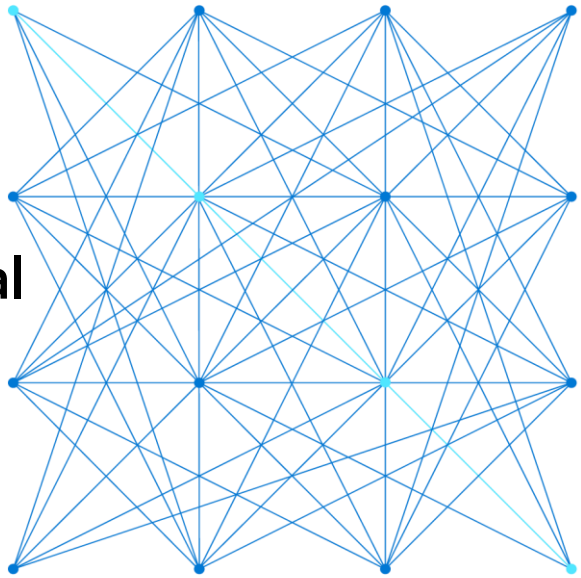
Azure Data Fundamental

Microsoft Azure



1

Tissana Tanaklang

Software and Solution Development Trainer
Iverson Training Center Co., Ltd.
tissana_t@hotmail.com

- Master of Science Program in Software Engineering King Mongkut's University of Technology Thonburi
- Bachelor of Science Program in Computer Science Naresuan University

- Microsoft Certified Trainer (MCT)
- Microsoft Certified Solutions Associate (MCSA) - Web Application Development
- Microsoft Certified Azure Fundamentals

2

Course Agenda

- 01 – Explore Core Data Concepts
- 02 – Explore Relational Data in Azure
- 03 – Explore Non-Relational Data in Azure
- 04 – Azure Machine Learning : No-Code with Designer

Microsoft

3

Azure Learning Path

Level	Category	Code	Course	Role
Beginner (Fundamentals)	-	AZ-900	Microsoft Azure Fundamentals	IT Professional and Non-IT Professional (All)
	Data	DP-900	Microsoft Azure Data Fundamentals	Data Engineer, Database Administrator
	AI	AI-900	Microsoft Azure AI Fundamentals	AI Engineer, Data Scientist, Developer, Solutions Architect
Intermediate (Associate)	DevOps	AZ-104	Microsoft Azure Administrator	Administrator, DevOps Engineer
		AZ-204	Developing solutions for Microsoft Azure	Developer, DevOps Engineer
	Security	AZ-500	Microsoft Azure Security Technologies	Security Engineer
	Data	DP-300	Administering Relational Databases on Microsoft Azure	Database Administrator
		DP-200	Implementing an Azure Data Solution	Data Engineer
		DP-201	Designing an Azure Data Solution	
		DP-100	Designing and Implementing a Data Science Solution on Azure	Data Scientist
	AI	AI-100	Designing and Implementing an Azure AI Solution	AI Engineer
Advance (Expert)	DevOps	AZ-400	Designing and Implementing Microsoft DevOps solutions	DevOps Engineer
	Solutions Architect	AZ-303	Microsoft Azure Architect Technologies	Solutions Architect
		AZ-304	Microsoft Azure Architect Design	
Specialty	Data	DA-100	Analyzing Data with Power BI	Data Analyst
	-	AZ-220	Microsoft Azure IoT Developer	Developer

Microsoft

4

Explore Core Data Concepts

5

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

What is data?

6

	Schema	Data relationships	Examples
Structured data	Adheres to a schema, with the same data fields or properties.	Storable in relational database tables, with rows and columns.	Sensor data and financial data.
Semi-structured data	Has an ad hoc schema with less organized fields and properties.	Non-relational or NoSQL data, not storable in tables, rows and column.	Books, blogs, JSON, HTML documents.
Unstructured data	Has no designated schema or data structure.	Non-relational or blob data, with no restrictions on the kinds of data blobs contain.	PDFs, JPGs, videos.

Azure Data Categories

7

- **Read-only** access means the users can read data but can't modify any existing data or create new data.
- **Read/write** access gives users the ability to view and modify existing data.
- **Owner** privilege gives full access to the data including managing the security like adding new users and removing access to existing users.

