



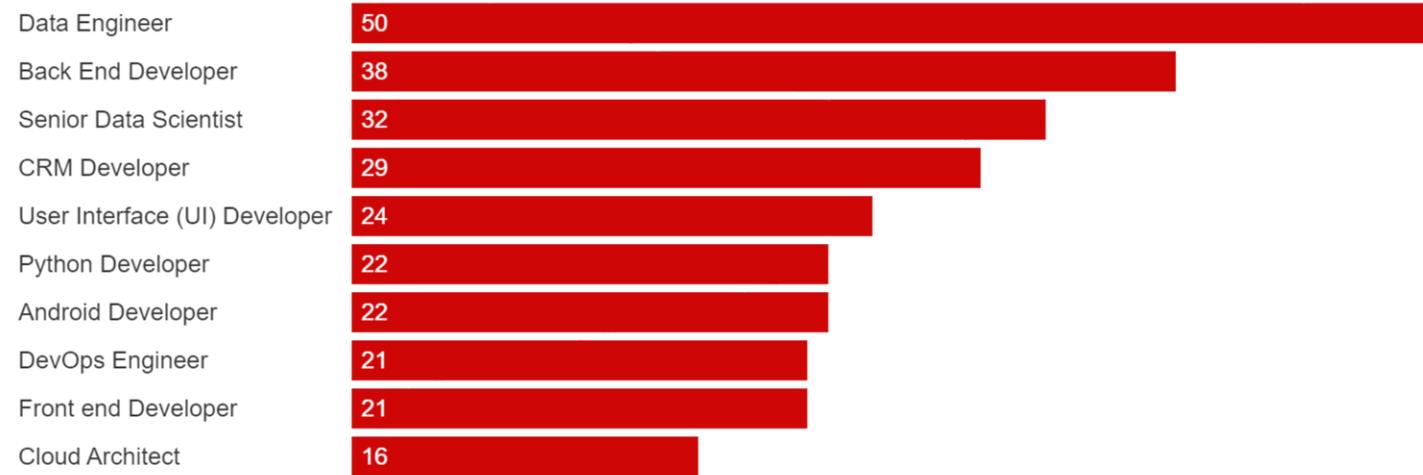
DP-900

Microsoft Azure Data Fundamentals



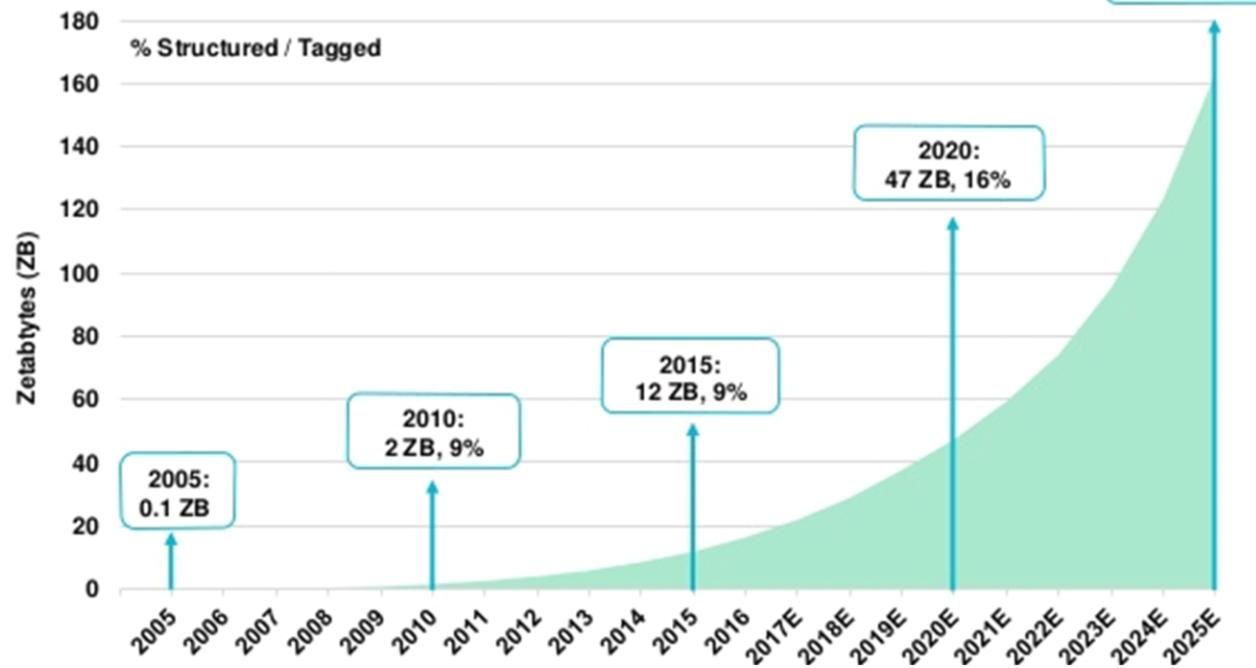
Fastest Growing Tech Occupations

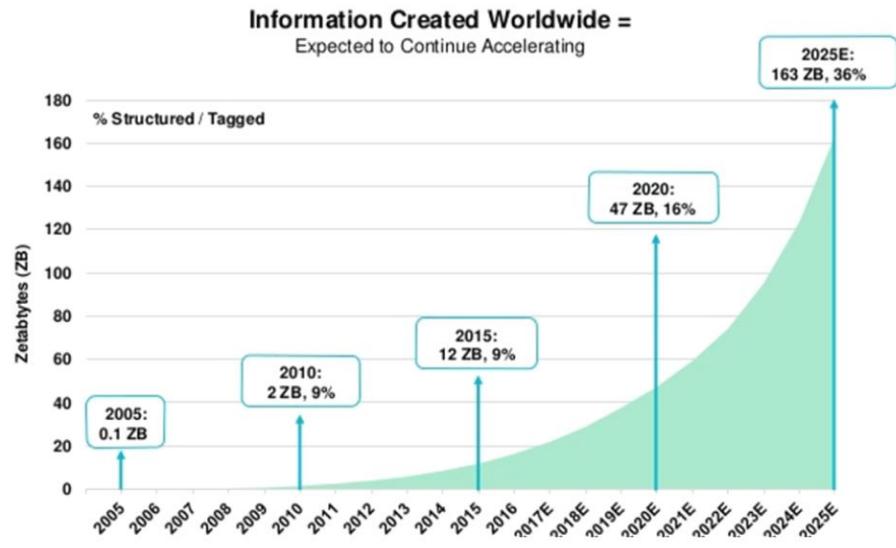
Year-Over-Year Growth



Information Created Worldwide =

Expected to Continue Accelerating



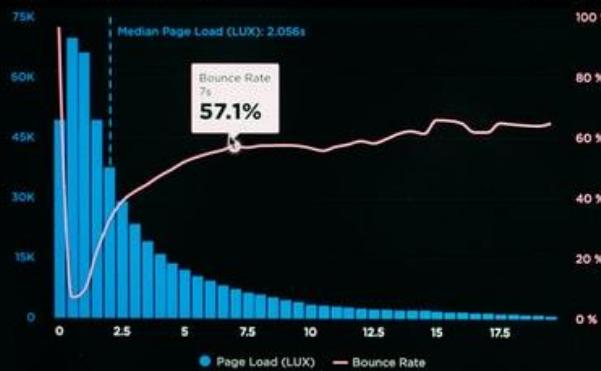


WHAT'S A ZETTABYTE?

1 kilobyte	1,000,000,000,000,000,000
1 megabyte	1,000,000,000,000,000,000
1 gigabyte	1,000,000,000,000,000,000
1 terabyte	1,000,000,000,000,000,000
1 petabyte	1,000,000,000,000,000,000
1 exabyte	1,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000

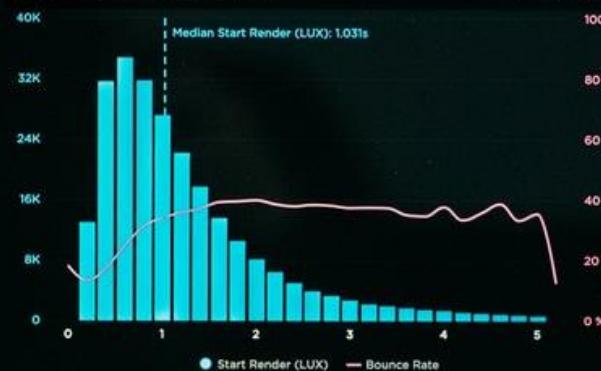
USERS: LAST 7 DAYS USING MEDIAN ▾

LOAD TIME VS BOUNCE RATE



O OPTIONS

START RENDER VS BOUNCE RATE



PAGE VIEWS VS ONLOAD

Page Load (LUX)

0.7s

Page Views (LUX)

2.7Mpv

O OPTIONS

Bounce Rate (LUX)

40.6%

SESSIONS

Sessions (LUX)

479K

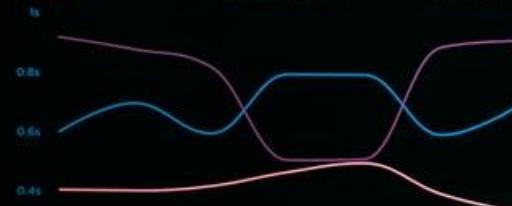
Session Length (LUX)

17min

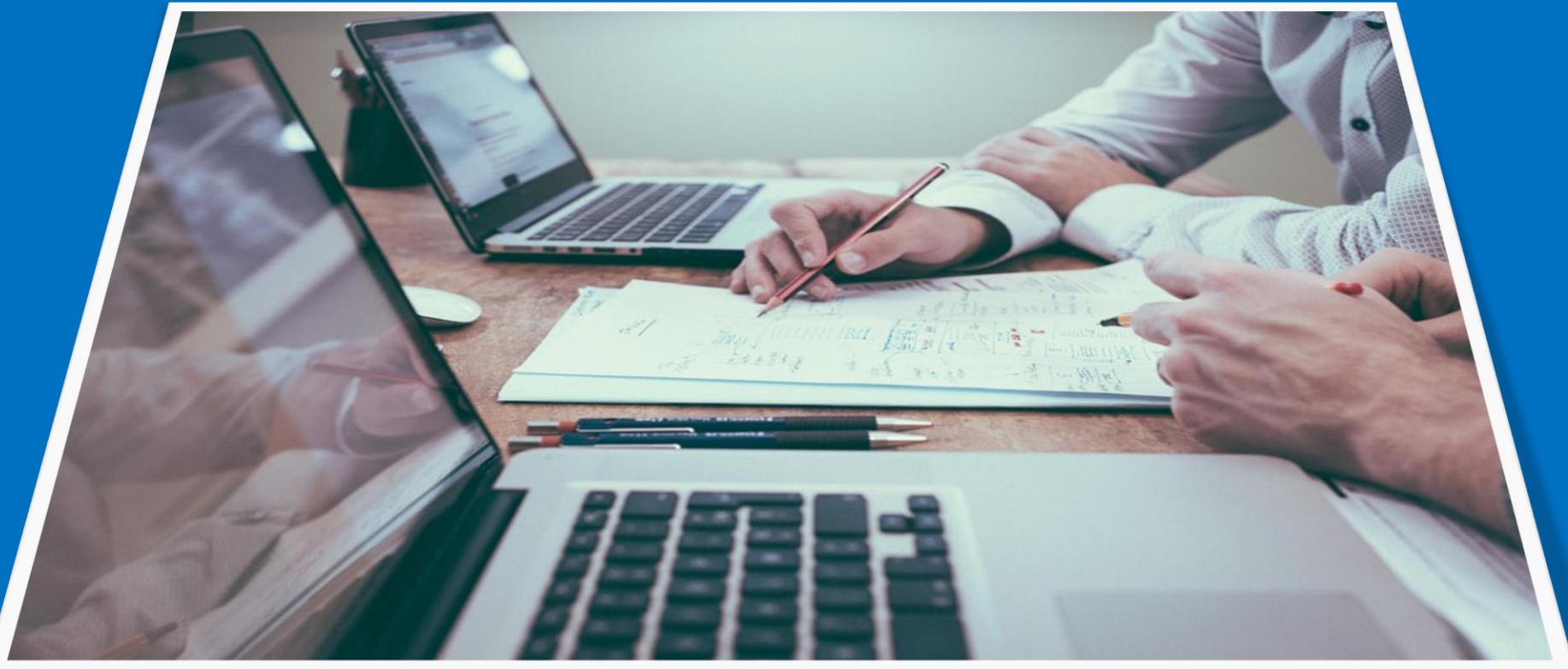
O OPTIONS

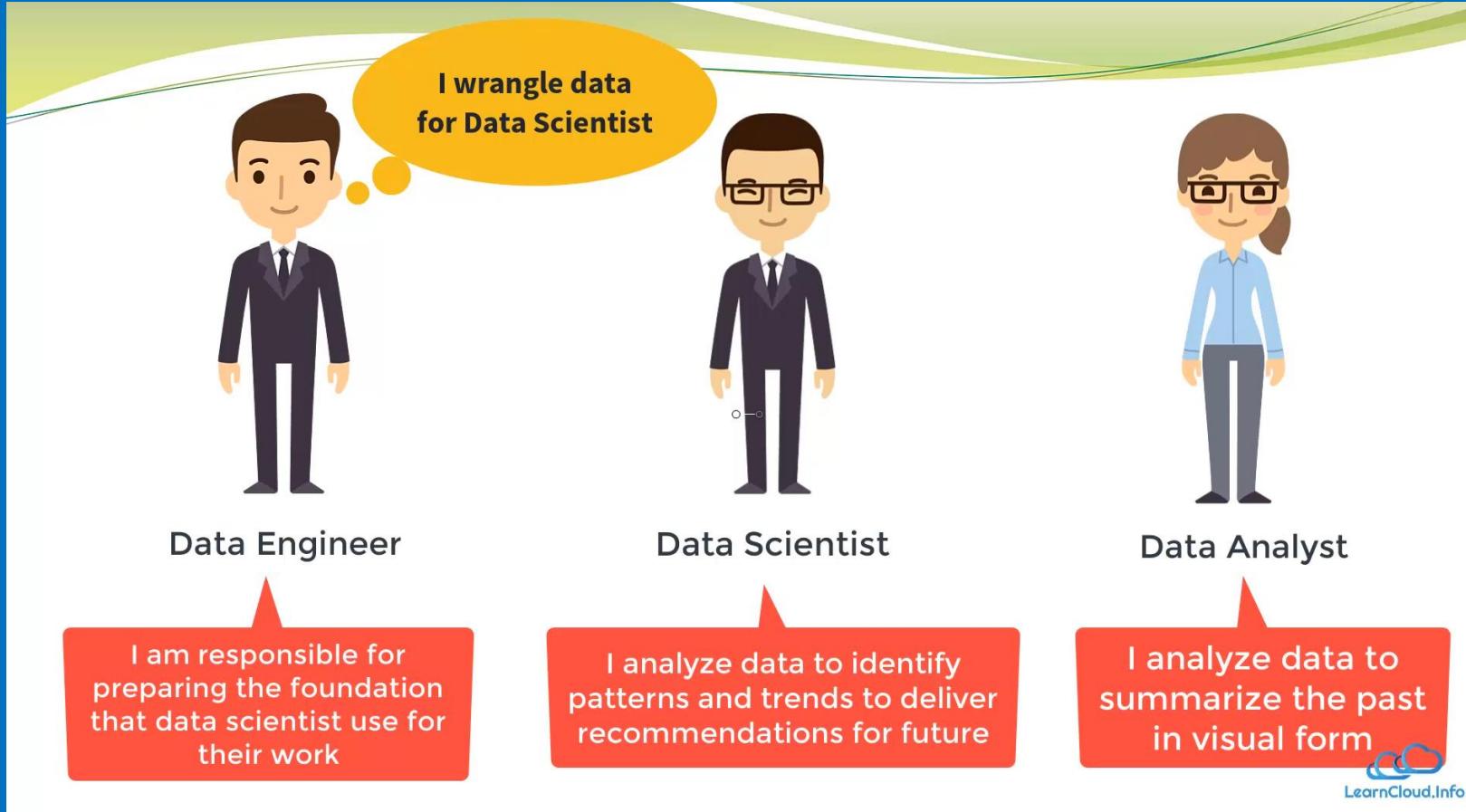
PVs Per Session (LUX)

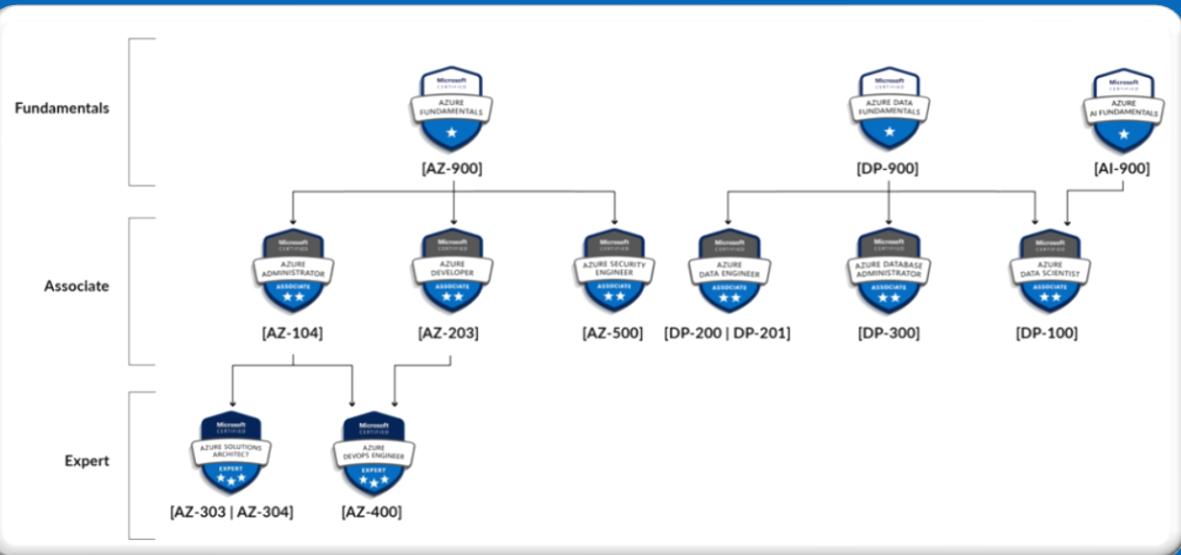
2pv



1

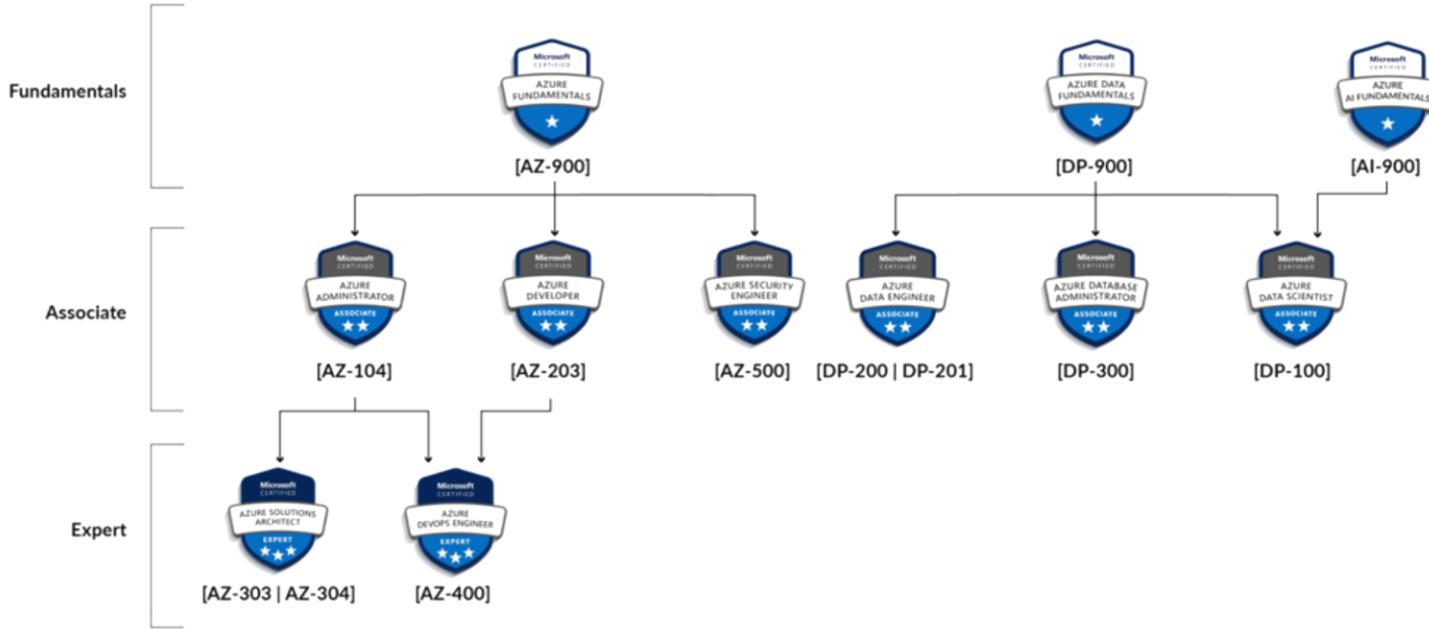






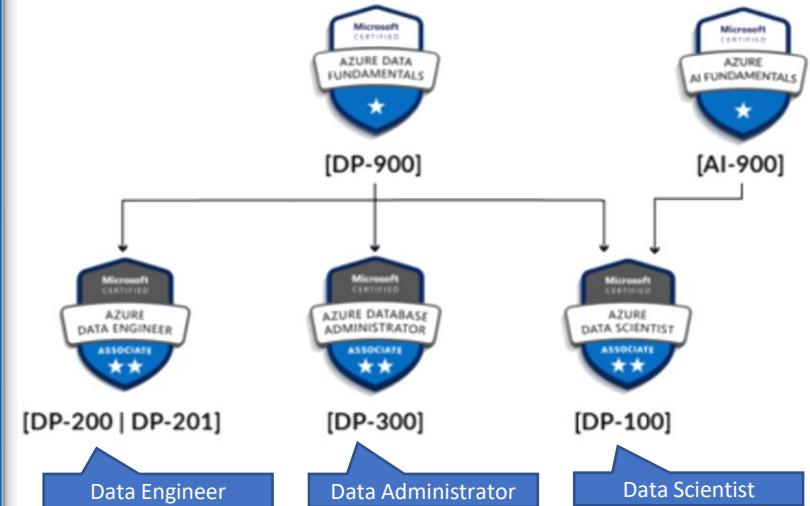
Technologies

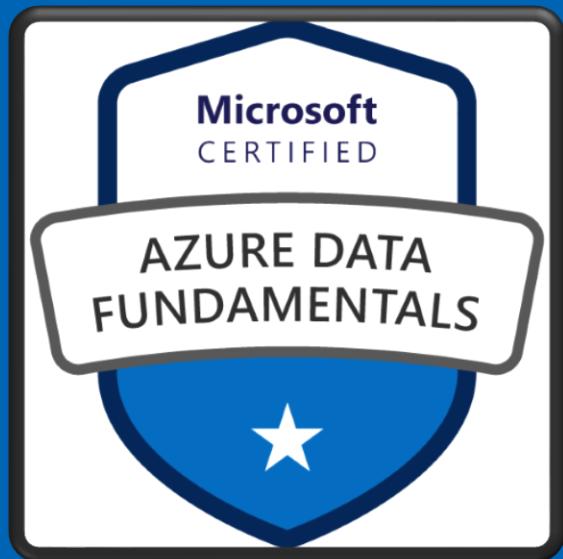




Expert

[AZ-303 | AZ-304] [AZ-400]







Syllabus

The learning objectives and Skills outline
for this exam

Modules

01

Core data concepts

Basics - types of data, processing, analytics and reports

02

Relational DB - Azure

Basics of Relational DB and corresponding offerings in Azure

03

Non-Relational DB - Azure

Basics of Non-Relational DB and corresponding offerings in Azure

04

Analytics workload - Azure

Basics of Data Analytics and corresponding offerings in Azure



Eshant Garg

Data Engineer, Architect, Advisor

LinkedIn: [in/eshantgarg](https://www.linkedin.com/in/eshantgarg)

Intended Audience

- Anyone who wants to clear DP-900 exam
- Anyone who wants to learn basics of database technologies in Azure

Prerequisite

- Absolutely no prerequisite

Language

- Comfortable in English and Indian accent

What includes?

- 6+ hrs. of content, Practice test, quizzes etc.
- PPT, Demo resources and other study material
- Full lifetime access
- Certificate of course completion
- 30-days Money-Back Guarantee



Exam Info and FAQ

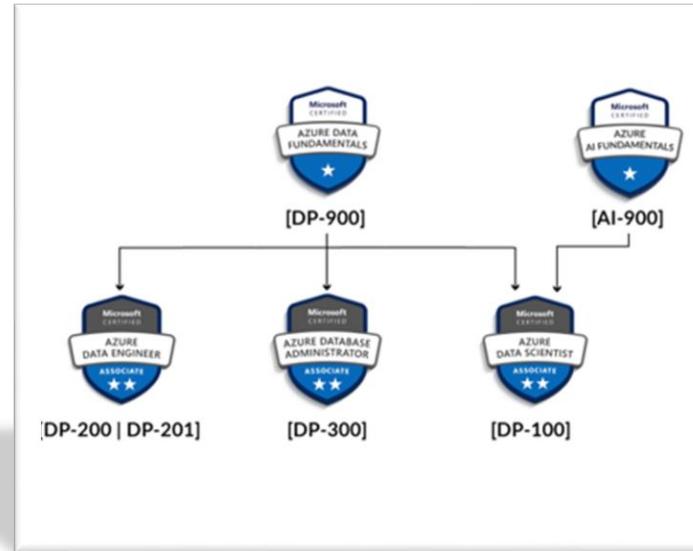


- ✓ Skill Measured
- ✓ Prerequisite for DP-900?
- ✓ Exam details?
- ✓ Question types?
- ✓ General tips for exam?
- ✓ How to prepare for exam? Or How to best use this course?
- ✓ Challenge – 4 badges

Exam Info

Prerequisite for DP-900

- NO Prerequisite
- NO prior knowledge of any language or in fact any technology



Exam Info

DP-900 Exam Details

- Exam Duration: 60 Minutes
- Number of Questions: 49 (40-60)
- Passing marks – 700/1000 points



Exam Info

Types of question

- Multiple choice question
- Multiple select question
- Drop down
- True/False

Do NOT have

- Case studies
- Labs



Exam Info

General Exam Tips

- Mark for review
- Eliminate wrong answers
- Look for keyword
- Marking system - 700 points out of 1000 points, NOT 70%
 - Partial correct answer also have points
- No negative marking
- Comment question



Exam Info

How to prepare for exam?

- Be confident that you are in right hands
- Go through all videos carefully and understand basic concepts
 - Refer “further studies” if required, not necessary
- Quizzes: further convey concepts on top of videos
- Practice Questions



COMPLETED

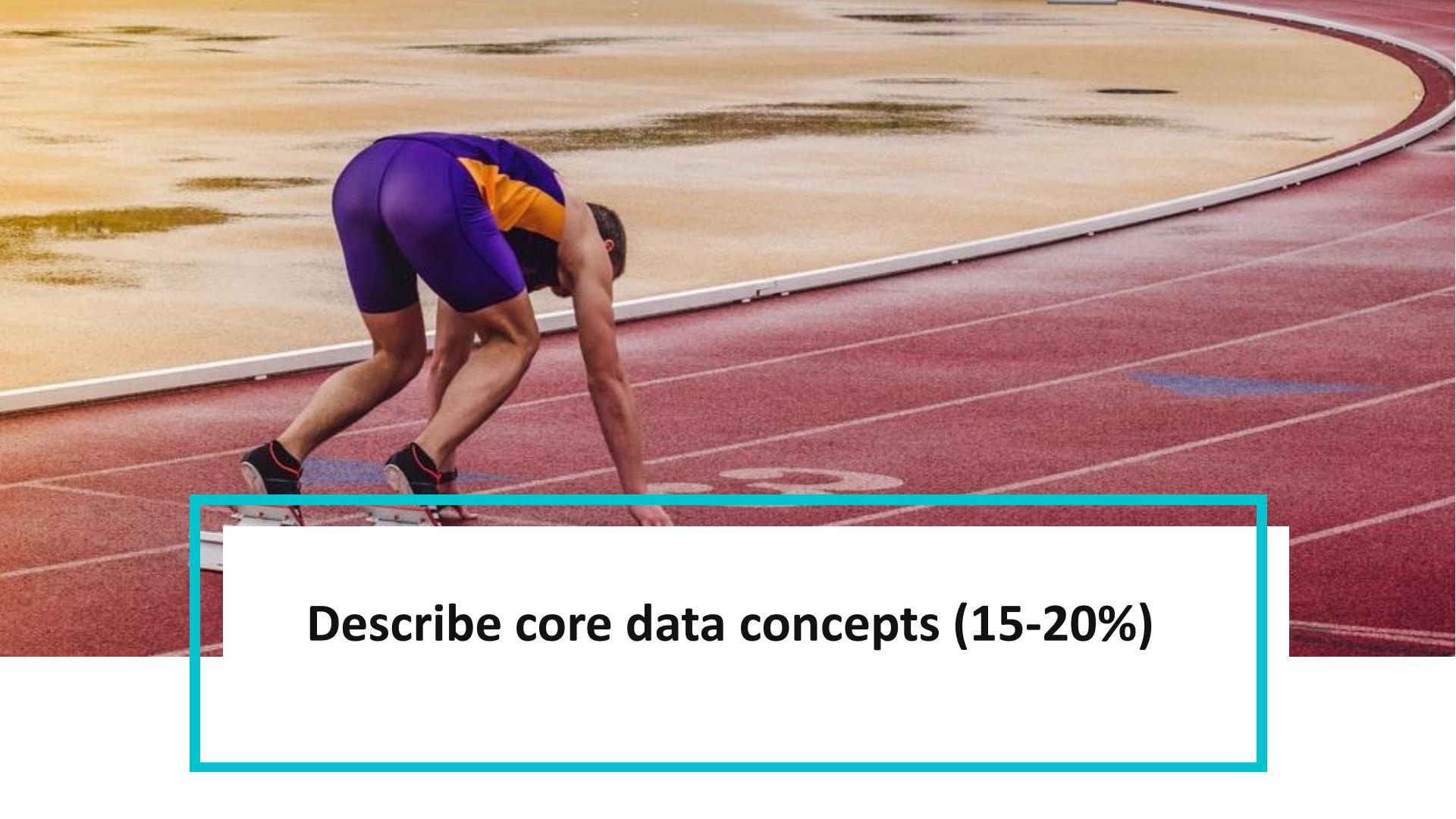
PENDING



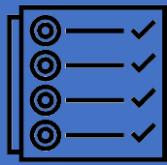
COMPLETED



PENDING

A photograph of a male athlete in mid-air, performing a long jump on a red running track. He is wearing purple and yellow athletic gear. The background shows a yellow sandpit and the curve of the track.

Describe core data concepts (15-20%)

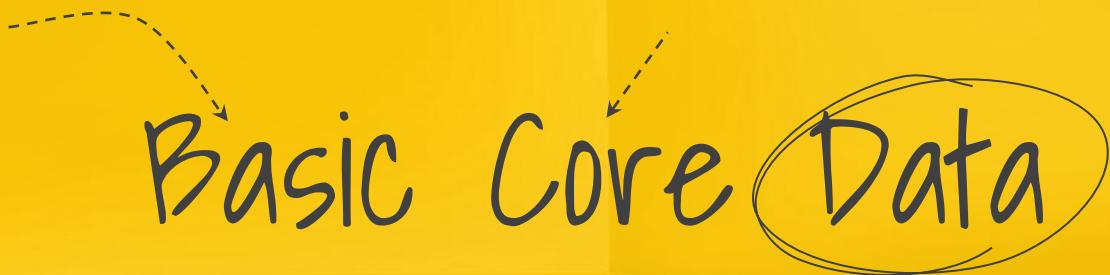


Learning Objectives

Core Data Workloads

- Batch vs Streaming data
- Characteristics of Relational Data

Basic Core Data Concepts



Agenda

- What is Data?
- Data Types?
- Data processing solutions
- OLTP vs OLAP

“

Data is a collection of facts such as numbers, descriptions, and observations used in decision making.

Data Classification

- Structured Data
- Semi-Structured Data
- Unstructured Data

Structured Data

CustomerKey	FirstName	MiddleName	LastName	BirthDate	MaritalStatus	Gender
11000	Jon	V	Yang	1971-10-06	M	M
11001	Eugene	L	Huang	1976-05-10	S	M
11002	Ruben	NULL	Torres	1971-02-09	M	M
11003	Christy	NULL	Zhu	1973-08-14	S	F
11004	Elizabeth	NULL	Johnson	1979-08-05	S	F
11005	Julio	NULL	Ruiz	1976-08-01	S	M
11006	Janet	G	Alvarez	1976-12-02	S	F
11007	Marco	NULL	Mehta	1969-11-06	M	M
11008	Rob	NULL	Verhoff	1975-07-04	S	F
11009	Shannon	C	Carlson	1969-09-29	S	M
11010	Jacquelyn	C	Suarez	1969-08-05	S	F
11011	Curtis	NULL	Lu	1969-05-03	M	M
11012	Lauren	M	Walker	1979-01-14	M	F
11013	Ian	M	Jenkins	1979-08-03	M	M
11014	Sydney	NULL	Bennett	1973-11-06	S	F
11015	Chloe	NULL	Young	1984-08-26	S	F
11016	Wyatt	L	Hill	1984-10-25	M	M
11017	Shannon	NULL	Wang	1949-12-24	S	F
11018	Clarence	D	Rai	1955-10-06	S	M
11019	Luke	L	Lal	1983-09-04	S	M
11020	Jordan	C	King	1984-03-19	S	M
11021	Destiny	NULL	Wilson	1984-03-02	S	F
11022	Ethan	G	Zhang	1984-04-10	M	M
11023	Seth	M	Edwards	1984-04-09	M	M
11024	Russell	NULL	Xie	1984-03-16	M	M
11025	Alejandro	NULL	Beck	1951-06-22	M	M

Structured data is typically tabular data that is represented by rows and columns in a database.

Example – Datawarehouse,
ERP, CRM

Semi-Structured Data

JSON

```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}

## Document 2 ##
{
  "customerID": "103249",
  "name":
  {
    "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB"
  },
  "address":
  {
    "street": "Another Street",
    "number": "202",
    "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK"
  },
  "ccOnFile": "yes"
}
```

It has some organizational framework but does not have the complete structure that is required to fit in a relational database.

Example – CSV, XML, JSON

unstructured Data



Unstructured data is data which is not organized in any predefined manner.

