



Data Engineering with Azure

Steve Young - Data & AI CSA (Microsoft)

styoung@Microsoft.com

<http://5minutebi.com>

<http://linkedin.com/in/steveyoung/>

Toronto PASS Event: January 23, 2020

Data Engineering with Azure



Steve Young, MCSE

Microsoft Azure Cloud Solution Architect (Data&AI) with 20+ years of hands-on consulting experience architecting and developing data solutions.



RSS
CAL
OUP

January 23, 2020

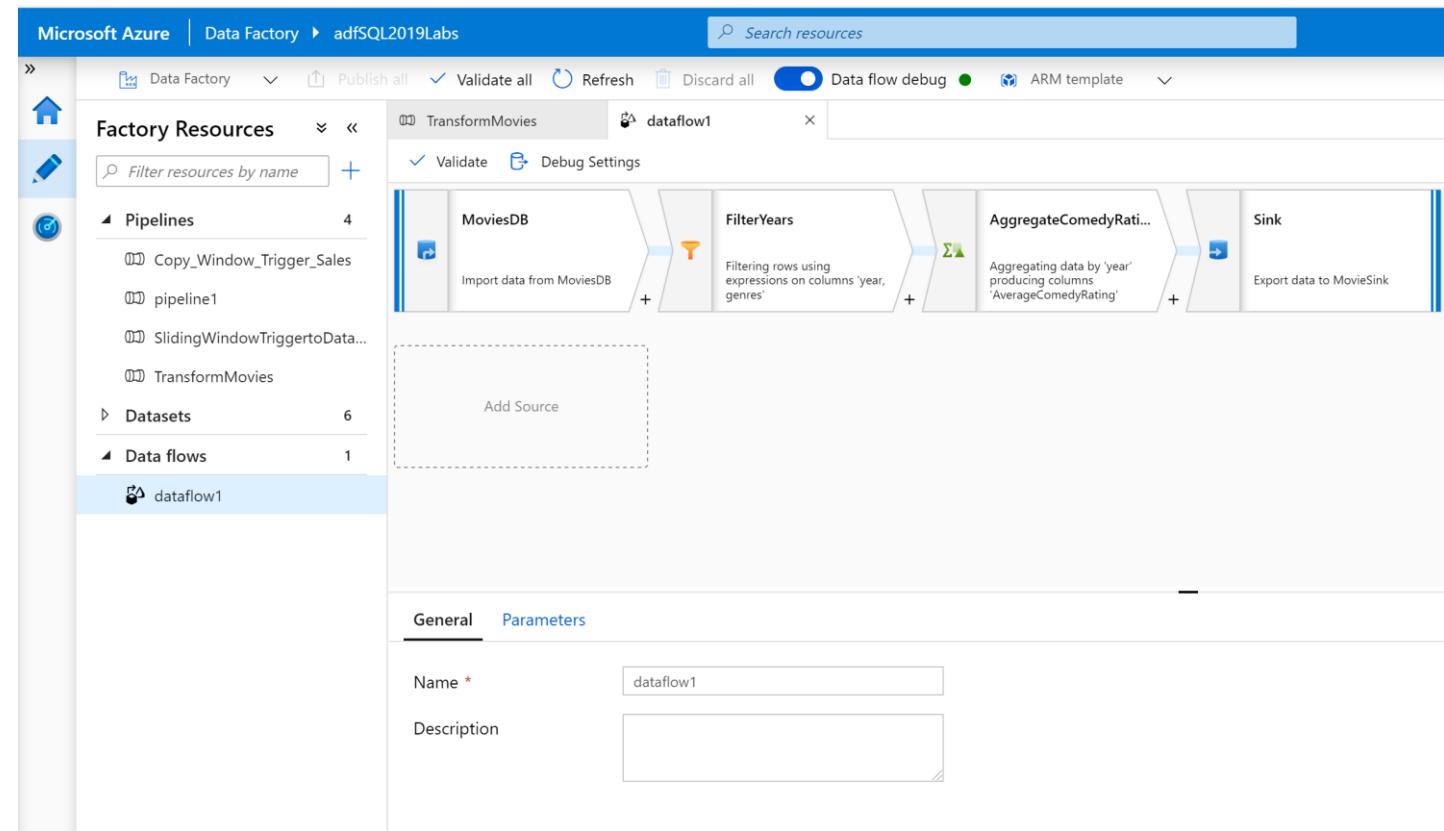
OBJECTIVES

- Introductions
- What is Data Engineering? – Definition & Intro
- Steve's 7 Principles of Data Engineering & Azure
- Walkthrough of Examples in Azure
 - SQL Notebooks
 - Azure Data Factory Sliding Window
 - Data Virtualization – Use Data Where it Resides
- Questions and hopefully answers ☺

The screenshot shows a Microsoft Learn course page for 'Azure for the Data Engineer'. At the top, there are navigation links: 'Docs' / 'Learn' / 'Browse' / 'Azure for the Data Engineer'. On the right, there are buttons for 'Bookmark', 'Theme', and 'Sign in', along with a '2500 XP' badge. The main content area features a blue hexagonal icon with a gear and code symbols. The title 'Azure for the Data Engineer' is displayed in bold, followed by '1 hr 51 min • Learning Path • 3 Modules'. Below the title are several tags: 'Beginner', 'Data Engineer', 'Azure', 'Azure Portal', 'Storage', and 'Virtual Machines'. A descriptive paragraph explains the course: 'Explore how the world of data has evolved and how the advent of cloud technologies is providing new opportunities for business to explore. You will learn the various data platform technologies that are available, and how a Data Engineer can take advantage of this technology to an organization benefit.' Underneath, 'Prerequisites' are listed as 'None'.

What is Data Engineering

<https://docs.microsoft.com/en-us/learn/patterns/azure-for-the-data-engineer/>



- Data Scientist
- Data Engineer
- Big Data Engineer
- BI Engineer
- BI Specialist
- Business Analytics Specialist
- Data Engineer
- Machine Learning Scientist
- Data Visualization Developer
- BI Solution Architect
- Analytics Manager

Security & Management

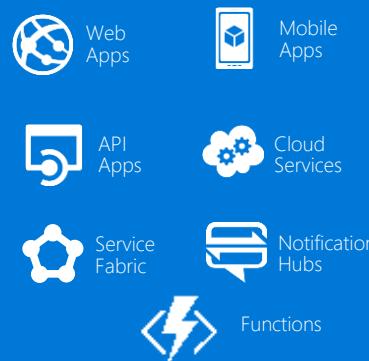
-  Security Center
-  Portal
-  Azure Active Directory
-  Azure AD B2C
-  Multi-Factor Authentication
-  Automation
-  Scheduler
-  Key Vault
-  Store/ Marketplace
-  VM Image Gallery & VM Depot

Platform Services

Media & CDN



Application Platform



Data



Integration



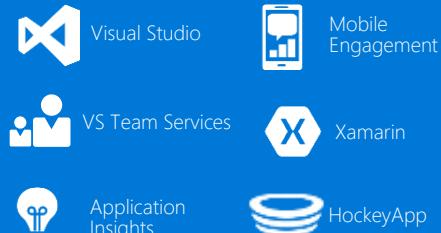
Intelligence



Compute Services



Developer Services



Analytics & IoT



Hybrid Cloud

-  Azure AD Health Monitoring
-  AD Privileged Identity Management
-  Domain Services
-  Backup
-  Operational Analytics
-  Import/Export
-  Azure Site Recovery
-  StorSimple

Compute



Storage



Networking



Datacenter Infrastructure

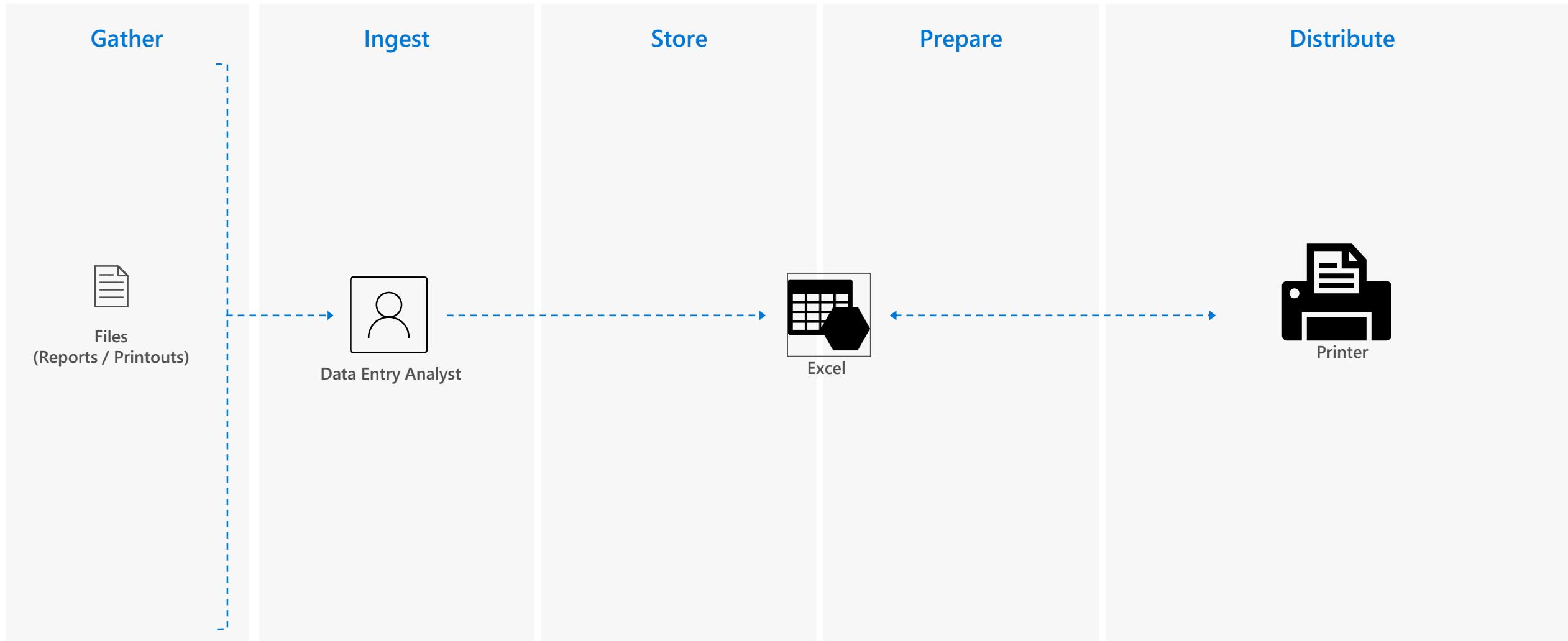


WHERE DID I START?

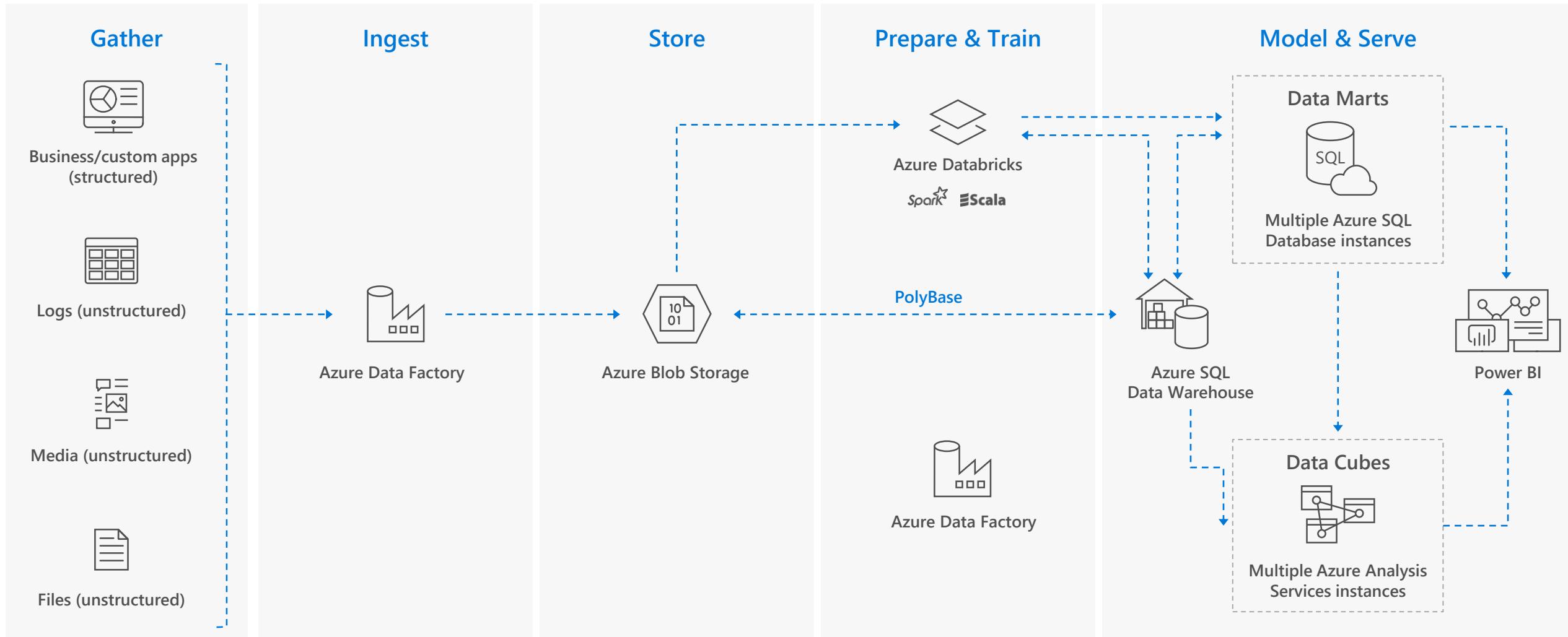
- Story about the Excel



WHERE DID I START? – HAVE WE REALLY CHANGED?