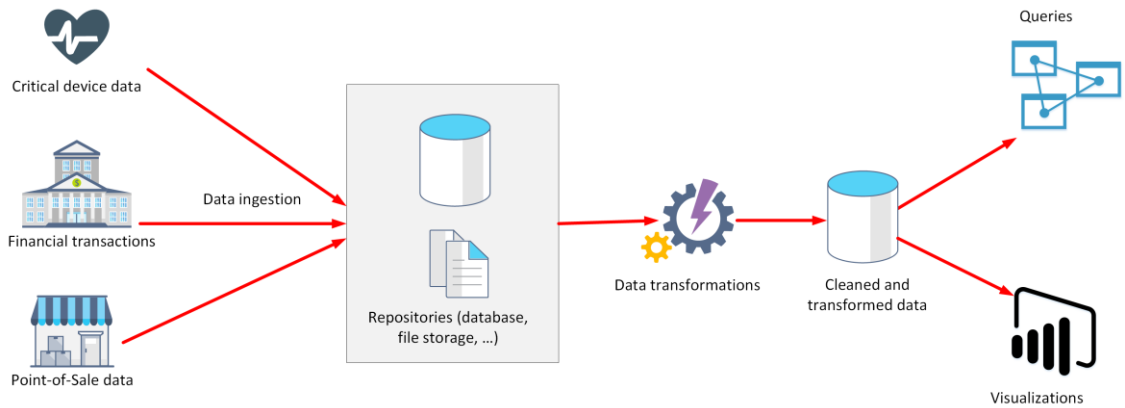
How is data defined and stored in cloud computing?

8

- A transactional system is often what most people consider the primary function of business computing.
- A transactional system records transactions.
- A transaction could be financial, such as the movement of money between accounts in a banking system, or it might be part of a retail system, tracking payments for goods and services from customers.

What is a Transactional System?

9



What is an Analytical System?

10

CustomerID	FirstName	LastName
1	Jay	Adams
2	Donna	Carreras
3	Linda	Burnett
4	Frances	Adams

CustomerID	AddressID
1	A
2	B
3	C
4	A

AddressID	LineNumber	Text
A	1	12
A	2	Park Street
A	3	Some City
B	1	The Big House
B	2	High Road
B	3	Another City
B	4	90210
C	1	Freepost
C	2	AAA 123

Characteristics of Relational Data

11

```
## Document for Jay Adams ##
{
  "customerID": "1",
  "name":
  {
    "firstname": "Jay",
    "lastname": "Adams"
  },
  "address":
  {
    "number": "12",
    "street": "Park Street",
    "city": "Some City",
  }
}
```

```
## Document for Frances Adams ##
{
  "customerID": "4",
  "name":
  {
    "firstname": "Francis",
    "lastname": "Adams"
  },
  "address":
  {
    "number": "12",
    "street": "Park Street",
    "city": "Some City",
  }
}
```

Characteristics of Non-Relational

12

Azure Database Administrator role

An Azure database administrator is responsible for the design, implementation, maintenance, and operational aspects of on-premises and cloud-based database solutions built on Azure data services and SQL Server. They are responsible for the overall availability and consistent performance and optimizations of the database solutions. They work with stakeholders to implement policies, tools, and processes for backup and recovery plans to recover following a natural disaster or human-made error.

The database administrator is also responsible for managing the security of the data in the database, granting privileges over the data, granting or denying access to users as appropriate.

What are the roles in the world of data?

13

Data Engineer role

A data engineer collaborates with stakeholders to design and implement data-related assets that include data ingestion pipelines, cleansing and transformation activities, and data stores for analytical workloads. They use a wide range of data platform technologies, including relational and nonrelational databases, file stores, and data streams.

They are also responsible for ensuring that the privacy of data is maintained within the cloud and spanning from on-premises to the cloud data stores. They also own the management and monitoring of data stores and data pipelines to ensure that data loads perform as expected.

What are the roles in the world of data?

14

Data Analyst role

A data analyst enables businesses to maximize the value of their data assets. They are responsible for designing and building scalable models, cleaning and transforming data, and enabling advanced analytics capabilities through reports and visualizations.

A data analyst processes raw data into relevant insights based on identified business requirements to deliver relevant insights.

What are the roles in the world of data?

15

- Installing and upgrading the database server and application tools.
- Allocating system storage and planning storage requirements for the database system.
- Modifying the database structure, as necessary, from information given by application developers.
- Enrolling users and maintaining system security.

Database Administrator tasks and responsibilities

16

- Ensuring compliance with database vendor license agreement.
- Controlling and monitoring user access to the database.
- Monitoring and optimizing the performance of the database.
- Planning for backup and recovery of database information.
- Maintaining archived data.