Example – audio and video files, and binary data files

Data Storage in Cloud

- Structured Data - Azure SQL Database
- Semi-Structured Data - Azure Cosmos DB
- Unstructured Data - Azure Blob Storage

Data Processing Solutions

Data Processing Solutions

- Transaction System (OLTP)
- Analytical System (OLAP)

OLTP vs OLAP

OLTP

Day-to-day handling of transactions that result from enterprise operation

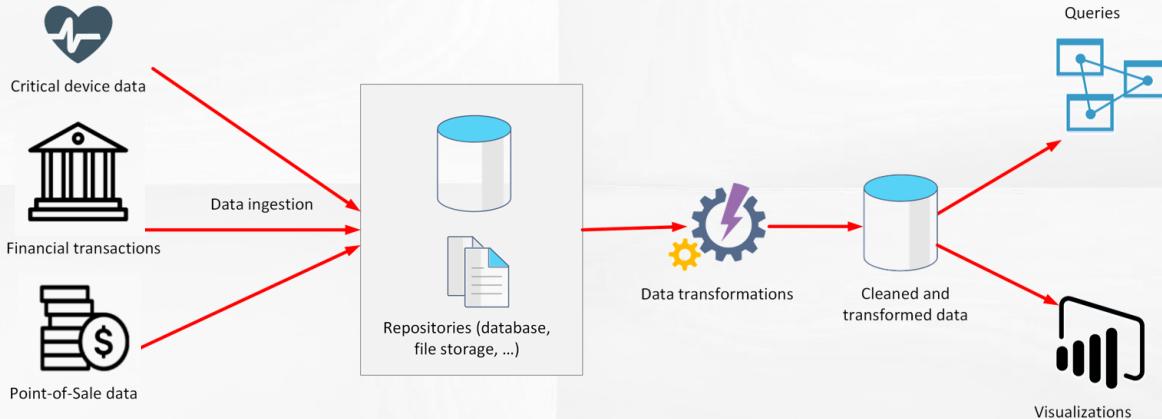
- Small, discrete, unit of work
- Often high-volume
- Data processed very quickly.

OLAP

Analysis of information in a database for the purpose of making management decisions

- Big picture view of the information held in a database.
- Generate insights to make business decisions

OLTP vs OLAP



Batch Processing

VS

Stream Processing

Batch vs Streaming

Batch processing

- Buffering and processing the data in groups
- Example – Credit Card Bill

Streaming processing

- Streaming handles data in real time. Each new piece of data is processed when it arrives.
- Stream processing is ideal for time-critical operations that require an instant real-time response.
- Example – Stock market, Heat alarm system, YouTube, Netflix, this course

Batch Processing Advantages

- Large volumes of data can be processed at a convenient time.
- Utilize system resources by running at the idle time

Stream Processing Disadvantages

- Can only process small volume of data, at real time
- Have to process at real time

Batch Processing Disadvantages

- Time delay between ingesting the data and getting the results.
- Minor data errors can affect the whole batch

Stream Processing Advantages

- No delay between event and result

Differences between batch and streaming data



Data Scope

- Batch: can process all the data in the dataset
- Stream: access to the most recent data received



Data Size

- Batch: suitable for handling large datasets efficiently
- Stream: intended for individual records or micro batches



Performance

- Batch: latency for few hours
- Stream: occurs immediately in seconds/milliseconds



Analysis

- Batch: used to perform complex analytics.
- Stream: used for simple response functions, aggregates, or calculations such as rolling averages.

Characteristics of Relational Data

Characteristics of relational data

Customers

Customer ID	Customer Name	Customer Address
C1	Fred	...
C2	Bert	...
C3	Jane	...

Products

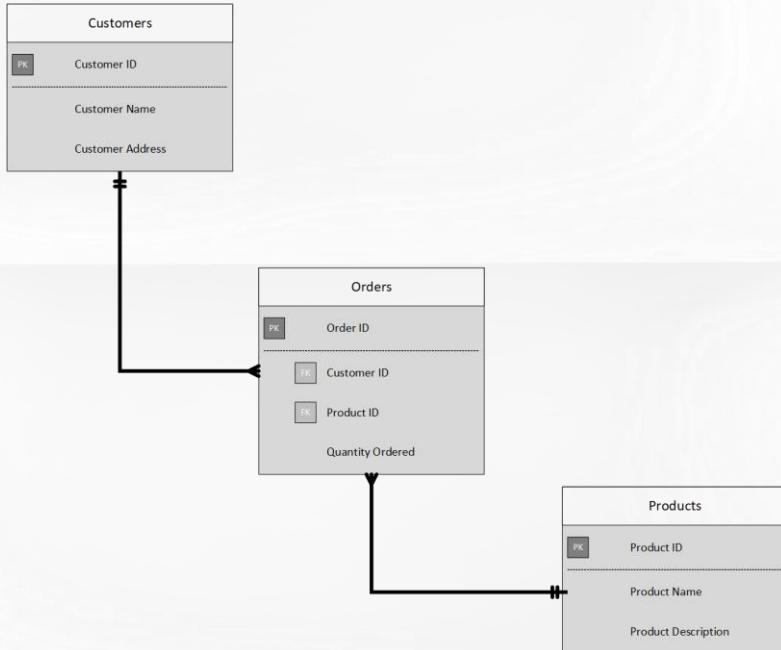
Product ID	Product Name	Description
P1	Shirt	...
P2	Tie	...
P3	Collar	...

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

- Table: Data is stored in a table
- Table consist of rows and columns
- Row: each row represents a single instance of an entity
- Column: define the properties of the entity
- Each column is defined by a datatype
- All rows have the same number of columns

Characteristics of relational data



- Some columns are used to maintain relationships between tables
- Model shows the structure of the entities
- Primary Key: uniquely identify each row
- Foreign Key: reference, or link to, the primary key of another table
 - used to maintain the relationships between tables.

Normalization

*Split an entity into more than one table. This process is called **normalization**.*

-  Reduce storage
-  Avoid data duplication
-  Improve data quality

Customers

CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Orders

OrderID	CustomerName	CustomerPhone
AD100	Noam Maoz	XXX-XXX-XXXX
AD101	Noam Maoz	XXX-XXX-XXXX
AD102	Noam Maoz	XXX-XXX-XXXX
AX103	Qamar Mounir	XXX-XXX-XXXX
AS104	Qamar Mounir	XXX-XXX-XXXX
AR105	Claude Paulet	XXX-XXX-XXXX
MK106	Muisto Linna	XXX-XXX-XXXX

Normalization

*Split an entity into more than one table. This process is called **normalization**.*

- Primary keys and Foreign keys are used to define relationship
- No data duplication exists other than key values
- Data is retrieved by joining tables together in a query

Customers			Orders		
CustomerID	CustomerName	CustomerPhone	OrderID	CustomerID	SalesPersonID
100	Muisto Linna	XXX-XXX-XXXX	AD100	101	200
101	Noam Maoz	XXX-XXX-XXXX	AD101	101	200
102	Vanja Matkovic	XXX-XXX-XXXX	AD102	101	200
103	Qamar Mounir	XXX-XXX-XXXX	AX103	103	201
104	Zhenis Omar	XXX-XXX-XXXX	AS104	103	201
105	Claude Paulet	XXX-XXX-XXXX	AR105	105	200
106	Alex Pettersen	XXX-XXX-XXXX	MK106	105	201

Structured Query Language (SQL)

- Most relational databases support Structured Query Language (SQL)
 - SQL Server – TSQL
 - Oracle – PL/SQL
- You use SQL to **create** tables, **insert**, **update**, and **delete** rows in tables, and to **query** data.

Structured Query Language (SQL)

```
SELECT CustomerID, CustomerName, CustomerAddress  
FROM Customers
```

```
SELECT OrderID, ProductID  
FROM Orders  
WHERE CustomerID = 'C1'
```

```
SELECT Customers.CustomerName, Orders.QuantityOrdered, Products.ProductName  
FROM Customers JOIN Orders  
ON Customers.CustomerID = Orders.CustomerID  
JOIN Products  
ON Orders.ProductID = Products.ProductID
```

Relational database use cases

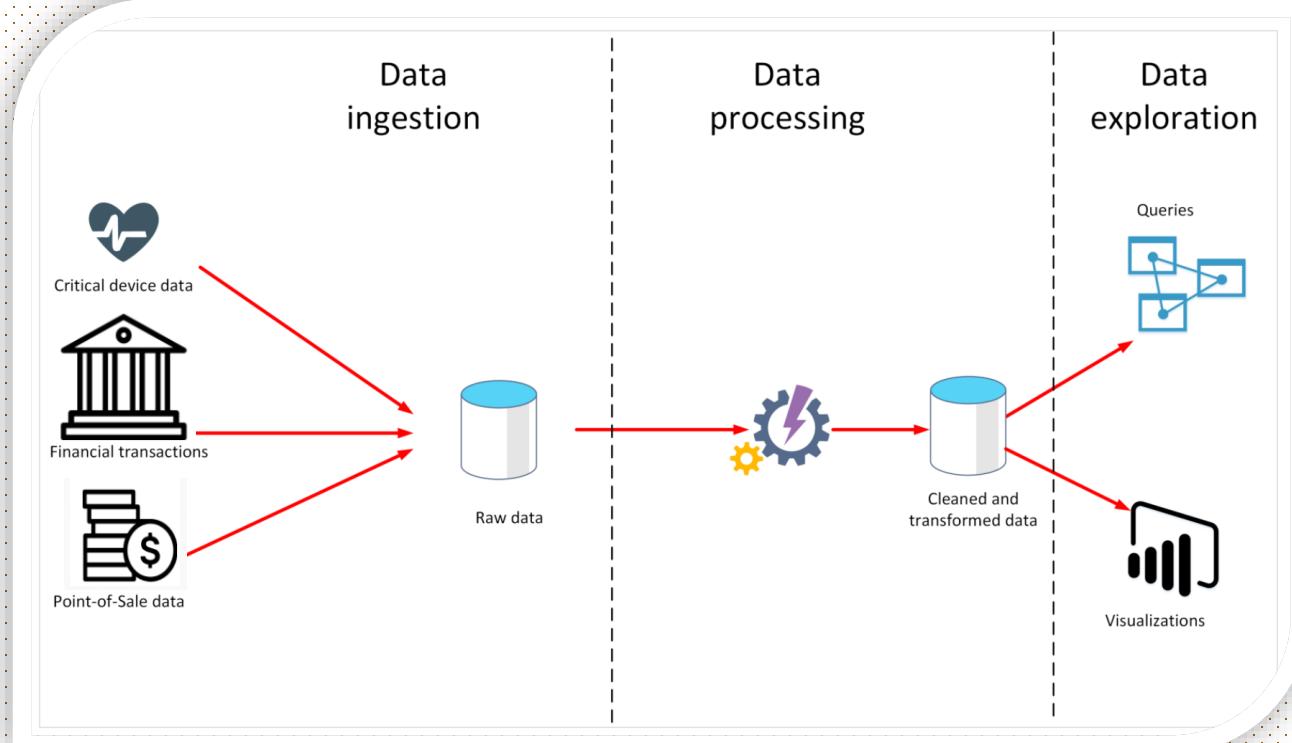
-  Relational databases are well suited for OLTP applications
-  OLTP applications are focused on transaction-oriented tasks that process a very large number of transactions per minute.
-  Examples of OLTP applications that use relational databases are
 -  Banking solutions
 -  Online retail applications,
 -  Flight reservation systems, and
 -  Most online purchasing applications.
-  Relational Database **NOT** use for
 -  Music, video, or other media files – Blob storage
 -  Social networking sites, highly complex web of relationships – Graph Database

Benefits of Relational Database

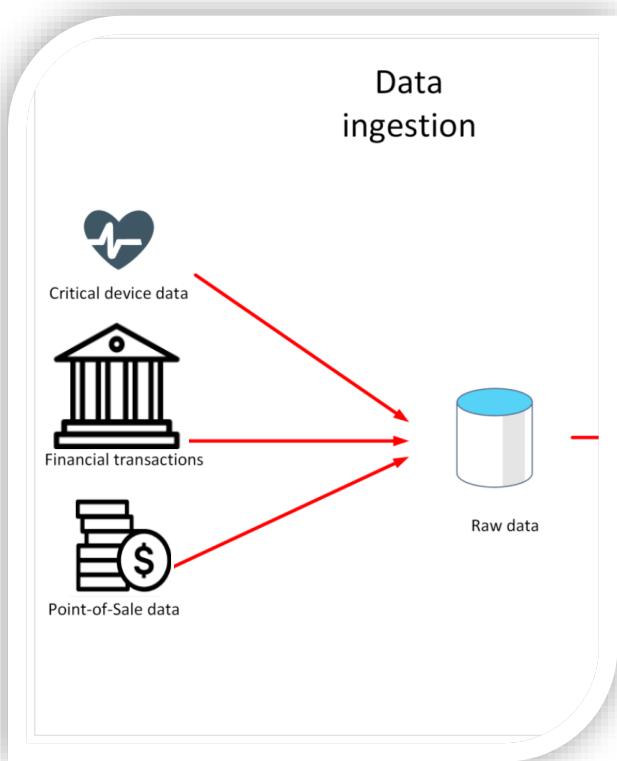
- **Consistency:** Maintains data consistency across applications and database queries.
- **Commitment and Atomicity:** Handles business rules and policies at a very granular level with strict policies
- **Locking and Concurrency:** Prevents other users and applications from accessing data while being updated.

Describe data ingestion and processing

Data Ingestion and Processing

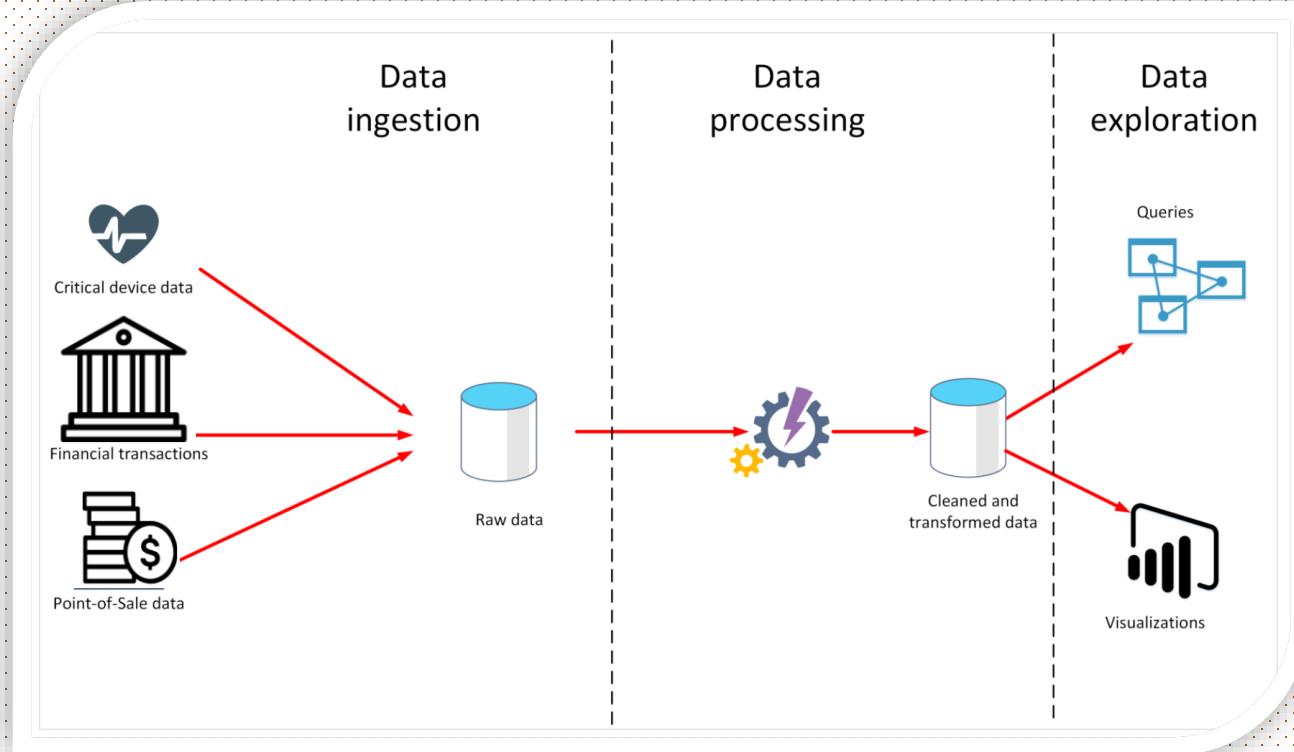


Data Ingestion

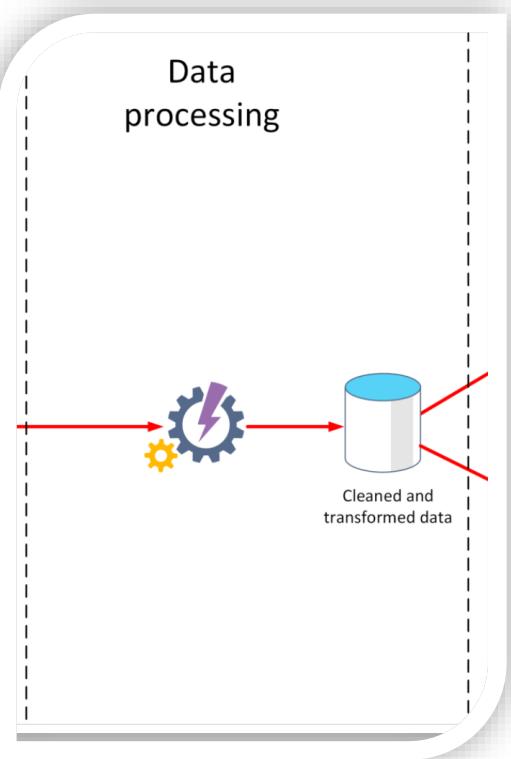


- Data ingestion is the process of obtaining and importing data
- Data can arrive as a continuous stream or batches
- Raw data can be stored at DBMS, a set of files, or some other type of fast, easily accessible storage.
- Ingestion might perform:
 - Filtering: Example reject suspicious, corrupt, or duplicated data
 - Simple transformation: converting data into a standard form, example: reformat all date and time

Data Ingestion and Processing

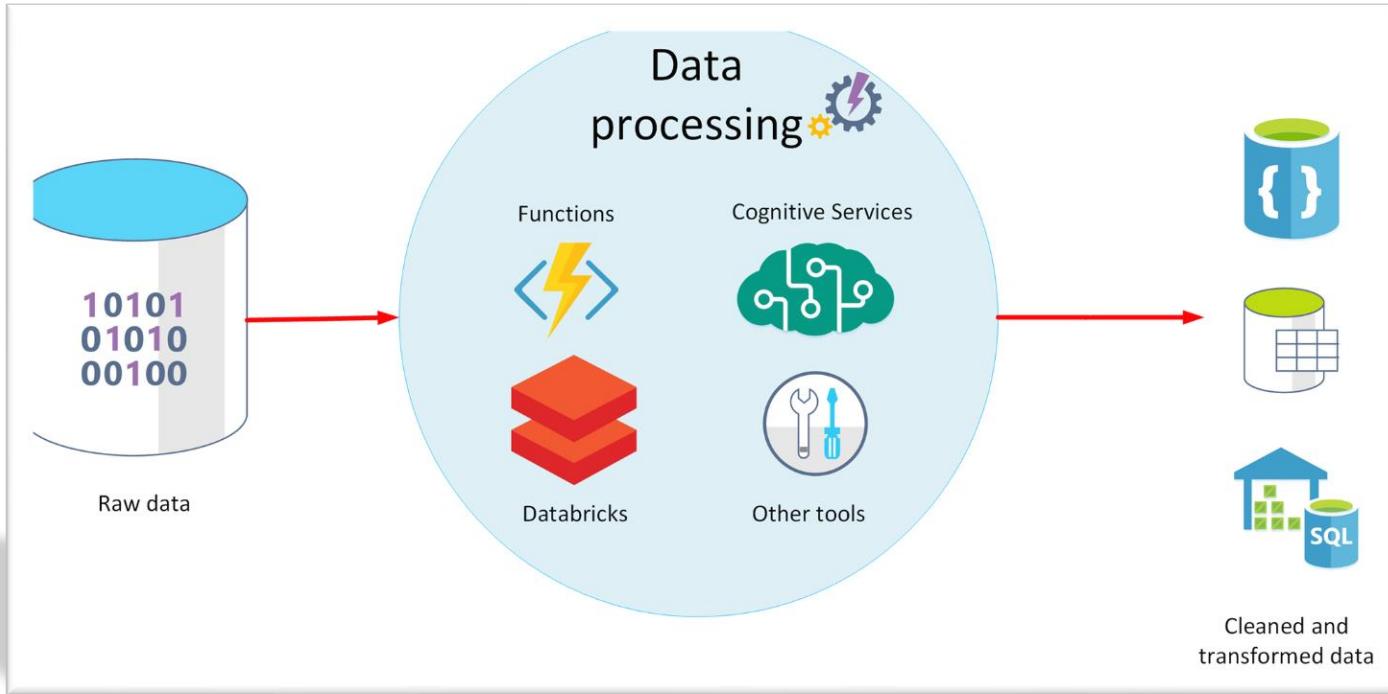


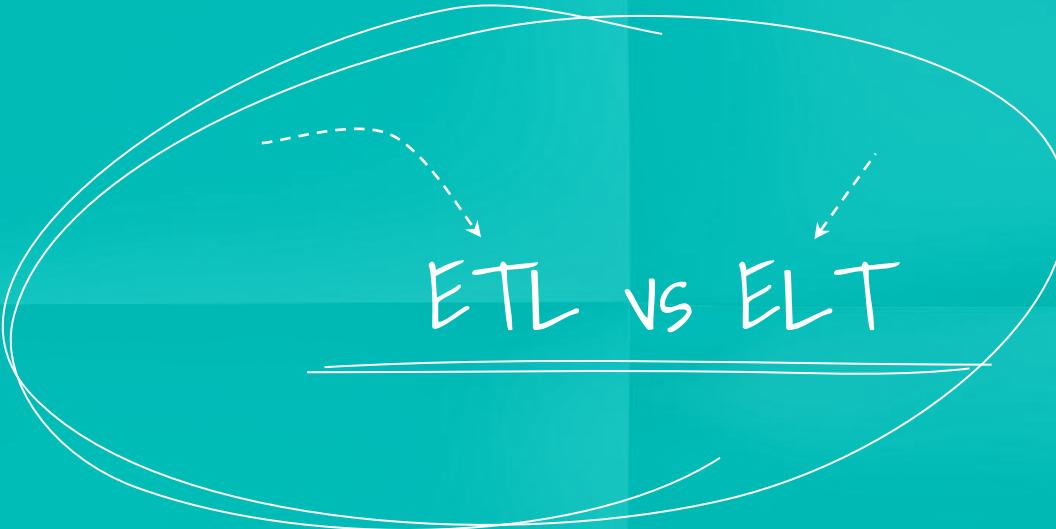
Data Processing



- Data processing takes the data in its raw form, cleans it, and converts it into a more meaningful format (tables, graphs, documents, and so on)
- The result is a database that you can use to perform queries and generate visualizations
- Data Cleaning: removing anomalies, and applying filters and transformations
- Data Wrangling: capture, filter, clean, combine, and aggregate data

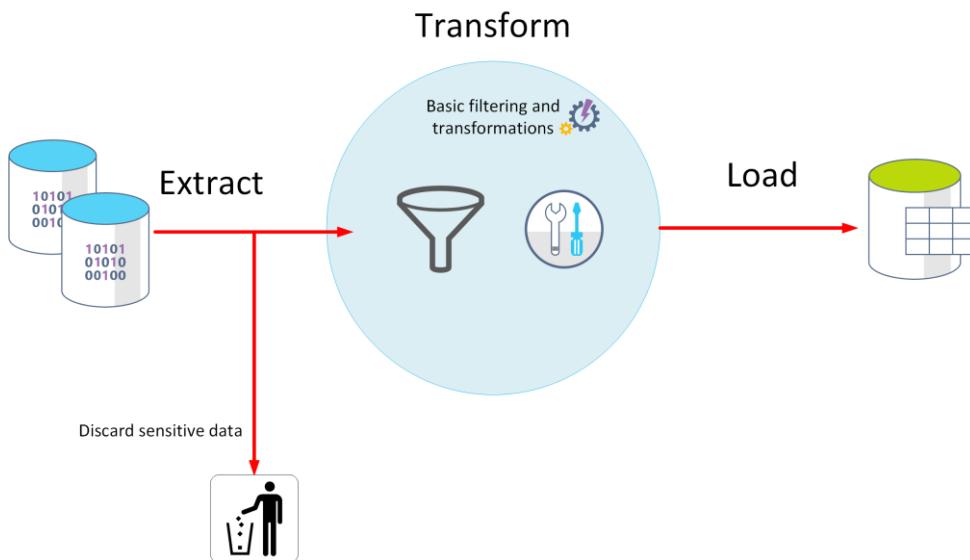
Data Processing





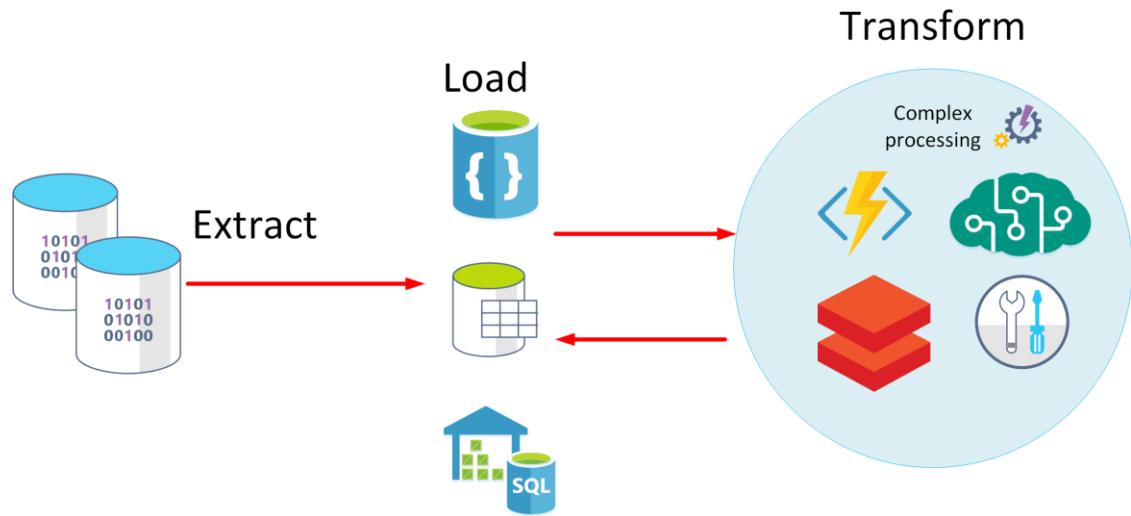
ETL vs ELT

ETL - Extract, Transform, and Load.



- Raw data is retrieved and transformed before being saved
- Suitable for systems that only require simple models
- Basic data cleaning tasks, deduplicating data, and reformatting the contents of individual fields.
- Stream-oriented approach - emphasis on throughput
- ETL can help with data privacy and compliance, removing sensitive data before it arrives in your analytical data models.
- Performed as a continuous pipeline of operations
- **SSIS** can be used

ELT - Extract, Load, and Transform



- Data is stored before being transformed
- More suitable for constructing complex models
- Iterative approach, often using periodic batch processing.
- Suitable for the cloud
- Azure Data Factory - create and schedule data-driven workflows, using compute services such as *Azure HDInsight Hadoop, Azure Databricks*



Data Visualization

Agenda

-  What is reporting?
-  What is business intelligence?
-  What is data visualization?
-  Common Charts of visualizations

What is Reporting?

- Process of organizing data into informational summaries
- Reporting shows you what has happened
- Analysis focuses on explaining why it happened and what you can do about it.

Territory	Product	Units Sold	Sales Revenue
Los Angeles	Posters	6,807	50,916
Sacramento	Posters	4,007	29,972
San Diego	Posters	1,535	11,482
San Francisco	Posters	4,738	35,440
Los Angeles	Frames	1,324	14,498
Sacramento	Frames	637	6,975
San Diego	Frames	234	2,562
San Francisco	Frames	640	7,008
Los Angeles	T-Shirts	6,567	98,177
Sacramento	T-Shirts	2,749	41,098
San Diego	T-Shirts	1,035	15,473
San Francisco	T-Shirts	2,299	34,370
Los Angeles	Hats	6,776	67,421
Sacramento	Hats	2,267	22,557
San Diego	Hats	1,272	12,656
San Francisco	Hats	3,961	39,412

What is Business Intelligence?

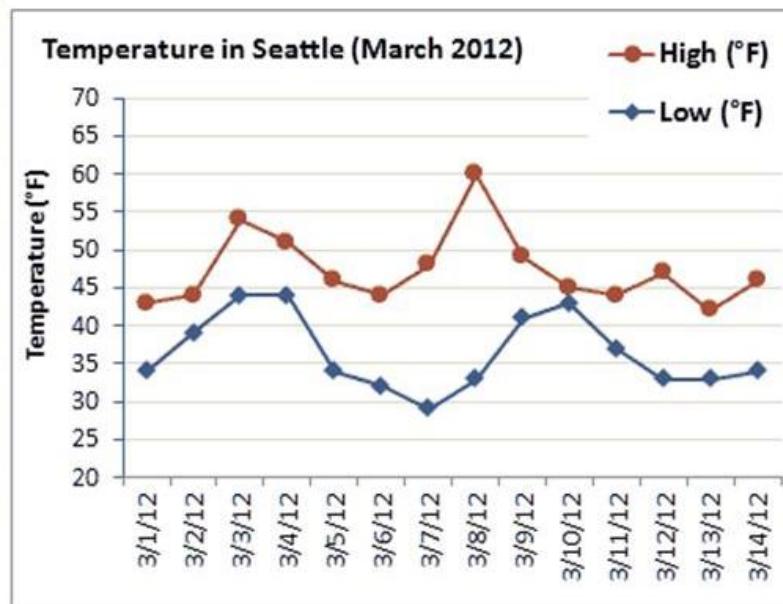
- Refers to technologies, applications, and practices for the collection, integration, analysis, and presentation of business information.
- Business intelligence systems provide historical, current, and predictive views of business operations
- A few ways that business intelligence can help companies make smarter, data-driven decisions:
 - Identify ways to increase profit
 - Analyze customer behavior
 - Compare data with competitors
 - Track performance
 - Optimize operations
 - Predict success
 - Spot market trends

Data Visualization

	A	B	C
1	Temperature Data for Seattle		
2	Date	High (°F)	Low (°F)
3	3/1/12	43	34
4	3/2/12	44	39
5	3/3/12	54	44
6	3/4/12	51	44
7	3/5/12	46	34
8	3/6/12	44	32
9	3/7/12	48	29
10	3/8/12	60	33
11	3/9/12	49	41
12	3/10/12	45	43
13	3/11/12	44	37
14	3/12/12	47	33
15	3/13/12	42	33
16	3/14/12	46	34

Data Visualization

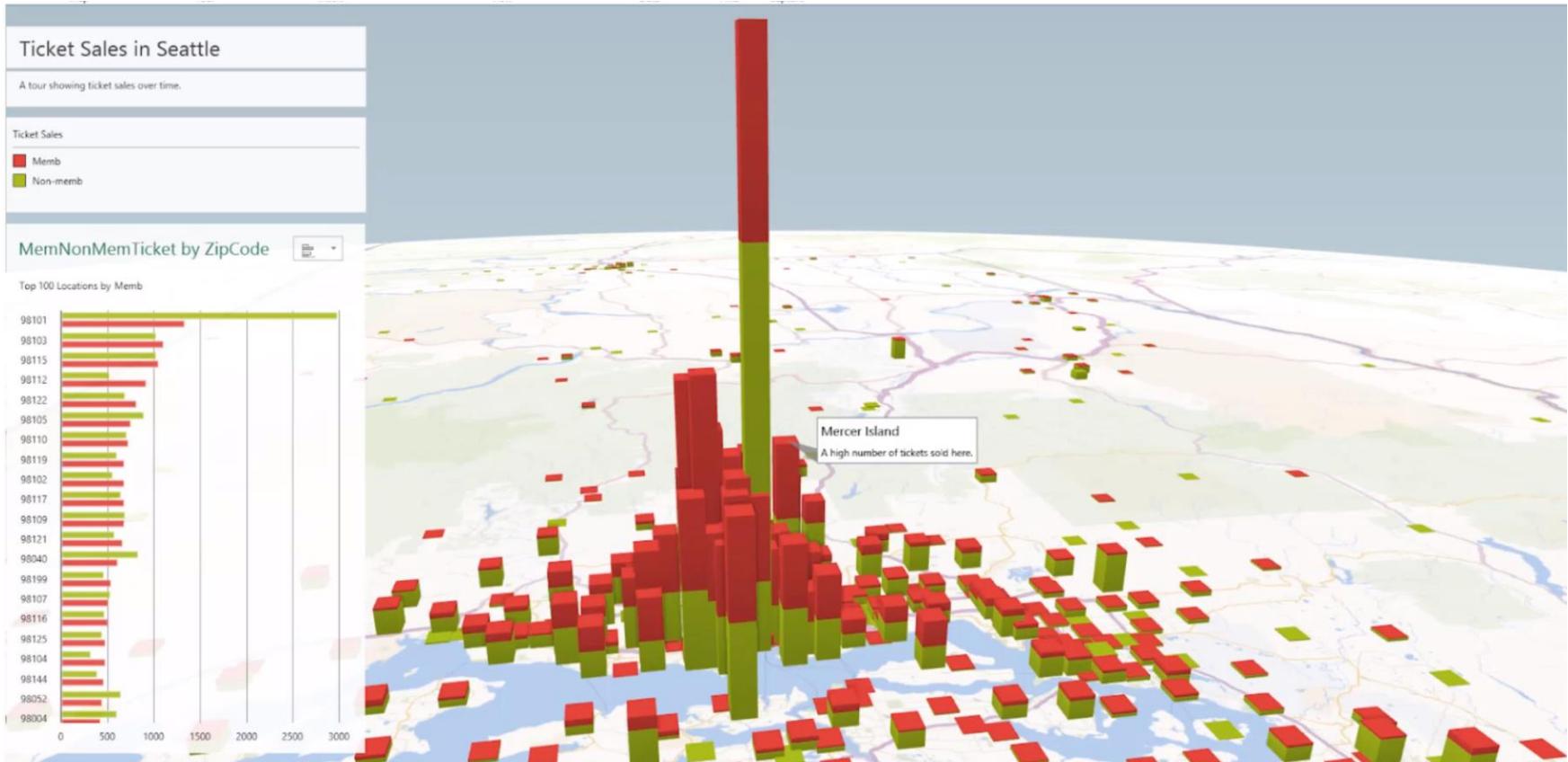
A	B	C	D	E	F	G	H	I	J	K
1 Temperature Data for Seattle										
2 Date	High (°F)	Low (°F)								
3 3/1/12	43	34								
4 3/2/12	44	39								
5 3/3/12	54	44								
6 3/4/12	51	44								
7 3/5/12	46	34								
8 3/6/12	44	32								
9 3/7/12	48	29								
10 3/8/12	60	33								
11 3/9/12	49	41								
12 3/10/12	45	43								
13 3/11/12	44	37								
14 3/12/12	47	33								
15 3/13/12	42	33								
16 3/14/12	46	34								
34										
35	Chart type in Excel:									
36										



Example of a line graph in Excel

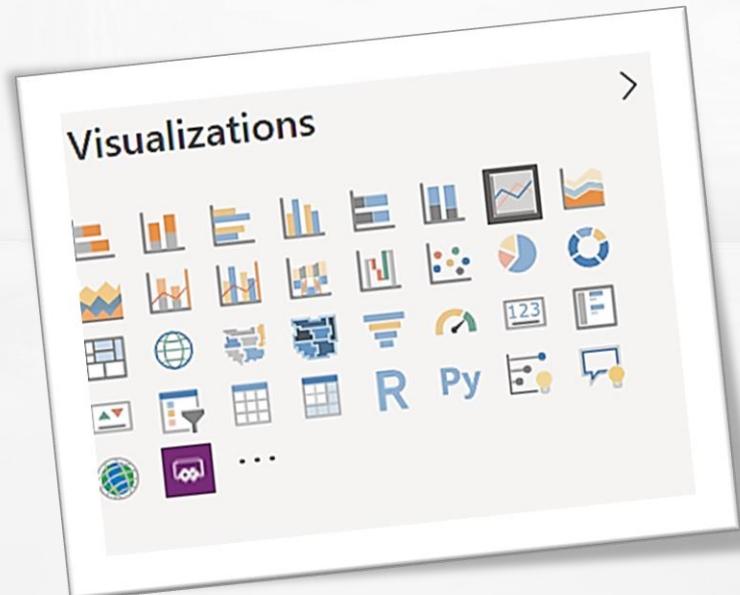
Data Source: <http://www.beautifulseattle.com/mthsum.asp>

Data Visualization



What is Data Visualization?