HUB & SPOKE ARCHITECTURE FOR BI – OUR JOBS ARE MORE COMPLEX

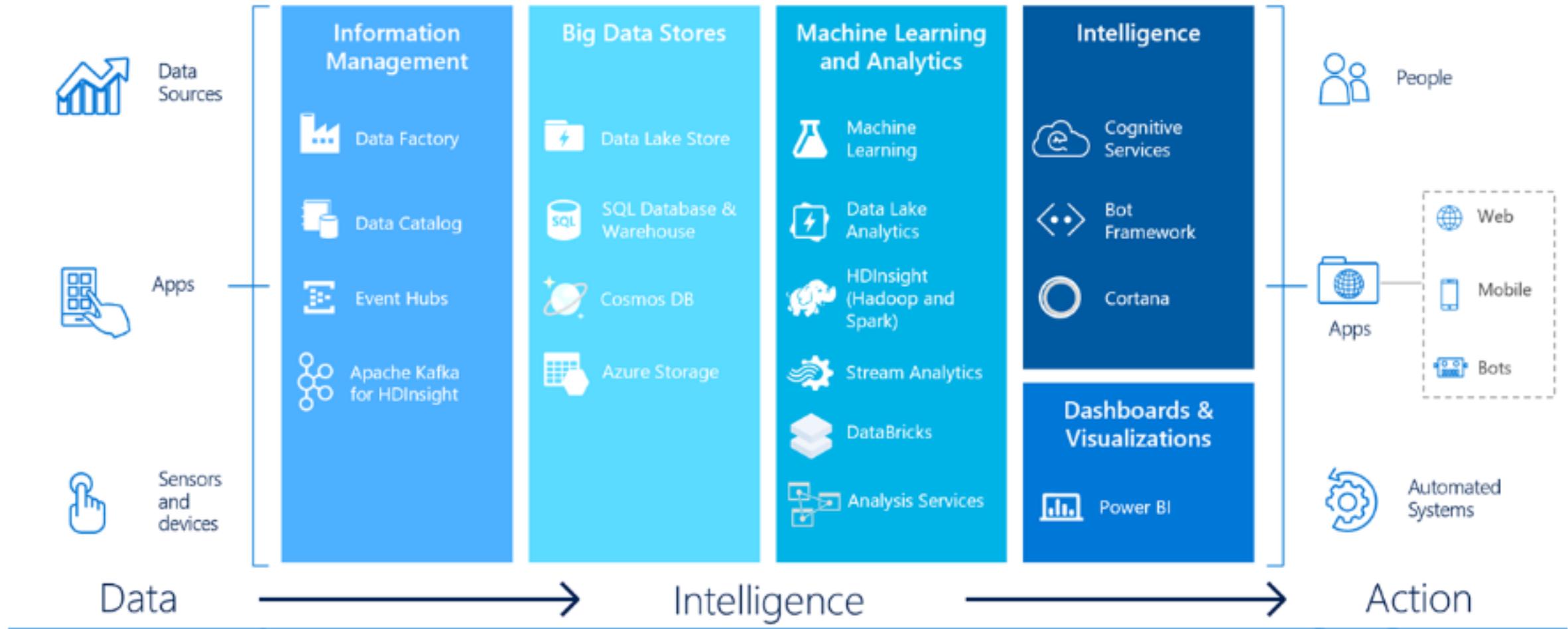


WHAT IS DATA ENGINEERING / A DATA ENGINEER?

- Boils down to being *responsible* for preparing data for retention and analysis
- Architect the data flow process
- Architect the technological environment
- Organize data flows
- Translates business requests into technical projects
- Out of all that what is the key word?



AZURE DATA PLATFORM



WE NOW HAVE MANY DIFFERENT TOOLS → BYOT!!