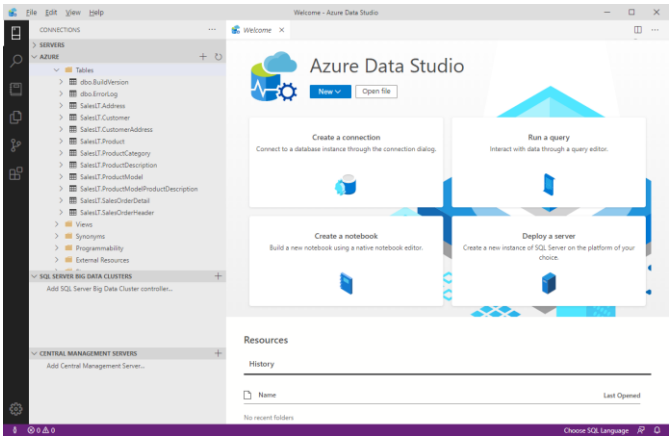
Database Administrator tasks and responsibilities

17

- Backing up and restoring databases.
- Contacting database vendor for technical support.
- Generating various reports by querying from database as per need.
- Managing and monitoring data replication.
- Acting as liaison with users.

Database Administrator tasks and responsibilities

18



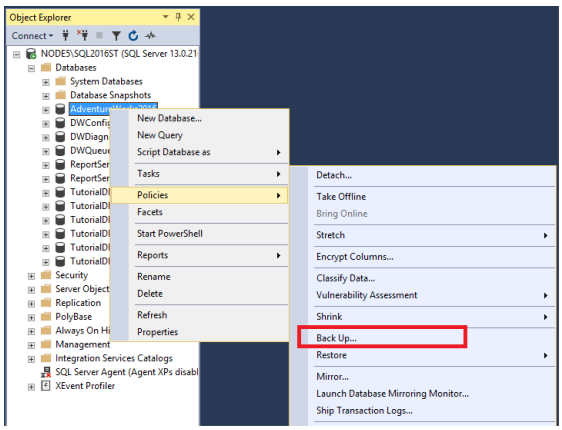
Azure Data Studio provides a graphical user interface for managing many different database systems.

It currently provides connections to on-premises SQL Server databases, Azure SQL Database, PostgreSQL, Azure SQL Data Warehouse, and SQL Server Big Data Clusters, amongst others.

It's an extensible tool, and you can download and install extensions from third-party developers that connect to other systems, or provide wizards that help to automate many administrative tasks.

Common Database Administrator Tools

19

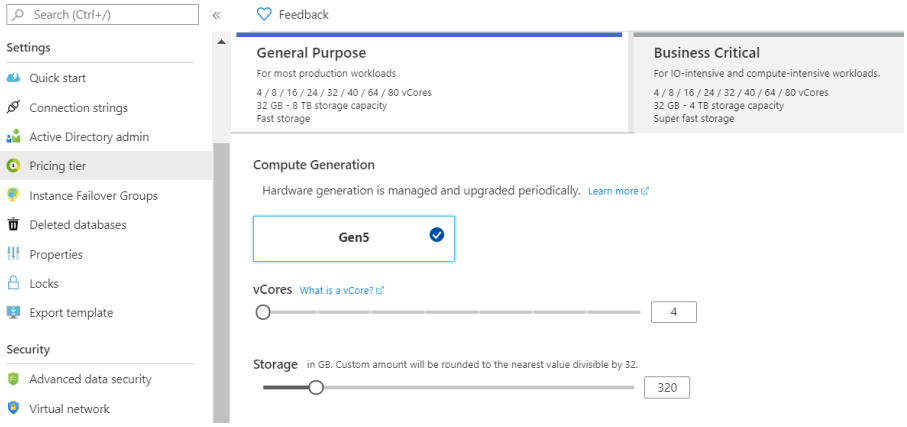


SQL Server Management Studio provides a graphical interface, enabling you to query data, perform general database administration tasks, and generate scripts for automating database maintenance and support operations.

The example below shows SQL Server Management Studio being used to back up a database.

Common Database Administrator Tools

20



Azure SQL database provides database services in Azure.

It's similar to SQL Server, except that it runs in the cloud. You can manage Azure SQL database using Azure portal.

Use the Azure portal to manage Azure SQL Database

21

- Developing, constructing, testing, and maintaining databases and data structures.
- Aligning the data architecture with business requirements.
- Data acquisition.
- Developing processes for creating and retrieving information from data sets.
- Using programming languages and tools to examine the data.