- Graphical representation of information and data
- Using visual elements like charts, graphs, and maps
- Helps you to focus on the meaning of data, rather than looking at the data itself
- Data visualization tools provide an accessible way to spot and understand trends, outliers, and patterns in data.
- In Azure we use Power BI



Data Visualization

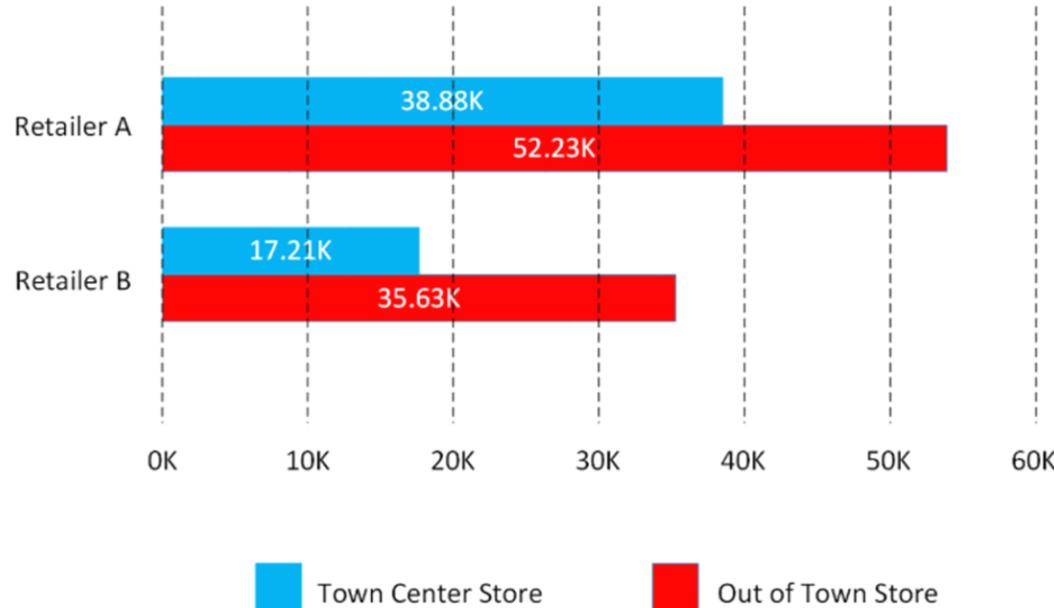


The most common forms of visualizations are:

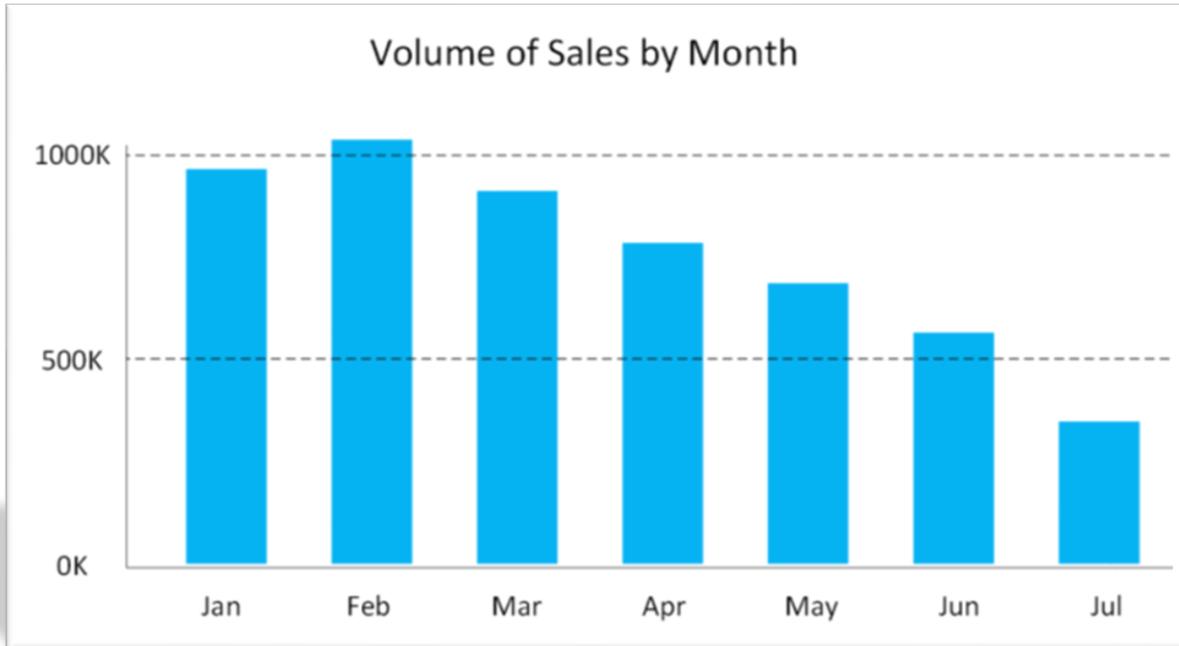
- Bar and column charts
- Line charts
- Matrix
- Key influencers
- Treemap
- Scatter
- Filled map

Bar and column charts

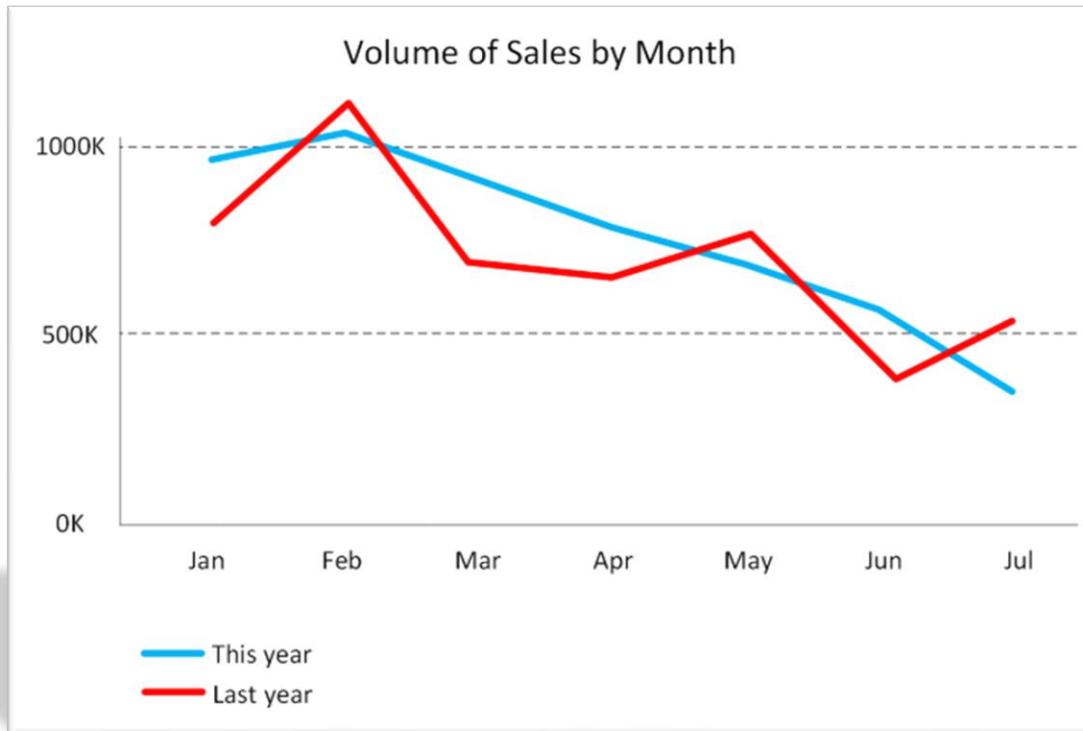
Volume of Sales by Retailer and Store Location



Bar and column charts



Line charts



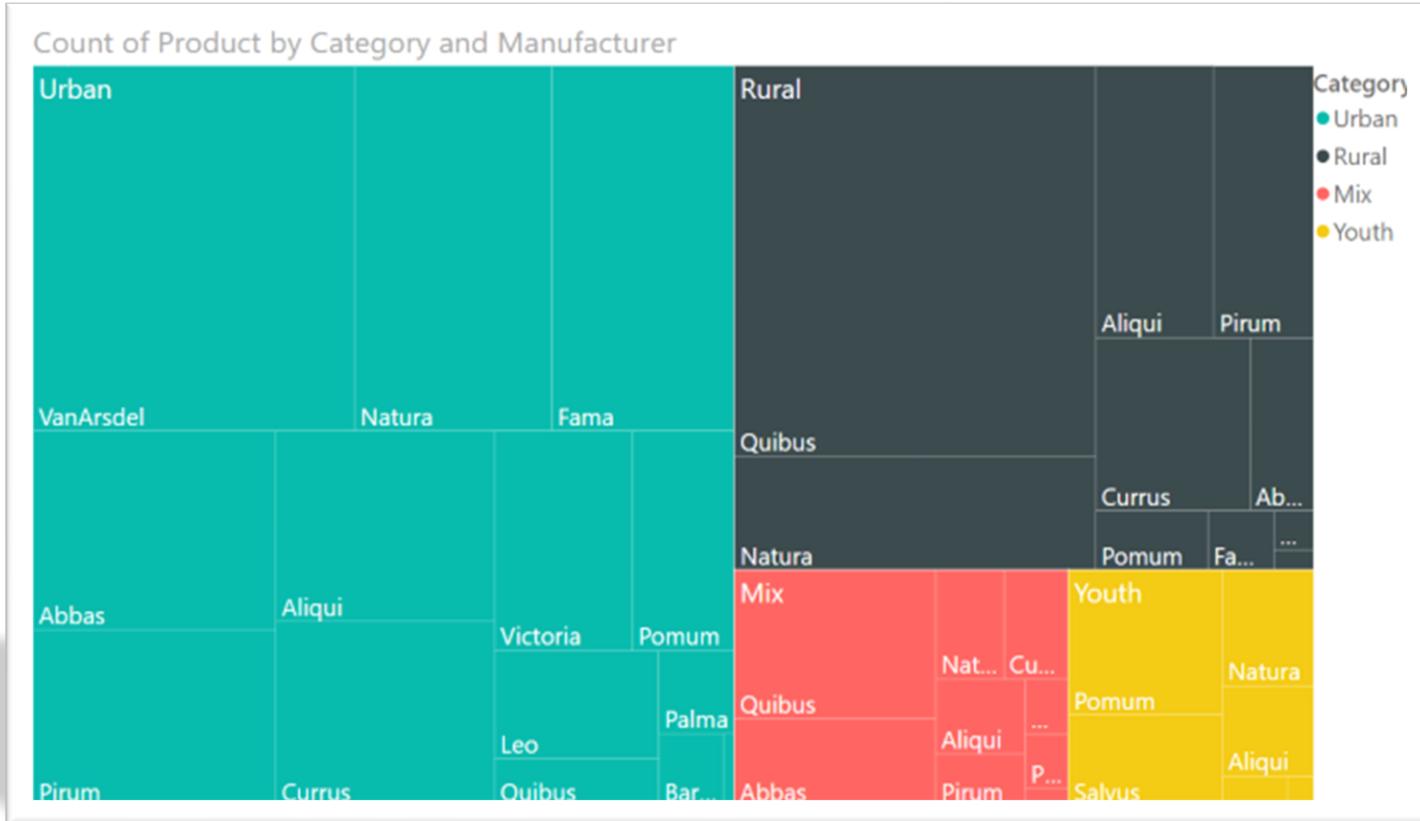
Matrix

Quarter	Q1		Q2	
Year	Revenue	YTD Revenue	Revenue	YTD Revenue
2015	\$45,186	\$45,186	\$70,609	\$115,795
2016	\$52,154	\$52,154	\$73,542	\$125,696
2017	\$51,388	\$51,388	\$68,149	\$118,537
2018	\$48,281	\$48,281	\$66,853	\$115,134
2019	\$53,145	\$53,145	\$49,135	\$102,280

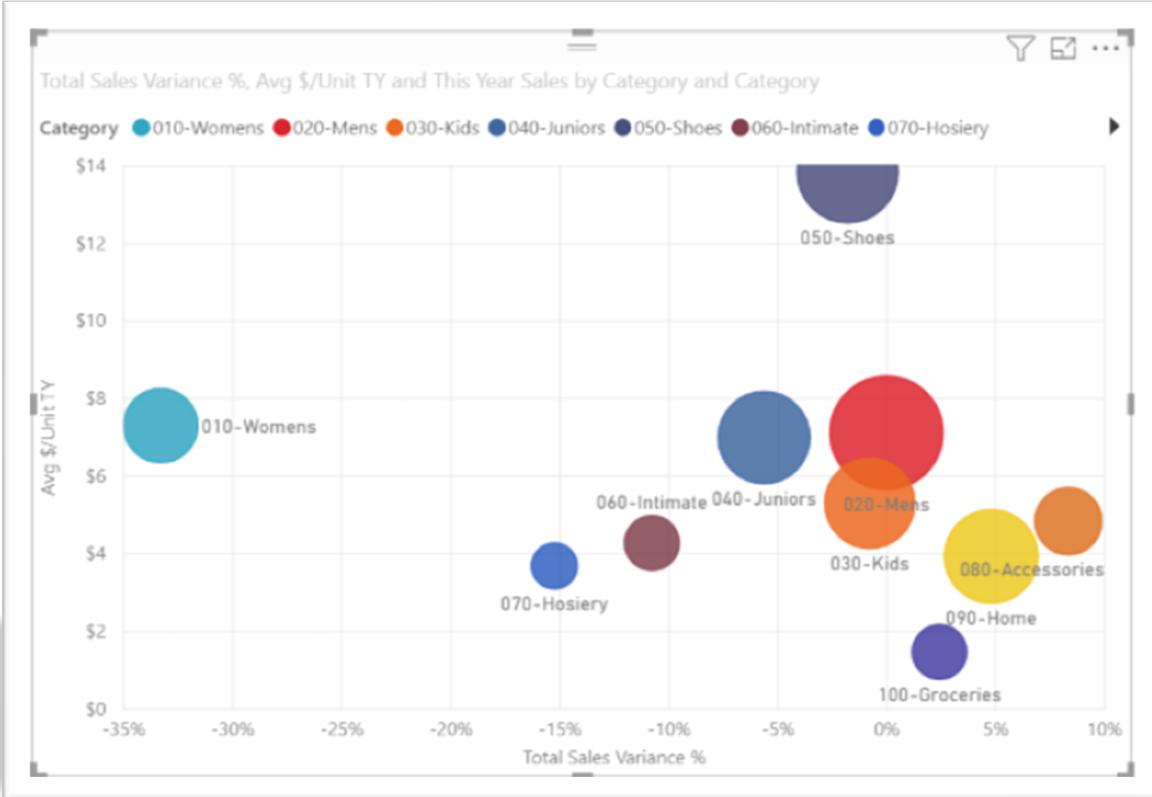
Key influencers



Treemap



Scatter



COMPARISON

To compare the magnitude of measures



CHANGE OVER TIME

To display the changing trend of measures



RANKING

To rank measures in an order



SPATIAL

To display measures over spatial maps



FLOW

To display a flow or dynamic relations



PART-TO-WHOLE

To identify the parts making up a measure total



DISTRIBUTION

To display the distribution of values



CORRELATION

To show correlations between measures



SINGLE

To present single values



FILTER

To control report filters



NARRATIVE

To tell a story with data





Data Analytics



Data Analytics

Data analytics is concerned with examining, transforming, and arranging data so that you can study it and **extract useful information.**

Data Analytics

-  Data analytics is a catch-all that covers a range of activities, each with its own focus and goals. You can categorize these activities as:
 -  Descriptive analytics
 -  Diagnostic analytics
 -  Predictive analytics
 -  Prescriptive analytics
 -  Cognitive analytics

Data Analytics

-  Data analytics is a catch-all that covers a range of activities, each with its own focus and goals. You can categorize these activities as:
 -  **Descriptive analytics:** [What is happening in your business?](#)
 -  **Diagnostic analytics:** [Why it is happening in your business?](#)
 -  **Predictive analytics:** [What likely to happen in the future based on previous trends and patterns?](#)
 -  **Prescriptive analytics:** [Helps you to determine the best course of action to eliminate future issues?](#)
 -  **Cognitive analytics:** [It combines a number of intelligent technologies like AI, ML, DL, etc. to apply human brain like intelligence to perform certain tasks.](#)

Data Analytics



Descriptive analytics

- What is happening in your business?
- What has happened, based on historical data?
- Summary of existing data record
- It gives us only insight about whether everything is going well or not in our business without explaining the root cause
- Metrics such as return on investment (ROI) are used in many industries.
- Examples: view of an organization's sales and financial data.
- Also called "Hindsight"

Data Analytics



Diagnostic analytics

- Why it is happening in your business? “Why things happened”
- Diagnostic Analytics explains the root cause behind the outcome of descriptive analytics,
- Supplement more basic descriptive analytics
- Three steps:
 - Identify anomalies in the data. These may be unexpected changes in a metric or a particular market.
 - Collect data that's related to these anomalies.
 - Use statistical techniques to discover relationships and trends that explain these anomalies.

Data Analytics



Predictive analytics

- What's likely to happen in the future based on previous trends and patterns?
- By utilizing various statistical and machine learning algorithms to provide recommendations and provide answers to questions related to what might happen in the future, that cannot be answered by BI
- Based on past trends
 - Projected sales reports
 - Azure Portal predicts your bill at the end of month
 - Weather prediction

Data Analytics



Prescriptive analytics

- Helps you to determine the best course of action to choose to bypass or eliminate future issues
- What actions should be taken to achieve a goal or target
- You can use Prescriptive analytics to advise users on possible outcomes and what should they do to maximize their key business metrics
- Advise on best approach for maximum success
 - Google Maps navigation
 - Recommendations - If you liked this movie, you might like that one
 - Search Engine Optimization tools

Data Analytics



Cognitive analytics

- It combines a number of intelligent technologies like artificial intelligence, machine-learning algorithms, deep learning etc. to apply human brain like intelligence to perform certain tasks
- Cognitive analytics helps you to learn what might happen if circumstances change, and how you might handle these situations.
- Analyzing data to come up with a “model” of how the world works
- Makes predictions based on that model
- Learn and improve over time
 - Reading Twitter to determine brand sentiment
 - Self-driving car

Data Analytics

-  Data analytics is a catch-all that covers a range of activities, each with its own focus and goals. You can categorize these activities as:
 -  **Descriptive analytics:** [What is happening in your business?](#)
 -  **Diagnostic analytics:** [Why it is happening in your business?](#)
 -  **Predictive analytics:** [What likely to happen in the future based on previous trends and patterns?](#)
 -  **Prescriptive analytics:** [Helps you to determine the best course of action to eliminate future issues?](#)
 -  **Cognitive analytics:** [It combines a number of intelligent technologies like AI, ML, DL, etc. to apply human brain like intelligence to perform certain tasks.](#)

COMPLETED

PENDING



COMPLETED



PENDING



Summary



Learning Outcome

- You have learnt about diff types of Data
- You have learnt about diff types of Data storage services
- You have learnt to choose right storage service for particular requirement
- We discussed about Azure Data Platform Architecture
- How to design partition distribution type

Describe Relational Data Structure



Characteristics of relational data

Customers

Customer ID	Customer Name	Customer Address
C1	Fred	...
C2	Bert	...
C3	Jane	...

Products

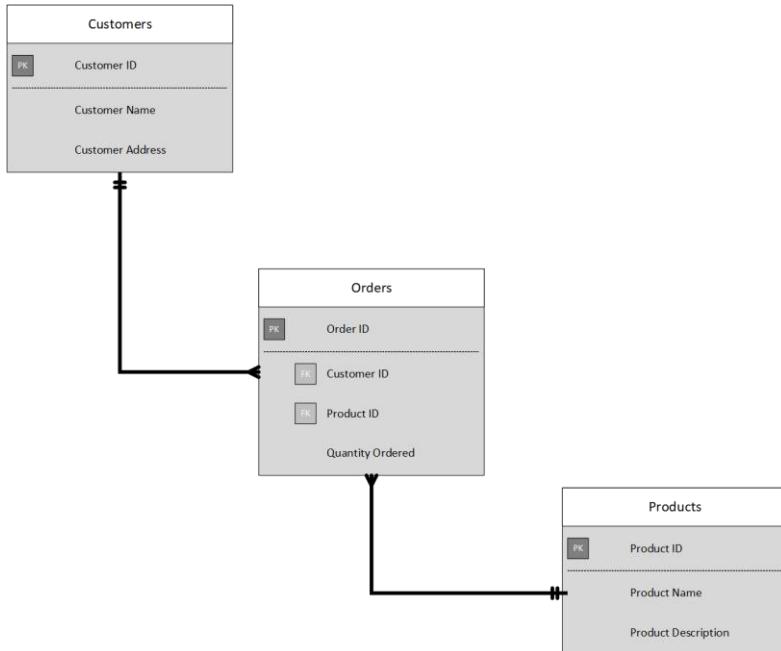
Product ID	Product Name	Description
P1	Shirt	...
P2	Tie	...
P3	Collar	...

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

- Table: Data is stored in a table
- Table consist of rows and columns
- Row: each row represents a single instance of an entity
- Column: define the properties of the entity
- Each column is defined by a datatype
- All rows have the same number of columns

Characteristics of relational data



- Some columns are used to maintain relationships between tables
- Model shows the structure of the entities
- Primary Key: uniquely identify each row
- Foreign Key: reference, or link to, the primary key of another table
 - used to maintain the relationships between tables.

What is an Index?

A

accordion, layouts
about 128
movie form, adding 131
nesting, in tab 128, 129
toolbar, adding 129-131
adapters, Ext
about 18
using 18, 20
Adobe AIR 285
Adobe Integrated Run time. *See* Adobe AIR
AJAX 12
Asynchronous JavaScript and XML.
See AJAX

B

built-in features, Ext
client-side sorting 86
column, reordering 86, 87
columns, hidden 86
columns, visible 86
button, toolbars
creating 63
handlers 67, 68
icon buttons 67
split button 64
buttons, form 53

C

cell renderers

lookup data stores, creating 83
two columns, combining 84
classes 254
ComboBox, form
about 47
database-driven 47-50
component config 59
config object
about 28, 29
new way 28, 29
old way 28
tips 26, 29
content, loading on menu item click 68, 69
custom class, creating 256-259
custom component, creating 264-266
custom events, creating 262-264

D

data, filtering
about 238
remote, filtering 238-244
data, finding
about 237
by field value 237
by record ID 238
by record index 237
data, formatting
about 278
date, formatting 279
other formatting 280, 281
string, formatting 278
data displaying `GridPanel`

What is an Index?

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

What is an Clustered Index?

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

Orders

Customer ID	Order ID	Product ID	Quantity
C1	1000	P1	1
C1	1002	P1	3
C1	1003	P3	1
C1	1005	P3	2
C2	1001	P2	4
C2	1004	P2	2
C3	1006	P3	1

Clustered Index

An index helps you search for data in a table.

What is an Non-Clustered Index?

Index

Orders

Customer ID	Order ID	Customer ID	Product ID	Quantity
C1	1000	C1	P1	1
C1	1001	C2	P1	3
C1	1002	C1	P3	1
C1	1003	C1	P3	2
C2	1004	C2	P2	4
C2	1005	C1	P2	2
C3	1006	C3	P3	1

What is an Index?

Index

Customer ID
C1
C1
C1
C1
C2
C2
C3

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

```
SELECT OrderID, ProductID  
FROM Orders  
WHERE CustomerID = "C1"
```



Index

Customer ID
C1
C1
C1
C1
C2
C2
C3

Orders

Order ID	Customer ID	Product ID	Quantity
1000	C1	P1	1
1001	C2	P1	3
1002	C1	P3	1
1003	C1	P3	2
1004	C2	P2	4
1005	C1	P2	2
1006	C3	P3	1

Non-Clustered Index - Good or Bad?

-  Non-Cluster Indexes consume additional storage space
-  Each time you insert, update, or delete data in a table, the indexes for that table must be maintained
-  Recommendations
 -  Read only table – more indexes will improve query performance
 -  Transaction table – more indexes on that table can slow your system down.
 -  You must strike a balance between having indexes that speed up your queries versus the cost of performing other operations.

What is a View?

-  View is a virtual table based on the result set of a query
 -  View does not store data but very much behave like a table
 -  You can query view like a table
-  A view can combine data from two or more table, using joins, and also just contain a subset of information. This makes them convenient to abstract, or hide complicated queries.



Azure Data Services



MySQL



My SQL

- Simple-to-use open-source database management
- Can be used for Linux, Apache, MySQL, and PHP (LAMP) stack apps
- Several editions; Community, Standard, and Enterprise.

Benefits of Azure Database for MySQL

- High availability
- Scalable
- Secure data, both at rest and in motion.
- Automatic backups and point-in-time restore for the last 35 days.
- Enterprise-level security and compliance with legislation.
- Azure Database for MySQL servers provides monitoring functionality to add alerts, and to view metrics and logs.

MariaDB



MariaDB

- New DBMS, created by the original developers of MySQL
- Compatibility with Oracle Database
- Optimized to improve performance
- built-in support for temporal data

Benefits of Azure Database for MariaDB

- Fully managed and controlled by Azure
- Built-in high availability with no additional cost.
- Predictable performance, using inclusive pay-as-you-go pricing.
- Scaling as needed within seconds.
- Secured protection of sensitive data at rest and in motion.
- Automatic backups and point-in-time-restore for up to 35 days.
- Enterprise-grade security and compliance.

PostgreSQL



PostgreSQL

- Hybrid relational-object database
- Enables you to store custom data types, with their own non-relational properties
- Code modules can be added
- Manipulate geometric data, such as lines, circles, and polygons
- `pgsql`

Benefits of Azure Database for PostgreSQL

- Provides the same availability, performance, scaling, security, and administrative benefits
- Some features of on-premises PostgreSQL databases are not available in Azure Database for PostgreSQL
- Continue to use pgAdmin tool

Migrate data to Azure

- Azure Database Migration Service (DMS)
 - Enables you to restore a backup of your on-premises databases directly to databases running in Azure
- Data Services
 - Replication from an on-premises database



Provisioning and Deployment

What is provisioning and Deployment

■ **Provisioning and deployment** means to execute series of steps to create and configure a service

- We need to provide parameters that provide estimate of size of workload we want to run
- Behind the scene Azure will create other required resources: Disks, memory, CPUs, network and so on
- You will be charged for these resources until you delete them
- We can Scale dynamically

Methods of Provisioning and Deployment

-  **The Azure portal**
 -  Convenient but manual
-  **The Azure command-line interface (CLI)**
 -  Set of commands to create and manage Azure resources specifically
 -  Can run from the operating system command prompt or the Cloud Shell in the Azure portal.
 -  Suitable if you need to automate service creation
-  **Azure PowerShell**
 -  This is a cross-platform task automation and configuration management framework.
 -  Command-line shell and scripting language that is built on top of .Net
 -  Azure provides a series of commandlets (Azure-specific commands) that you can use in PowerShell to create and manage Azure resources.
-  **Azure Resource Manager templates**
 -  JSON (JavaScript Object Notation) file template that describes the service, and can be used to create resources.

PowerShell example

```
$SubscriptionId = 'XXXXXXXXXXXXXXXXXXXXXX'  
$resourceGroupName = "rg-name"  
$location = "West US"  
$adminSqlLogin = "admin-userid"  
$password = "pass"  
$databaseName = "db-name"  
$serverName="svrname"
```

```
Set-AzContext -SubscriptionId $subscriptionId
```

```
$server = [New-AzSqlServer] -ResourceGroupName $resourceGroupName -ServerName $serverName -Location $location -  
SqlAdministratorCredentials $(New-Object -TypeName System.Management.Automation.PSCredential -ArgumentList $adminSqlLogin,  
$(ConvertTo-SecureString -String $password -AsPlainText -Force))
```

```
$database = [New-AzSqlDatabase] -ResourceGroupName $resourceGroupName -ServerName $serverName -DatabaseName  
$databaseName -RequestedServiceObjectiveName "S0"
```

Why SQL Server in Azure?



Fully Managed



Predictable
performance
and pricing



Elastic pool for
unpredictable
workloads



99.99%
availability
built-in



Geo-replication
and restore
services



Supports existing SQL
Server tools, libraries,
and APIs



Scalability with no
downtime



Secure and compliant
for your sensitive data

Azure IaaS vs PaaS Database offerings?



SQL Server on Azure VMs
SQL Server inside a
fully-managed VM in Azure



Azure SQL Database
Database-as-a-service (DBaaS)
hosted in Azure

Responsibility comparison



SQL Server on Azure VMs
SQL Server inside a
fully-managed VM in Azure



Azure SQL Database
Database-as-a-service (DBaaS)
hosted in Azure

Benefits comparison



SQL Server on Azure VMs
SQL Server inside a
fully-managed VM in Azure



Azure SQL Database
Database-as-a-service (DBaaS)
hosted in Azure

Limitations comparison

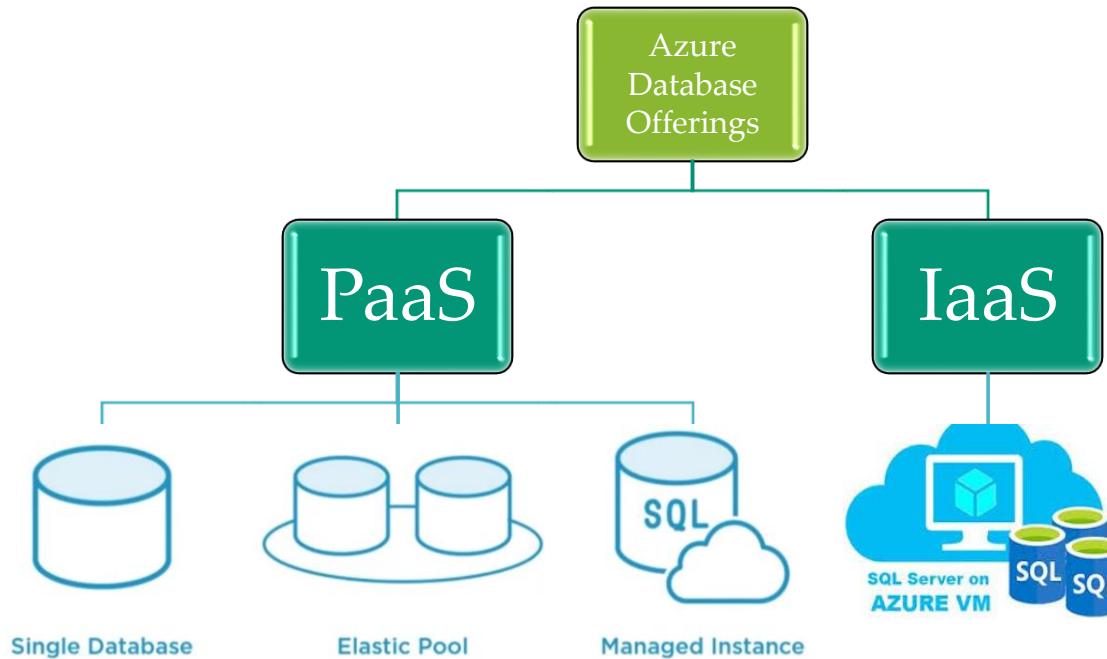


SQL Server on Azure VMs
SQL Server inside a
fully-managed VM in Azure



Azure SQL Database
Database-as-a-service (DBaaS)
hosted in Azure

Azure Database Deployment options



SQL Server(PaaS) Deployment Options



Single database

Each DB with its own guaranteed compute, memory, and storage



Elastic pool

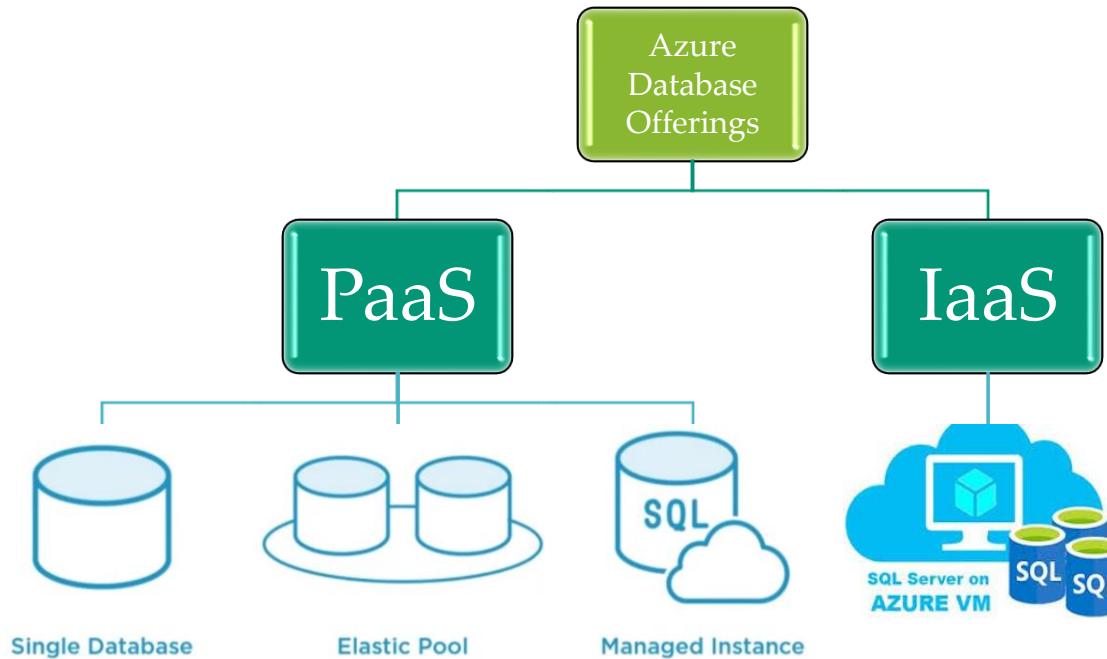
Fixed resources will be shared by all databases in the pool



Managed instance

Each managed instance has its guaranteed resources

Azure Database Deployment options



SQL Server on Virtual machine



Full Version of SQL
Server in Cloud



Geographic
regions



Variety of
Configuration to
Choose

SQL Server on Virtual machine



Automatic
Updates



Automatic
Updates



High
Availability



Automatic
Updates

Security



Network security

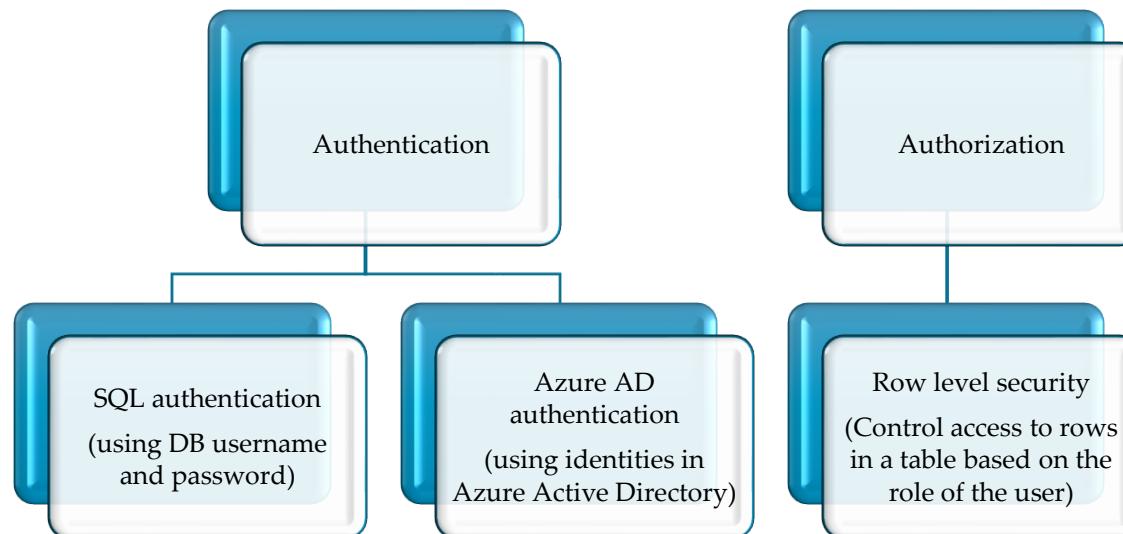
IP firewall rules

- Grant access to databases based on the originating IP address of each request.