DevOps

Nagios®



Management

CHEF™



Applications



Pivotal



App frameworks & tools



nodeJS



APACHE
Spark™



Databases & middleware



cloudera



mongoDB



Couchbase



Infrastructure



SUSE



ORACLE
LINUX



docker



WE NOW HAVE MANY DIFFERENT TOOLS → OPEN SOURCE

SQL 2019 – Not Just On-Prem Anymore

DevOps

Nagios®



Management

SCALR
CLOUD MANAGEMENT

Applications



App frameworks & tools

AZURE CONTAINER SERVICE

Databases & middleware



cloudera



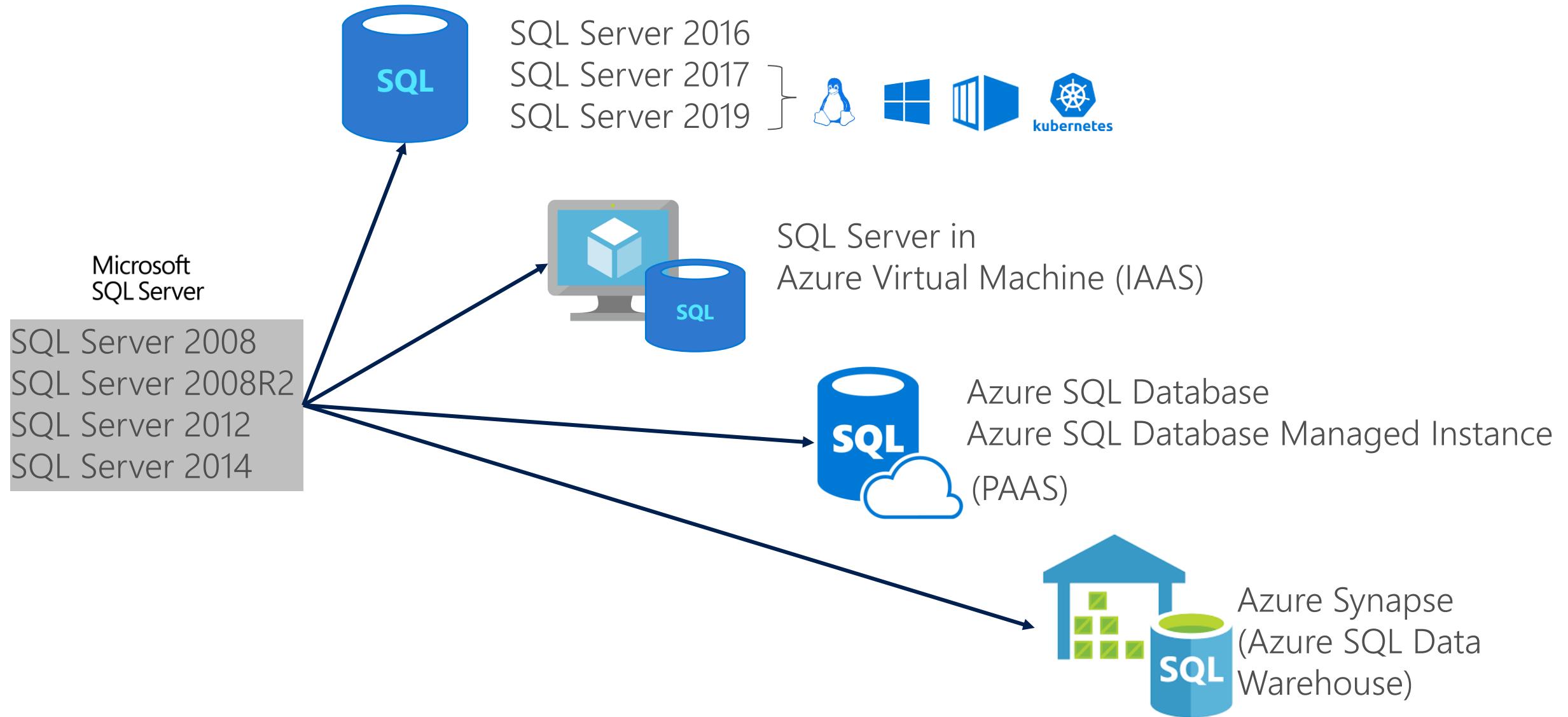
Infrastructure



ORACLE
LINUX



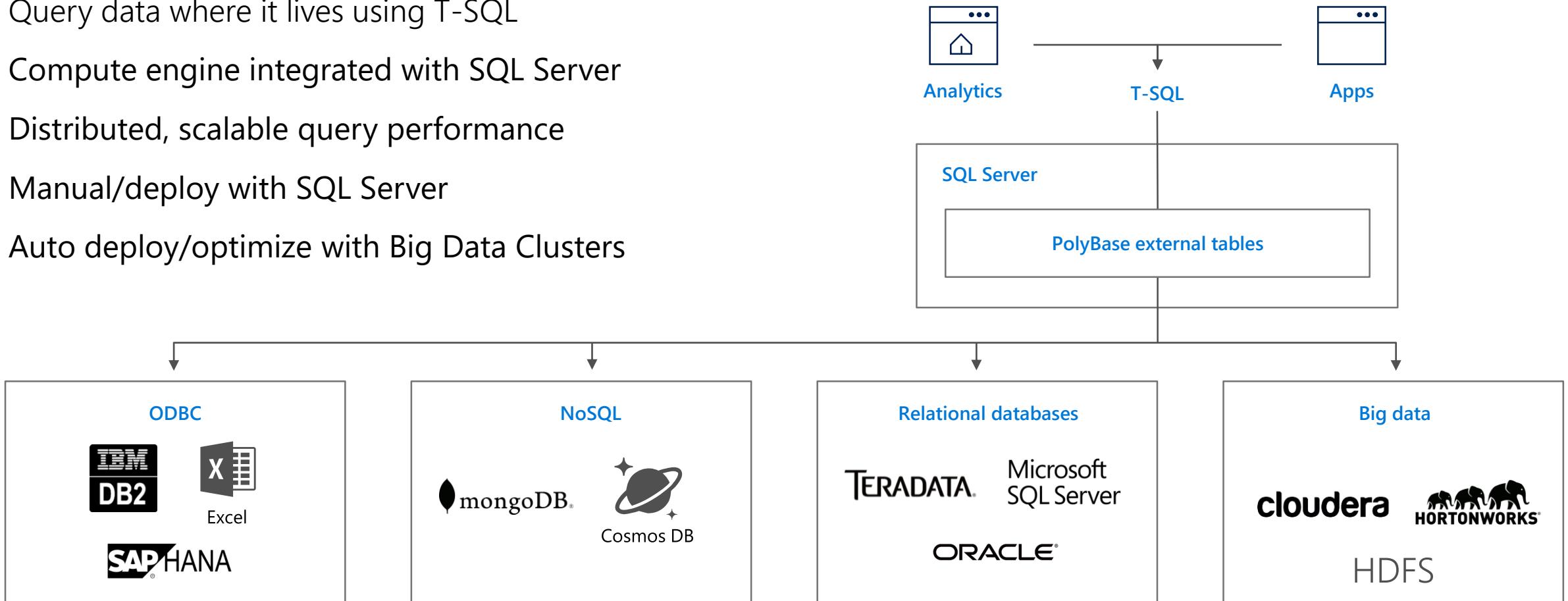
AZURE DATA MODERNIZATION CHOICES



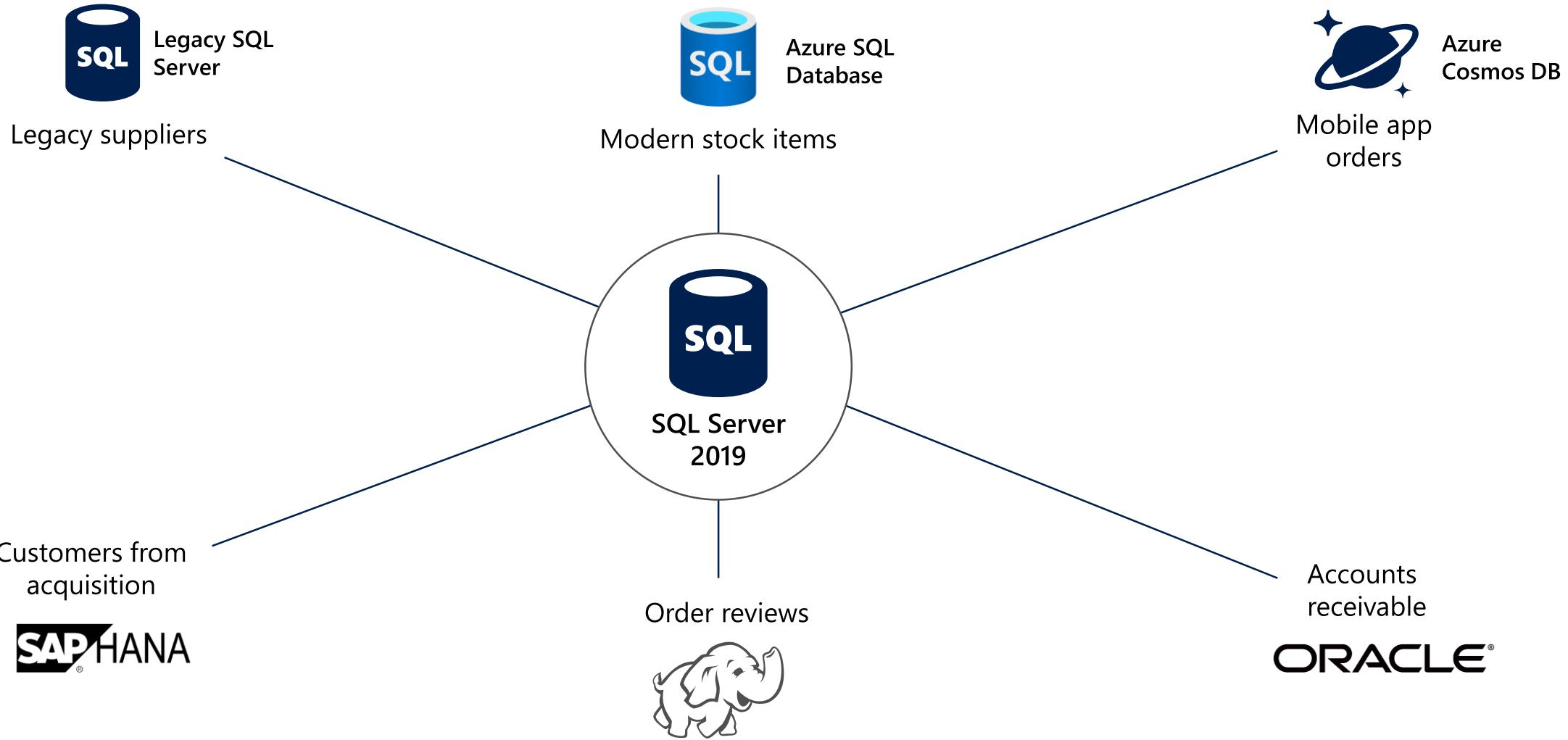
YOU DON'T EVEN HAVE TO MOVE DATA

"It's all about
Data Virtualization"

- ✓ It all started with T-SQL against Hadoop
- ✓ Query data where it lives using T-SQL
- ✓ Compute engine integrated with SQL Server
- ✓ Distributed, scalable query performance
- ✓ Manual/deploy with SQL Server
- ✓ Auto deploy/optimize with Big Data Clusters



SQL SERVER 2019 SERVES AS THE HUB FOR ALL YOUR DATA



Principles of Data Engineering

(Steve's)



PRINCIPLES ARE ALWAYS A BALANCE

- Need to evolve – things change
- Cost vs Principles
- Am I going to spend 2 weeks automating A 10 MINUTE JOB?
- Am I going to generate 2 TB log files for a process or generate alerts?
- Am I really going to keep everything?
- Use Built in tools when possible
- Do I really need to document this? (Hit, Heck ya)

DATA ENGINEERING DESIGN PRINCIPLES / BEST PRACTICES

1. Worry about security – Avoid the front page of the NY Times
2. The Data Must Always be correct and auditable
3. Document your code and design decisions
4. Design for Failure
5. Design for Process Insights
6. Design for Scale
7. Design for Change

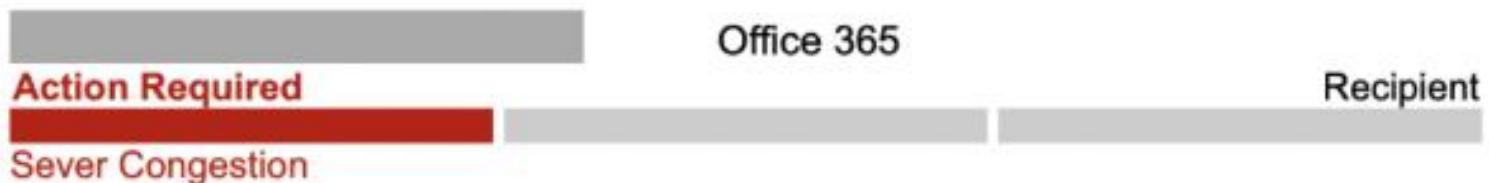
PRINCIPLES - 1 WORRY ABOUT SECURITY



Your messages couldn't be delivered.

Would You
click on this?

Microsoft found Several Undelivered Messages



How to Fix It

Retype the recipient's address, then resend the message - If you're using Outlook, open this non-delivery report message and click **Send Again**. In Outlook on the web, select this message, and then click the "**Send Again**" link located just below the message preview

[Send Again](#)

Was this helpful? [Send feedback to Microsoft](#).

PRINCIPLES - 1 WORRY ABOUT SECURITY

PSA!!
This was
interesting.
What is your
policy for
database
passwords?

The image shows two windows side-by-side. On the left is a 'Recommended based on your password settings' card with the title 'Prevent unneeded password changes'. It states: 'Making people change their passwords periodically may do more harm than good, leading to more predictable passwords and disrupted work. We recommend setting passwords to never expire.' A blue 'View recommendation' button is at the bottom. On the right is an 'SQL Server' connection dialog box. It shows 'Server type: Database Engine', 'Server name: [REDACTED].database.windows.net', 'Authentication: Windows Authentication' (selected), 'User name: [REDACTED]', and 'Password: [REDACTED]'. The 'Azure Active Directory - Universal with MFA' option is highlighted in the dropdown menu. Buttons at the bottom include 'Connect', 'Cancel', 'Help', and 'Options >'. The top of the dialog has a close button 'X'.

[O365 - Set the password expiration policy for your organization](#)

[O365 - Set up multi-factor authentication](#)

[Configure multi-factor authentication for SQL Server Management Studio and Azure AD](#)

[Using Multi-factor AAD authentication with Azure SQL Database and Azure SQL Data Warehouse \(SSMS support for MFA\)](#)

PRINCIPLES - 1 WORRY ABOUT SECURITY

- Avoid the front page of the NY Times
- Your biggest threat is from inside !!

Options

- Involve your security team at start of a project
- Volunteer for internal audits
- Use the systems Available –
Advanced Data Security Assessment

Notifications (1) Features (6)

All Security (4) Performance (1) Recovery (1)

Active Directory admin Allows you to centrally manage identity and access to your Azure SQL databases. NOT CONFIGURED	Advanced data security Data Discovery & Classification, Vulnerability Assessment and Advanced Threat Protection. CONFIGURED
Auditing Track database events and writes them to an audit log in Azure storage. CONFIGURED	Transparent data encryption Encryption at rest for your databases, backups, and logs. SERVICE-MANAGED KEY

Advanced Threat Protecti...

[Learn more - Advanced Threat Protection alerts](#)

- All
- SQL injection ⓘ
- SQL injection vulnerability ⓘ
- Data exfiltration ⓘ
- Unsafe action ⓘ
- Brute Force ⓘ
- Anomalous client login ⓘ

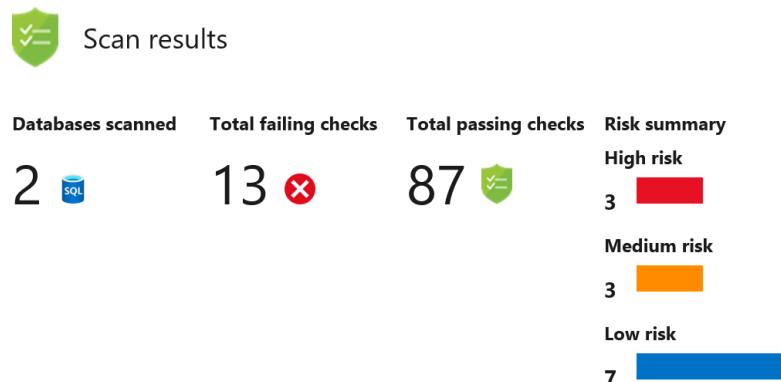


Turn on Advanced Data Security for all databases on this server, at the cost of 15 USD/server/month. This includes Data Discovery & Classification, Vulnerability Assessment and Advanced Threat Protection for the server. We invite you to a trial period for the first 30 days, without charge.

PRINCIPLES - 1 WORRY ABOUT SECURITY – AZURE ASSESSMENT EMAIL

Database	Scan result	Failing checks	Passing checks	Details
AzureDem oDB	Findings	7	43	View results
Adventure worksLT	Findings	6	44	View results

A Vulnerability Assessment scan has completed on your server



Database Scan result Failing checks Passing checks Details

Database	Scan result	Failing checks	Passing checks	Details
AzureDemoDB	Findings	7 ✗	43 🛡️	View results
AdventureworksLT	Findings	6 ✗	44 🛡️	View results

Home > Vulnerability Assessment

Vulnerability Assessment

Scan Export Scan Results Scan History Feedback

Total failing checks: 6 ✗ Total passing checks: 44 ✓ Risk summary:

Risk Level	Count	Color
High Risk	1	Red
Medium	2	Orange
Low Risk	3	Blue

Last scan time: Sun, 19 Jan 2020 00:38:19 UTC

Learn more SQL Security Center Best Practices for SQL Security

[Failed \(6\)](#) [Passed \(44\)](#)

Filter by ID or security check Category: All selected Risk: All selected

ID	SECURITY CHECK	APPLIES TO	CATEGORY	RISK	ADDITIONAL IN...
VA2065	Server-level firewall rules should be tracked and maintained at a strict minimum	master	Surface area reduction	High	Should set an initial ...
VA1143	'dbo' user should not be used for normal service operation	AdventureworksLT	Surface area reduction	Medium	
VA1288	Sensitive data columns should be classified	AdventureworksLT	Data protection	Medium	
VA1145	Feature restrictions should be applied to all principals that are used to access the...	AdventureworksLT	Surface area reduction	Low	
VA2130	Track all users with access to the database	AdventureworksLT	Authentication & Au...	Low	Should set an initial ...
VA2130	Track all users with access to the database	master	Authentication & Au...	Low	Should set an initial ...

PRINCIPLES – 2 DATA MUST ALWAYS BE CORRECT AND AUDITABLE

- Who has Sign off?
- Sentiment Analysis - Do I worry about missing some records in 5million Twitter messages?
- Financial reporting – Do sales by department match? Balance Sheet Totals

Options

- ✓ Get summary file from source owner – Auto Compare
- ✓ Put a Power BI Summary Report Online, make it easy
- ✓ Alerts etc – Be Proactive
- ✓ Put Summary file in same place in Azure data lake, PowerBI will update from there

PRINCIPLES – 2 DATA MUST ALWAYS BE CORRECT AND AUDITABLE

This screenshot illustrates a Power BI report designed to audit employee dimension data. The report includes a data flow diagram, a fact table view, and a commentary section.

5MinuteBI logo is in the top left corner.

Employee Dimension Audit Report title bar with a red notification badge showing '4'.

Pay Fact With Missing Employees table (3) showing one record:

PayFactID	EmployeeID	NameFull	DateValue	HoursWorked	Per Hour	SalaryCost	Facility Name
100567	-99	Unknown Employee	Wednesday, January 13, 2016	7	1.83	0.00	Unknown
Total				7	1.83	0.00	

Year Month filter pane (6) set to 'All'.

Commentary section (5) stating: "There have been issues in getting the correct Employees loaded before the facts are built".

Data Flow Diagram (1) showing relationships between **Employee**, **Facility**, and **FactPay** tables.

Employee table fields: EmployeeID, FacilityID, OrganizationLevel, SalesPersonFlag, JobTitle, BirthDate, MaritalStatus, Gender, HireDate, SalariedFlag, VacationHours, SickLeaveHours, CurrentFlag, ModifiedDate, NameFull, PayGradeID, Facility Name.

Facility table fields: FacilityID, FacilityName, Latitude, Longitude.

FactPay table fields: EmployeeID, PayGradeID, DateValue, CalendarID, SalariedFlag, SalaryCost, HourlyCost, HoursWorked, Per Hour.