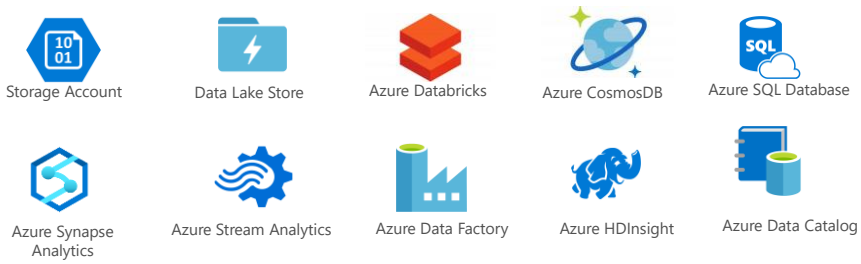
Data Engineer tasks and responsibilities

22

- Identifying ways to improve data reliability, efficiency, and quality.
- Conducting research for industry and business questions.
- Deploying sophisticated analytics programs, machine learning, and statistical methods.
- Preparing data for predictive and prescriptive modeling.
- Using data to discover tasks that can be automated.

Data Engineer tasks and responsibilities

23



Common Data Engineer Tools

24

- Making large or complex data more accessible, understandable, and usable.
- Creating charts and graphs, histograms, geographical maps, and other visual models that help to explain the meaning of large volumes of data, and isolate areas of interest.
- Transforming, improving, and integrating data from many sources, depending on the business requirements.

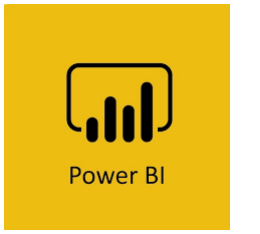
Data Analyst tasks and responsibilities

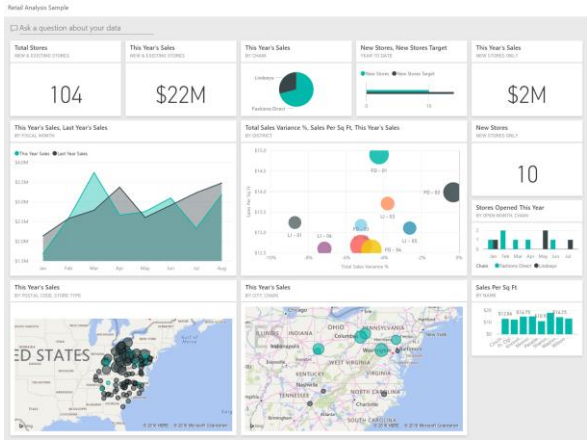
25

- Combining the data result sets across multiple sources. For example, combining sales data and weather data provides a useful insight into how weather influenced sales of certain products such as ice creams.
- Finding hidden patterns using data.
- Delivering information in a useful and appealing way to users by creating rich graphical dashboards and reports.

Data Analyst tasks and responsibilities

26





Common data visualization tools

27

Explore Relational Data in Azure

28

Customers

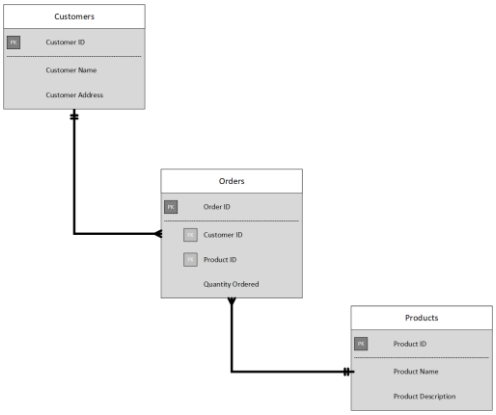
Customer ID	Customer Name	Customer Address
C1	Fred	…
C2	Bert	…
C3	Jane	…

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



Understand the characteristics of relational data

29

- All data is tabular. Entities are modeled as tables, each instance of an entity is a row in the table, and each property is defined as a column.
- All rows in the same table have the same set of columns.
- A table can contain any number of rows.

Characteristics of a Relational Database

30

- A primary key uniquely identifies each row in a table. No two rows can share the same primary key.
- A foreign key references rows in another, related table. For each value in the foreign key column, there should be a row with the same value in the corresponding primary key column in the other table.