Virtual network firewall rules

- Enable Azure SQL Database to only accept requests originating from subnets inside a virtual network.

Access Management



Threat Protection

Monitoring
in Azure
Monitor logs
and Event
Hub

Advanced
Threat
Protection

- Tracks database activities and helps maintaining compliance with security standards

- Analyzes your SQL Server logs to detect unusual behavior and potentially harmful attempts

Information Protection

Transport Layer Security TLS

- Always enforces encryption for all connections

Transparent Data Encryption

- (Protects data at rest from offline access to raw files or backups)

Dynamic Data masking

- (Protects sensitive data by masking it for non-privileged users)

Information Protection

Transport Layer Security TLS

- Always enforces encryption for all connections

Transparent Data Encryption

- (Protects data at rest from offline access to raw files or backups)

Dynamic Data masking

- (Protects sensitive data by masking it for non-privileged users)

Security Management

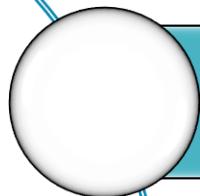
Vulnerability assessment

- Discover track and remediate potential database vulnerabilities.

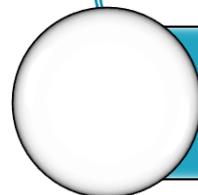
Data discovery & Classification

- Identify and label sensitive data for monitoring and alerting

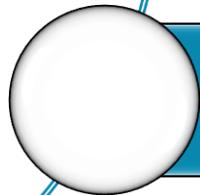
Managed Instance Advance Security



Native virtual network implementation and connectivity to your on-premises environment using Azure Express Route or VPN Gateway.

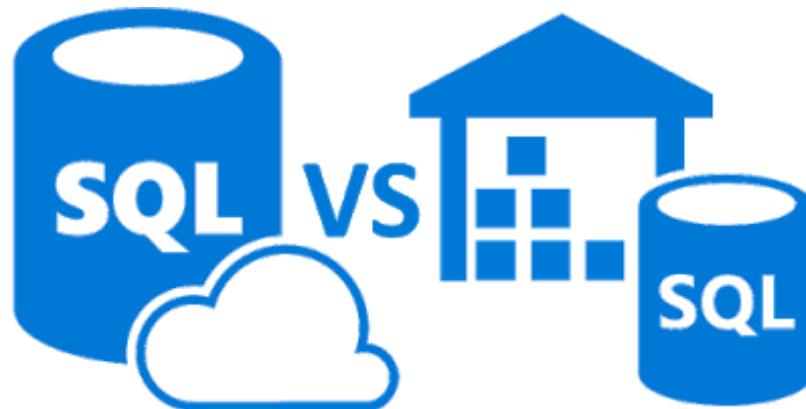


In a default deployment, SQL endpoint is exposed only through a private IP address, allowing safe connectivity from private Azure or hybrid networks.



Single-tenant with dedicated underlying infrastructure (compute, storage).

Azure SQL DB vs. Azure SQL DW



Azure
SQL Database

Azure
SQL Data Warehouse

Identify Query Tools

Identify Query Tools

-  Azure Portal
-  SQL Server Management Studio (SSMS)
-  Azure Data Studio
-  Sqlcmd utility
-  Visual Studio Code

Azure Portal

- Query editor built in to the portal
- Execute queries against your database in Azure SQL Database or data warehouse in Azure Synapse Analytics.
- Doesn't support connecting to the master database
- 5-minute timeout for query execution
- No support for IntelliSense for database tables and views

SQL Server Management Studio (SSMS)

- Integrated environment for managing any SQL infrastructure
- Use SSMS to access, configure, manage, administer, and develop all components of SQL Server, Azure SQL Database, and Azure Synapse Analytics
- Support database administration tasks
 - User management, vulnerability assessment, security features
 - Performance tuning advisors
 - Import and Export of DACPAC files

Azure Data Studio

- Cross-platform database tool
- Can be used for both on-premises and cloud data platforms
- Works with Windows, Mac OS and Linux
- Intellisense, code snippets, source control integration and integrated terminal
- Charting of query results, customizable dashboards
- Open source, free

Azure Data Studio vs SSMS

Use Azure Data Studio if you:

- Are mostly editing or executing queries.
- Need the ability to quickly chart and visualize result sets.
- Can execute most administrative tasks via the integrated terminal using sqlcmd or PowerShell.
- Have minimal need for wizard experiences.
- Do not need to do deep administrative or platform related configuration.
- Need to run on macOS or Linux

Use SQL Server Management Studio if you:

- Are doing complex administrative or platform configuration.
- Are doing security management, including user management, vulnerability assessment, and configuration of security features.
- Need to make use of performance tuning advisors and dashboards.
- Use database diagrams and table designers.
- Need access to Registered Servers.
- Make use of live query stats or client statistics.

sqlcmd

- Command line utility
- Execute T-SQL statements, stored procs, and script files
- Uses ODBC to execute Transact-SQL batches
- Windows and Linux options (RHEL, Ubuntu, SUSE)
- MacOS in preview
- Can run in a docker container

Introduction to SQL



SQL - Structured Query Language



Structured Query Language (SQL)

- Used to communicate with a relational database.
- First developed as SEQUEL in the early 1970s
- SQL Standard first formalized in 1986 as SQL-86, now up to SQL:2019

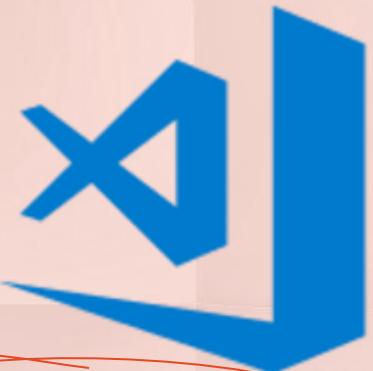


Database Providers

- Each database engine has its own version of SQL that extends the standard for its own purpose, usually to add programming elements (stored procs!)
- Most vendor version of SQL are not 100% compatible with the standard
- SQL Server - Transact-SQL (T-SQL)
- Oracle - PL/SQL [Procedural Language/SQL]
- MySQL - SQL/PSM
- PostgreSQL - PL/pgSQL
- They are not often compatible - an Oracle PL/SQL command would not work unaltered on SQL Server

DDL vs DML vs DCL

- Data Definition Language (DDL)
 - CREATE, ALTER, DROP, TRUNCATE, RENAME
- Data Manipulation Language (DML)
 - DML statements to manipulate the rows in a relational table
 - SELECT, INSERT, UPDATE, DELETE
- Data Control Language (DCL)
 - GRANT and REVOKE



Visual Studio Code

Visual Studio Code

-  Visual Studio environment
-  Thousands of extensions
-  Debug
-  Language support
-  Source Control
-  Cross platform
-  Simple to use
-  FREE

Azure Database for MariaDB



Azure Database for MariaDB

- When MySQL was acquired by Oracle in 2009, a copy of the MySQL code was made and developed further as MariaDB.
- Optimized to improve performance
- Built-in support for temporal data
- FREE and Open Source
- “Azure Database for MariaDB” is PaaS for MariaDB database engine
- Azure manage Infrastructure, and provide all PaaS features like high availability, dynamic scalability and automated backups

Azure Database for PostgreSQL



Azure Database for PostgreSQL

- Hybrid relational-object database
- Enables you to store custom data types, with their own non-relational properties
- Manipulate geometric data, such as lines, circles, and polygons
- Language: pgsql
- “Azure Database for PostgreSQL” is PaaS for PostgreSQL database engine
- Provides the same availability, performance, scaling, security, and administrative benefits
- High availability – Built in failure detection and failover mechanisms
- Azure.sys.qs_view - record information about the queries run against databases on the server
- Some features of on-premises PostgreSQL databases are not available in Azure Database for PostgreSQL
- Continue to use pgAdmin tool
- Use psql to query a database
 - Azure Cloud Shell
 - Desktop computer - download and install the psql client from postgresql.org website.
- Azure Data Studio



Azure Database for MySQL



Azure Database for MySQL

- Simple-to-use open-source database management
- Can be used for Linux, Apache, and PHP (LAMP) stack apps
- Several editions; Community, Standard, and Enterprise
- “Azure Database for MySQL” is PaaS for MySQL database engine
- Provides the same availability, performance, scaling, security, and administrative benefits
- Secure data, both at rest and in motion.
- Automatic backups and point-in-time restore for the last 35 days.
- Enterprise-level security and compliance with legislation.
- Azure Database for MySQL servers provides monitoring functionality to add alerts, and to view metrics and logs.
- MySQL Workbench to query a database

Identify the right data offering



Identify the right data offering



Azure SQL Server

- Needs to perform high-speed online transaction processing (OLTP) operations
- Highly normalized with enforced schemas
- Requires high integrity and strong consistency
- Relationships are maintained between data tables
- Example - Inventory management and Finance and accounting system



Azure Synapse Analytics (Azure Data Warehouse)

- Enterprise data analytics system workload
- Provides a set of tools to organize and analyze enterprise data
- MPP Architecture



Azure Cosmos DB

- Mainly targeted for non-relational data stores



Azure Blob Storage / Azure Data Lake

- Store binary data – Images, videos

Identify the right data offering for a relational workload



Azure SQL Server – Single Database

- Up to date relational database service
- Not fully compatible with on-prem



Azure SQL Server – Managed instance

- Most of PaaS benefits + high compatibility
- Lift and Shift your on-prem databases



Azure SQL Server – Virtual Machine

- Install specific version of SQL Server on virtual machine
- Allow operating system level access

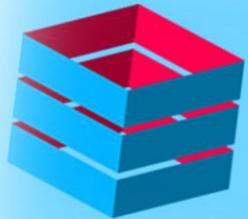


PostgreSQL, MySQL, MariaDB

- Deploy existing on-prem database to Azure

Why NoSQL DB?

What traditional databases were lacking?



An Introduction to
NoSQL

RDBMS were lacking



Scalability

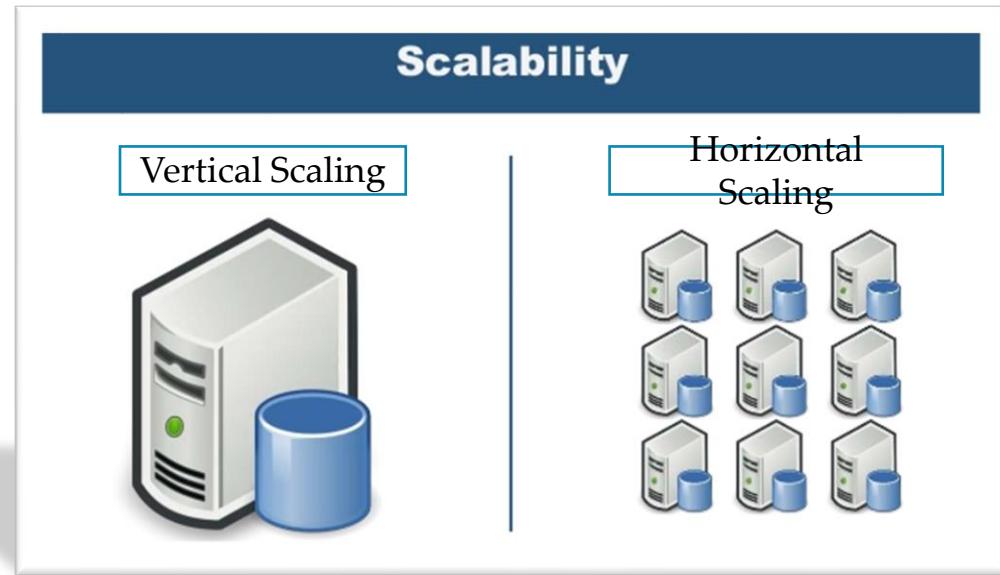


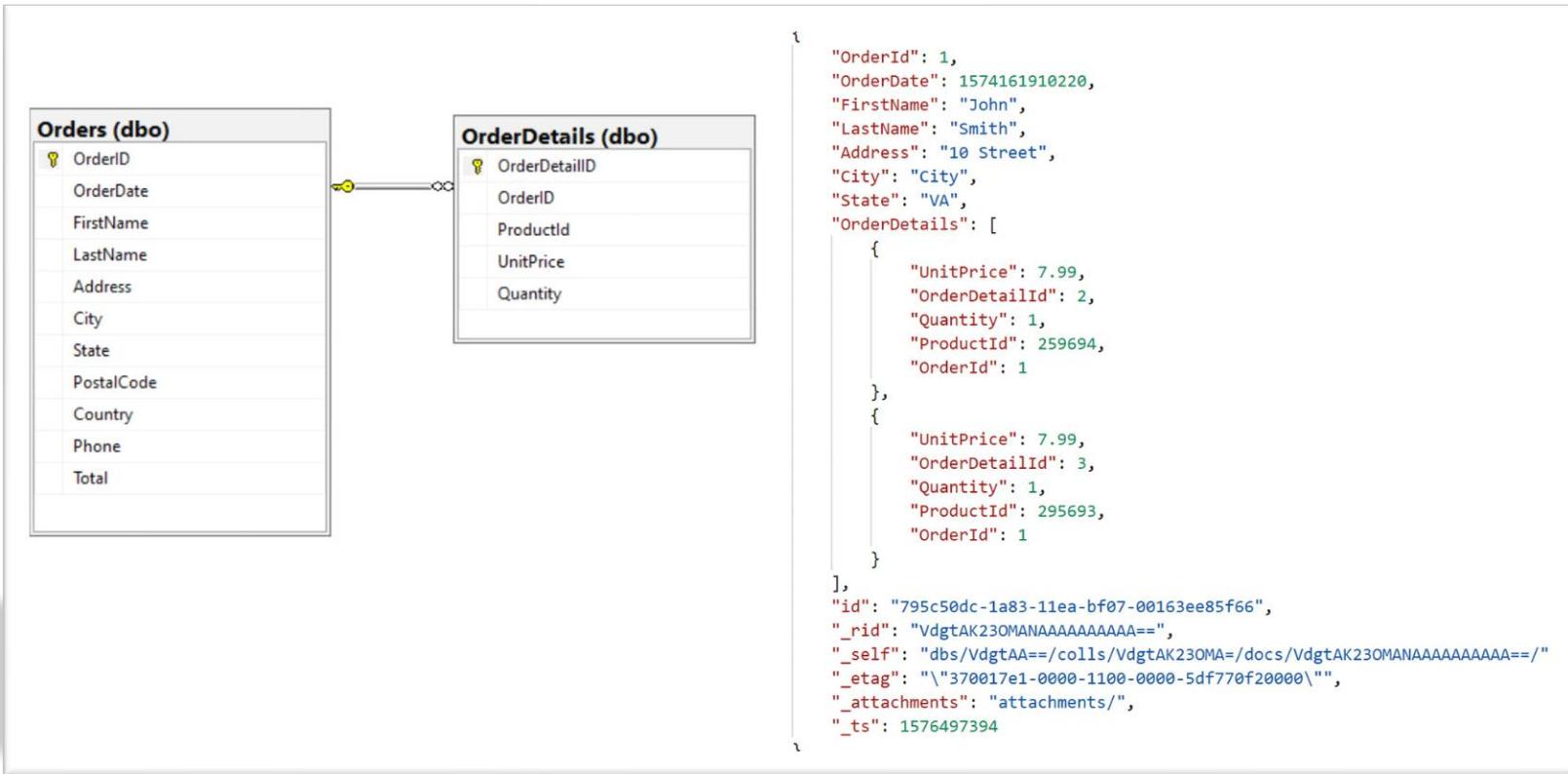
Flexibility

What is NoSQL

- NoSQL {
 - “Not Only SQL” or no structure query language
- Scalability {
 - Horizontal scaling possible
- Flexibility {
 - No Schema enforced

- **Vertical scaling**
 - Add more CPU, RAM, HDD in same system
- **Horizontal Scaling**
 - Add more commodity machines in system





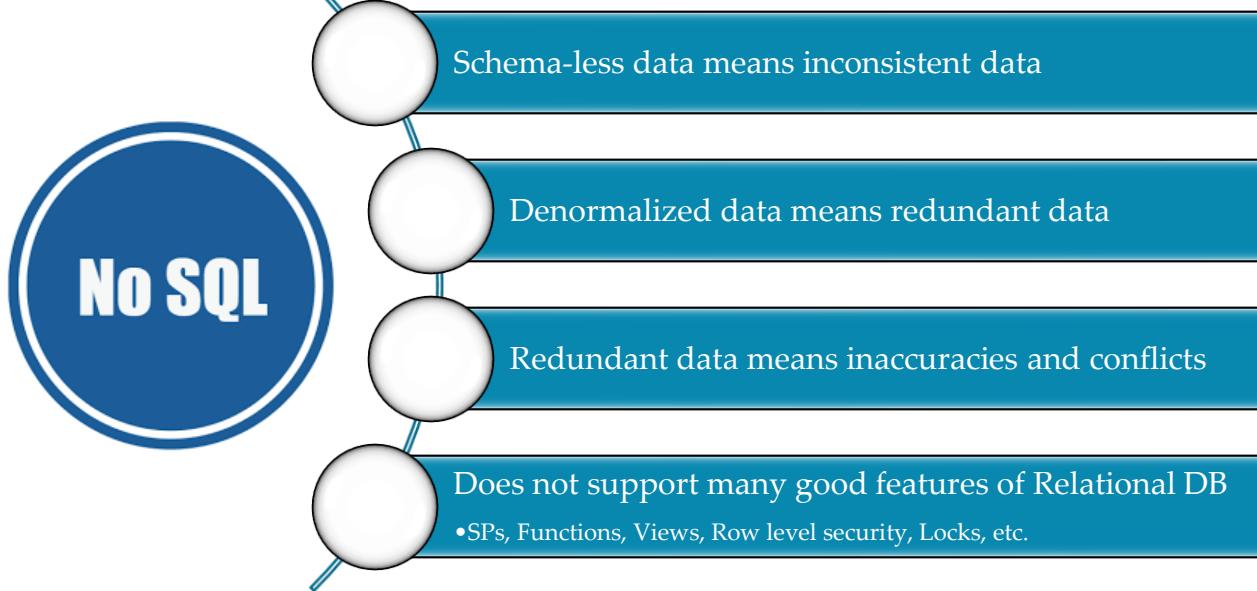
```
{  
    "orderid": 12212,  
    "orderdate": "12/4/2020",  
    "customer":  
        { "name": "Bob Smith", "email": "bobsmith@email.bob" },  
    "status": "in process",  
    "paymentmethod": "invoice",  
    "products": [  
        { "name": "Product 1", "quantity": 1 },  
        { "name": "Product 2", "quantity": 1, status: 3 }  
    ]  
}
```

NoSQL Use Cases



- Big data and real-time web applications.
- Relationship b/w data is not important
- Data change frequently

NoSQL Limitations

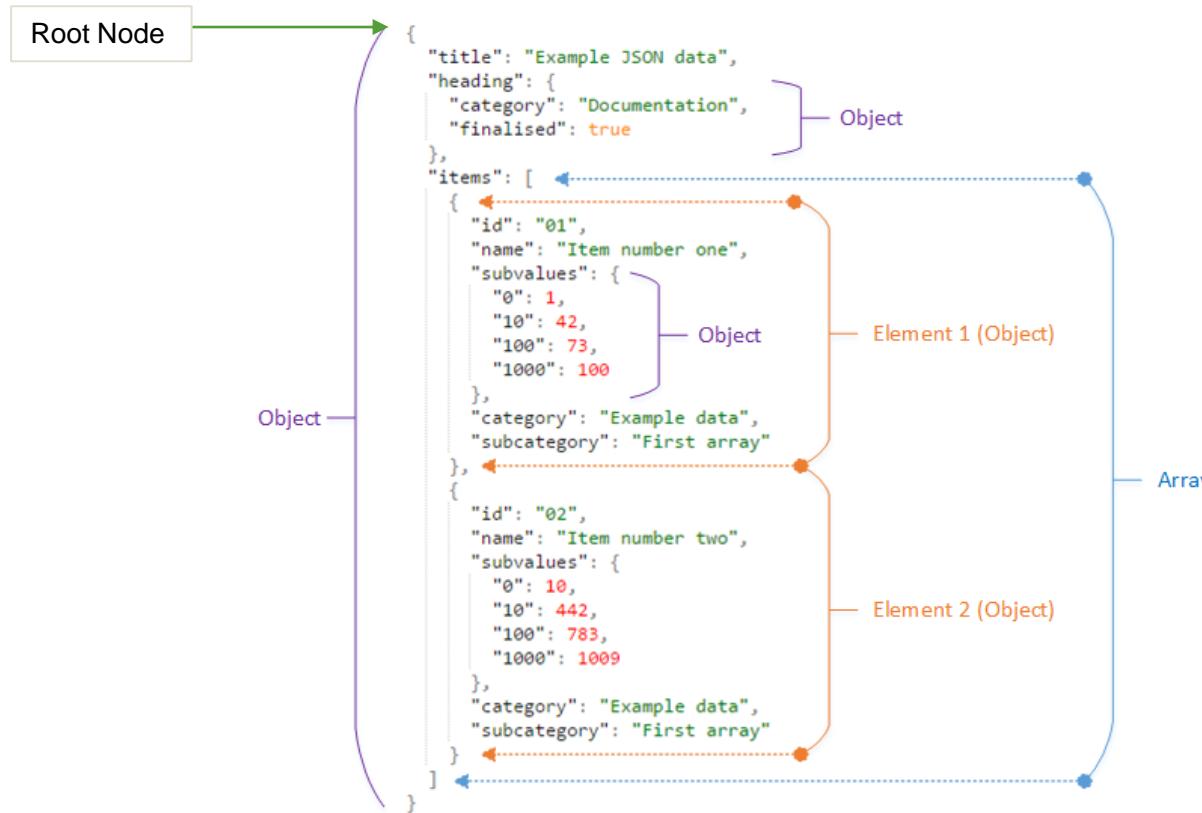


JSON file

- JSON — short for JavaScript Object Notation
- format for sharing data like XML
- JSON offers a good alternative to XML - Very readable and lightweight
- A JSON object is a key-value data format that is typically rendered in curly braces
- At the granular level, values can be one of 6 simple data types:
 - strings
 - numbers
 - Booleans (true or false)
 - null
 - objects
 - arrays

```
{  
  "first_name" : "Sammy",  
  "last_name" : "Shark",  
  "location" : "Ocean",  
  "online" : true,  
  "followers" : 987  
}
```

JSON Structure



SQL vs NoSQL

SQL

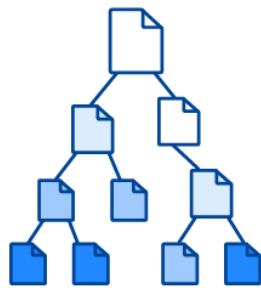
- Relational database
- Fixed schema
- Designed for complex queries
- SQL, MySql, Oracle, Postgres
- Vertical scaling
- Row Oriented
- Tables
- Limited for big data

NoSQL

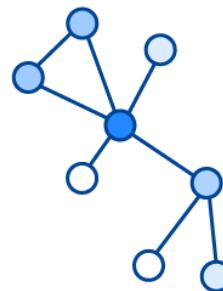
- Non-relational or distributed
- Dynamic
- Not for complex queries
- MongoDB, Redis, Hbase
- Horizontal scaling
- Multi-model oriented
- Collections
- Great for big data

4 Types of NoSQL Databases

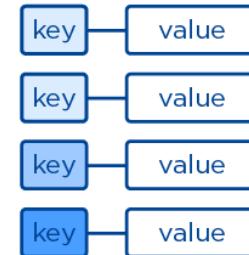
Document



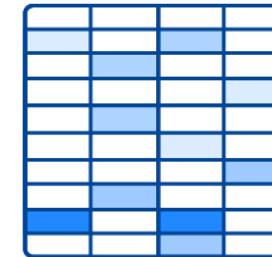
Graph



Key-Value

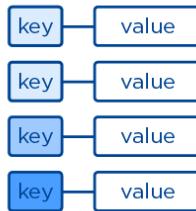


Wide-column



Key-value store

Key-Value



- Uses a simple key/value to store data
- Quick to query due to its simplicity
- Value can be JSON, BLOB, String etc.

- **Use Cases:**

- User profiles and session info on a website, blog comments, telecom directories, IP forwarding tables, shopping cart contents on e-commerce sites, and more.

Phone Directory

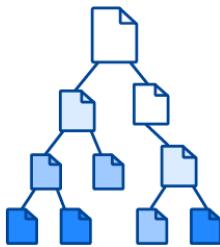
Key	Value
Bob	(123) 456-7890
Jane	(234) 567-8901
Tara	(345) 678-9012
Tiara	(456) 789-0123

- **Examples**

- Cosmos DB Table API, Redis, Table Storage, Oracle NoSQL Database, Voldemort, Aerospike, Oracle Berkeley DB

Document store

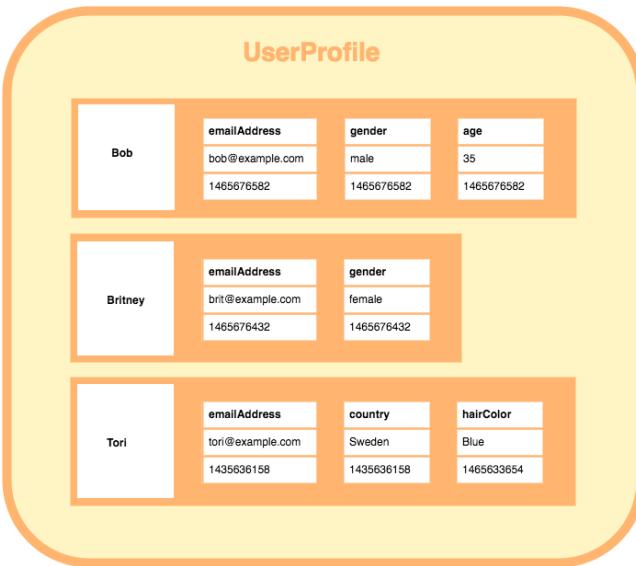
Document



- Document-oriented model to store data
- Similar to key/value store, difference is that, the value in a document store database consists of semi-structured data.
- Each record and its associated data within a single document.
- Document stores are usually XML, JSON, BSON, YAML, etc.
- **Use Cases:**
 - Content management systems, blogging platforms, and other web applications, blog comments, chat sessions, tweets, ratings, etc.
- **Examples**
 - Cosmos DB, MongoDB, DocumentDB, CouchDB, MarkLogic, OrientDB

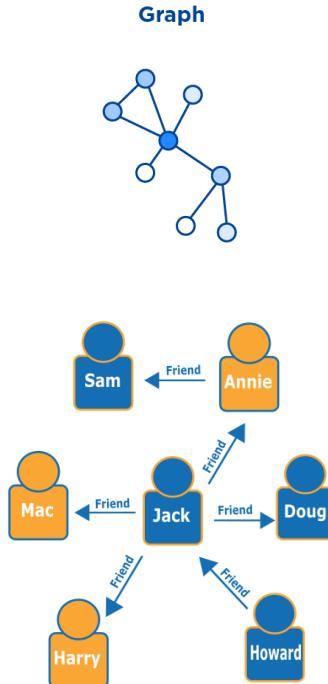
```
{  
  "orderid": 12212,  
  "orderdate": "12/4/2020",  
  "customer":  
    { "name": "Bob Smith", "email": "bobsmith@email.bob" },  
  "status": "in process",  
  "paymentmethod": "invoice",  
  "products": [  
    { "name": "Product 1", "quantity": 1 },  
    { "name": "Product 2", "quantity": 1, status: 3 }  
  ]  
}
```

Column store



- Stores data using a column oriented model
- Columns in each row are contained within that row
- Each row can have different columns to the other rows.
- Extremely quick to load and query
- **Use Cases:**
 - Sensor Logs [Internet of Things (IOT)], User preferences, Geographic information, Reporting systems, Time Series Data, Logging and other write heavy applications
- **Examples**
 - Cosmos DB, Bigtable, Cassandra, Hbase, Vertica, Druid, Accumulo, Hypertable

Graph store

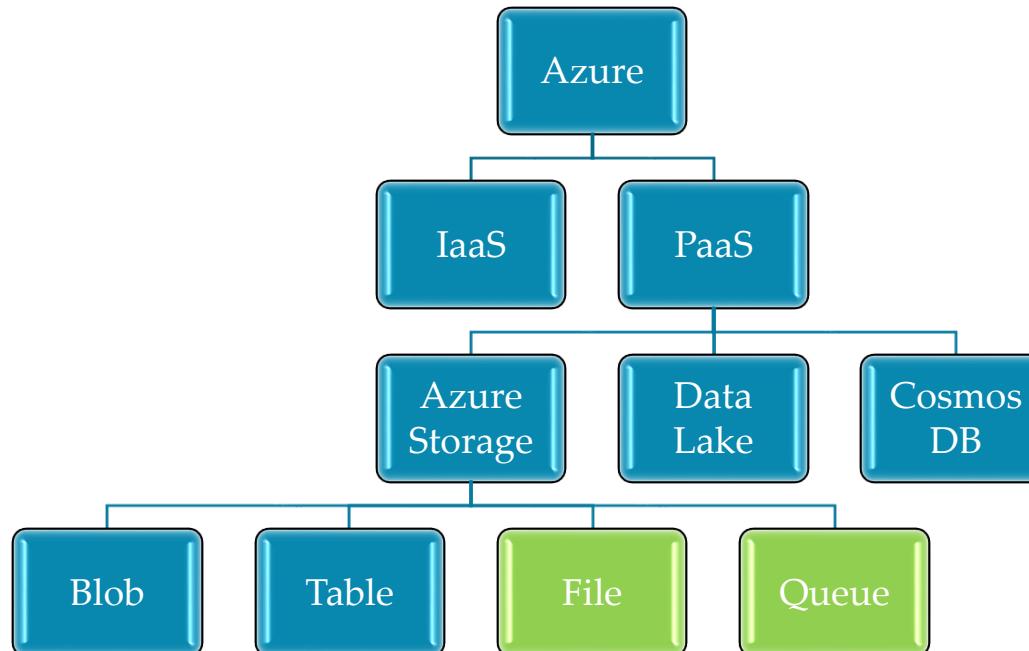


- Focuses on how data relates to other data points.
- A **node** is a specific entity or piece of information
- **Edge** simply specifies the relationship between two nodes.
- **Use Cases:**
 - Social networks, realtime product recommendations, network diagrams, fraud detection, access management, and more.
- **Examples**
 - Cosmos DB Gremlin API, Neo4j, Blazegraph, and OrientDB.

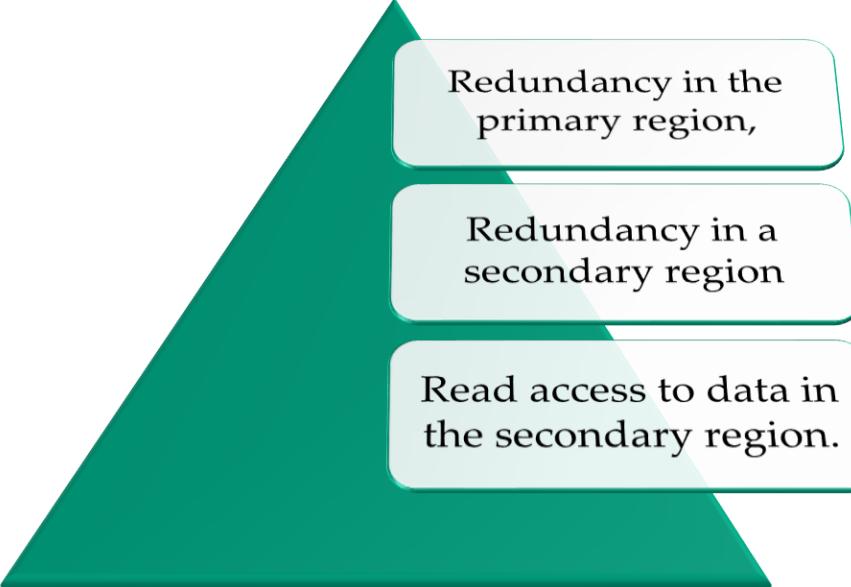
Multi-model

- Include features/characteristics of more than one data model.
- **Example:**
 - **OrientDB:** OrientDB combines a graph model with a document model.
 - **ArangoDB:** Uses key/value, document, and graph models.
 - **Virtuoso:** Combines relational, graph, and document models.