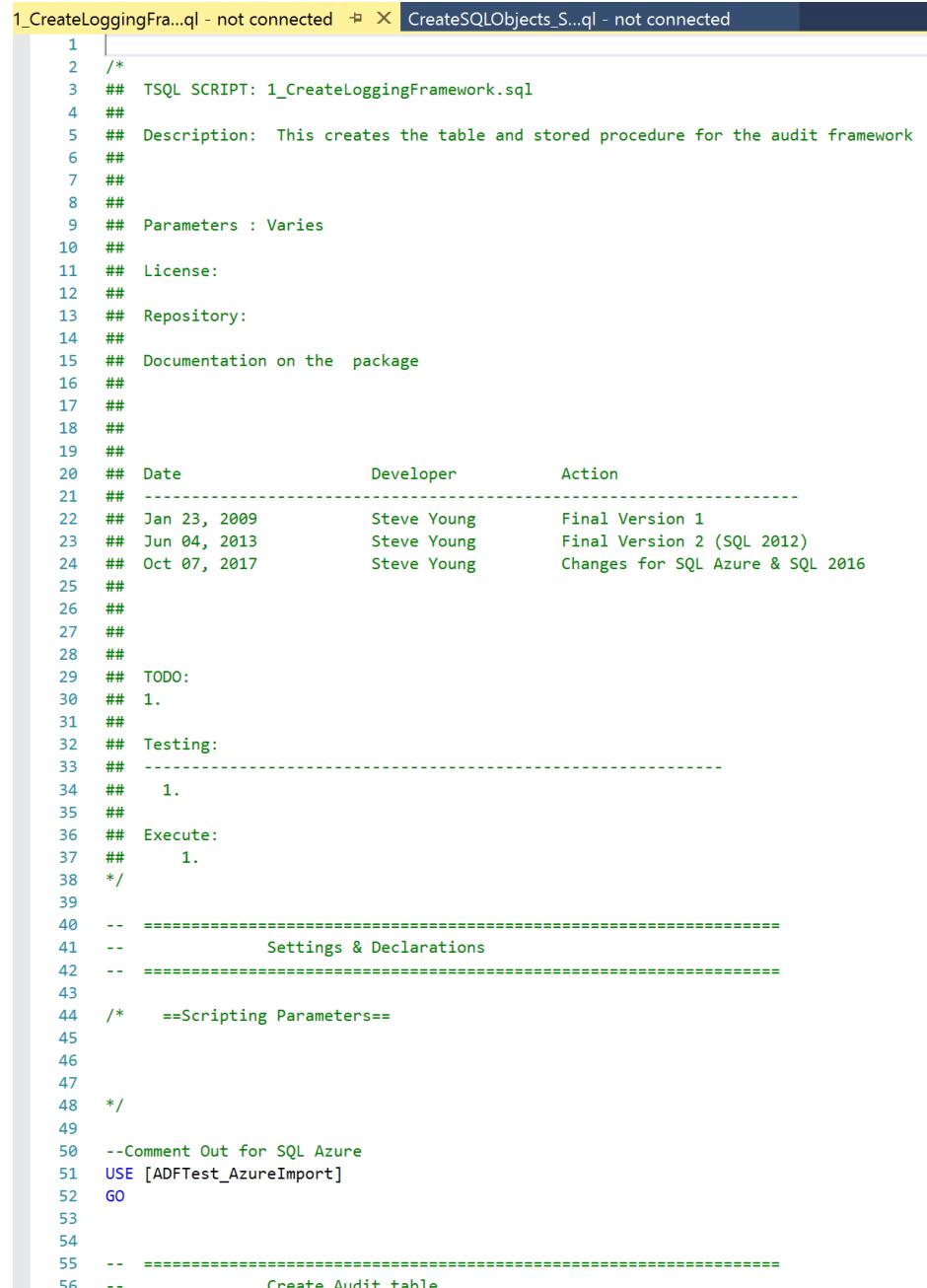
PRINCIPLES – 3 DOCUMENT YOUR CODE AND DESIGN DECISIONS

- Be kind, Please Rewind – If not for you, the person that comes after you.
- (Which might be you ☺)

Options

- ✓ Inline documentation
- ✓ Build Books
- ✓ Run Books
- ✓ Operational Books
- ✓ On the platform as notes, documentation

Q. Which is better, No Documentation or of date documentation?



```
1_CreateLoggingFra...ql - not connected X CreateSQLObjects.S...ql - not connected
1
2 /*
3 ##  TSQL SCRIPT: 1_CreateLoggingFramework.sql
4 ##
5 ##  Description: This creates the table and stored procedure for the audit framework
6 ##
7 ##
8 ##
9 ##  Parameters : Varies
10##
11##  License:
12##
13##  Repository:
14##
15##  Documentation on the package
16##
17##
18##
19##
20##  Date           Developer      Action
21##  -----          -----        -----
22##  Jan 23, 2009   Steve Young   Final Version 1
23##  Jun 04, 2013   Steve Young   Final Version 2 (SQL 2012)
24##  Oct 07, 2017   Steve Young   Changes for SQL Azure & SQL 2016
25##
26##
27##
28##
29##  TODO:
30##  1.
31##
32##  Testing:
33##  -----
34##  1.
35##
36##  Execute:
37##  1.
38*/
39
40-- =====
41-- Settings & Declarations
42-- =====
43
44/* ==Scripting Parameters==
45
46
47*/
48
49
50--Comment Out for SQL Azure
51USE [ADFTest_AzureImport]
52GO
53
54
55-- =====
56-- Create Audit table
57
```

PRINCIPLES – 3 DOCUMENT YOUR CODE AND DESIGN DECISIONS

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation pane lists 'Factory Resources' with sections for Pipelines, Datasets, and Data flows. A pipeline named 'TransformMovies' is selected. The main workspace displays the 'Activities' section, which includes a 'Mapping Data Flow' component named 'TransformMovies'. The pipeline details pane at the bottom shows the following configuration:

- Name:** TransformMovies (marked with a red circle labeled 1)
- Description:** This will allow us to summarize the movie database and output the results to ADL (marked with a red circle labeled 2)
- Timeout:** 7.00:00:00
- Retry:** 0
- Retry interval:** 30
- Secure output:**
- Secure input:**

- Which is better?

Linked services		Integration runtimes
New		
Showing 1 - 5 of 5 items		
NAME ↑↓		TYPE ↑↓
AzureDataLakeStorage1		Azure Data Lake Storage Gen2
AzureDataLakeStorage2		Azure Data Lake Storage Gen2
AzureDataLakeStorage2_WindowTrigger		Azure Data Lake Storage Gen2
AzureSqlDatabase1		Azure SQL Database
AzureSqlDatabase_WindowTriggerDemo		Azure SQL Database

PRINCIPLES – 4 DESIGN TO SURVIVE FAILURE

- What happens if your process Fails, stops, is not complete.
 - Rollback, Restart,
 - What about Disaster Recovery –
 - What about the Roll Back?
 - Is it a failure or an unexpected result?
 - Is your plan stored in the failed region?
- You can have a failure and still meet your SLA.

Availability %	Downtime per year	Downtime per month*	Downtime per week
90% ("one nine")	36.5 days	72 hours	16.8 hours
99% ("two nines")	3.65 days	7.20 hours	1.68 hours
99.9% ("three nines")	8.76 hours	43.2 minutes	10.1 minutes
99.99% ("four nines")	52.56 minutes	4.32 minutes	1.01 minutes
99.999% ("five nines")	5.26 minutes	25.9 seconds	6.05 seconds
99.9999% ("six nines")	31.5 seconds	2.59 seconds	0.605 seconds

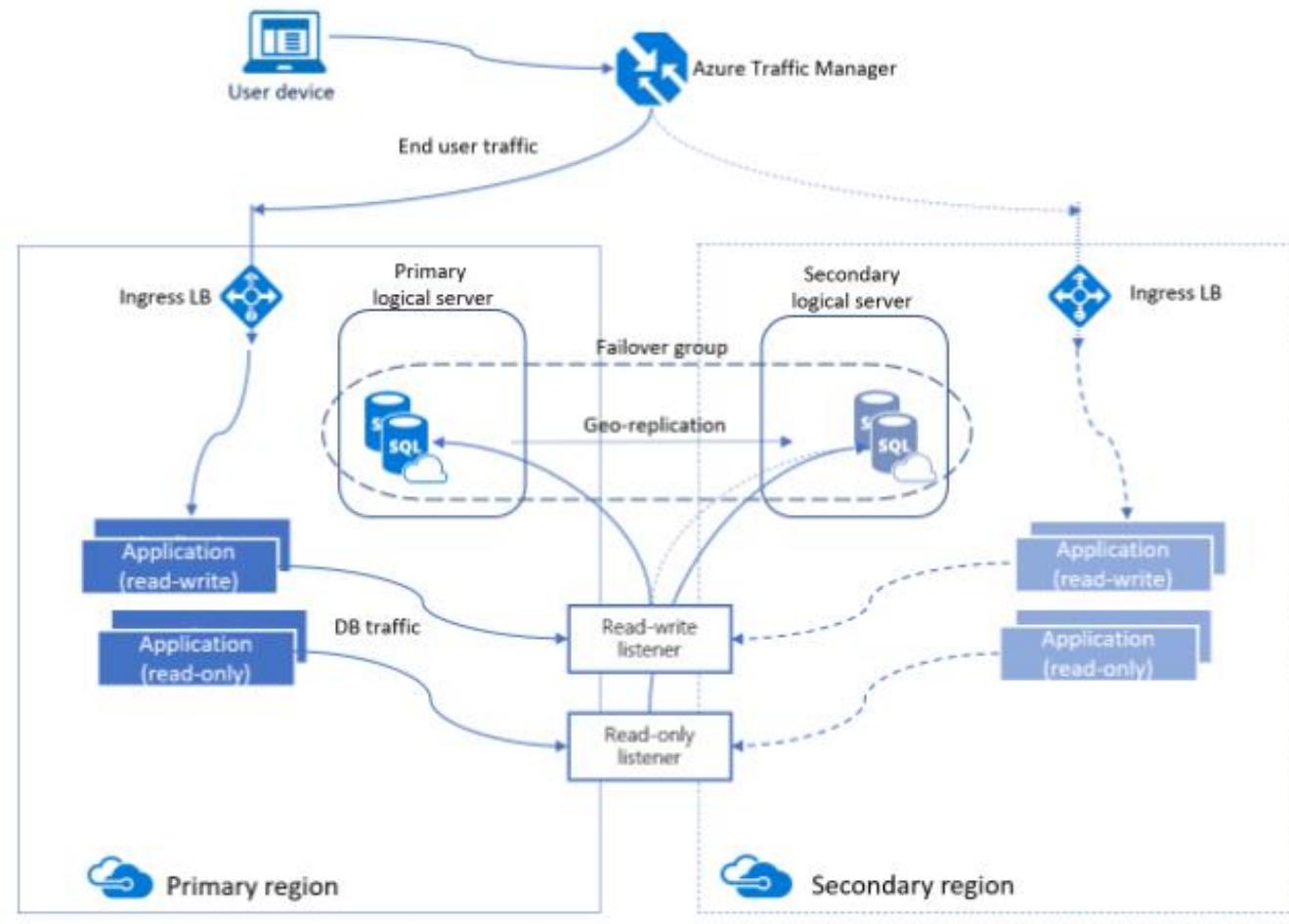
- ✓ Build a prototype – Test it !!
- ✓ Use Source Control, copy to another region !!!!
- ✓ Don't be afraid to redesign
- ✓ How do you communicate?
- ✓ Test Solutions – minimize failure, test failures
- ✓ Do you have Test Data
- ✓ Story boards are useful to talk out and plan options.

- 3 components have 99.9 SLA
- Your composite SLA <> 99.9%



PRINCIPLES – 4 DESIGN TO SURVIVE FAILURE

- The following diagram illustrates a typical configuration of a geo-redundant cloud application using multiple databases and auto-failover group.