Characteristics of a Relational Database

31

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

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

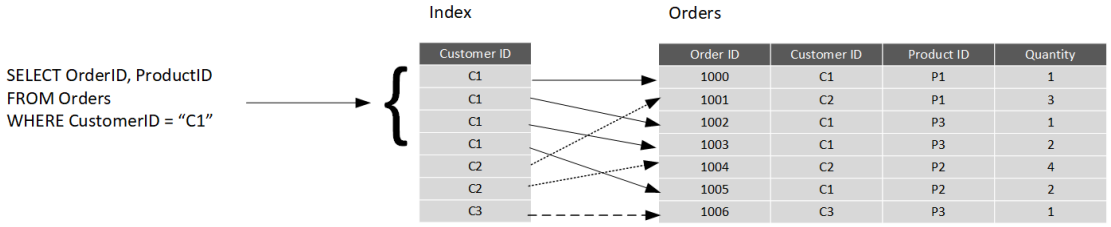
Query - SQL

32

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

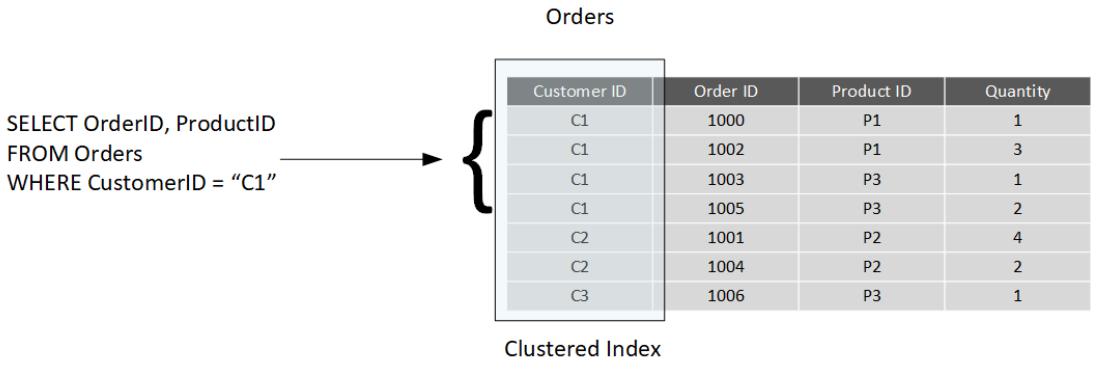
Query - SQL

33



What is an index?

34



Clustered indexes.

35

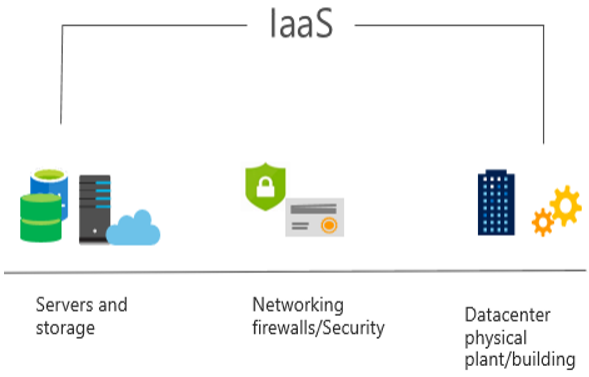
A view is a virtual table based on the result set of a query. In the simplest case, you can think of a view as a window on specified rows in an underlying table.

```
CREATE VIEW P1Orders AS
SELECT CustomerID, OrderID, Quantity
FROM Orders
WHERE ProductID = "P1"
```

What is a view?

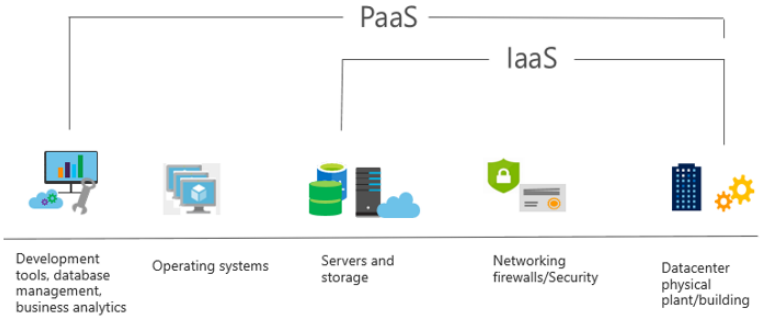
36

Build pay-as-you-go IT infrastructure by renting servers, virtual machines, storage, networks, and operating systems from a cloud provider.



Infrastructure as a Service (IaaS)

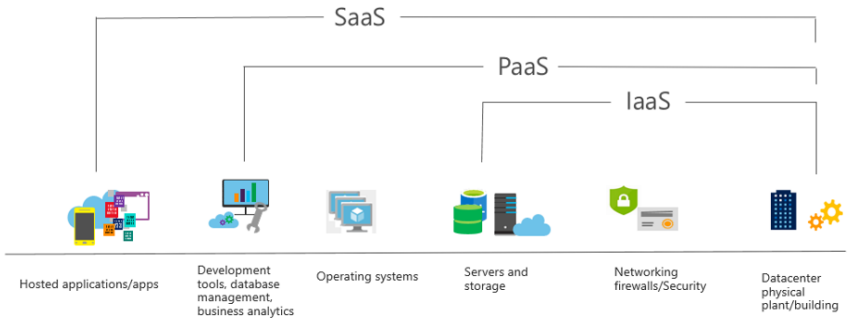
37



Provides environment for building, testing, and deploying software applications; without focusing on managing underlying infrastructure.