NoSQL Offerings by Microsoft Azure



Three categories of replication options

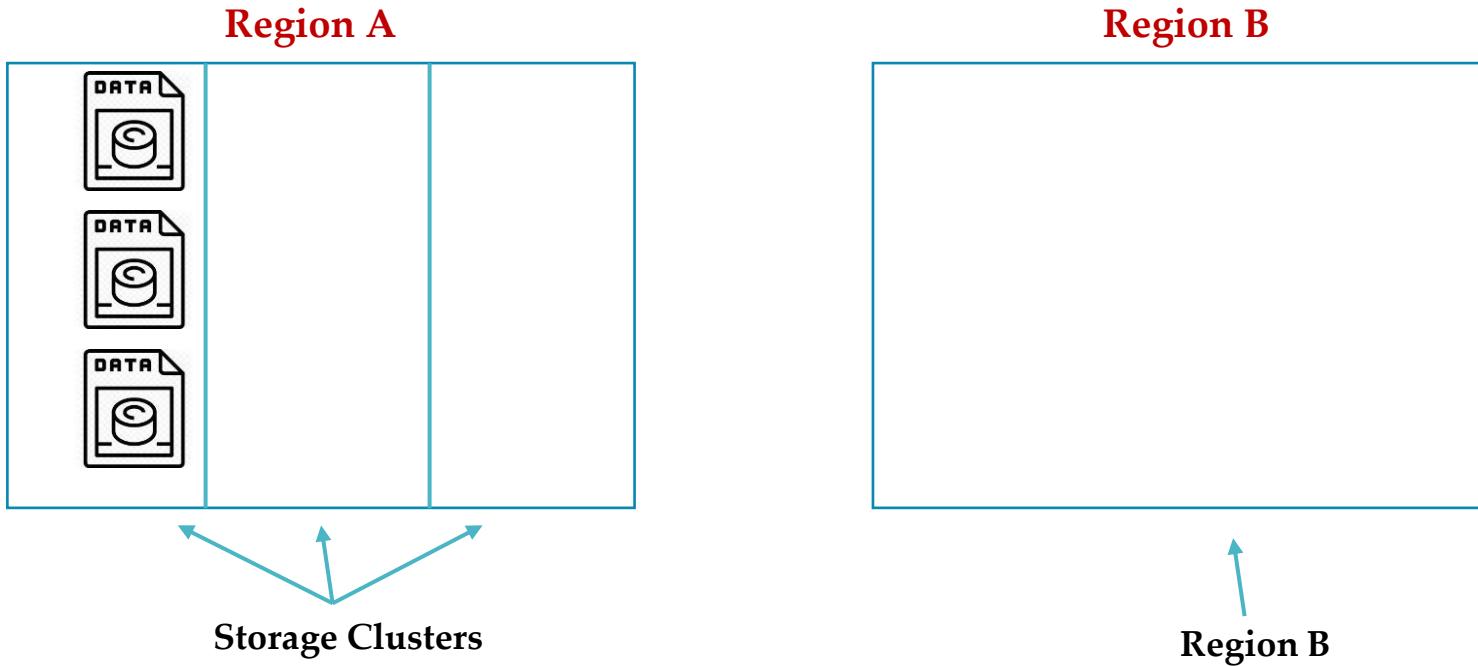


Redundancy in the primary region,

Redundancy in a secondary region

Read access to data in the secondary region.

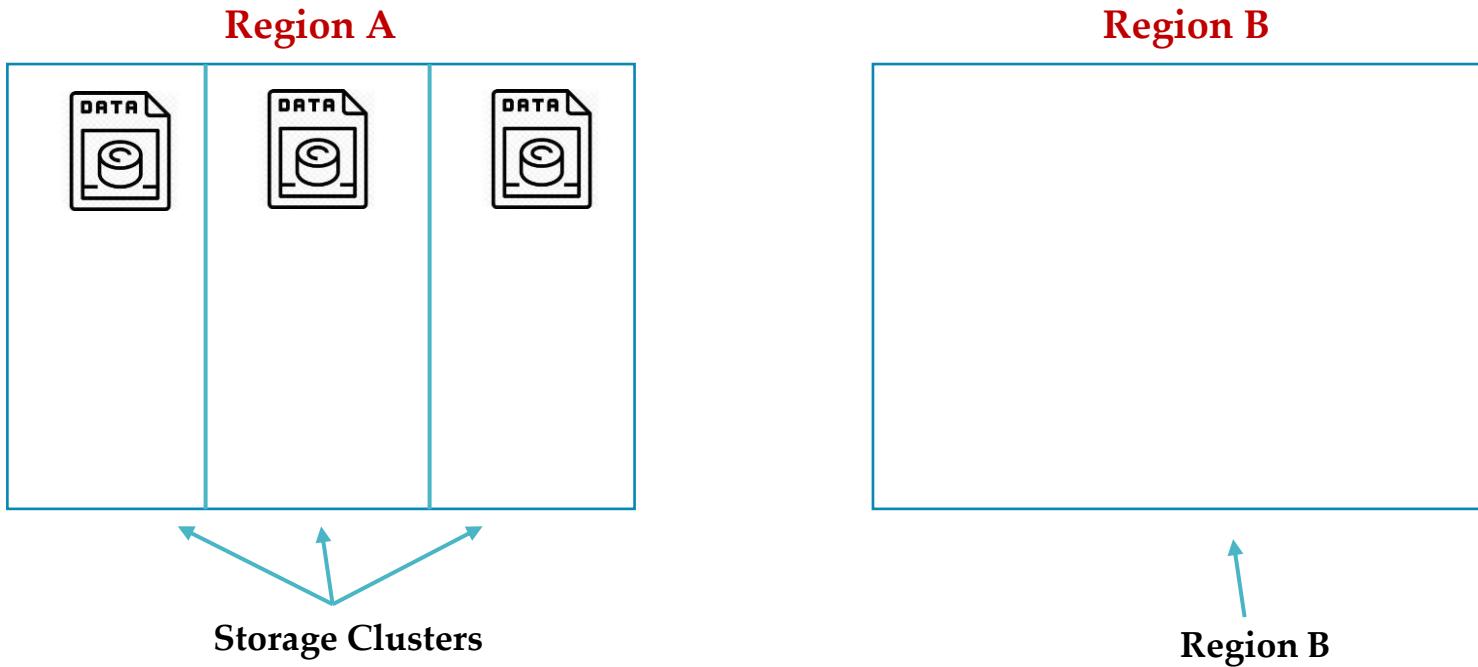
Locally Redundant Storage (LRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

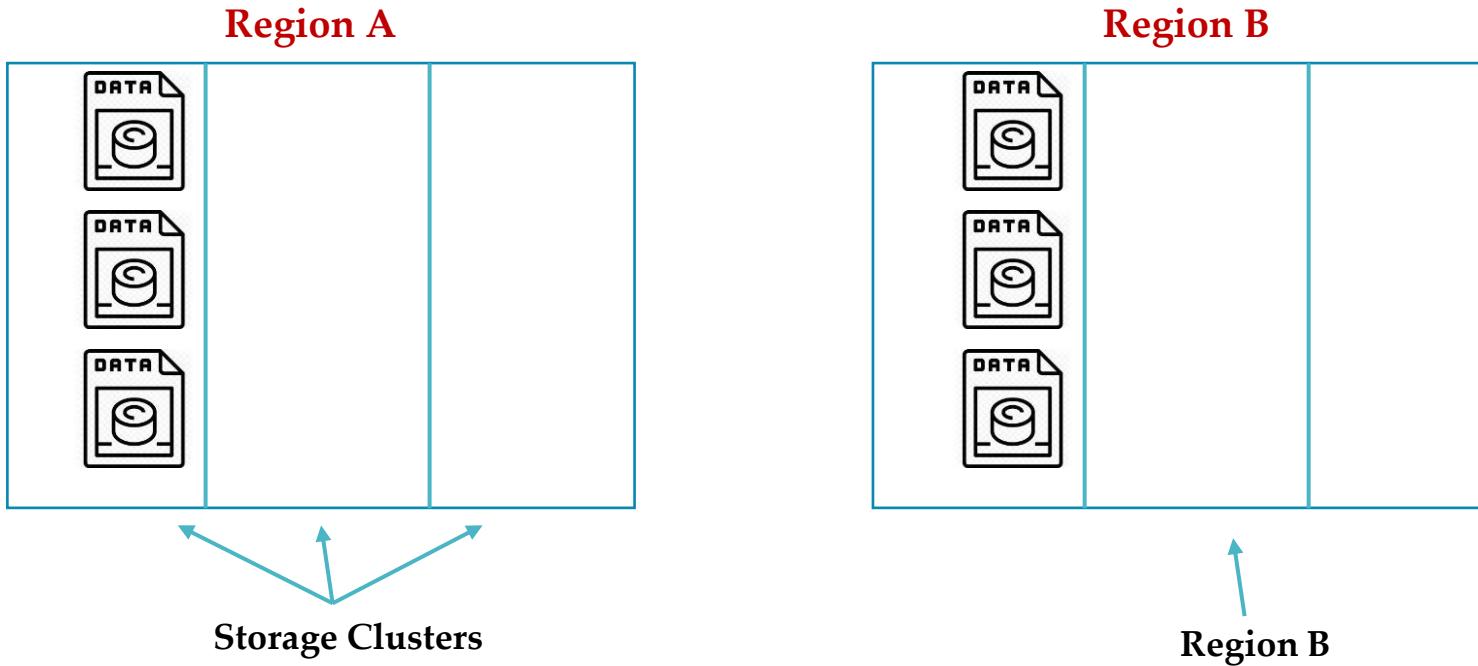
Zone Redundant Storage (ZRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

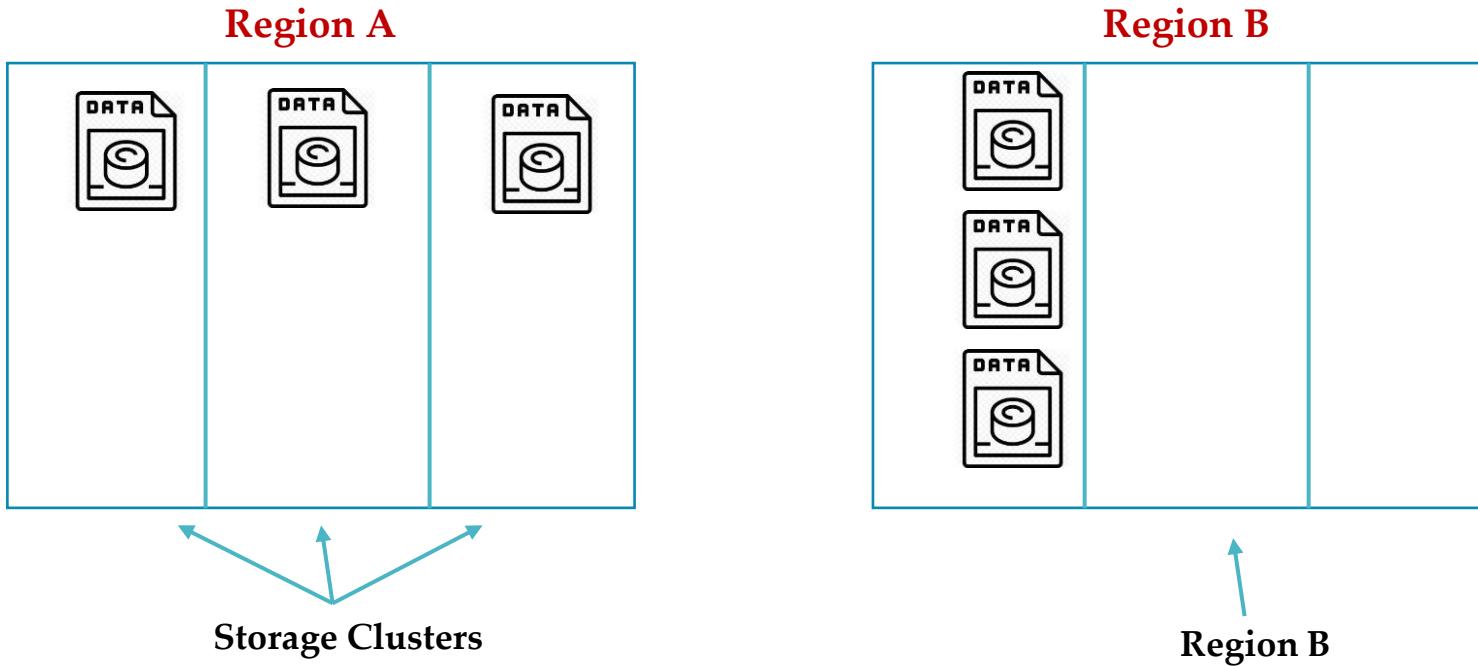
Geo Redundant Storage (GRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

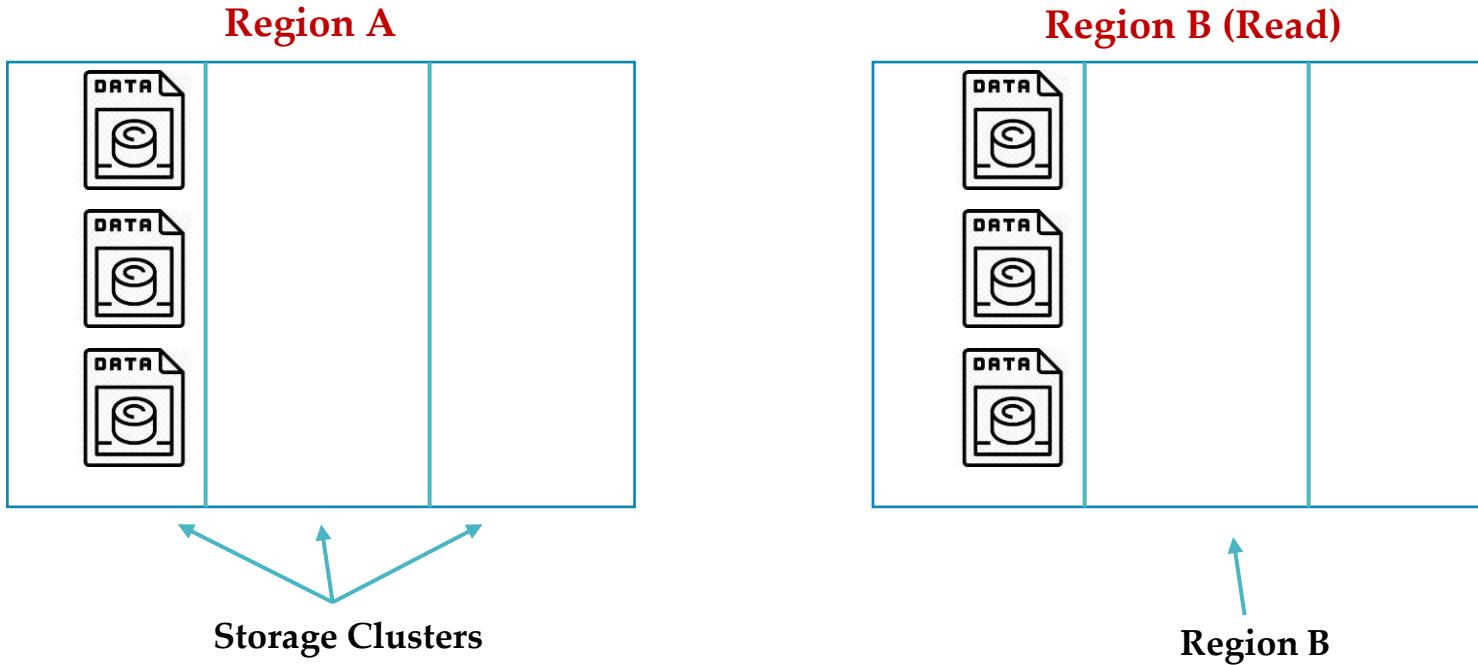
Geo Zone Redundant Storage (GZRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

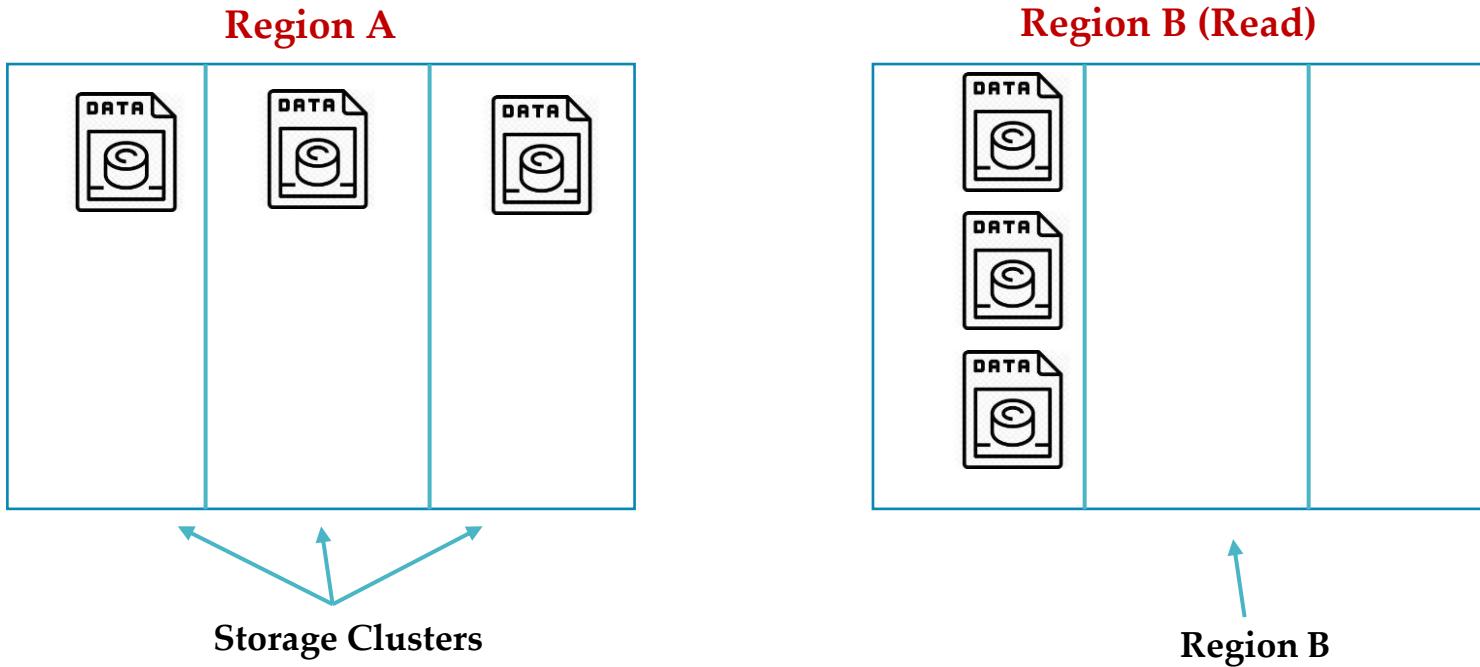
Read access geo Redundant Storage (RA-GRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

Read access Geo Zone Redundant Storage (RA-GZRS)



Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.

Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

Azure Table Storage



Key	Value (fields)		
AA	Data for AA
BB	Data for BB
CC	Data for CC
...			
ZZ	Data for ZZ

Azure Table Storage

- NoSQL key-value Storage
- Items are referred to as rows, and fields are known as columns
- All rows in a table must have a key
- No concept of relationships, stored procedures, secondary indexes, or foreign keys
- Data will usually be denormalized
- To help ensure fast access, Azure Table Storage splits a table into partitions
- Support very large volume of Data
- High-availability guarantees in a single region, create tables in geo-redundant storage
- Can configure security and role-based access control

	Key (Customer ID)	Value (Customer Data)
C1	AAAAA BBB	101 Block Street YY 999 888
C2	MM NN	21 A Street 5 B Avenue
C3	DDD EEE FFF	111 222 66 C Road

Advantages/Disadvantages - Azure Table Storage

Advantages

- It's simpler to scale
- A table can hold semi-structured data
- No complex relationships
- Data insertion and retrieval is fast

Disadvantages

- Consistency
- No Referential Integrity
- Difficult to filter and sort on non-key data

Good to use for:

- Storing TBs of structured data capable of serving web scale applications
- Storing datasets that don't require complex joins, foreign keys, or stored procedures, and that can be denormalized for fast access.
- Capturing event logging and performance monitoring data.

Key
(Customer ID) Value
(Customer Data)

C1	AAAAA	BBB	101 Block Street	YY	999	888
C2	MM	NN	21 A Street	5 B Avenue		
C3	DDD	EEE	FFF	111	222	66 C Road

Azure File Storage



Azure File Storage

-  Azure File storage enables you to create file shares in the cloud, and access these file shares from anywhere with an internet connection
-  Server Message Block 3.0 (SMB) protocol
-  Authentication using Azure Active Directory Domain Services
-  Share up to 100 TB of data, Single file up to 1 TB
-  Upload file using Portal or AzCopy tool
-  Fully managed service
-  Encryption at rest by default, and can be configured for encryption at transit
-  Use Cases:
 -  Migrate existing applications to the cloud
 -  Share server data across on-premises and cloud
 -  Integrate modern applications with Azure File Storage
 -  Simplify hosting high Availability (HA) workload data

Methods of Provisioning and Deployment Non-SQL database



The Azure portal

- Convenient but manual



The Azure command-line interface (CLI)

- Set of commands to create and manage Azure resources specifically
- Can run from the operating system command prompt or the Cloud Shell in the Azure portal.
- Suitable if you need to automate service creation



Azure PowerShell

- This is a cross-platform task automation and configuration management framework.
- Command-line shell and scripting language that is built on top of .Net
- Azure provides a series of commandlets (Azure-specific commands) that you can use in PowerShell to create and manage Azure resources.



Azure Resource Manager templates

- JSON (JavaScript Object Notation) file template that describes the service, and can be used to create resources
- Provision a set of Azure services to multiple environments
- Save the configuration for future deployments

Learning objective



- **Authentication**
 - Storage Account keys
 - Shared access signature (SAS)
 - Azure Active Directory (Azure AD)
- **Access Control**
 - Role based access control (RBAC)
 - Access control list (ACL)
- **Network access**
 - Firewall and virtual network
- **Data Protection**
 - Data encryption in transit
 - Data encryption at rest
- **Advanced threat Protection**

Storage Account Access Keys

Authentication

Shared Access Signature (SAS)

Authentication



Shared Access Signature (SAS)



Shared Access Signature

Security token string

“SAS Token”

Contains permission like start and end time

Azure doesn't track SAS after creation

To invalidate, regenerate storage account
key used to sign SAS

Stored access policy



Stored access policy

Reused by multiple SAS

Defined on a resource container

Permissions + validity period

Service level SAS only

Stored access policy can be revoked

Azure Active Directory

Authentication



Azure Active Directory



Azure Active Directory

Azure Active Directory (AD)

- Grant access to Azure Active directory (AD) **Identities**
- AD is an enterprise identity provider, Identity as a Service (IDaaS)
- Globally available from virtually any device
- Identities – user, group or application principle
- Assign role at Subscription, RG, Storage account, container level.
- No longer need to store credentials with application config files
- Similar to IIS Application pool identity approach



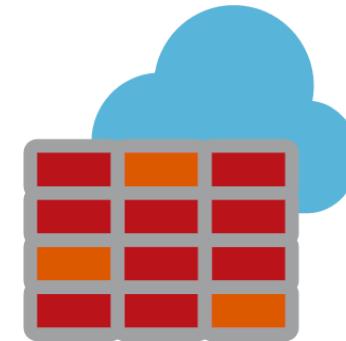
Azure Active Directory

Role based access control (RBAC)

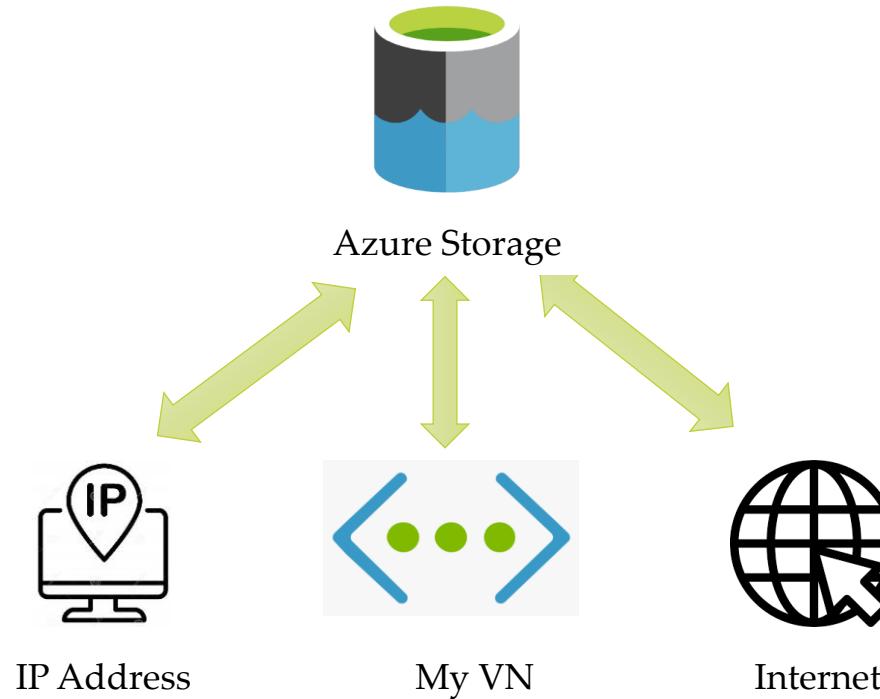
Access Control



Firewalls and Virtual Networks



Firewalls and Virtual Networks



Why Cosmos DB?

What traditional databases were lacking?



Challenges with globally distributed Databases

- Long time
- Lot of effort
- Need own infrastructure
 - Teams
 - Data centers etc.



How Cosmos DB evolved?

In 2010

Microsoft realized they need to build something very different to handle global distribution need.

In 2015

Microsoft released Azure Document DB
Supports SQL queries over JSON documents

In 2017

Rebranded Document DB as Cosmos DB
Globally distributed and scalable database

Why Cosmos DB?



FULLY MANAGED

- Database as a service (DaaS)
 - Serverless architecture
 - No operational overhead
- No schema or Index management

GLOBALLY DISTRIBUTED

- Turnkey global distribution

MULTIMODEL & MULTI-LANGUAGE

- Supports Jason documents, table graph and columnar data models
 - Java, .NET, Python, Node.js, JavaScript, etc.

CONSISTENCY CHOICES

- Azure Cosmos DB's support for consistency levels like strong, eventual, consistent prefix, session, and bounded-staleness.

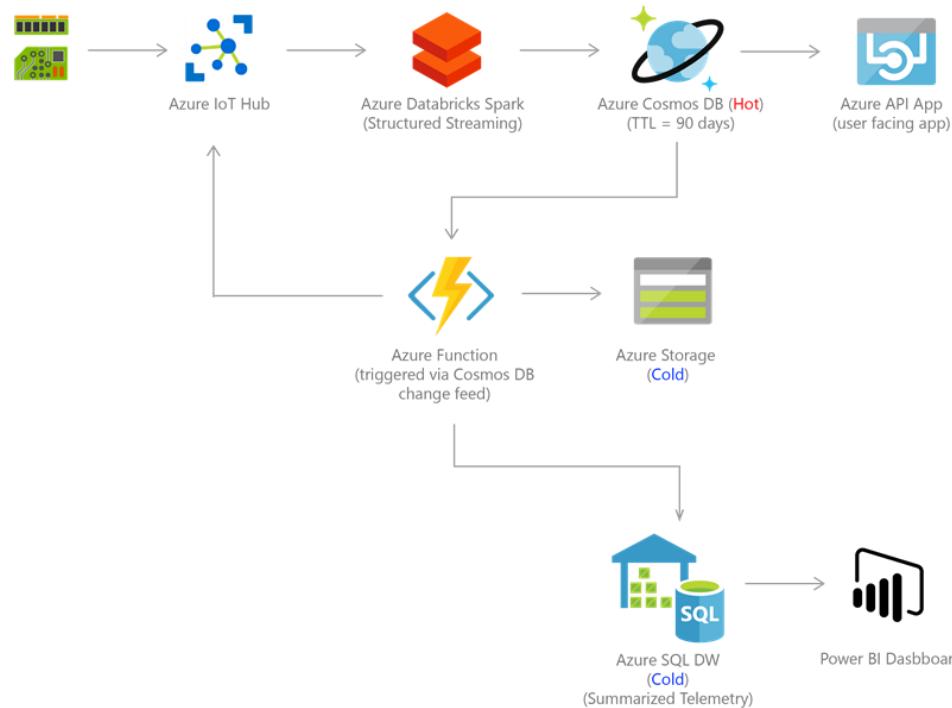
SCALABLE

- Unlimited scale for both storage and throughput.

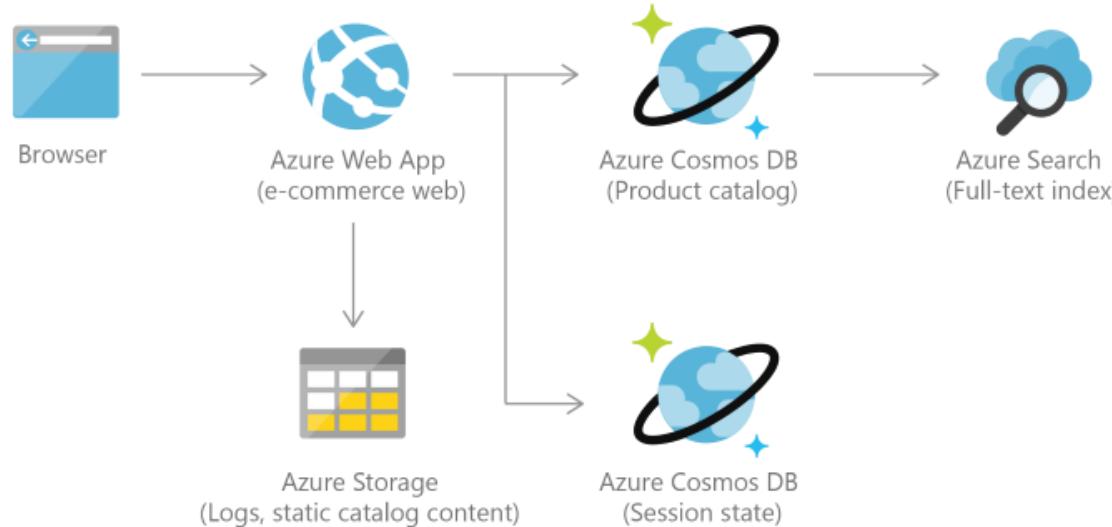
HIGHLY AVAILABLE, RELIABLE & SECURE

- Always on
- 99.999% SLA
- < 10ms latency

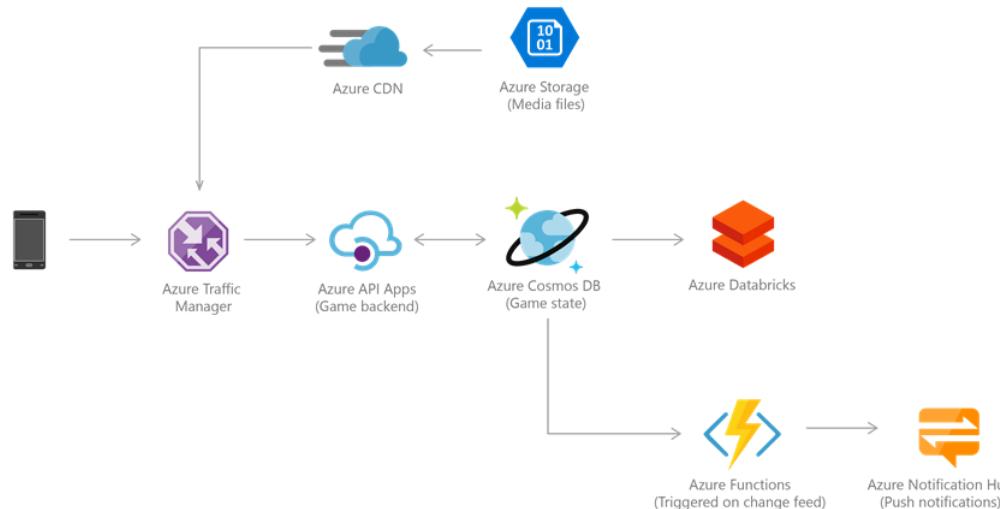
Use case - IoT



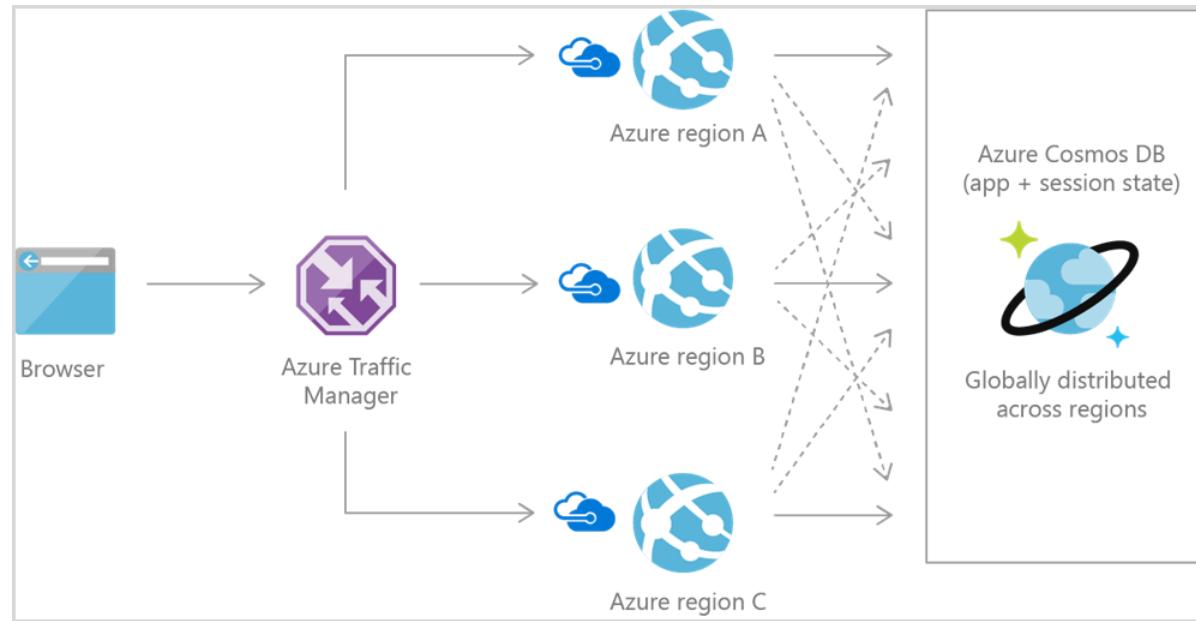
Use case – Retail and Marketing



Use case – Gaming



Use case – Web and mobile



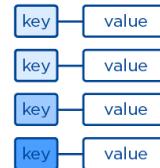
NoSQL



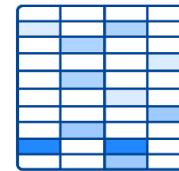
Table



Key-Value



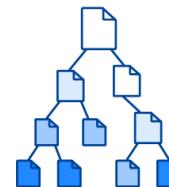
Wide-column



Graph



Document



SQL API vs MongoDB API

SQL(CORE) API

JSON Documents

Microsoft original Document DB platform

Supports server side programming model

You can use SQL like language to query JSON documents.