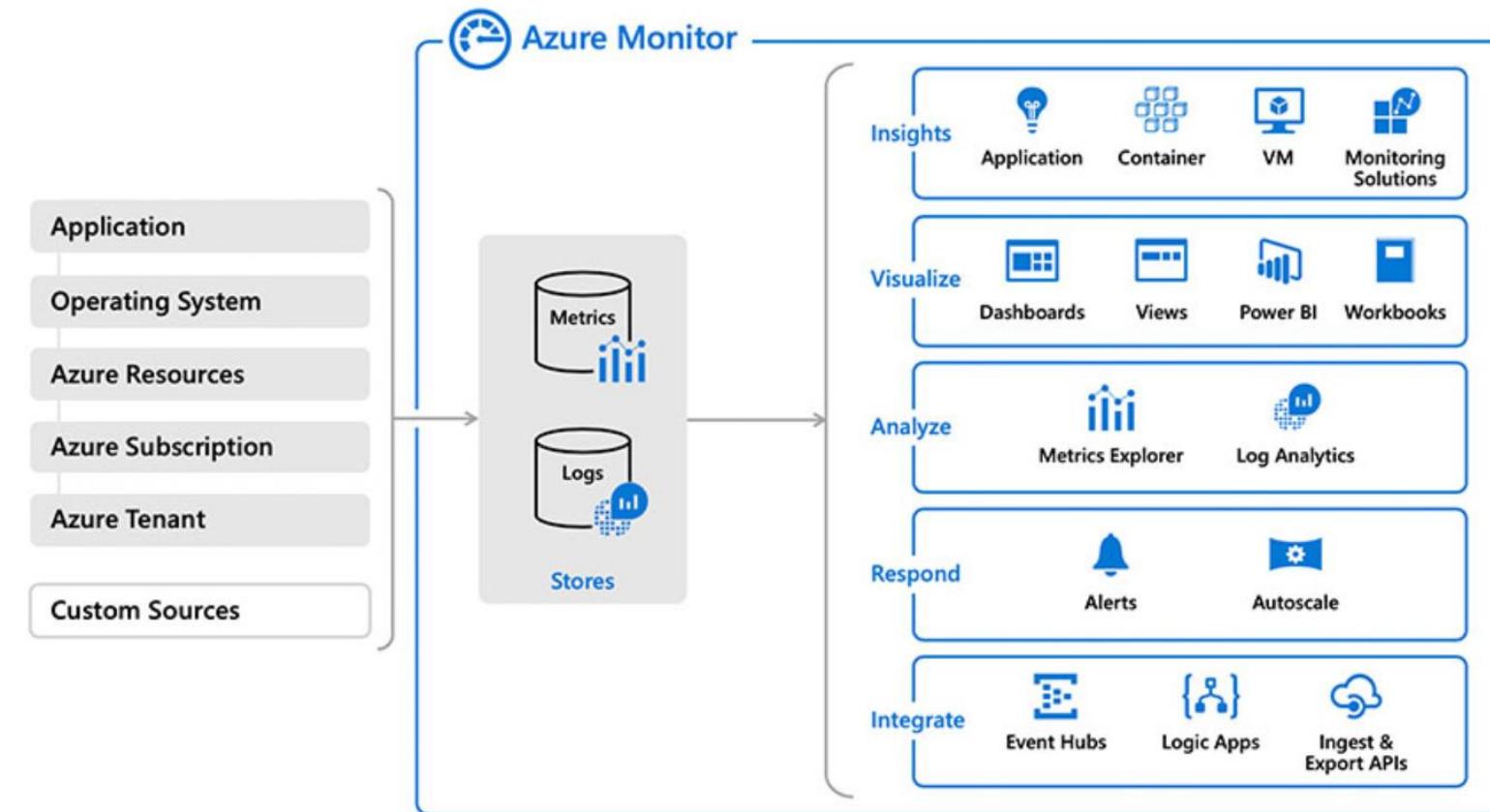


<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-auto-failover-group?tabs=azure-powershell>

PRINCIPLES – 5 DESIGN FOR PROCESS INSIGHTS

- Users complain the process takes too long!!!!
- More than 1 user needs to know what happened

- ✓ Azure Monitor
- ✓ Azure Log Analytics
- ✓ Azure Data Factory Monitor
- ✓ Custom logging and reporting



PRINCIPLES – 5 DESIGN FOR PROCESS INSIGHTS - WALKTHROUGH

- Azure Data Factory monitor.

Microsoft Azure | Data Factory > adfSQL2019Labs

Pipeline runs

Time : Last 24 hours (1/17/20 8:13 PM - 1/18/20 8:13 PM) Time zone : Eastern Time (US & Canada) (UT...) Runs : Latest runs Refresh

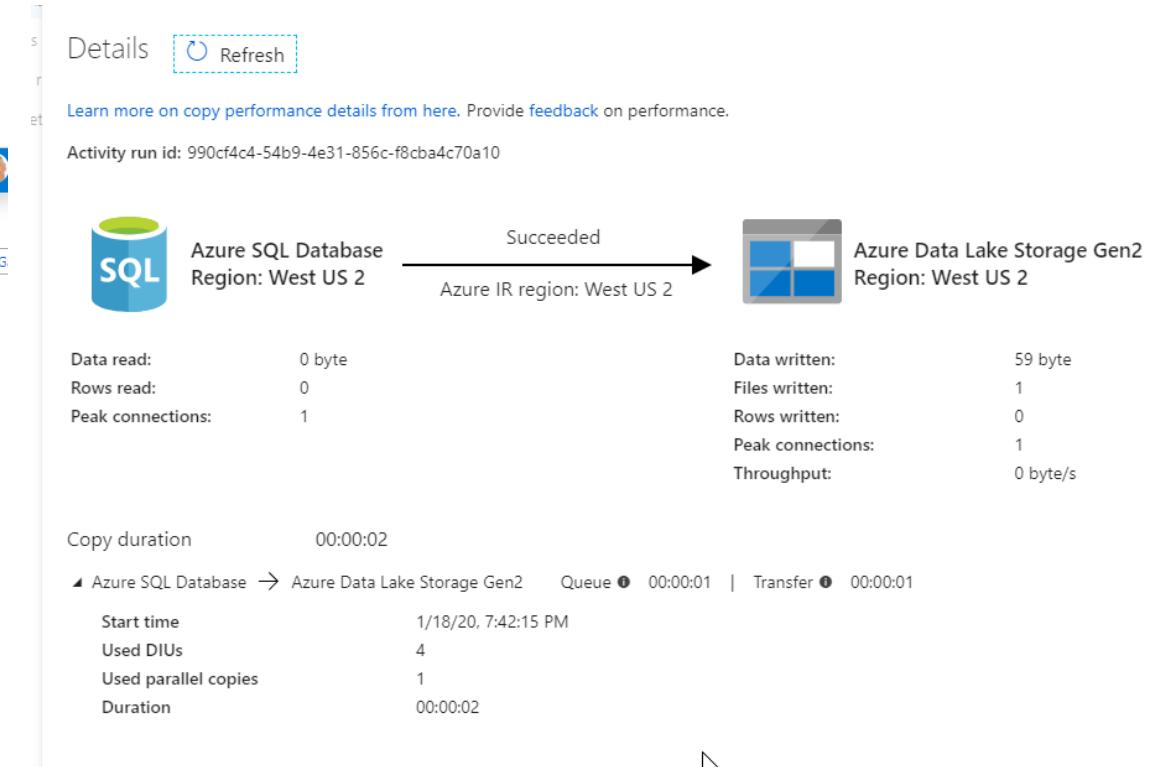
All status Rerun Cancel Refresh Edit columns

Showing 1 - 24 items

Pipeline Name	Run Start	Duration	Triggered By	Status	Parameters
SlidingWindowTrigertoDatal...	1/18/20, 7:41:59 PM	00:00:21	Trigger_wrf	Succeeded	[@]
SlidingWindowTrigertoDatal...	1/18/20, 6:41:59 PM	00:04:16	Trigger_wrf	Succeeded	[@]
SlidingWindowTrigertoDatal...	1/18/20, 5:42:00 PM	00:00:21	Trigger_wrf	Succeeded	[@]
SlidingWindowTrigertoDatal...	1/18/20, 4:42:00 PM	00:00:23	Trigger_wrf	Succeeded	[@]
SlidingWindowTrigertoDatal...	1/18/20, 3:42:00 PM	00:00:23	Trigger_wrf	Succeeded	[@]

Parameters

NAME	VALUE	EDIT COLUMNS
windowStart	1/18/2020 11:42:00 PM	+ Add column
windowEnd	1/19/2020 12:42:00 AM	+ Add column

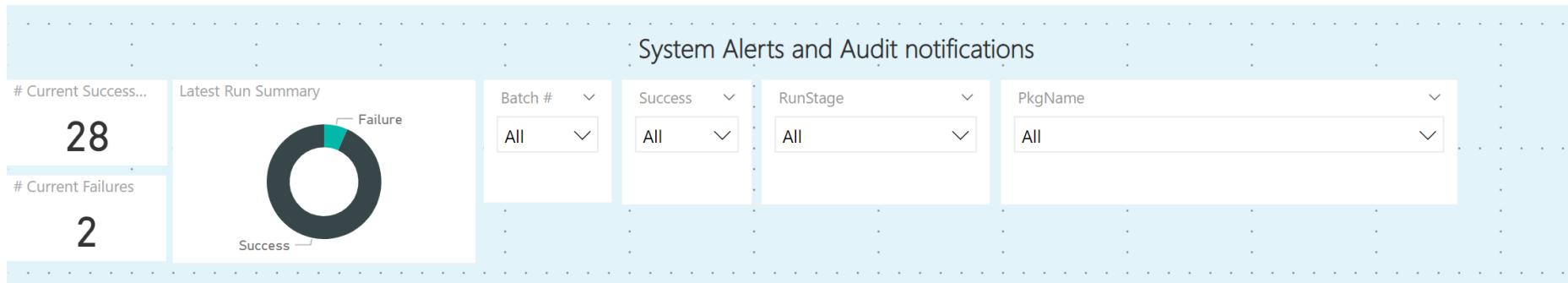


PRINCIPLES – 5 DESIGN FOR PROCESS INSIGHTS

- Azure Log Analytics - Data Factory stores pipeline-run data for only 45 days. Use Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis. You can also keep them in a storage account so that you have factory information for your chosen duration.
- Save your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.
- Stream the logs to Azure Event Hubs. The logs become input to a partner service or to a custom analytics solution like Power BI.
- Analyze the logs with Log Analytics.

<https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor>

PRINCIPLES – 5 DESIGN FOR PROCESS INSIGHTS



Custom monitoring

Package Run Details

Batch # Success RunStage PkgName ExecStartDT

All All All All 11/8/2017 11/30/2017

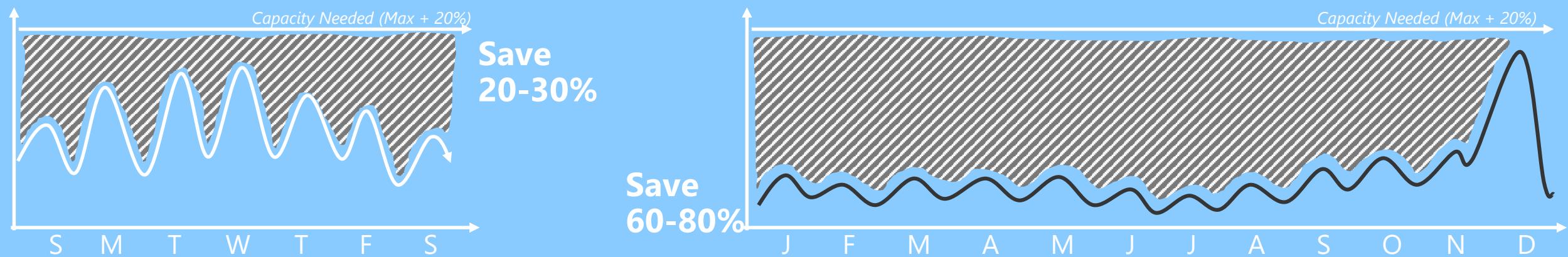
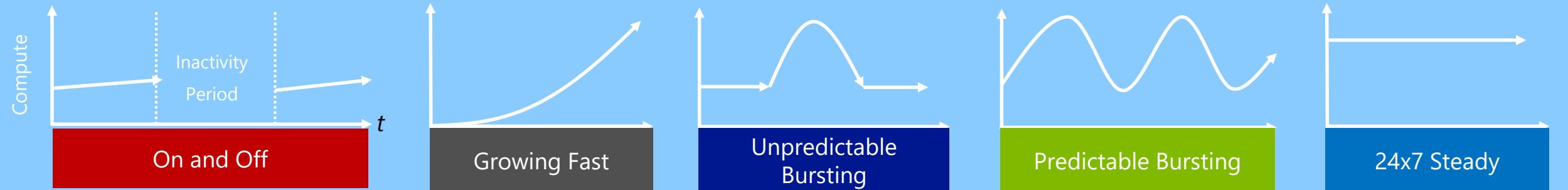
Details

Success	MB#	AuditID	PkgName	RunStage	ExecStartDT	Notes	SQLErrorMessage	CommandTxt
N	95	163	[usp_Stage_To_Prod_LeapKnee_1YR]	Begin Procedure	11/30/2017 12:52:54 PM	Stage to Production		
N	95	165	[usp_Stage_To_Prod_LeapKnee_1YR]	Begin Transforms	11/30/2017 12:52:54 PM	Stage to Production		
N	95	164	[usp_Stage_To_Prod_LeapKnee_1YR]	Insert Success	11/30/2017 12:52:54 PM	Stage to Production		
Y	95	166	[usp_Stage_To_Prod_LeapKnee_1YR]	Process Complete	11/30/2017 12:52:54 PM	Stage to Production		

PRINCIPLES – 6 DESIGN FOR SCALE

- Incremental Load / Historical Load
 - Archive data? How much history do you need to keep.
-
- ✓ Azure data Lake and Polybase comes in handy
 - ✓ Scale SQL up and down

Saving cost



On and Off
(30%)

Growing Fast
(15%)

Unpredictable Bursting
(25%)

Predictable Bursting
(20%)

24x7 Steady
(10%)

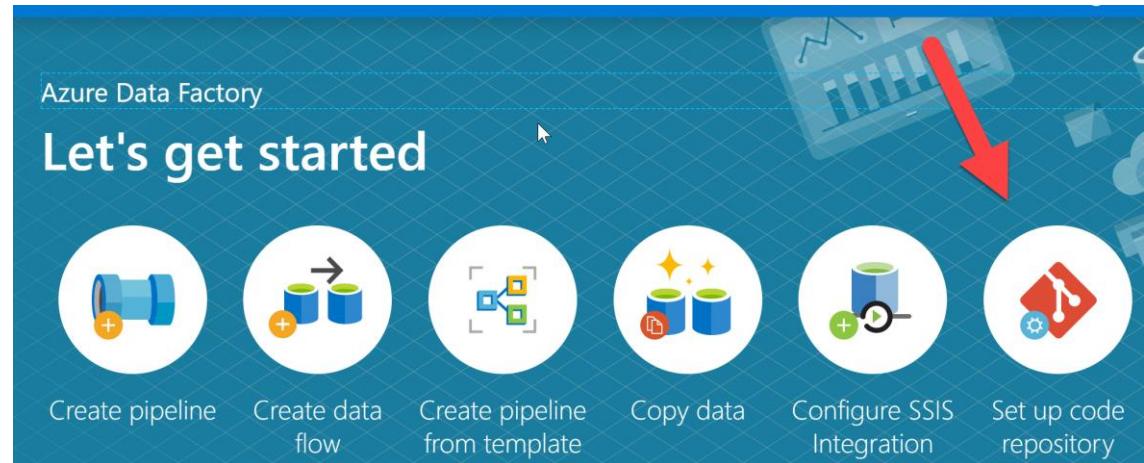
Your Application Portfolio – What Does it Look Like..?