Platform as a Service (PaaS)

38



Users connect to and use cloud-based apps over the internet: for example, Microsoft Office 365, email, and calendars.

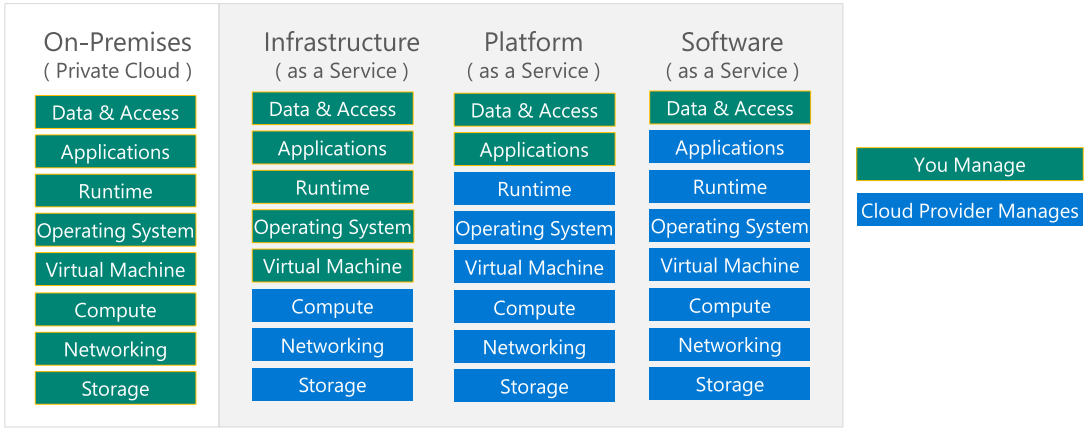
Software as a Service (SaaS)

39

IaaS	PaaS	SaaS
• The most flexible cloud service. • You configure and manage the hardware for your application.	• Focus on application development. • Platform management is handled by the cloud provider.	• Pay-as-you-go pricing model. • Users pay for the software they use on a subscription model.

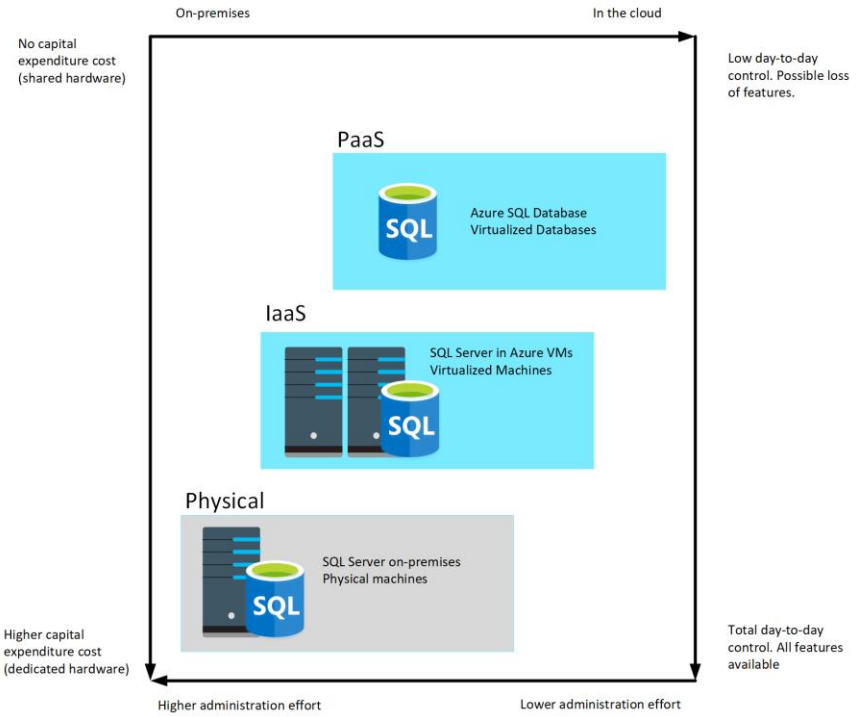
Compare cloud services

40



Shared responsibility model

41

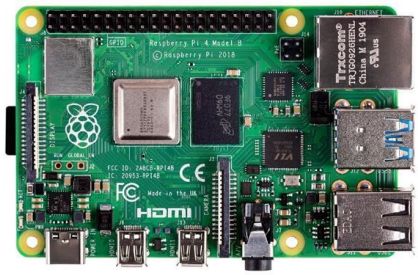


42

Explore concepts of Non-Relational Data

43

- IoT and telematics.
- Retail and marketing.
- Web and mobile applications.
- Gaming



