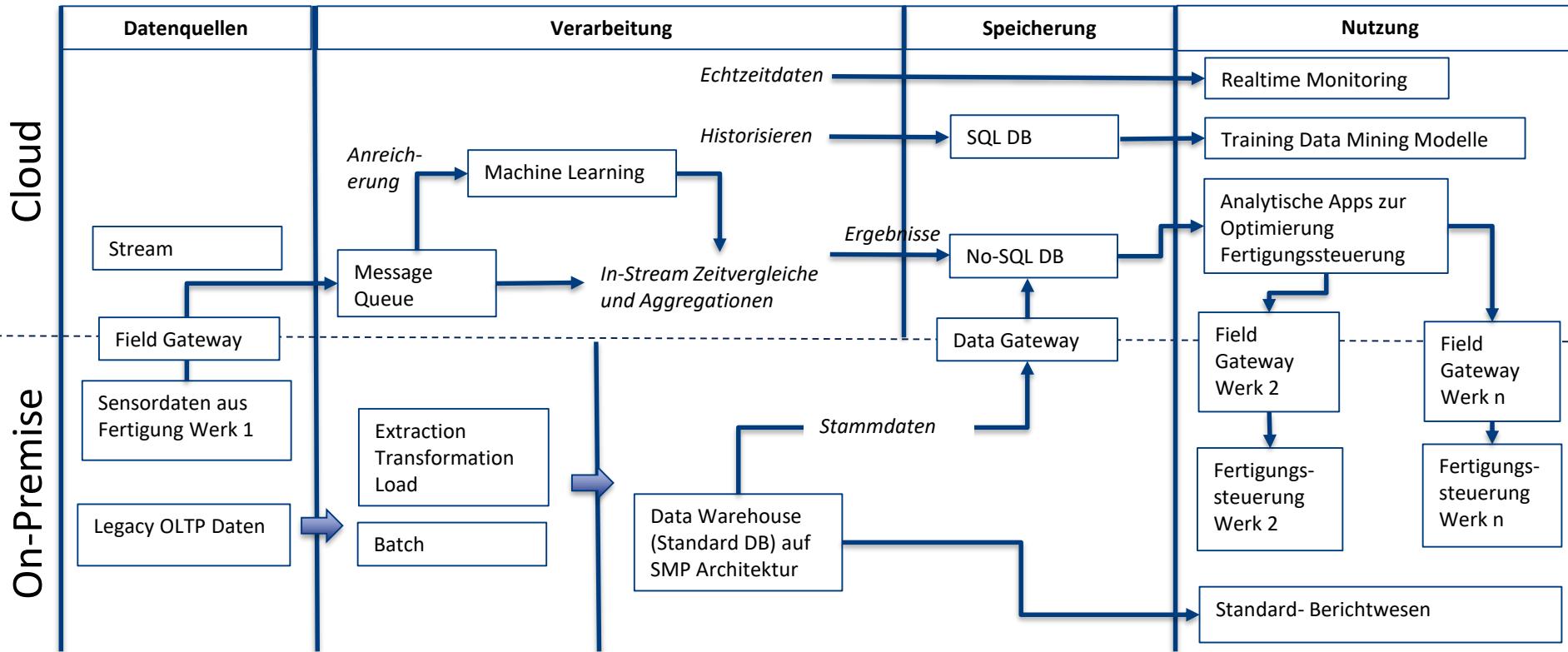
In the Mix – SQL Server on Azure IaaS & Power BI



Alle Daten im Zugriff – Stretch DB



Technologievielfalt für die Optimierung der Prozesse



Fazit

- › Elastizität, Modernität und Innovationsgeschwindigkeit sprechen für Azure Cloud!
- › Aber: Höherer Reifegrad SQL Server Services, hohe Akzeptanz in Rechenzentren und natürlich die entfallende Datenschutzdiskussion sind klare Vorteile. Release-Zyklen werden kürzer ;-)
- › Hybriden Architekturen zeigen Kombination und Integration der beiden Welten...best of breed is coming!



Durchstarten bei inovex

Wir suchen Verstärkung in den Bereichen

- Product Ownership
- Web Frontend & Backend
- Replatforming
- Apps & Smart Devices
- Big Data
- Data Science
- Search
- Business Intelligence
- DevOps
- Cloud
- Operations

jobs@inovex.de

www.inovex.de/jobs

+49 721 619 021 50