PRINCIPLE – 7 DESIGN FOR CHANGE

- Everything Changes
- Speed and Control are not exclusive
- Be able to roll back, no matter how difficult

- ✓ Don't box yourself in
- ✓ You can also use GITHUB for the source control etc.
- ✓ Used instead of Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface for the 'adfSQL2019Labs' resource group. At the top, there's a navigation bar with 'Microsoft Azure', 'Data Factory', and 'adfSQL2019Labs'. Below it is a toolbar with various icons for saving, publishing, validating, refreshing, discarding, and toggling data flow debug mode. A search bar labeled 'Search resources' is also present. On the left, a sidebar titled 'Factory Resources' lists 'Pipelines' (4), 'Datasets' (6), 'Data flows' (1), and 'Templates' (0). The main area displays a list of data factory resources. A red arrow points to the GitHub integration section at the top of the page, which includes a dropdown menu set to 'GitHub' and a branch selector showing 'master branch'. Other options like 'Save All', 'Publish', 'Validate all', 'Refresh', 'Discard all', 'Data flow debug', and 'ARM template' are visible.



This screenshot shows the 'Repository settings' configuration dialog. It includes fields for 'Repository type' (set to 'GitHub'), 'Use GitHub Enterprise' (unchecked), 'GitHub Account' (set to 'steveyoungca'), 'Git repository name' (set to 'AzureDataFactoryLabsql2019'), 'Collaboration branch' (set to 'master'), 'Root folder' (set to '/'), and a checked checkbox for 'Import existing Data Factory resources to repository'. Below this, there's a section for 'Branch to import resources into' with radio buttons for 'Use Collaboration' (unchecked), 'Create new' (unchecked), and 'Use existing' (checked, with 'master' selected).

1TB Restore Story

PRINCIPLE – 7 DESIGN FOR CHANGE

✓ Azure Data Factory vs GitHub

The screenshot shows a comparison between the Microsoft Azure Data Factory interface and a GitHub repository for an Azure Data Factory project named 'adfSQL201'.

Azure Data Factory Interface:

- Left sidebar: Home, Data Factory, Pipelines, Datasets, Data flows.
- Top bar: Microsoft Azure, Data Factory, adfSQL201.
- Center: Factory Resources section with Pipelines (4), Datasets (6), and Data flows (1).
- Bottom: Publish all button.

GitHub Repository Comparison:

- Top right: Your recently pushed branches: adf_publish (4 minutes ago) (highlighted in yellow).
- Branch dropdown: master.
- Buttons: New pull request, Create new file, Upload files, Find file, Clone or download.
- Commit list:
 - steveyoungca Adding dataflow: dataflow1 ... (Latest commit 1b72a47 12 minutes ago)
 - dataflow Adding dataflow: dataflow1 (12 minutes ago)
 - dataset Adding dataflow: dataflow1 (12 minutes ago)
 - linkedService Adding dataflow: dataflow1 (12 minutes ago)
 - pipeline Adding dataflow: dataflow1 (12 minutes ago)
 - trigger Adding dataflow: dataflow1 (12 minutes ago)
 - README.md Initial commit (16 minutes ago)
- Bottom: README.md file entry.

<https://github.com/steveyoungca/AzureDataFactoryLabsql2019>

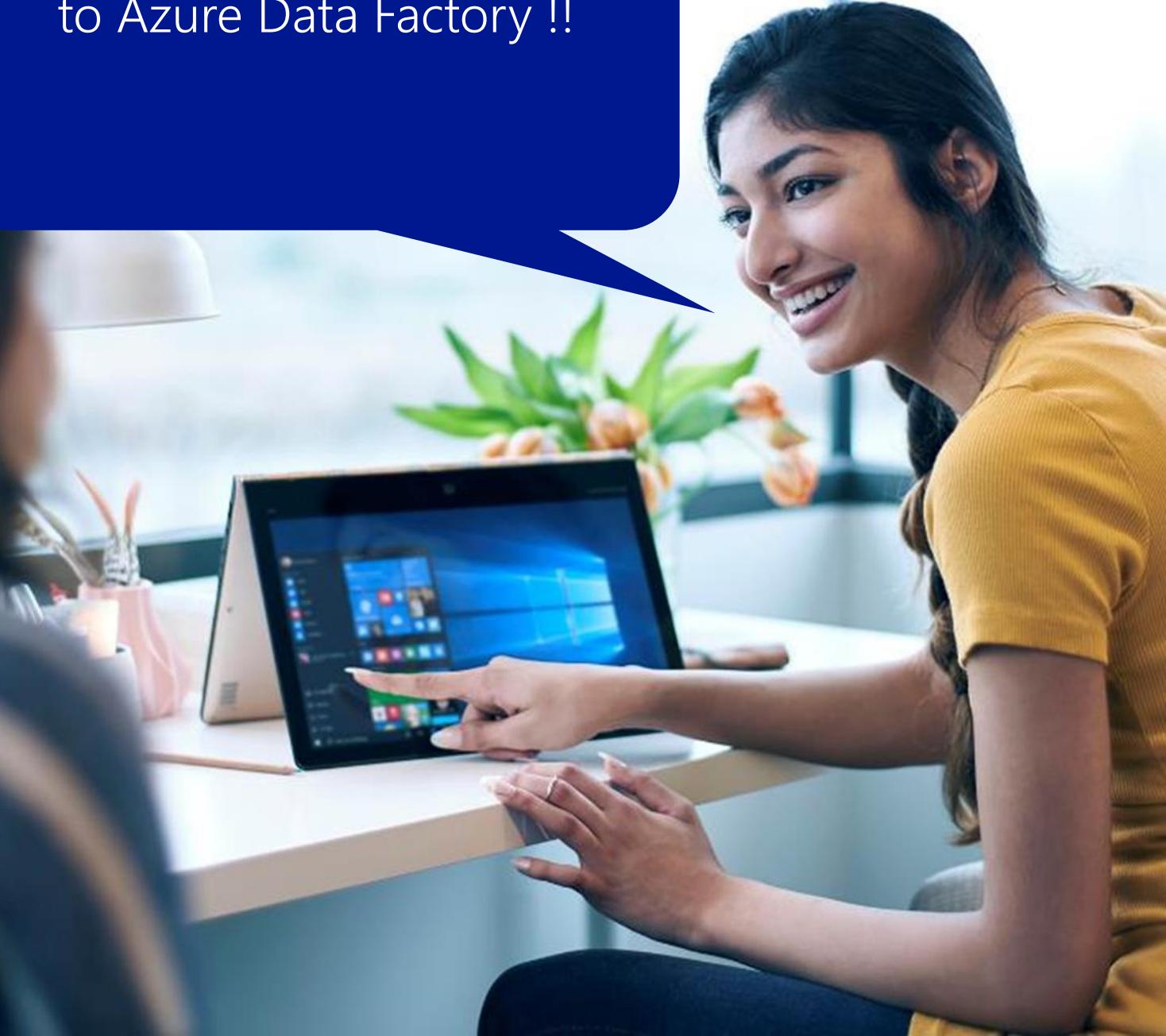
1TB Restore Story

Demo 1

Azure Data Factory

Sliding Window

Capture sequential updates
to Azure Data Factory !!



CLOUD SCALE ANALYTICS

Ingest



Azure Data Factory

Copy data from over 75 data sources

Drag-n-drop UI

Execute your SQL Server Integration Services packages in Azure

Store



Azure Data Lake Storage

Industry leading performance

POSIX-compliant, fine grained access control

Tiered multi-modal storage

Prep



Azure Databricks

Fast, easy, collaboration Apache® Spark-based™ analytics platform

Native integration with Azure services

Enterprise-grade Azure security

Polybase

Serve



Azure SQL Data Warehouse

Unlimited scale & performance

Flexibility to fit your needs

Enterprise-grade DW at the click of a button

Reporting



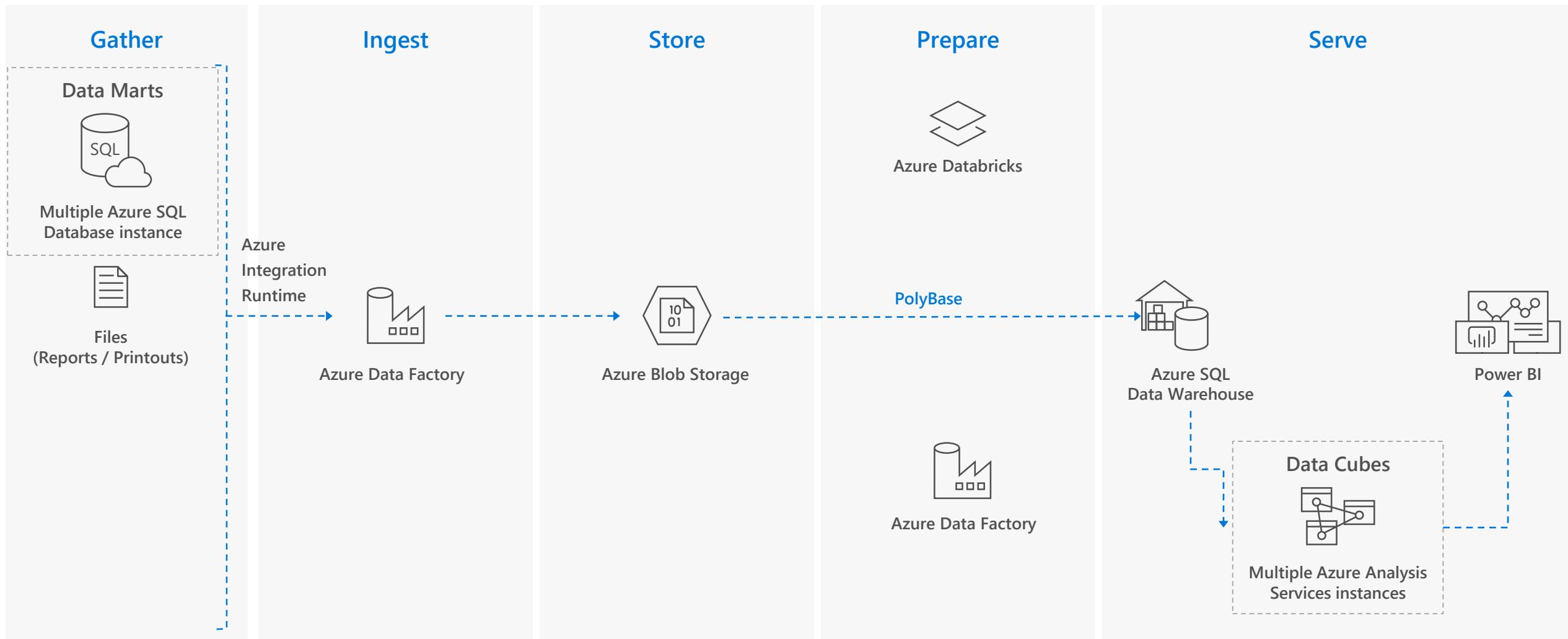
Power BI

Get started in seconds

Ask questions, get answers

Access intelligence anywhere

WHERE DID I START?



DEMO 1

Results Messages

	FactKey	DateTimeEvent	Dimension1Key	Dimension2Key	MyNotes
1	1	2019-12-06 20:38:38.833	1	2	FirstLoad1
2	2	2019-12-06 20:38:38.843	1	2	FirstLoad1
3	3	2019-12-06 20:38:39.893	1	2	FirstLoad1
4	4	2019-12-06 20:38:39.900	1	2	FirstLoad1
5	5	2019-12-06 20:38:40.797	1	2	FirstLoad1
6	6	2019-12-06 20:38:40.800	1	2	FirstLoad1
7	7	2019-12-06 21:10:17.890	1	2	FirstLoad1
8	8	2019-12-06 21:10:17.903	1	2	FirstLoad1
9	9	2019-12-06 21:10:18.890	1	2	FirstLoad1
10	63	2020-01-12 19:05:08.530	1	2	FirstLoad1
11	64	2020-01-12 19:05:08.537	1	2	FirstLoad1



Microsoft Azure Storage Explorer

File Edit View Help

EXPLORER One or more accounts requires reauthentication.