Non-Relational Database use case

44

```
{"latitude":37.8267,"longitude":-122.4233,"timezone":"America/Los_Angeles","currently":{"time":1598191217,"summary":"Partly Cloudy","icon":"partly-
cloudy-
day","nearestStormDistance":5,"nearestStormBearing":58,"precipIntensity":0,"precipProbability":0,"temperature":58.63,"apparentTemperature":58.63,"dew
Point":52.42,"humidity":0.8,"pressure":1011.8,"windSpeed":5.08,"windGust":7.73,"windBearing":210,"cloudCover":0.54,"uvIndex":0,"visibility":9.933,"oz
one":291.2},"minutely":{"summary":"Partly cloudy for the hour.","icon":"partly-cloudy-day","data":
[{"time":1598191200,"precipIntensity":0,"precipProbability":0},{"time":1598191260,"precipIntensity":0,"precipProbability":0},
{"time":1598191320,"precipIntensity":0,"precipProbability":0},{"time":1598191380,"precipIntensity":0,"precipProbability":0},
{"time":1598191440,"precipIntensity":0,"precipProbability":0},{"time":1598191500,"precipIntensity":0,"precipProbability":0},
{"time":1598191560,"precipIntensity":0,"precipProbability":0},{"time":1598191620,"precipIntensity":0,"precipProbability":0},
{"time":1598191680,"precipIntensity":0,"precipProbability":0},{"time":1598191740,"precipIntensity":0,"precipProbability":0},
{"time":1598191800,"precipIntensity":0,"precipProbability":0},{"time":1598191860,"precipIntensity":0,"precipProbability":0},
{"time":1598191920,"precipIntensity":0,"precipProbability":0},{"time":1598191980,"precipIntensity":0,"precipProbability":0},
{"time":1598192040,"precipIntensity":0,"precipProbability":0},{"time":1598192100,"precipIntensity":0,"precipProbability":0},
{"time":1598192160,"precipIntensity":0,"precipProbability":0},{"time":1598192220,"precipIntensity":0,"precipProbability":0},
{"time":1598192280,"precipIntensity":0,"precipProbability":0},{"time":1598192340,"precipIntensity":0,"precipProbability":0},
{"time":1598192400,"precipIntensity":0.0026,"precipIntensityError":0.0004,"precipProbability":0.01,"precipType":"rain"},
{"time":1598192460,"precipIntensity":0.0026,"precipIntensityError":0.0005,"precipProbability":0.01,"precipType":"rain"},
{"time":1598192520,"precipIntensity":0,"precipProbability":0},
{"time":1598192580,"precipIntensity":0.0026,"precipIntensityError":0.0005,"precipProbability":0.01,"precipType":"rain"},
{"time":1598192640,"precipIntensity":0.0026,"precipIntensityError":0.0005,"precipProbability":0.01,"precipType":"rain"},
{"time":1598192700,"precipIntensity":0.0027,"precipIntensityError":0.0005,"precipProbability":0.01,"precipType":"rain"},
{"time":1598192760,"precipIntensity":0.0027,"precipIntensityError":0.0005,"precipProbability":0.02,"precipType":"rain"},
{"time":1598192820,"precipIntensity":0.0026,"precipIntensityError":0.0005,"precipProbability":0.02,"precipType":"rain"},
{"time":1598192880,"precipIntensity":0,"precipProbability":0},
```


Non-Relational Database use case

https://api.darksky.net/

45

Open API :) สำหรับนักพัฒนา

แสดงค่าประจำวัน :
//covid19.th-stat.com/api/open/today

ข้อมูลสรุปตามช่วงเวลา (เริ่มตั้งแต่วันที่ 01/01/20) :
//covid19.th-stat.com/api/open/timeline

ข้อมูลแต่ละเคส :
//covid19.th-stat.com/api/open/cases

ข้อมูลสรุปจากเคส :
//covid19.th-stat.com/api/open/cases/sum

แจ้งเตือนพื้นที่ตามคำประกาศ :
//covid19.th-stat.com/api/open/area

กรมควบคุมโรค
DEPARTMENT OF DISEASE CONTROL

Non-Relational Database use case

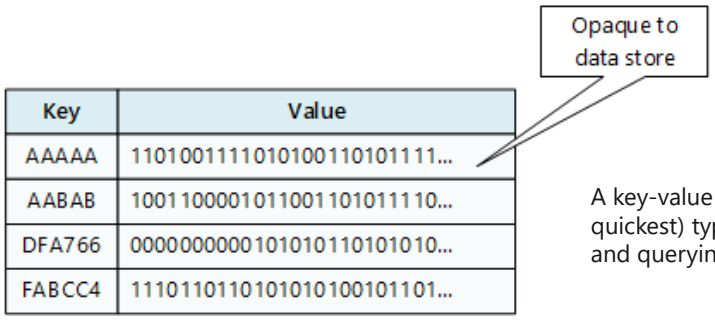
46

- You might see the term *NoSQL* when reading about non-relational databases.
- NoSQL is a rather loose term that simply means non-relational.
- NoSQL (non-relational) databases generally fall into four categories:

  - key-value stores
  - document databases
  - column family databases
  - graph databases.