MongoDB API

BSON Documents

Implement Wire protocol

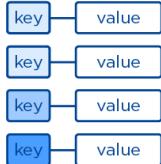
Fully compatible with Mongo DB application code

Migrate existing Cosmos DB without much change of logic

Use SQL(CORE) API for new development

Cosmos DB Table API

Key-Value

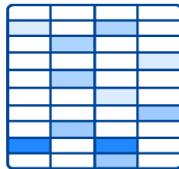


- Key-Value store
- Premium offering for Azure Table Storage
- Existing Table Storage customers will migrate to Cosmos DB Table API
- Row value can be simple like number or string
- Row cannot store object



Cosmos DB Cassandra API

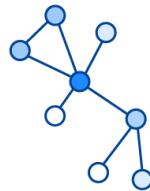
Wide-column



- Wide column No SQL Database
- Name and format of column can vary from row to row.
- Simple migrate your Cassandra application to Cosmos Cassandra API and change connection string.
- Interact
 - Cassandra based tools
 - Data Explorer
 - Programmatically, using SDK (CassandraCSharpdriver)

Cosmos DB Gremlin API

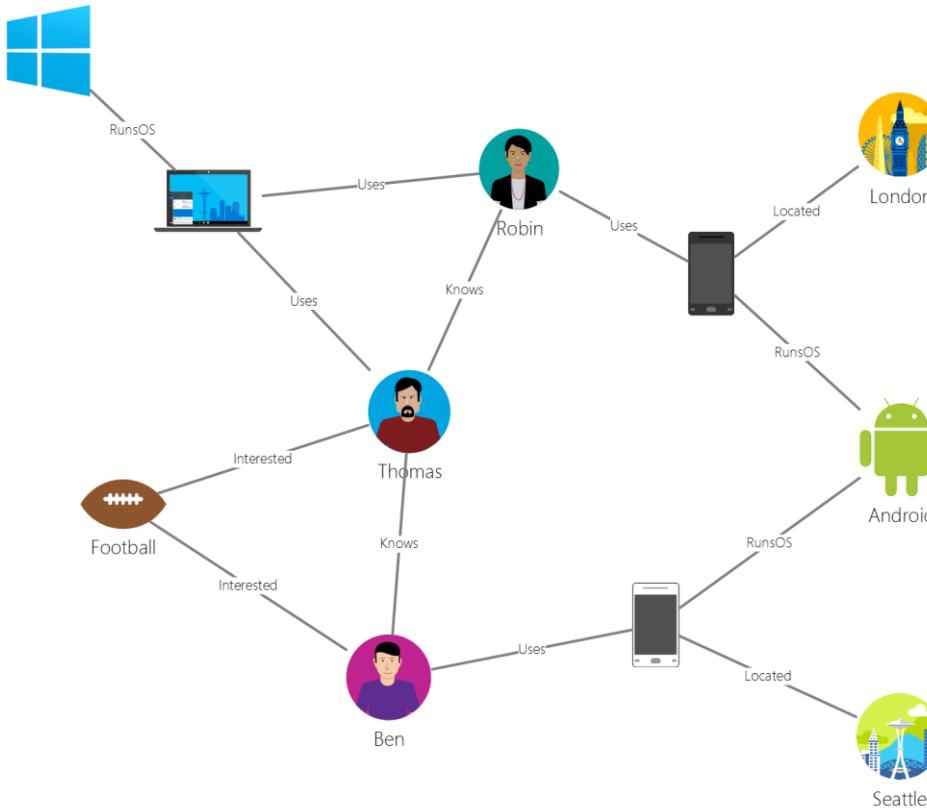
Graph



- Graph Data Model
- Real world data connected with each other
- Graph database can persist relationships in the storage layer



Graph Model



Cosmos DB Gremlin API

Graph

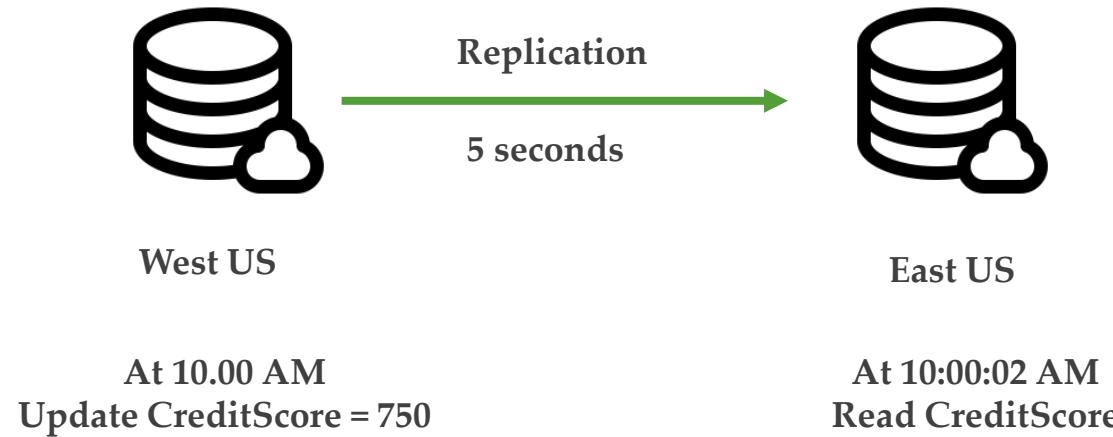


- Graph Data Model
- Real world data connected with each other
- Graph database can persist relationships in the storage layer
- Use cases
 - Social networks
 - Recommendation engines
 - Geospatial
 - Internet of things
- Migrate existing apps to Cosmos DB Gremlin API
- Graph traverse a language

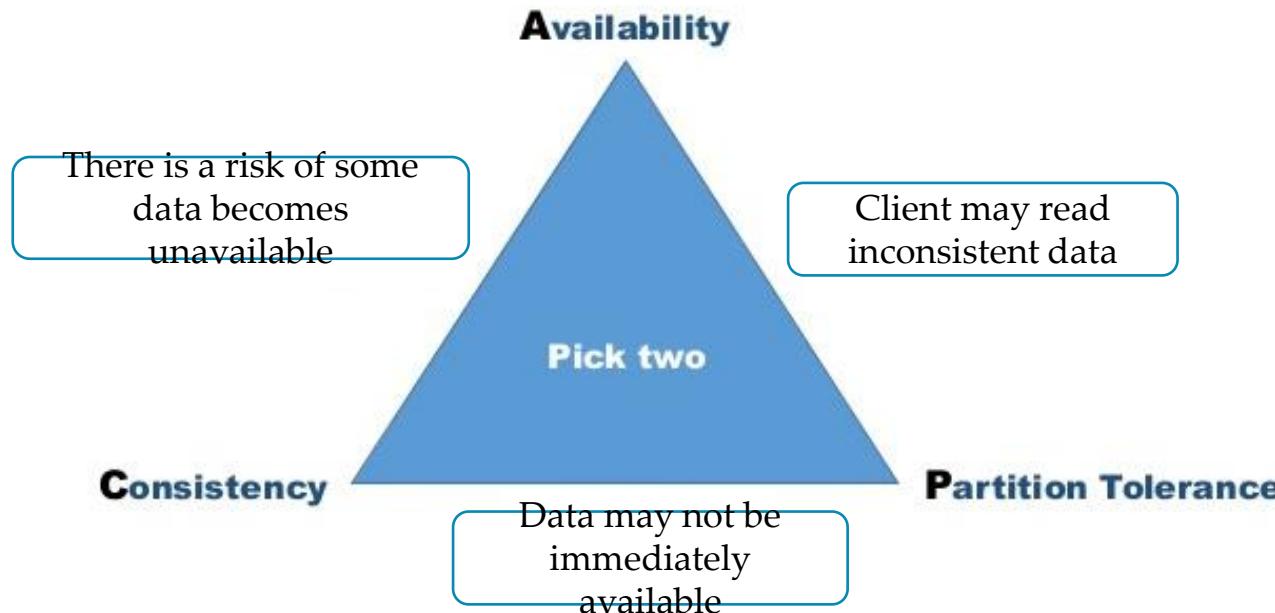
Analyze the decision criteria

	Core (SQL)	MongoDB	Cassandra	Azure Table	Gremlin
New projects being created from scratch	✓				
Existing MongoDB, Cassandra, Azure Table, or Gremlin data		✓	✓	✓	✓
Analysis of the relationships between data					✓
All other scenarios	✓				

Data consistency



CAP Theorem



Five consistency Levels



Strong: No dirty reads, high latency, cost highest, closest to RDBMS

Bounded staleness: Dirty reads possible, bounded by time and updates

Session: No dirty reads for writers (within same session), dirty read possible for other users

Consistency prefix: Dirty reads possible but sequence maintain, reads never see out-of-order writes

Eventual: No guaranteed order, but eventually everything gets in order

Setting the consistency level

Set default for entire account

Can be changed any time

Override at request level

Any request can weaken the default consistency level



Cosmos DB



- **Role based access control (RBAC)**
- **Network security**
- **Access Security Keys**
- **CORS (Cross-Origin Resource Sharing)**
- **Azure Private Endpoint**
- **Advance Security Option**

Management tools for Non-Relational Database



Azure Cosmos explorer

- lets you set up temporary or permanent read or read-write access to your database
- Use to run queries, store procedures and triggers, and view their results
- Share query results with other users who do not have access to Azure portal or subscription.



Azure Storage Explorer

- Azure Storage Explorer is a multi-platform graphical user interface (GUI) application that runs on Windows, Linux, and MacOS
- lets you manage storage entities including Cosmos DB, Blob storage, and Queue storage
- Use to manipulate data and manage entities such as stored procedures
- Does NOT provide a way to manage database access.



Azure Data Studio

- Data Studio is a cross-platform tool for authoring and select management operations.



AzCopy

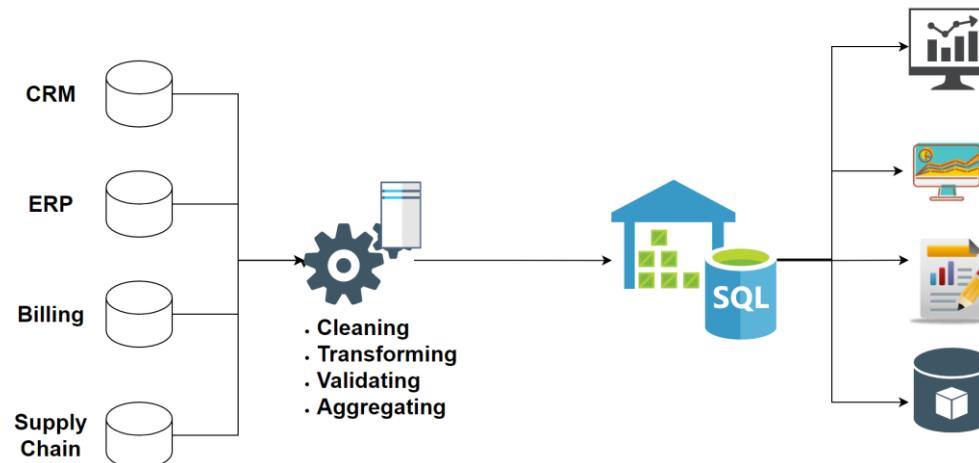
- CLI utility to copy files to Azure Storage
- Use AzCopy to copy files from a local path, other Azure Storage accounts, and AWS S3 buckets



Cosmos DB Data Migration Tool

- standalone tool that allows you to easily import data to Cosmos DB.
- You can use this tool to import data from: JSON files, MongoDB, SQL Server, CSV files, Azure Cosmos DB collections

What is Data Warehouse



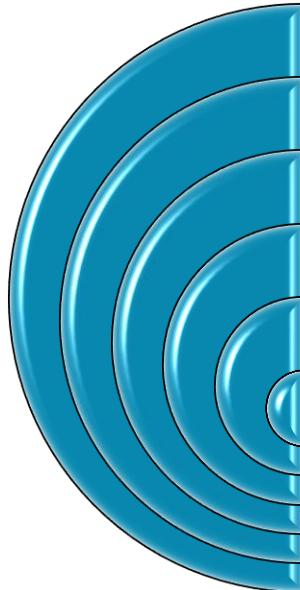
Concerns About a Data Warehouse

Extra piece of software that needs maintenance

The data is already present in the operational systems

However, extracting information from operational data is complicated

Problems a Data Warehouse Can Solve



We have so much data, but we can't make anything of it

I only want to know what is important

We need to slice and dice the data

Business people need to access data easily

Numbers between departments don't match and we don't know who's right

We want people to make decisions based on facts

OLTP vs OLAP

OLTP

- Many small transaction
- Current data
- Used to run the business
- Highly detailed
- Typically in the GB scale
- Processing performance limit

OLAP

- Low volume but complex queries
- Historic, non-volatile data
- Used to analyze the business
- Consolidated and summarized
- TB and above scale
- No limit, pause/resume compute

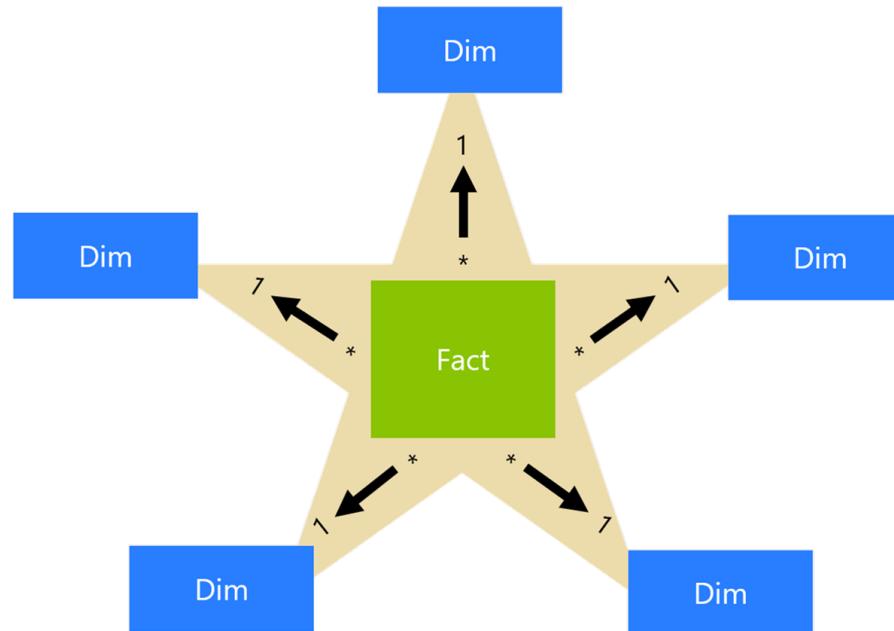
Star Schema

A Data Mart design pattern consisting of fact and dimension tables resembling a star like shape.

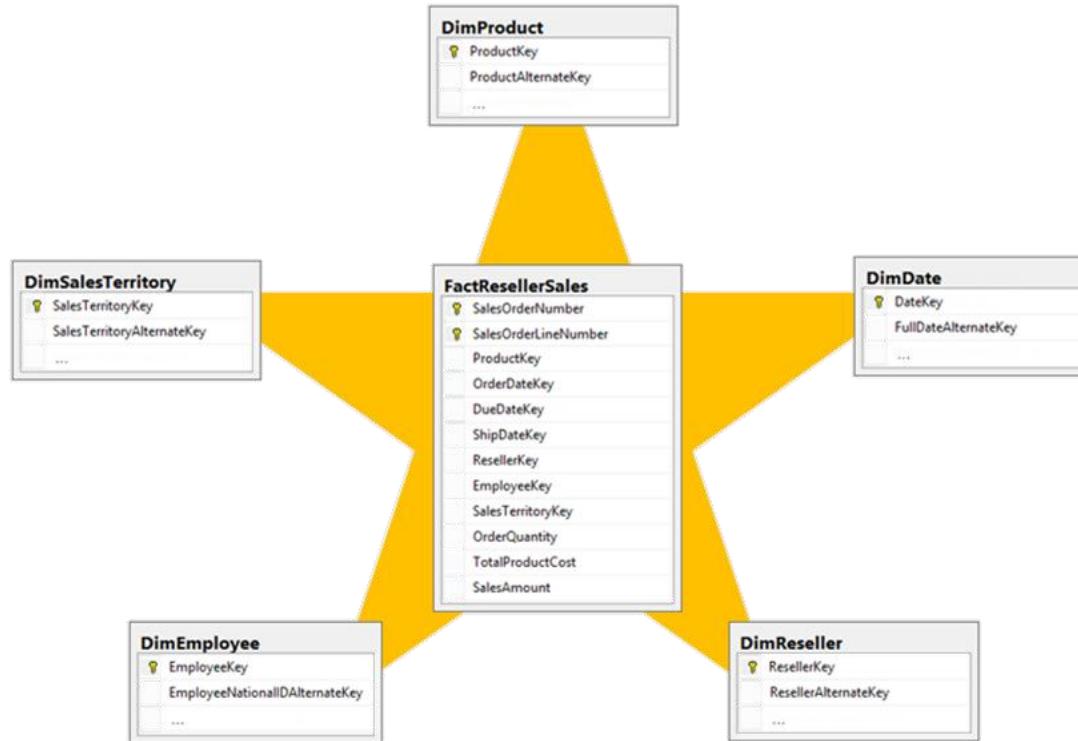
Facts vs Dimension

FACT	Dimension
Store observations or events	Describe business entities – things you model
Subject of Analysis	Context of Analysis
Numerical	Textual (most often)
Additive and Semi-additive	Contain most attributes
sales orders, stock balances, exchange rates, temperatures, etc.	Example: products, people, places, and concepts including time itself.

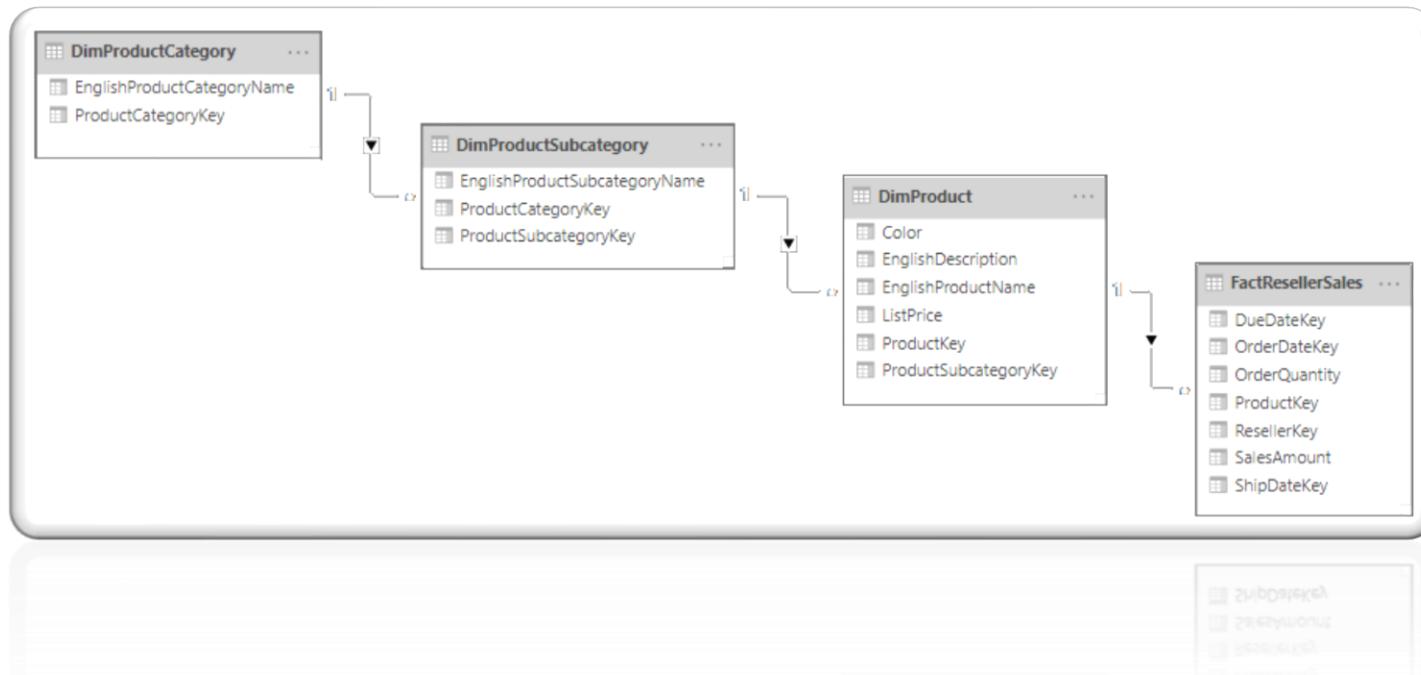
Star Schema



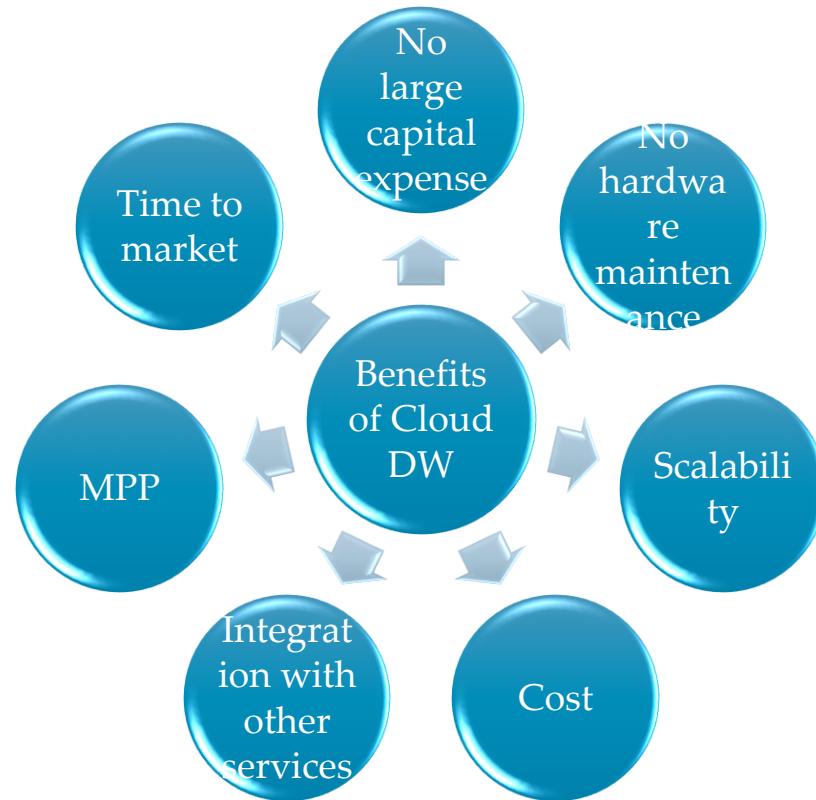
Star Schema

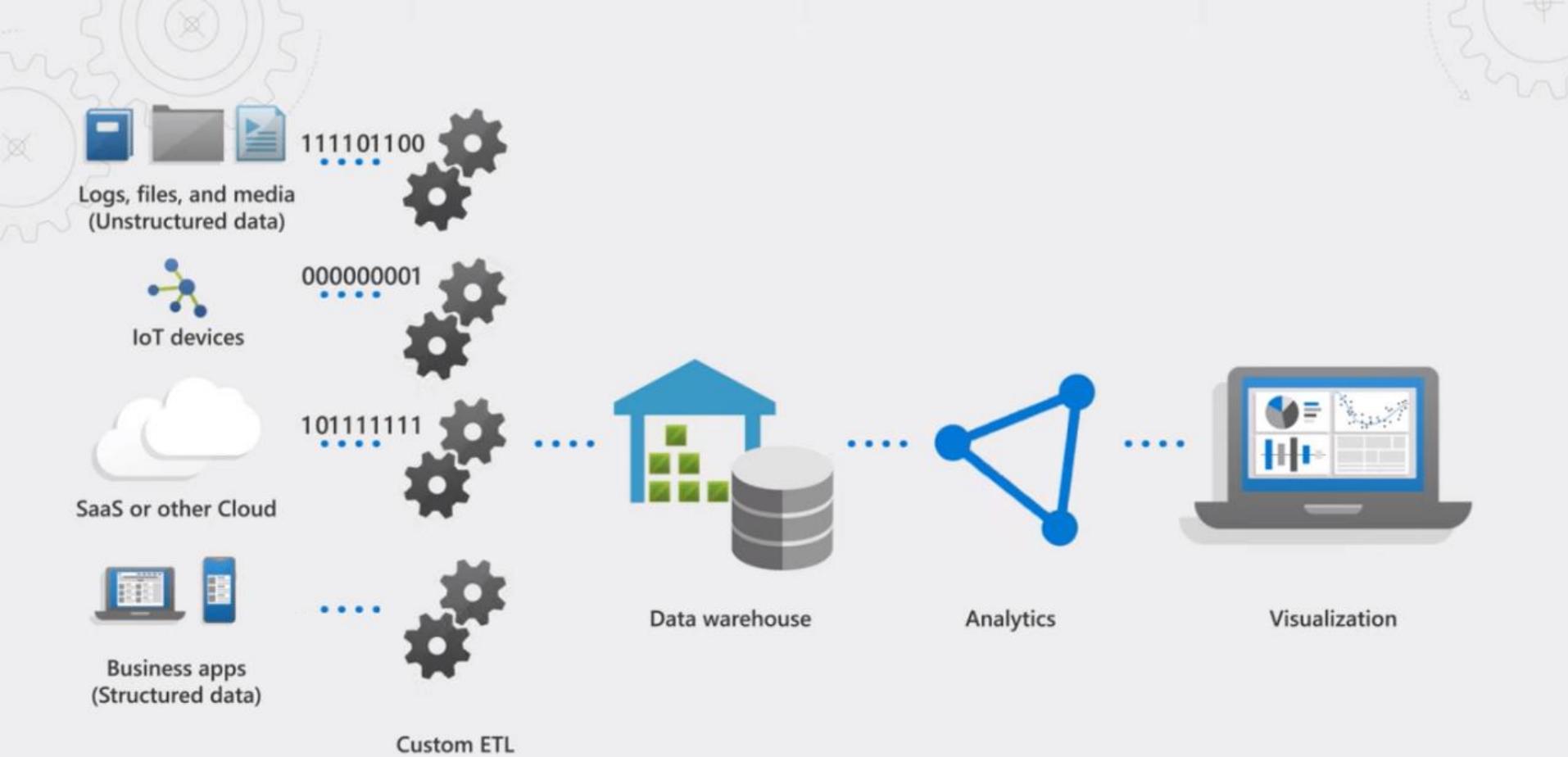


Snowflake dimensions



Why Warehousing in Cloud?

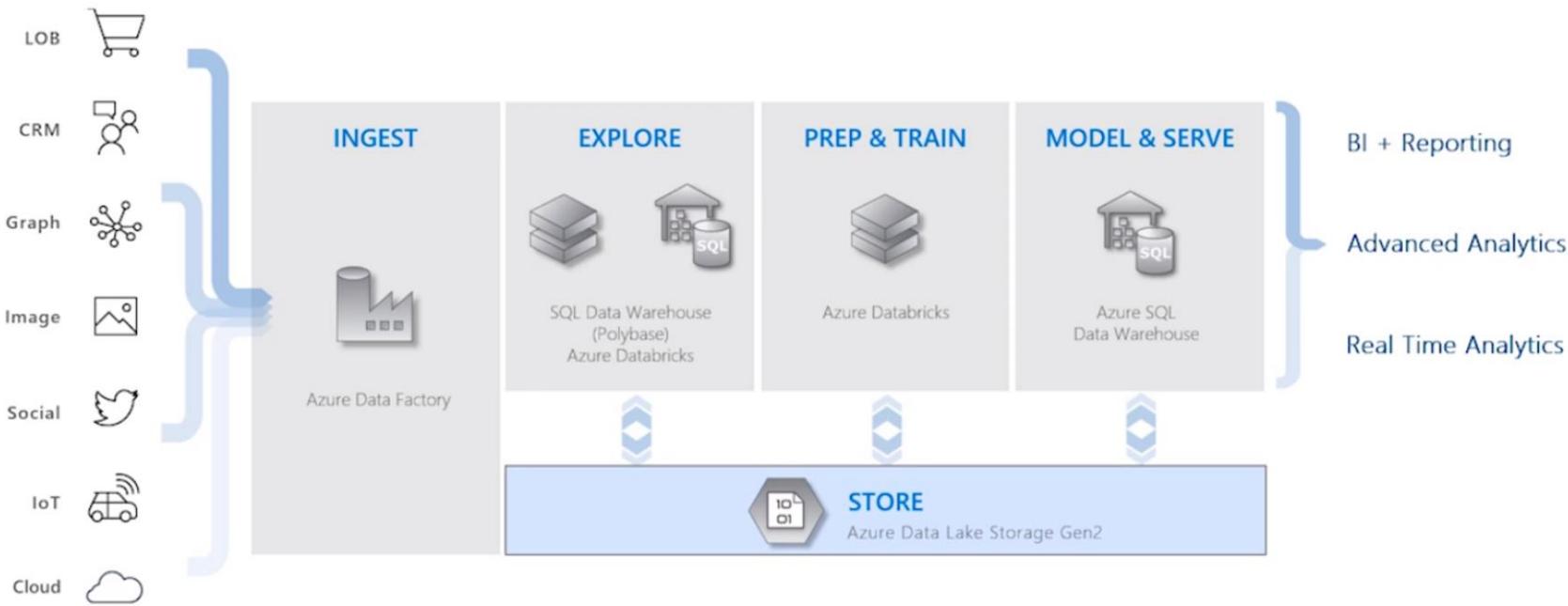




Old data warehouse

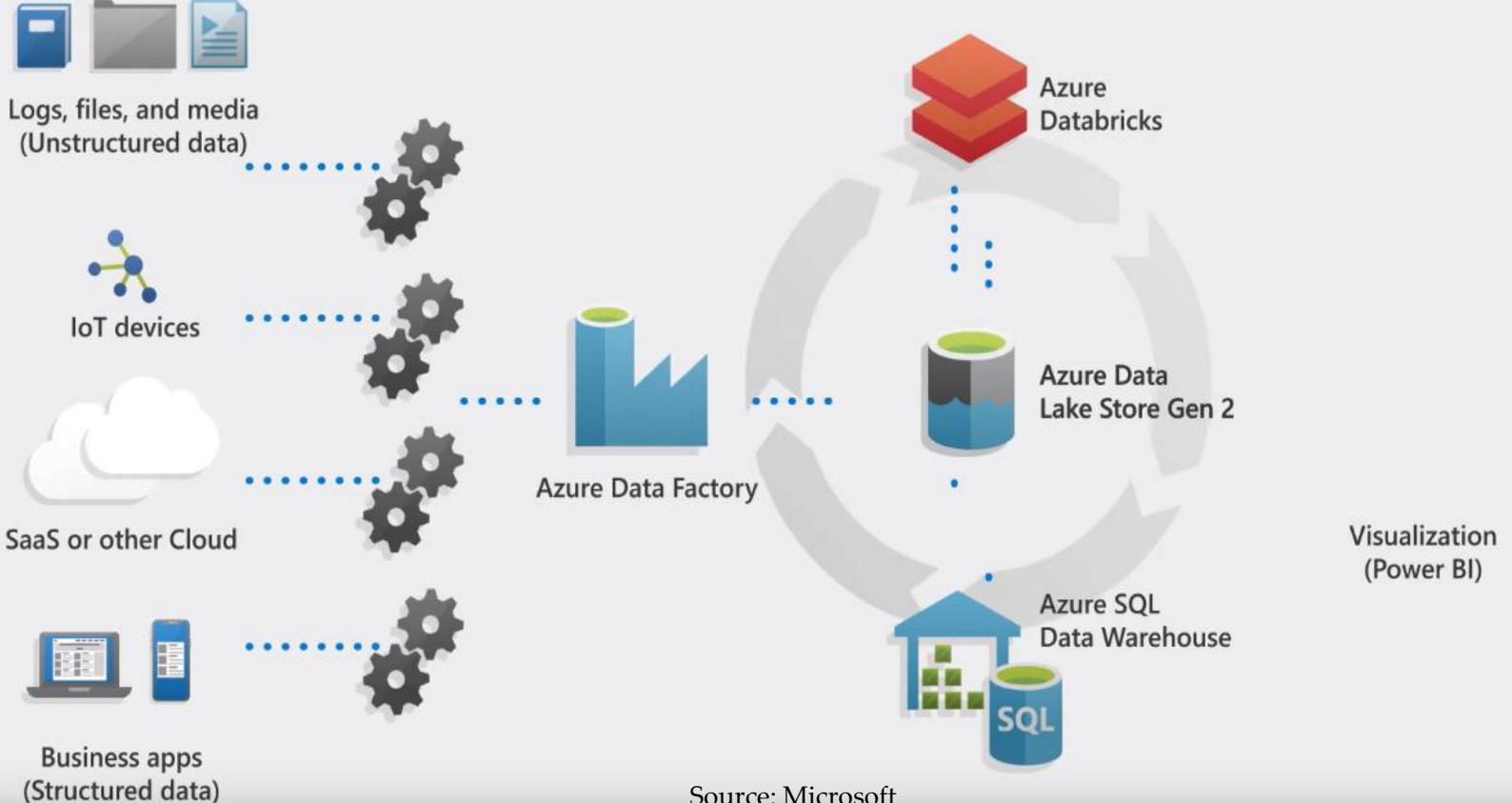
Source: Microsoft

Modern Data Warehousing

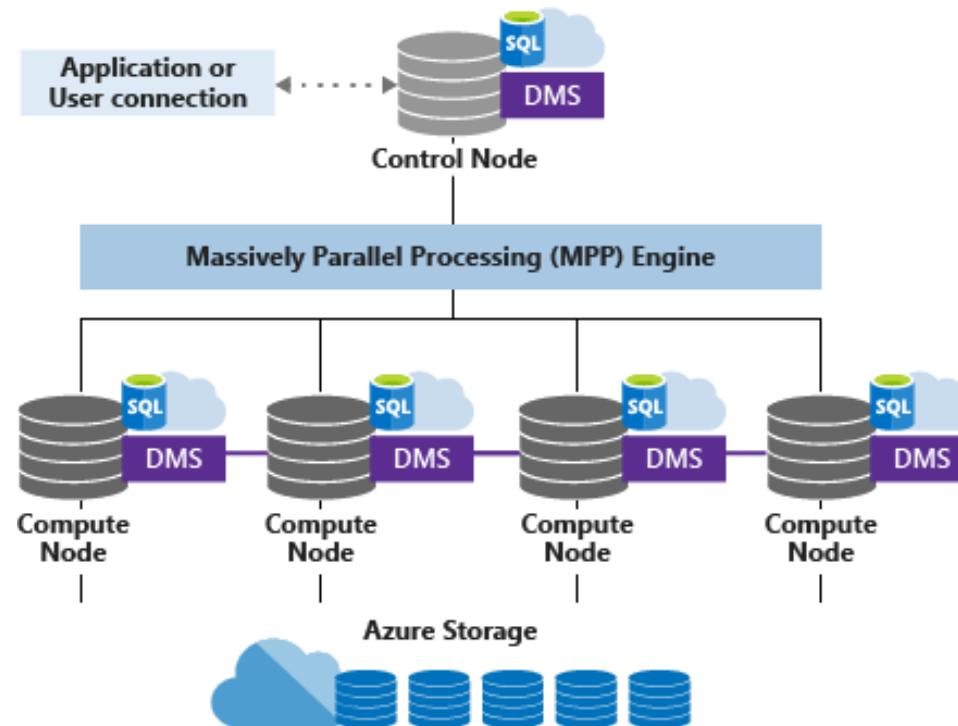


Source: Microsoft

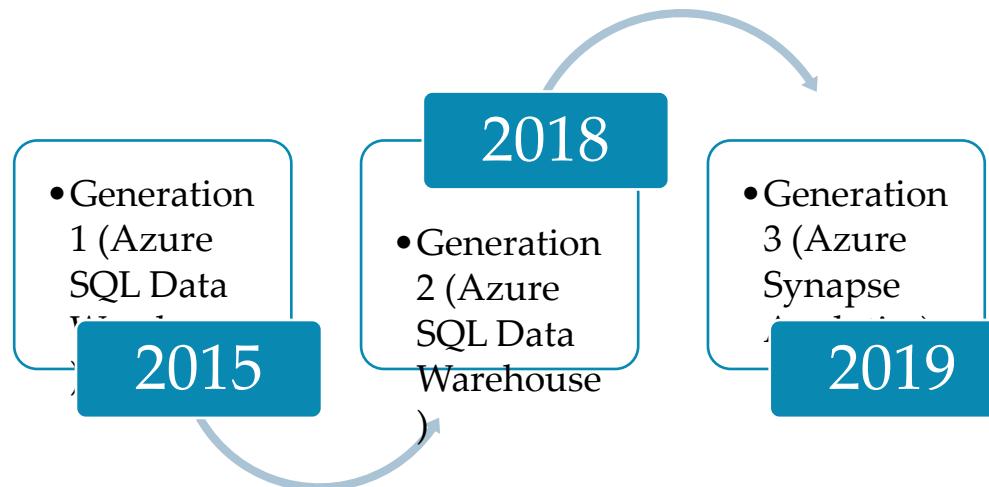
Modern Data Warehouse



MPP Architecture – Synapse Analytics Service



Azure SQL Data warehouse



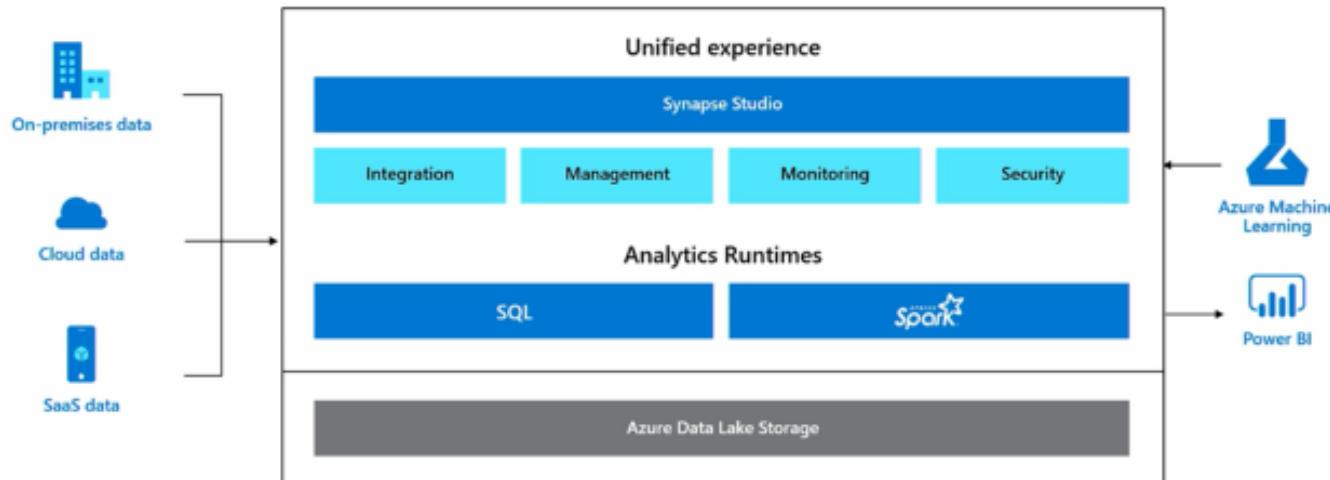
Azure Synapse Analytics

"Synapse is the next generation of Azure SQL Data Warehouse, blending big data analytics, data warehousing, and data integration into a single unified service that provides end-to-end analytics with limitless scale."