Upload Download New Folder Select All Copy Paste Rename Manage Access More

← → ↑ windowtrigger > 2020 > 01 > 01

Name	Access Tier	Access Tier Last Modified	Last Modified	Blob Type	Content Type	Size
00				Folder		
01				Folder		
02				Folder		
03				Folder		
04				Folder		
05				Folder		
06				Folder		
07				Folder		
08				Folder		
09				Folder		
10				Folder		
11				Folder		

Showing 1 to 24 of 24 cached items

Data Lake Store

Filter New folder Upload Access Rename folder Folder properties Delete folder Refresh

2018 > 05 > 30 > 04

NAME	SIZE	LAST MODIFIED
FactWindowHour.csv	194 bytes	5/30/2018, 1:15:13 AM

File preview FactWindowHour.csv

Format Download Rename file Access Properties Set expiry Delete file

FactKey	DateTimeEvent
28	2018-05-30 04:25:00.0000000
52	2018-05-30 04:25:00.0000000
76	2018-05-30 04:25:00.0000000

Demo 2 Data Virtualization

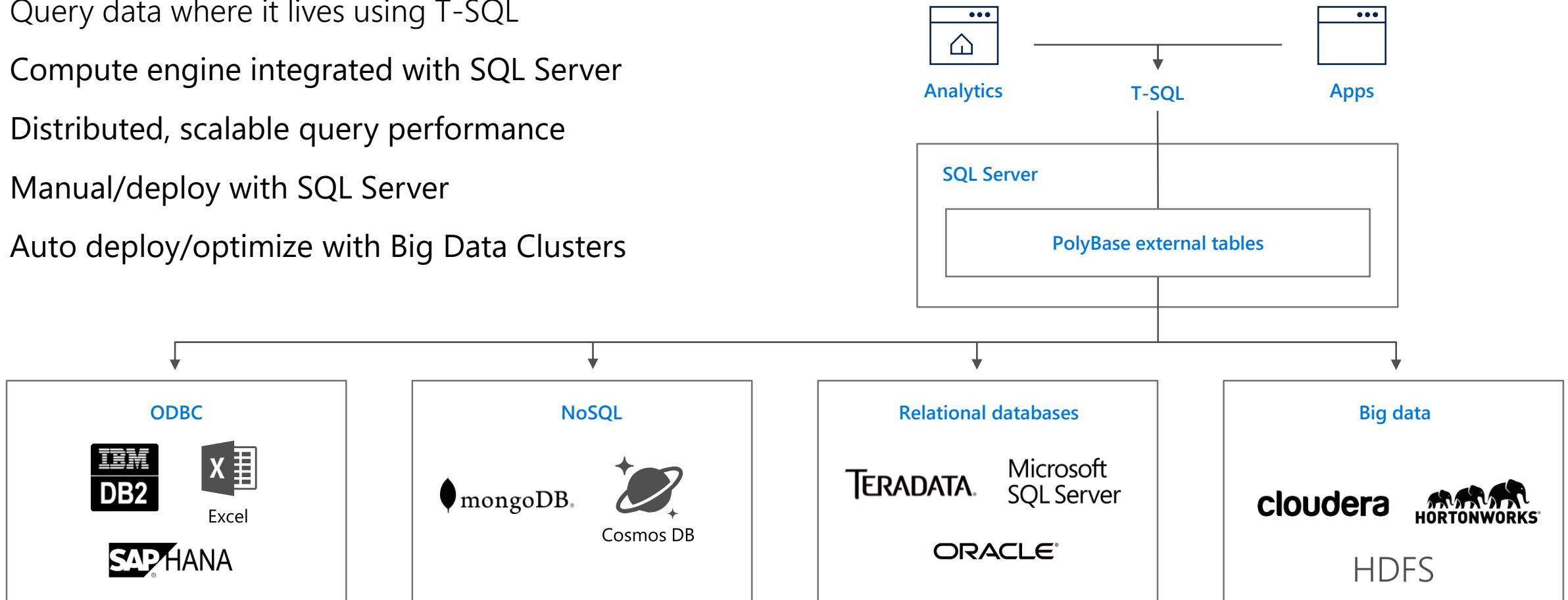
User data where it resides !!!



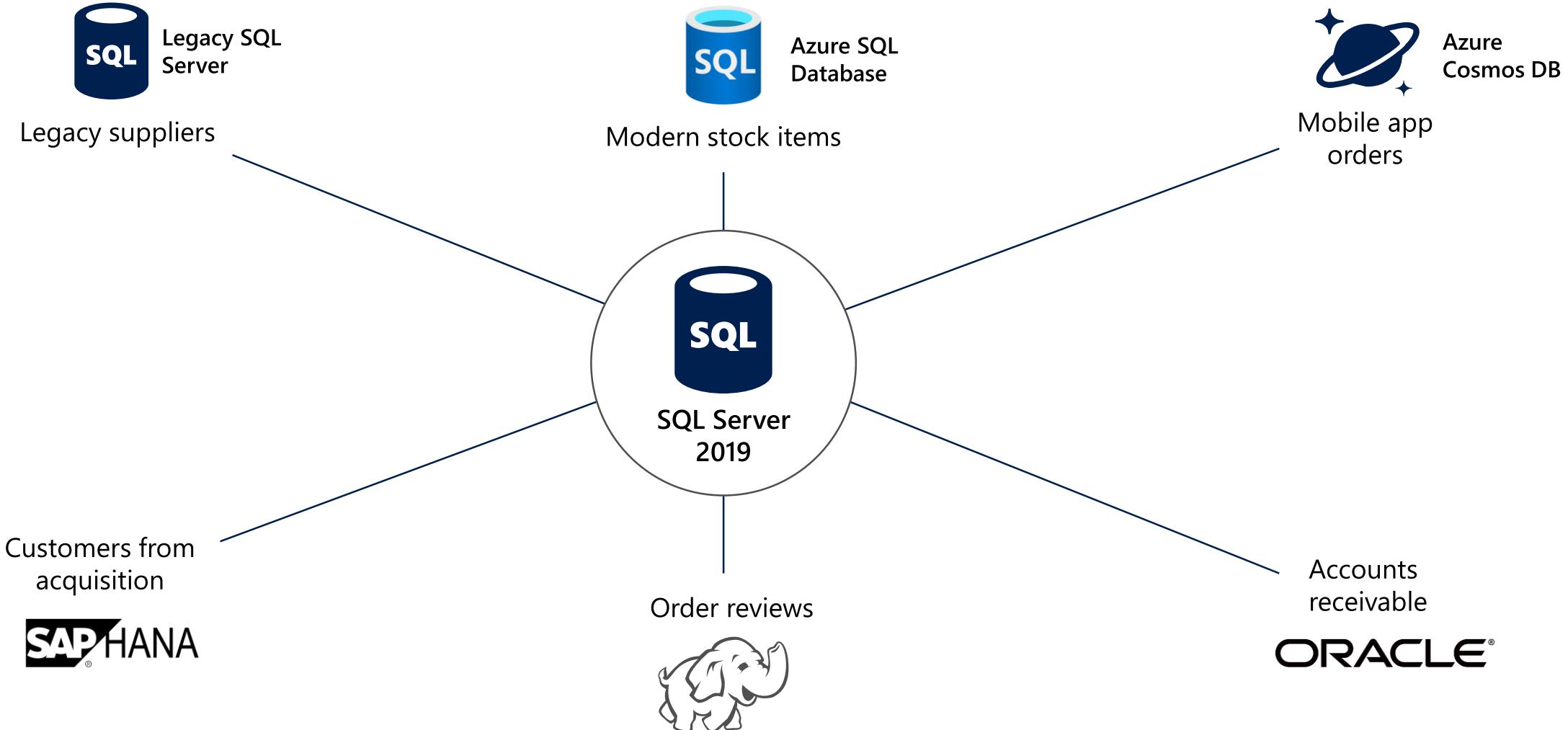
YOU DON'T EVEN HAVE TO MOVE DATA

"It's all about
Data Virtualization"

- ✓ It all started with T-SQL against Hadoop
- ✓ Query data where it lives using T-SQL
- ✓ Compute engine integrated with SQL Server
- ✓ Distributed, scalable query performance
- ✓ Manual/deploy with SQL Server
- ✓ Auto deploy/optimize with Big Data Clusters



SQL SERVER 2019 SERVES AS THE HUB FOR ALL YOUR DATA



Demo 3

Azure Data Studio

Notebooks

I've left the
best Data Engineering tool
to the last !!!!



AZURE DATA STUDIO

1. Cross Platform – Windows, macOS, Linux

<https://docs.microsoft.com/en-us/sql/azure-data-studio/download-azure-data-studio?view=sql-server-ver15>

AZURE DATA STUDIO

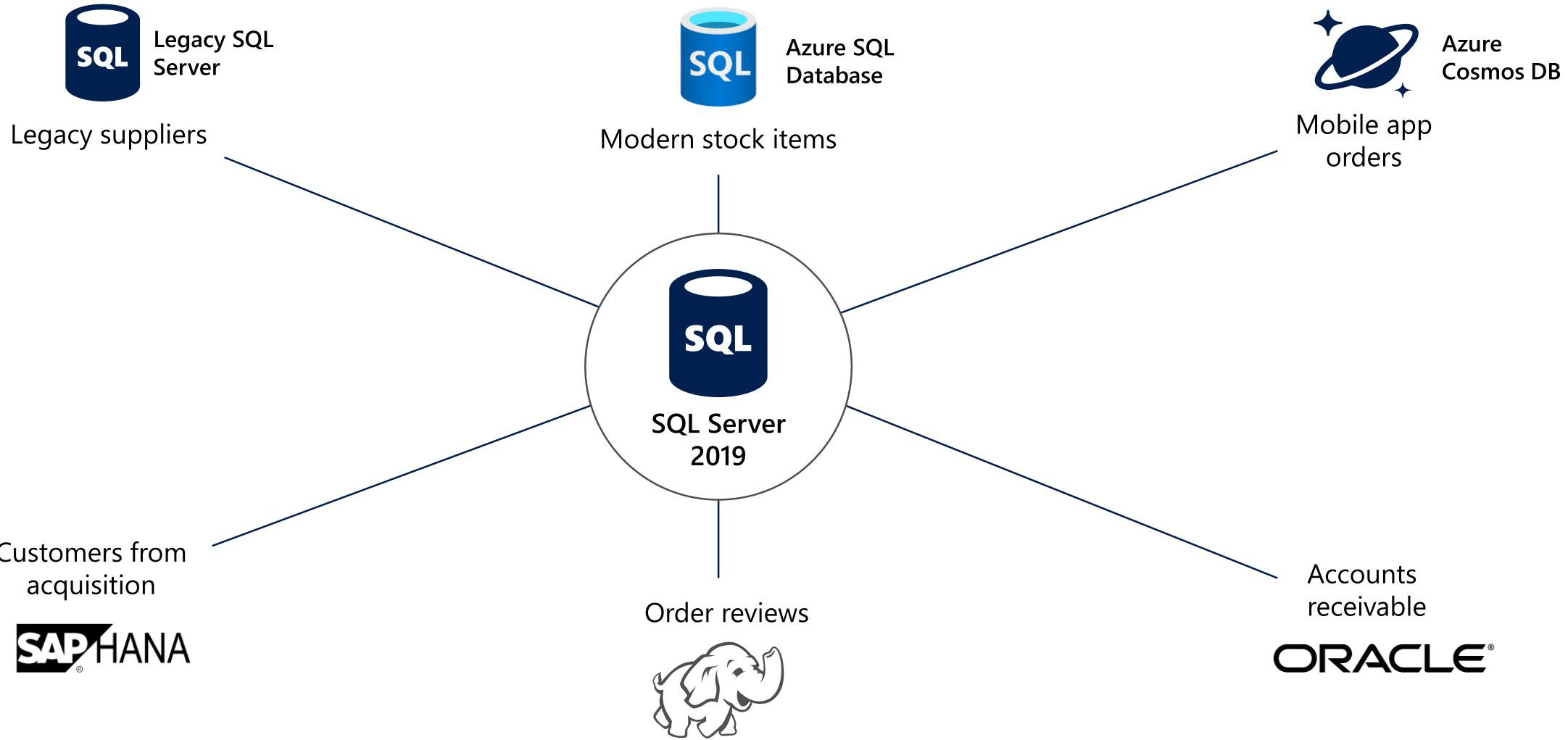
References

Quickstart: Use Azure Data Studio to connect and query Azure SQL database -

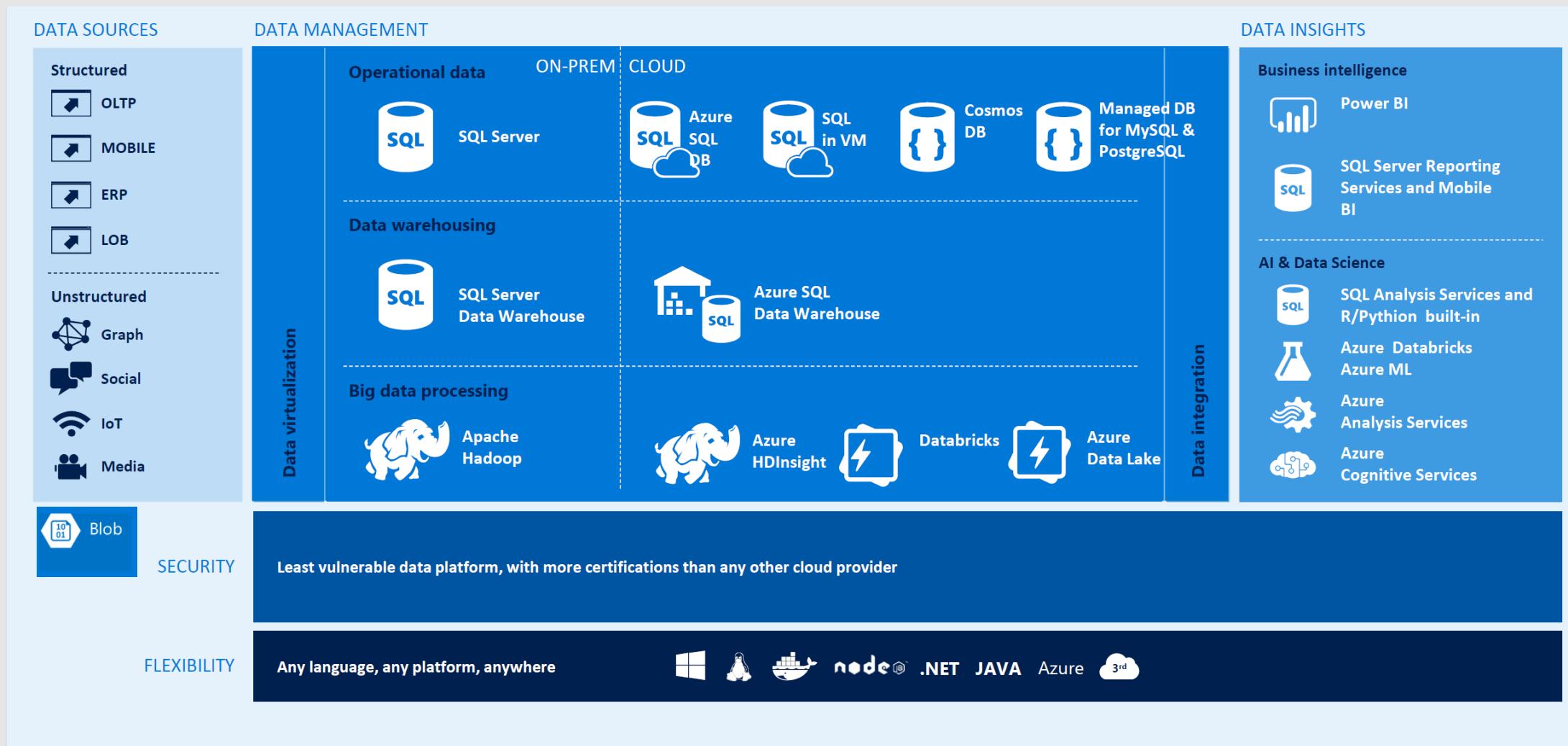
<https://docs.microsoft.com/en-us/sql/azure-data-studio/quickstart-sql-database?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/azure-data-studio/download-azure-data-studio?view=sql-server-ver15>

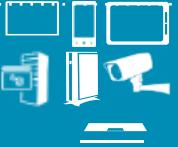
SQL SERVER 2019 SERVES AS THE HUB FOR ALL YOUR DATA



Azure and Data Estate Data View



Microsoft Azure BI & Analytics services portfolio

Devices	Device Connectivity	Storage	Analytics	Presentation & Action
	 IoT Hub	 SQL Database	 Machine Learning	 App Service
	 Event Hub	 Table/Blob Storage	 Stream Analytics	 Power BI
	 Service Bus	 DocumentDB	 HDInsight	 Notification Hubs
	External Data Sources	 3 rd party Databases	 Data Factory	 Mobile Services
			 Data Lake	 BizTalk Services

TOP 10 FEATURES IN SQL SERVER 2019

1. Harness the power of big data



Big data clusters with scalable compute and storage composed of SQL Server, Spark, and HDFS. Cache data in scale-out data marts.

2. Bring AI to your workloads



A complete AI platform to train and operationalize models in SQL Server ML Services or Spark ML using Azure Data Studio notebooks.

3. Eliminate the need for data movement



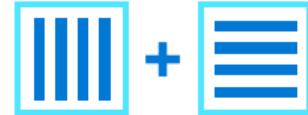
Data virtualization allows queries across relational and non-relational data without movement or replication.

4. Explore and interact with visual data



Visual data exploration and interactive analysis using SQL Server BI tools and Power BI Report Server.

5. Run real-time analytics on operational data



In-memory technologies for analytics on operational data using HTAP. Higher concurrency and scale through persistent memory.

6. Automatically tune SQL Server



Intelligent Query Processing improves scaling of queries and Automatic Plan Correction resolves performance problems.

7. Reduce database maintenance and increase business uptime



Greater uptime with more online indexing operations. Now run Always On availability groups on containers using Kubernetes.

8. Boost security and protect data in use



SQL Server enables layers of security including protection of computations in Always Encrypted secure enclaves.

9. Track compliance with sophisticated resources



Data Discovery & Classification labeling for GDPR and Vulnerability Assessment tool to track compliance.

10. Optimize with choice and flexibility



Support for your choice of Windows, Linux, and containers. Run Java code on SQL Server and store and analyze graph data.

SQL SERVER 2019

Industry-leading performance and security, with intelligence over all your data

Intelligence over
any data



AI and Machine Learning
over all data with the power
of SQL and Apache Spark

Choice of platform
and language



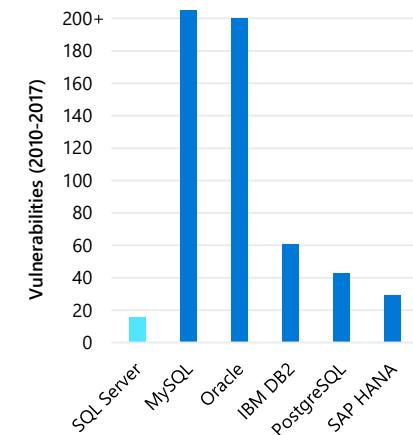
T-SQL
Java
C/C++
PHP
Node.js
C#/VB.NET
Python
Ruby

Industry-leading
performance



#1 OLTP performance¹
#1 DW performance on
1TB², 10TB³, and 30TB⁴
Intelligent Query Processing

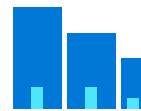
Most secure
over the last 8 years⁵



Insights in minutes
and rich reports



The best of Power BI and
SQL Server Reporting Services
with Power BI Report Server



In-memory across all workloads

Private cloud

Most consistent data platform

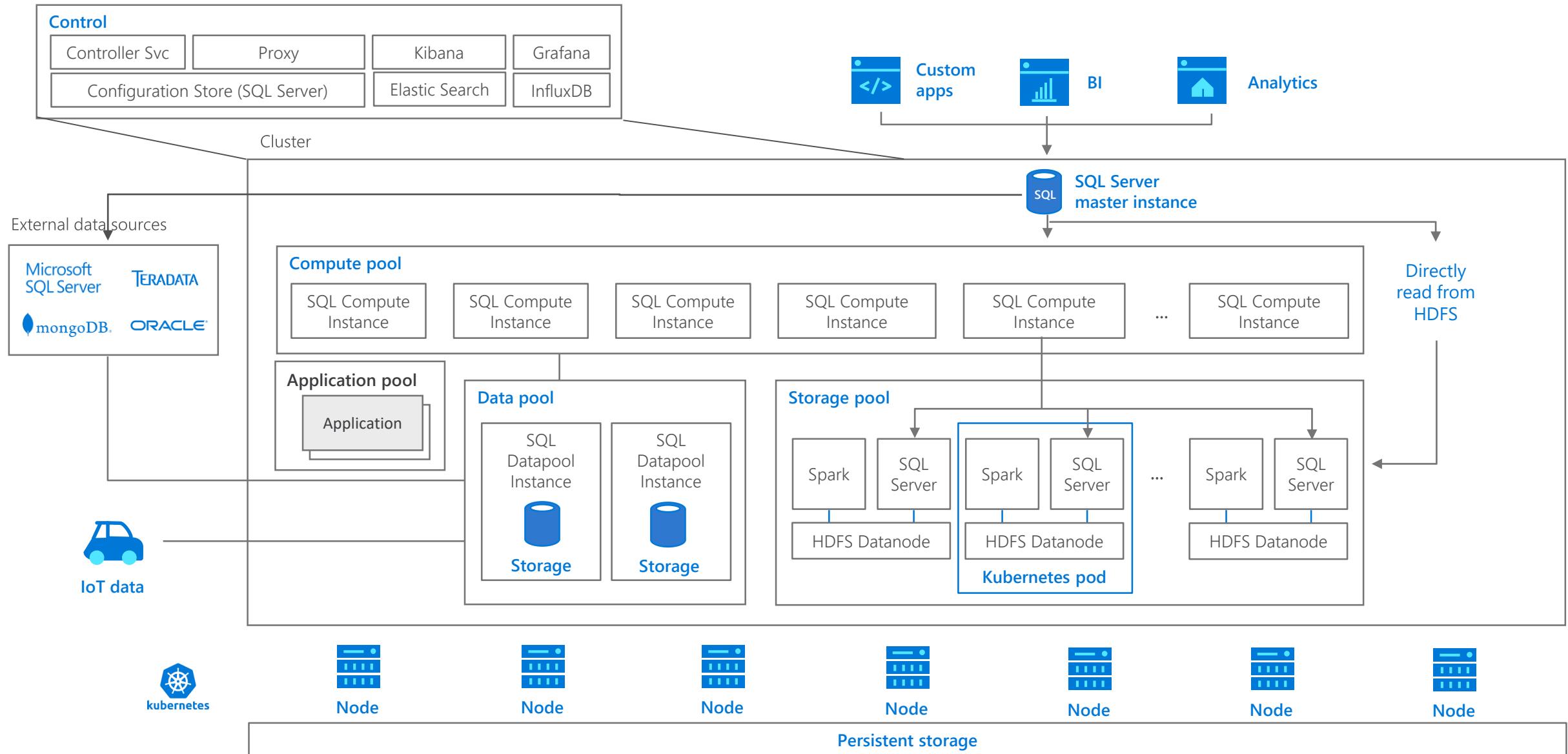


1/10th the cost of Oracle

All TPC Claims as of 1/19/2018.

¹<http://www.tpc.org/4081>; ²<http://www.tpc.org/3331>; ³<http://www.tpc.org/3326>; ⁴<http://www.tpc.org/3321>; ⁵National Institute of Standards and Technology Comprehensive Vulnerability Database

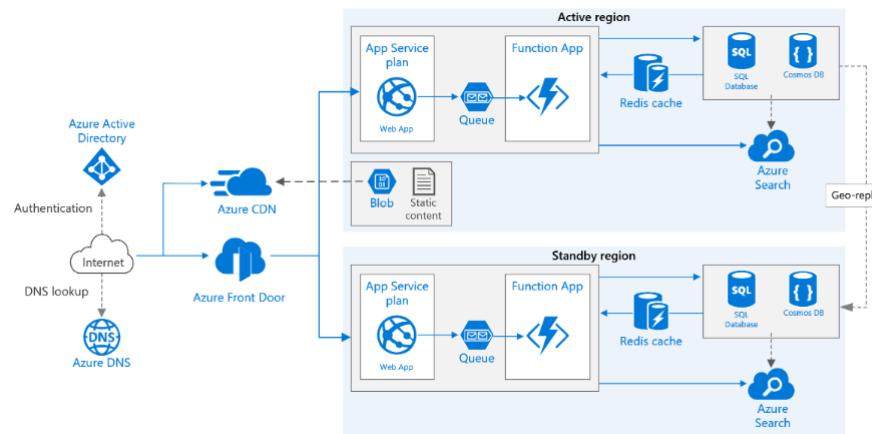
BIG DATA CLUSTERS DEPLOYED ON KUBERNETES, OPENSHIFT, AKS



PRINCIPLES -

- This reference architecture shows how to run an Azure App Service application in multiple regions to achieve high availability.

<https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/app-service-web-app/multi-region>



Attribution

Some pictures used are from
<http://www.Pixels.com/about> - Pexels
provides high quality and completely
free stock photos licensed under the
Pexels license.