Azure Synapse Analytics



Azure Synapse Analytics



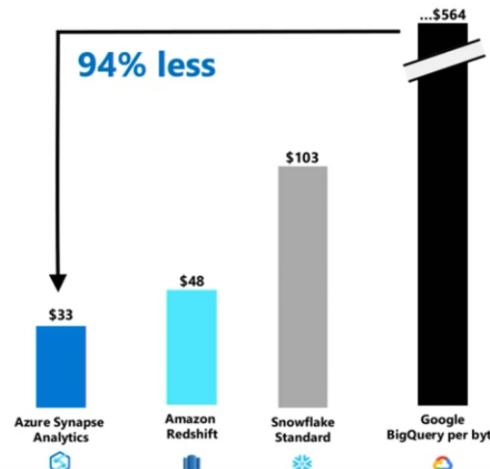
Unparalleled price-performance

Up to 14x faster

Costs 94% less than other cloud providers

TPC-H Benchmark

Price-performance | Lower is better



Azure Synapse



**Structured and
unstructured data**



**Data warehousing and
big data analytics**



Cloud-native



Azure Synapse Analytics



Limitless Scale



Powerful Insight



Unified Experience



Security

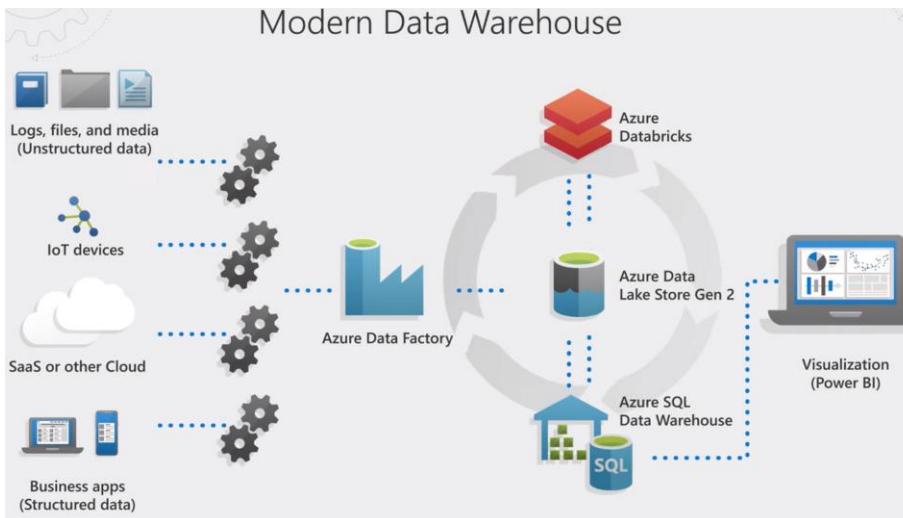


Code free data
orchestration

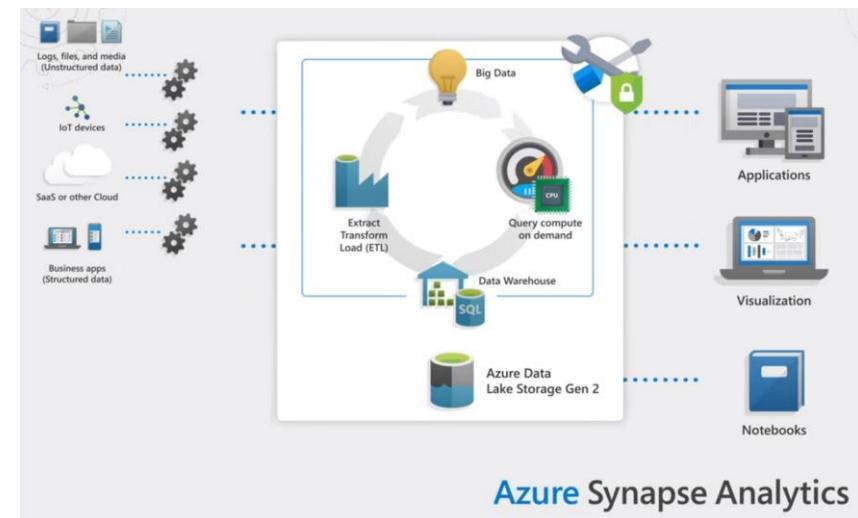


Choice of Language

Modern vs Synapse Architecture

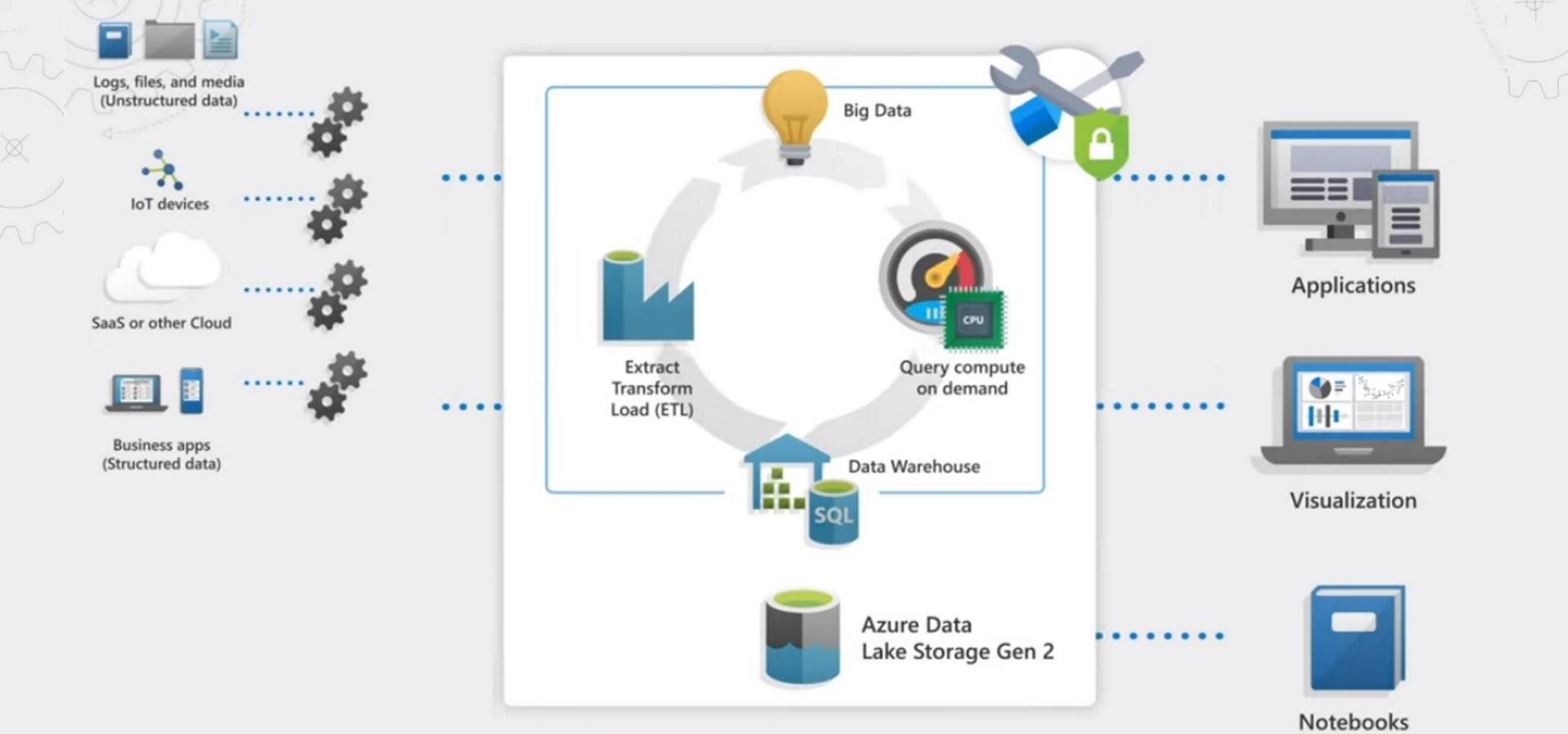


Modern Data Warehouse Architecture

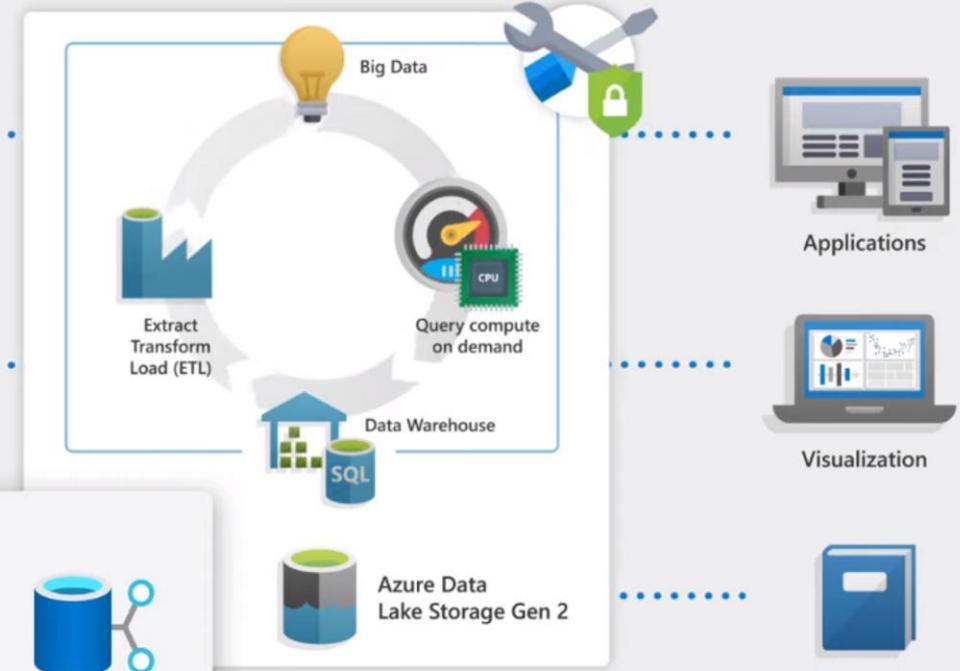
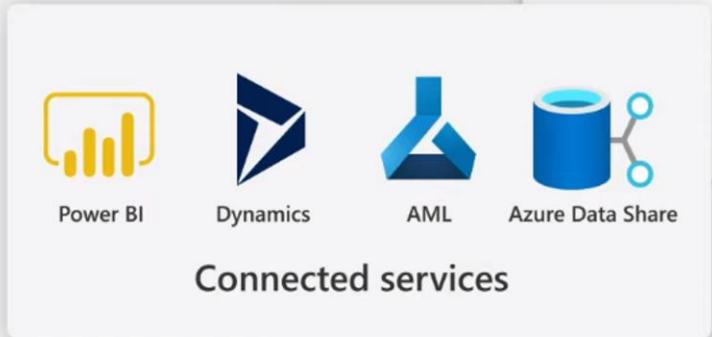


Azure Synapse Analytics

Synapse Analytics Architecture



Azure Synapse Analytics



Azure Synapse Analytics

Azure Synapse Analytics



Loading Methods

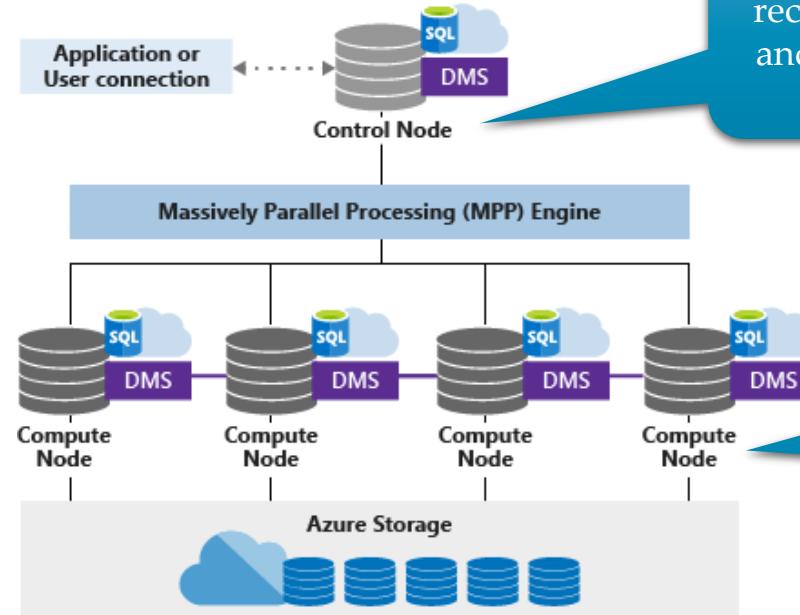
Single client loading methods

- SSIS
- Azure Data Factory
- BCP
- Can add some parallel capabilities but are bottlenecked at the control node

Parallel readers loading methods

- PolyBase
- Reads from Azure blob Storage and loads the contents into Azure SQL DW
- Bypasses the Control Node and loads directly into the Compute Nodes

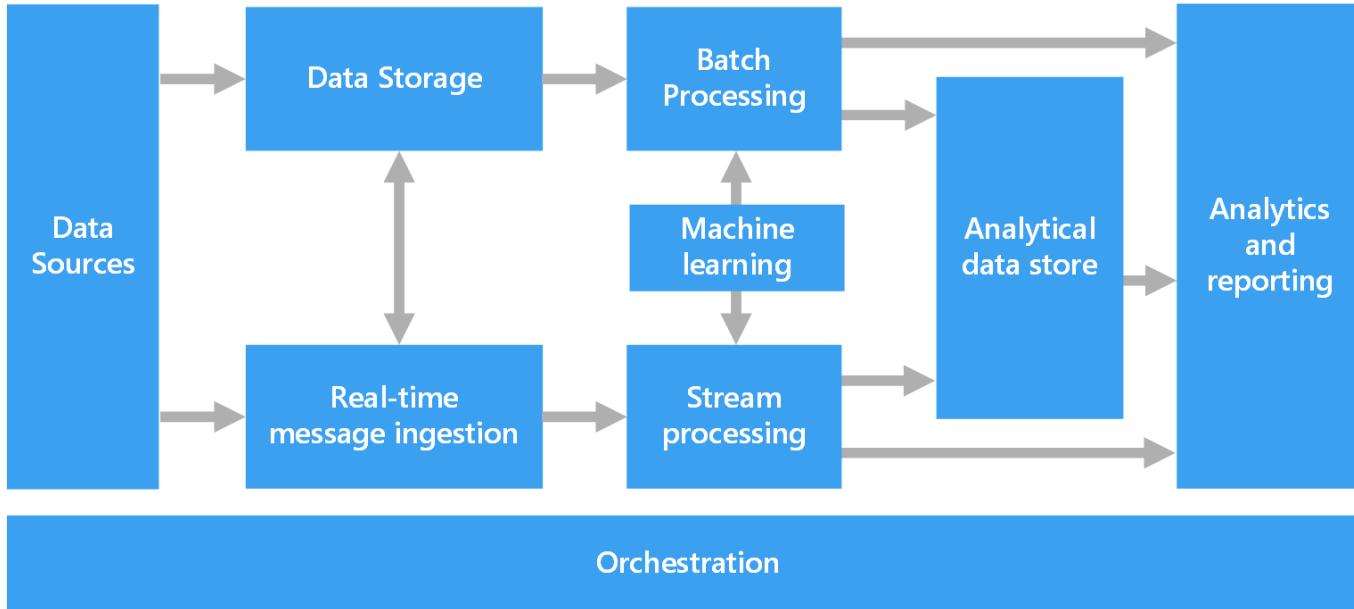
Control Node



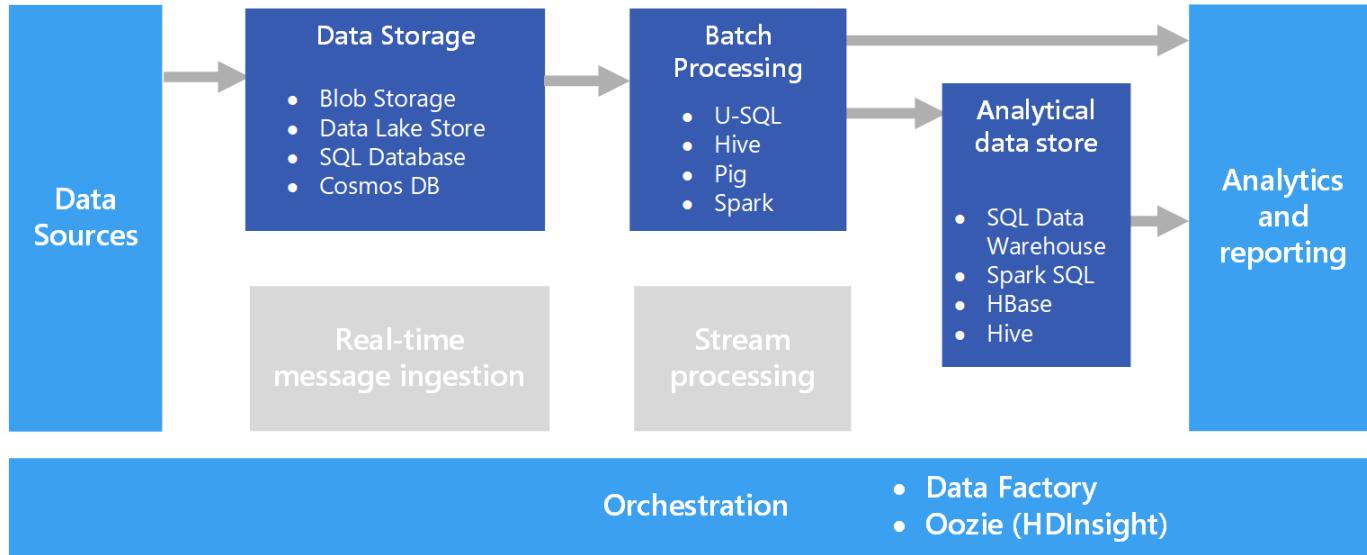
The **Control** node receives connections and orchestrates the queries

The **Compute** nodes do processing on the data and scale with the DWUs.

Big Data Architecture



Batch Processing



Batch Processing:

- Data at rest
- Operate on very large dataset
- Computation takes significant time

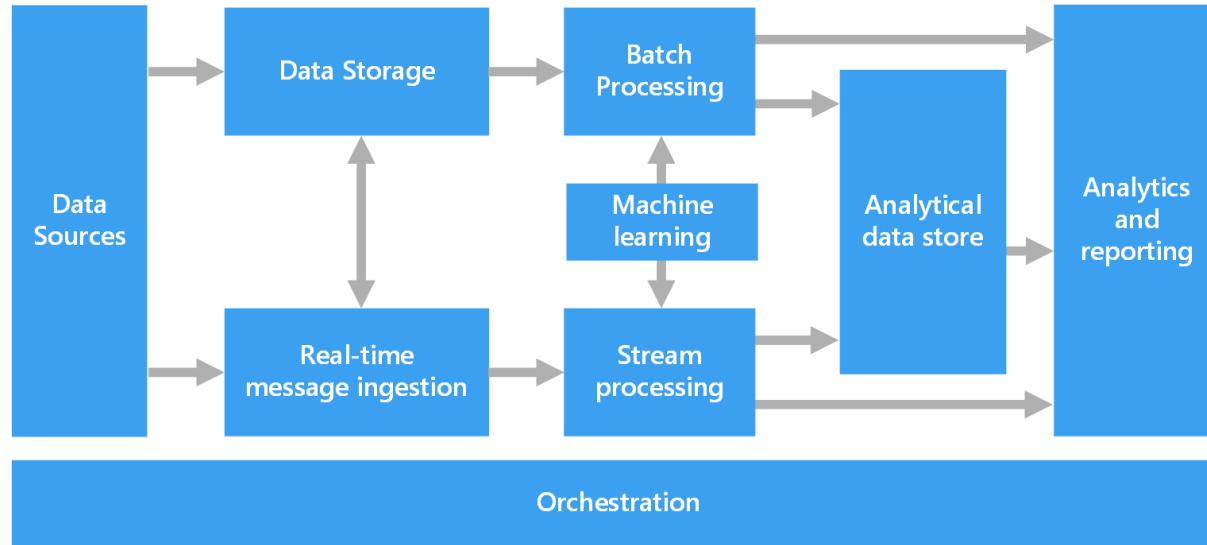
Use Cases:

- Example – Web Server logs to Report

Challenges:

- Data Format and encoding
- Orchestration time slices

Real Time Processing Architecture



Real time processing:

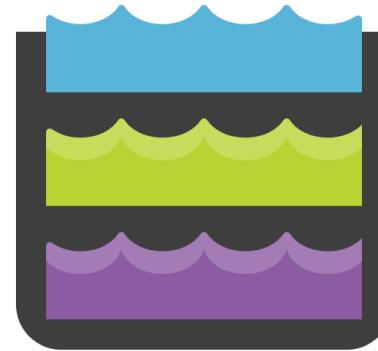
- Deals with streams of data that are captured in real-time
- Processed with minimal latency
- Incoming data typically arrives in an unstructured or semi-structured format, such as JSON
- Generate real-time (or near-real-time) reports or automated responses.

For example:

- Use sensor data to detect high traffic volumes

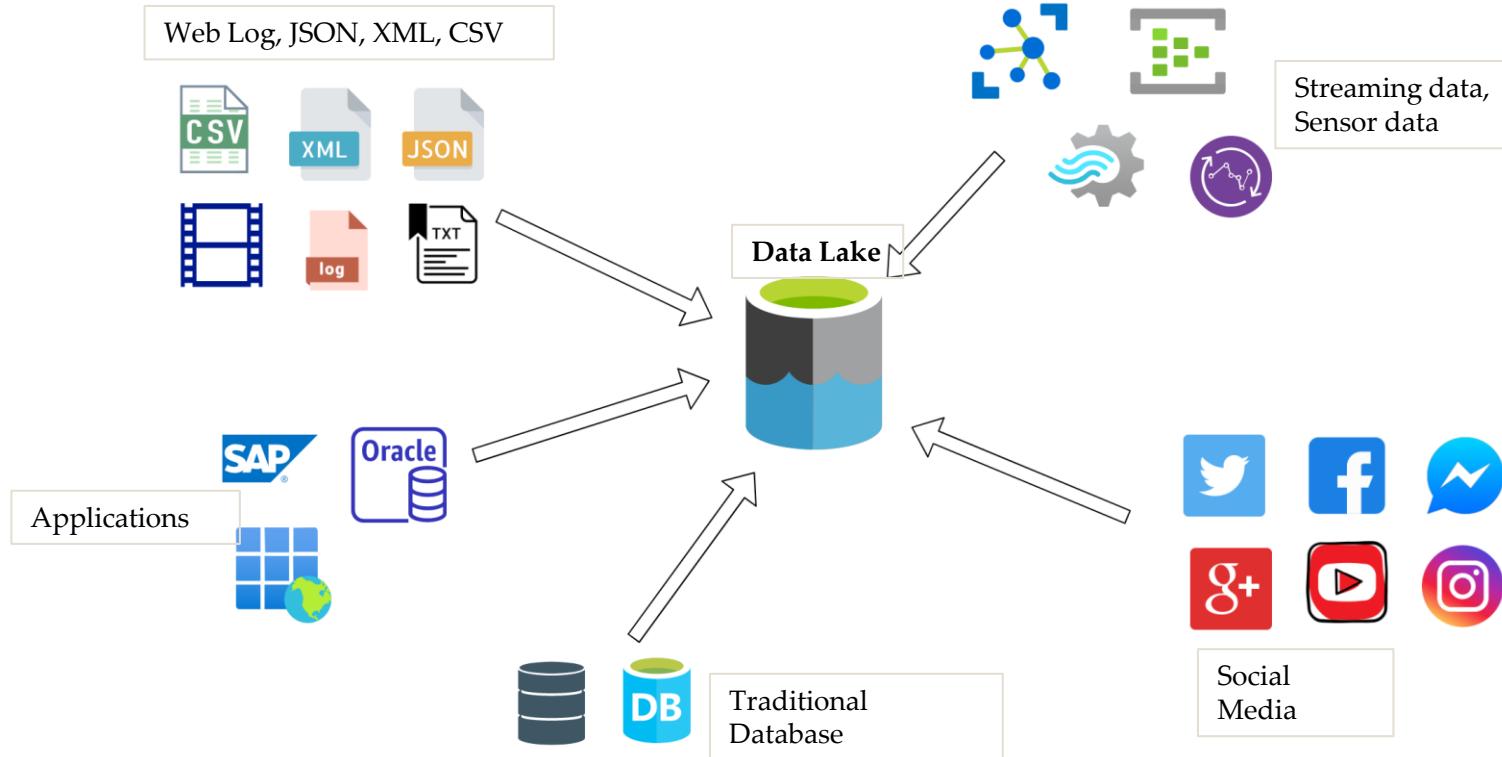
Challenges:

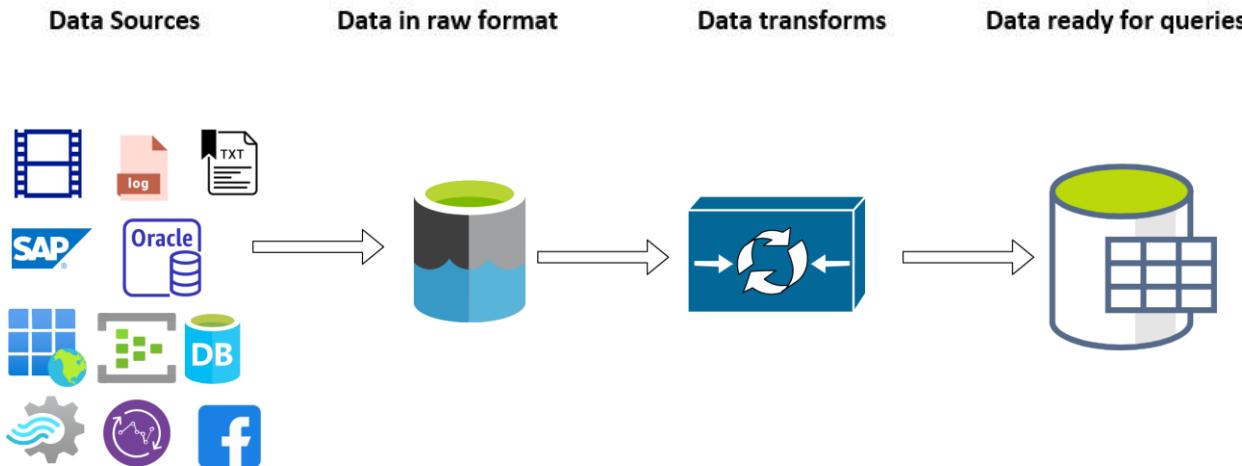
- Ingest, process, and store messages in real time, especially at high volumes



Data Lake is a big container to store data.

Data Lake Sources





What is Data Lake?

"If you think of a DataMart as a store of bottled water – clean and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples."



James Dixon
CTO, Pentaho



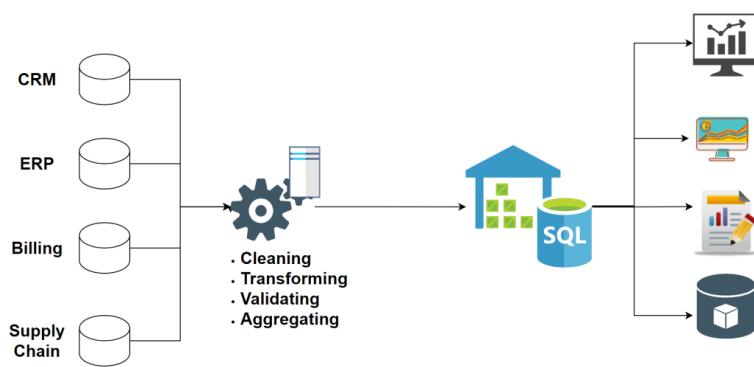
Data Warehouse



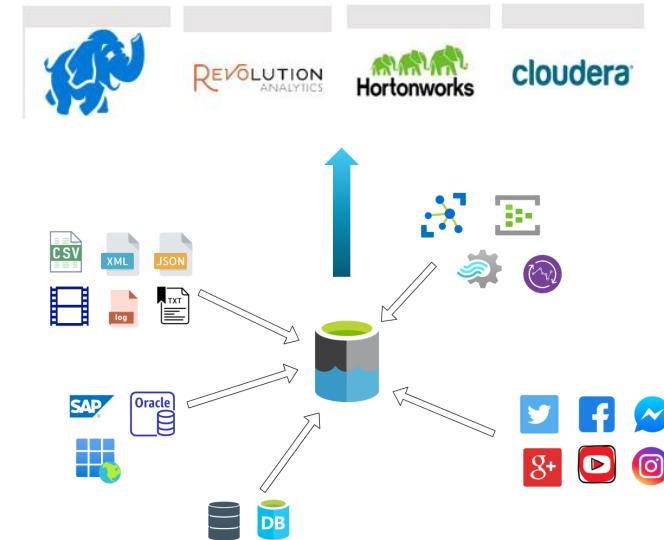
Data Lake



Data Warehouse vs Data Lake



Data Warehouse



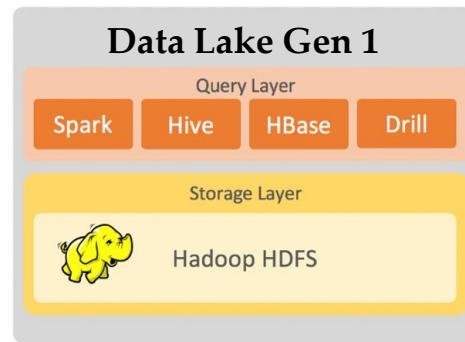
Data Lake

Data Lake Gen2 Hierarchical namespace

Azure Data lake Gen 2



Blob Storage



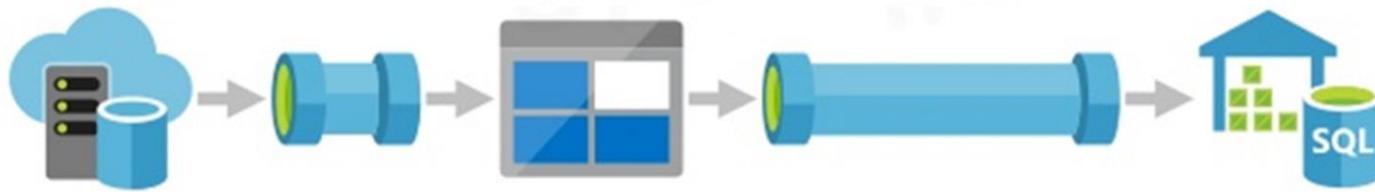
Azure Data Lake Storage Gen2



Azure Data Factory

Cloud version of SSIS

What can you do in Azure Data Factory?



Copy Data



Copy Data

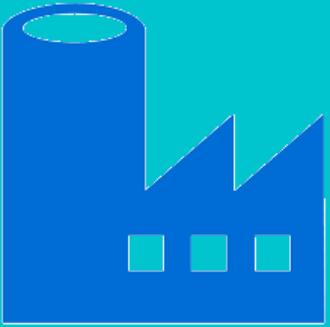
More than 80 connectors to different services are available

Transform Data



Transform Data

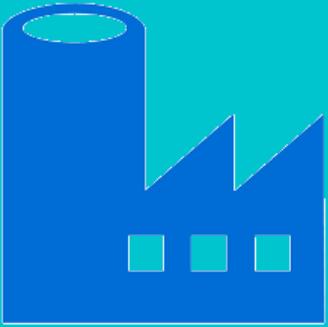
Using newly added Data Flow, now Data Factory is complete cloud based ETL tool.



Azure Data Factory

Definition:

Azure Data Factory (ADF) is a hybrid data integration service that enables you to quickly and efficiently create automated data pipelines – without having to write any code!



Azure Data Factory

- Hybrid Data Integration Service
- Simplifies ETL at scale
- Enables modern data integration
- Drag and drop interface
- Over 80 connectors available
- Move, transform and save data
- Managed Service
- Create Data Driver workflows
- Orchestrate and automate data movement
- Transform and store data
- Operationalize the process
- ETL or ELT scenarios



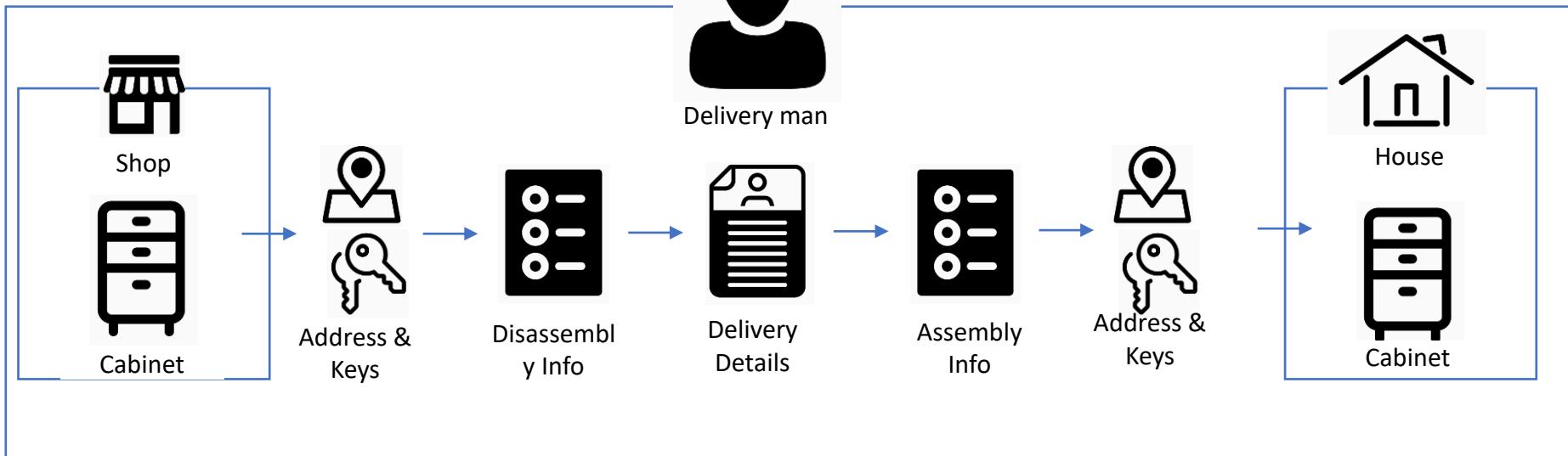
Azure Data Factory Components



Delivery Manager



Delivery man





Data Factory Pipeline



Integration Runtime



Blob Storage

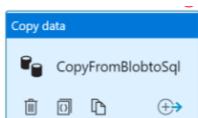


Order.csv

Linked Service



Dataset



Copy Activity



Dataset



Linked Service

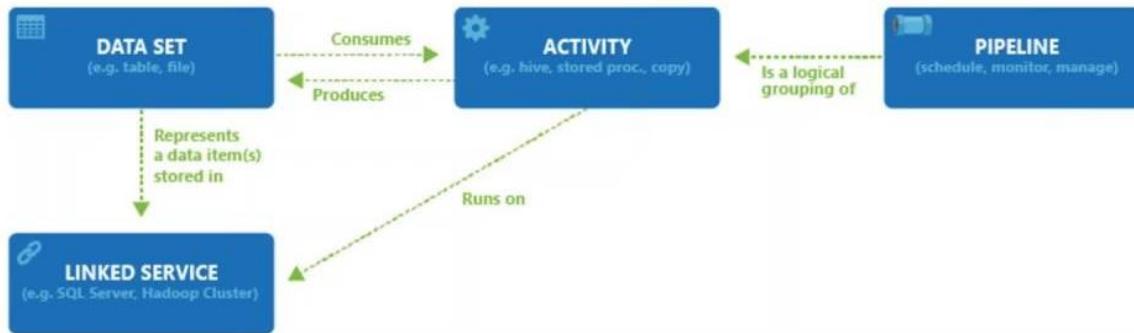


Order Table

Data Factory vs SSIS

Azure Data Factory	SSIS
Pipeline	Package
Linked Service	Connection manager
Source	Source
Sink	Destination
Activity	Control flow task
Data Flow	Data flow

Data Factory

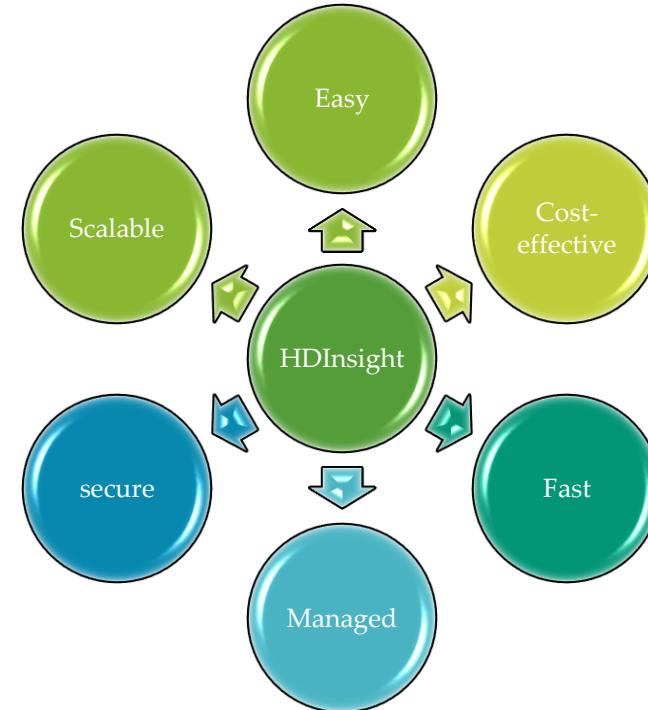


Data Factory Triggers



What is Azure HDInsight?

HDInsight is a
cloud distribution of
Hadoop components



HDInsight makes Hadoop easy



No HW costs

HDInsight makes Hadoop easy



No HW costs



Unlimited scale

HDInsight makes Hadoop easy



No HW costs



Unlimited scale



Pay what
you need

HDInsight makes Hadoop easy



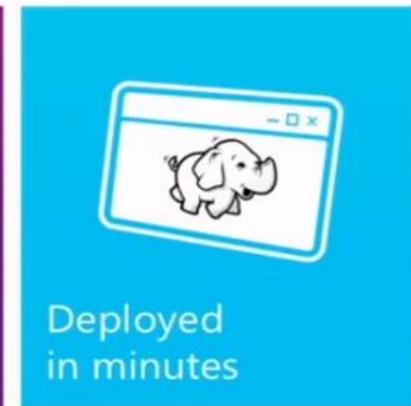
No HW costs



Unlimited scale



Pay what
you need



Deployed
in minutes

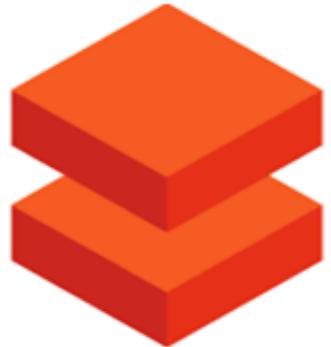
Microsoft Azure Databricks

A fast, easy, and collaborative Apache Spark™ based analytics platform optimized for Azure



Azure Databricks





Azure Databricks

Managed 1st Party Azure Service

Native integration with Azure & Its services;
Azure SLA and support

Transparency

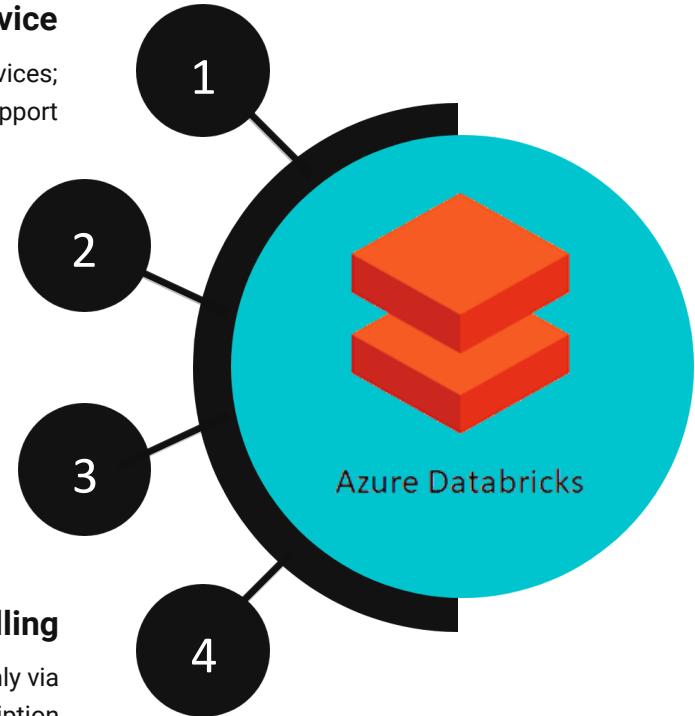
Deploys Databricks workspace and
clusters in customer subscription

Security

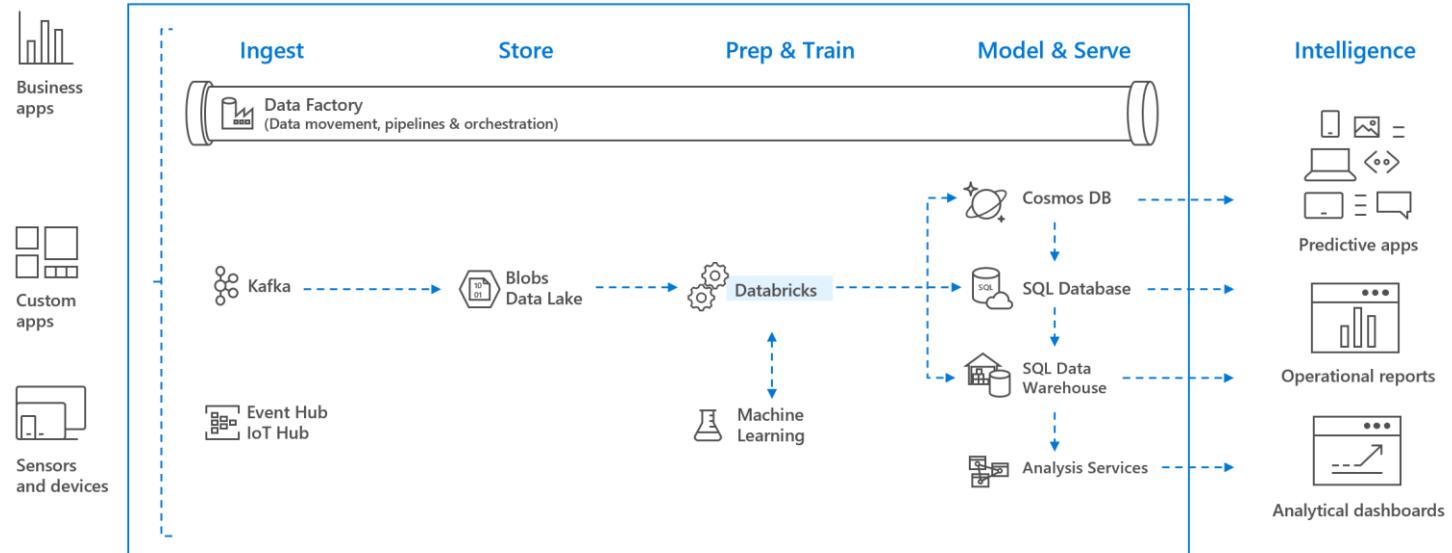
Natively integrates with Azure
Active Directory & Providers RBAC

United Billing

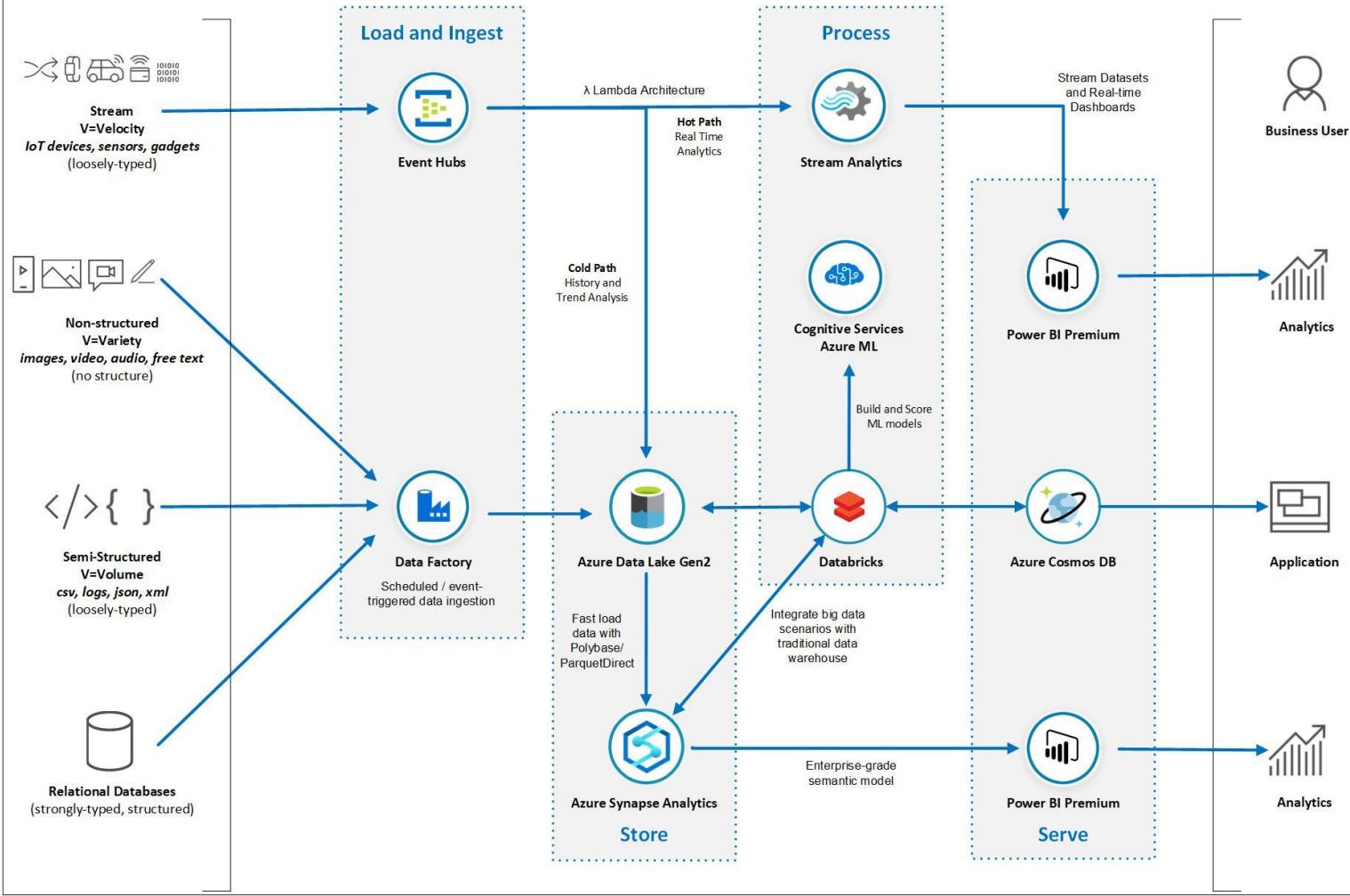
Pay for what you use only via
Azure subscription



Azure Databricks Architecture



Modern Data Platform Reference Architecture





Power BI



What is Power BI



Cloud Based Service



Cloud and On-premises data



Data Models



Reports and
Dashboard



Natural Language Queries
and Quick Insights



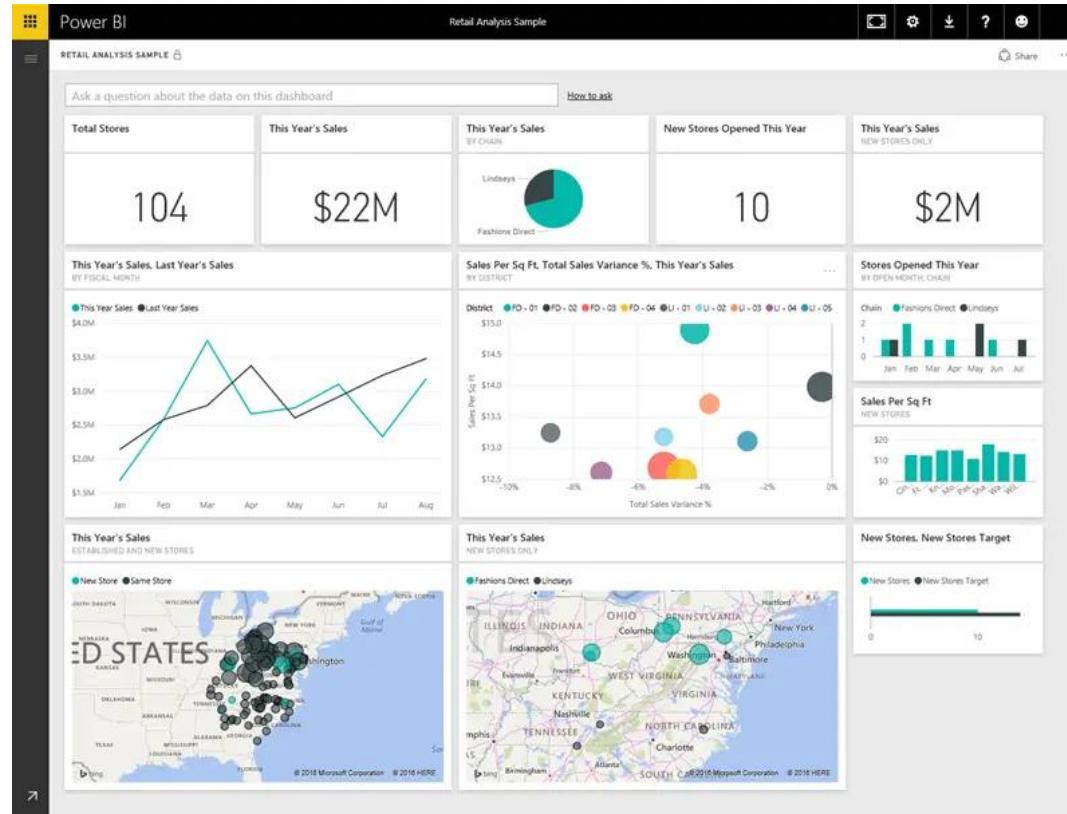
Mobile Apps

Building Blocks of Power BI

Building Blocks of Power BI

-  Visualizations
-  Datasets
-  Reports
-  Dashboards
-  Tiles

Visualizations



Datasets

- Collection of data that power BI uses to create it's visualization.
- Can also be combination of many sources
- Should filter data before bringing in to Power BI
- Data Connectors

C2132	B	C	D	E	F	G	H
1	Year	Month	Month Name	Calendar Month	Births	Births Per Day	Births (Normalized)
2119	2004	1	January	1/1/2004	2,937	94.7	2842
2120	2004	2	February	2/1/2004	2,824	97.4	2921
2121	2004	3	March	3/1/2004	3,128	100.9	3027
2122	2004	4	April	4/1/2004	2,896	96.5	2896
2123	2004	5	May	5/1/2004	3,008	97.0	2911
2124	2004	6	June	6/1/2004	3,047	101.6	3047
2125	2004	7	July	7/1/2004	2,981	96.2	2885
2126	2004	8	August	8/1/2004	3,079	99.3	2980
2127	2004	9	September	9/1/2004	3,219	107.3	3219
2128	2004	10	October	10/1/2004	3,547	114.4	3433
2129	2004	11	November	11/1/2004	3,365	112.2	3365
2130	2004	12	December	12/1/2004	3,143	101.4	3042
2131	2005	1	January	1/1/2005	2,921	94.2	2827
2132	2005	2	February	2/1/2005	2,699	96.4	2892
2133	2005	3	March	3/1/2005	3,024	97.5	2926
2134	2005	4	April	4/1/2005	3,037	101.2	3037
2135	2005	5	May	5/1/2005	3,231	104.2	3127
2136	2005	6	June	6/1/2005	3,163	105.4	3163
2137	2005	7	July	7/1/2005	3,119	100.6	3018
2138	2005	8	August	8/1/2005	3,156	101.8	3054
2139	2005	9	September	9/1/2005	3,439	114.6	3439

Reports

Human Resources Sample PBIX - Power BI Desktop

File Home View Modeling Help

Cut Copy Format Painter Get Data Recent Sources Enter Data Refresh Edit Queries New Page New Visual Ask A Question Insert Text box Shapes From Marketplace From File Switch Theme Manage Relationships New Measure New Column New Quick Measure Publish Clipboard External data Themes Relationships Calculations Share

New Hires

New Hires and Actives by Region and Ethnicity

New Hires by Month and FPDesc

New Hires by Gender

obviEnce

Filters on this page

- AgeGroup is (All)
- Ethnicity is (All)
- FPDesc is (All)
- Gender is (All)
- Month is not Dec
- Region is (All)
- VP is (All)
- Year is 2014

Add data fields here

Values

Drillthrough

Cross-report

Off

Keep all filters

On

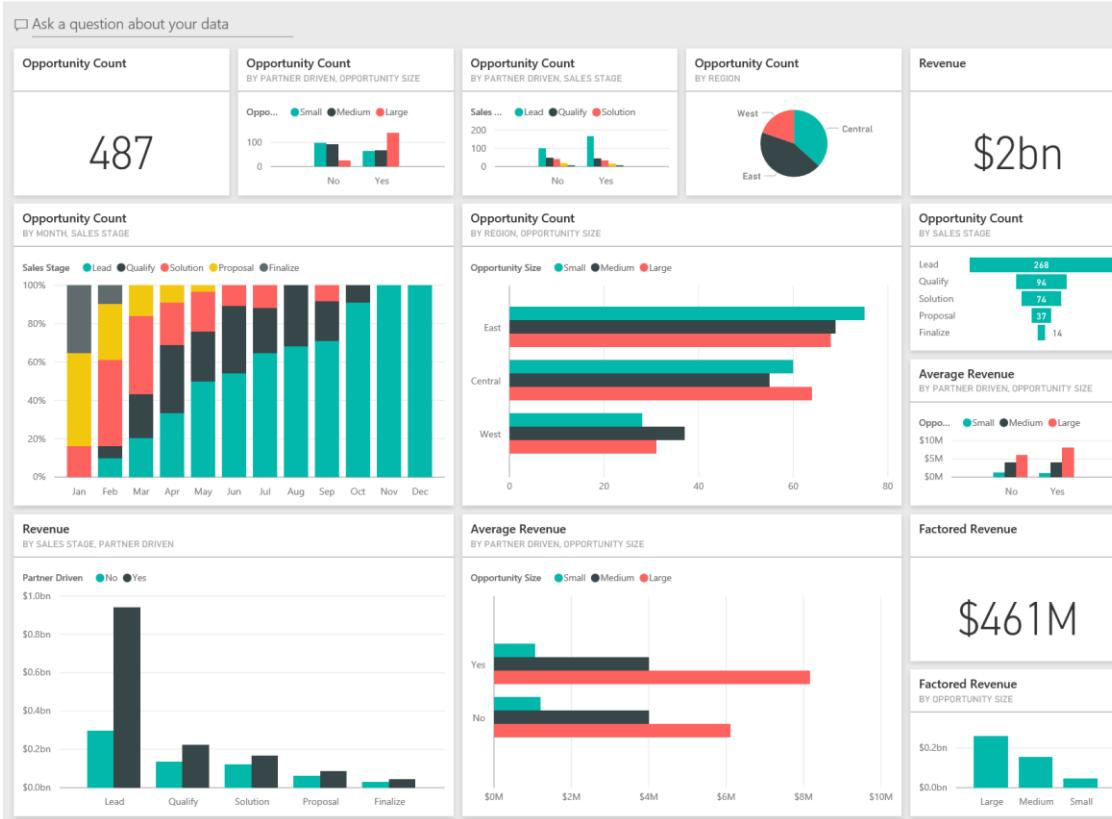
Add drillthrough fields here

Filters on all pages

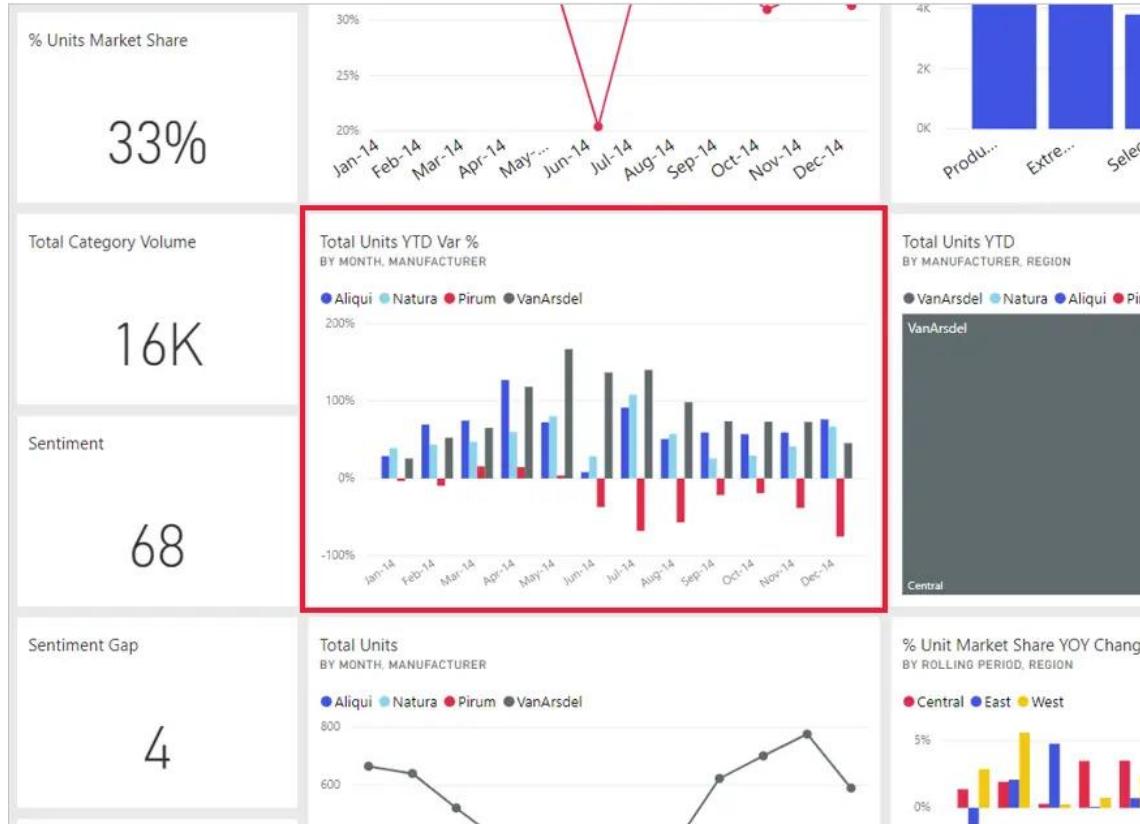
PAGE 2 OF 5

Dashboard

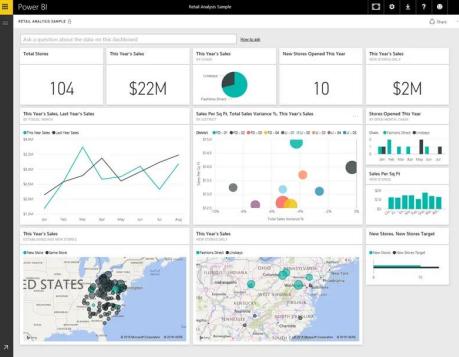
Opportunity Analysis Dashboard



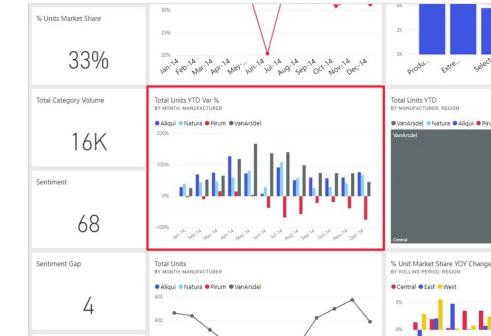
Tiles



Building Blocks of Power BI



B	C	D	E	F	G	H
Year	Month	Month Name	Calendar Month	Births	Births Per Day	Births (Normalized)
2119	2004	1 January	1/1/2004	2,937	94.7	2842
2120	2004	2 February	2/1/2004	2,824	97.4	2921
2121	2004	3 March	3/1/2004	3,128	100.9	3027
2122	2004	4 April	4/1/2004	2,896	96.5	2896
2123	2004	5 May	5/1/2004	3,008	97.0	2911
2124	2004	6 June	6/1/2004	3,047	101.6	3047
2125	2004	7 July	7/1/2004	2,981	96.2	2885
2126	2004	8 August	8/1/2004	3,079	99.3	2980
2127	2004	9 September	9/1/2004	3,219	107.3	3219
2128	2004	10 October	10/1/2004	3,547	114.4	3433
2129	2004	11 November	11/1/2004	3,365	112.2	3365
2130	2004	12 December	12/1/2004	3,143	101.4	3042
2131	2005	1 January	1/1/2005	2,921	94.2	2827
2132	2005	2 February	2/1/2005	2,699	96.4	2892
2133	2005	3 March	3/1/2005	3,024	97.5	2926
2134	2005	4 April	4/1/2005	3,037	101.2	3037
2135	2005	5 May	5/1/2005	3,231	104.2	3127
2136	2005	6 June	6/1/2005	3,163	105.4	3163
2137	2005	7 July	7/1/2005	3,119	100.6	3018
2138	2005	8 August	8/1/2005	3,156	101.8	3054
2139	2005	9 September	9/1/2005	3,439	114.6	3439





Paginated reports

Paginated reports

- Paginated reports are designed to be printed or shared
- They're called “paginated” because they're formatted to fit well on a page
- A major advantage of paginated reports is their ability to print all the data in a table, no matter how long
- Power BI report – Print only what you see on page
- Paginated report – Print many pages to include all the rows
- You can preview your report in **Report Builder**, then publish it to the **Power BI service**, app.powerbi.com

The screenshot shows the Power BI service interface with a paginated report titled "Buying Group Account Statement". The report details an invoice from "TailsSpin Toys" to "TailsSpin Toys (Head Office)" dated May 31, 2016. The report includes a header with the company logo and address, and a table of contents indicating "Invoice No: 73507-70431" and "Invoice 2 of 3". The main content area displays a table of items purchased, including "DBA joke mug - you might be a DBA if (White)", "Ogre battery-powered slippers (Green) XL", "Packing knife with metal insert blade (Yellow) 18mm", and "'The Gu' red shirt XML tag t-shirt (White) 3XS". The table also shows a "TOTAL" row with a total value of \$1,434.05.

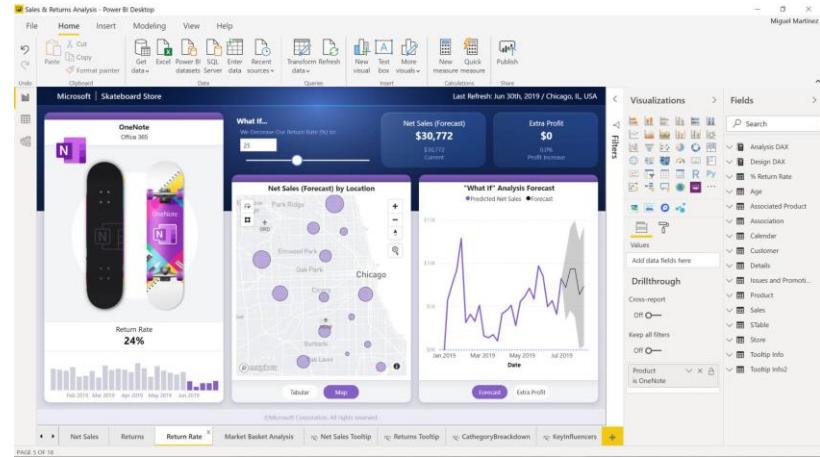
Item	Quantity	Unit Price	Tax	LineTotal
DBA joke mug - you might be a DBA if (White)	7	\$13.00	\$13.65	\$104.65
Ogre battery-powered slippers (Green) XL	8	\$32.00	\$38.40	\$294.40
Packing knife with metal insert blade (Yellow) 18mm	15	\$2.40	\$5.40	\$41.40
'The Gu' red shirt XML tag t-shirt (White) 3XS	48	\$18.00	\$129.60	\$993.60
			TOTAL	\$187.05
				\$1,434.05



Interactive reports

Interactive reports

- User can “Interact” with the report
- Designed to be viewed on screen
- Provides end users with the capability to do various manipulations to a report such as, drilling through various levels of data, filtering, and sorting.
- Make use of “hover”
- Report can change design/ layout if based on user action
- User Power BI Server
- Part of Power BI Premium

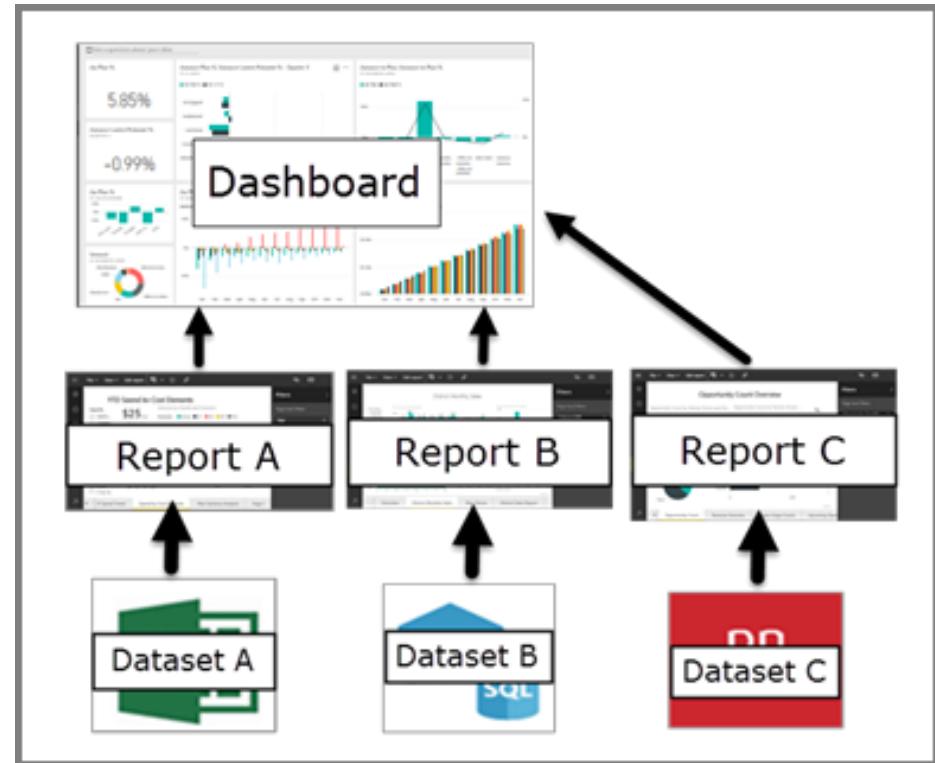


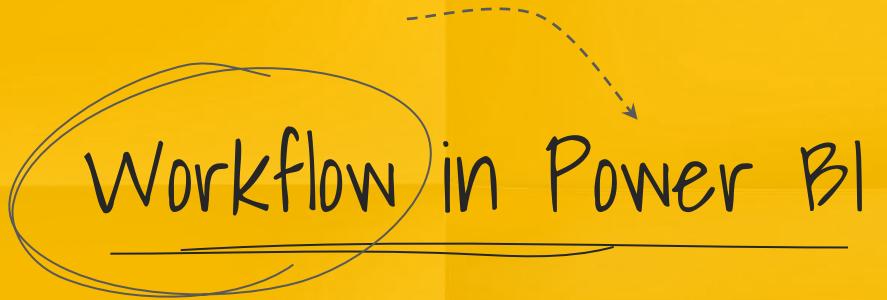


Dashboard

Dashboard

- Dashboards are a wonderful way to monitor your business, to look for answers, and to see all of your most-important metrics at a glance
- The visualizations on a dashboard come from reports and each report is based on one dataset
- Entryway into the underlying reports and datasets
- Dashboards can be viewed and shared on mobile devices
- Can share dashboards with other users or groups, who can then interact with your dashboards when they're in the Power BI service or on their mobile device.





Workflow in Power BI

Workflow in Power BI