What is NoSQL?

47



A key-value store is the simplest (and often quickest) type of NoSQL database for inserting and querying data.

Key-Value Stores

48

Key	Document
1001	{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }
1002	{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }

A document database represents the opposite end of the NoSQL spectrum from a key-value store. In a document database, each document has a unique ID, but the fields in the documents are transparent to the database management system. Document databases typically store data in JSON format,

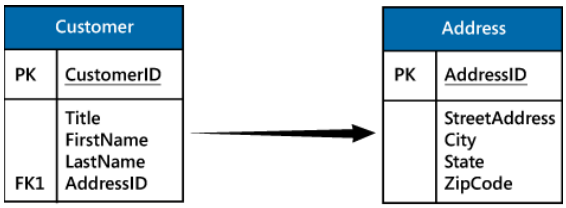
Document Databases

49

RDBMS	MongoDB
Database	Database
Table	Collection
Tuple/Row	Document
column	Field
Table Join	Embedded Documents
Primary Key	Primary Key (Default key _id provided by mongodb itself)

Document Databases

50



Customer Table

CustomerID	Title	FirstName	LastName	AddressID
1	Mr	Mark	Hanson	500
2	Ms	Lisa	Andrews	501
3	Mr	Walter	Harp	500

Address Table

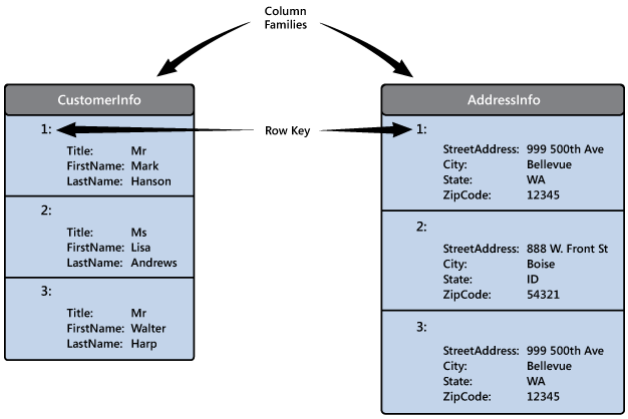
AddressID	StreetAddress	City	State	ZipCode
500	999 500th Ave	Bellevue	WA	12345
501	888 W. Front St	Boise	ID	54321

RDBMS is Row-based oriented

Column Family Databases

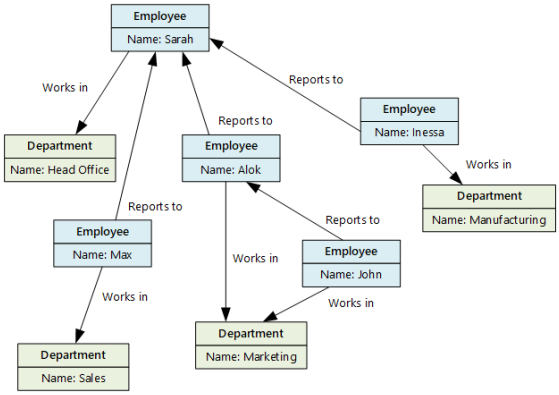
51

Row Key	Column Families			
CustomerID	CustomerInfo		AddressInfo	
1	CustomerInfo:Title CustomerInfo:FirstName CustomerInfo:LastName	Mr Mark Hanson	AddressInfo:StreetAddress AddressInfo:City AddressInfo:State AddressInfo:ZipCode	999 500th Ave Bellevue WA 12345
2	CustomerInfo:Title CustomerInfo:FirstName CustomerInfo:LastName	Ms Lisa Andrews	AddressInfo:StreetAddress AddressInfo:City AddressInfo:State AddressInfo:ZipCode	888 W. Front St Boise ID 54321
3	CustomerInfo:Title CustomerInfo:FirstName CustomerInfo:LastName	Mr Walter Harp	AddressInfo:StreetAddress AddressInfo:City AddressInfo:State AddressInfo:ZipCode	999 500th Ave Bellevue WA 12345



Column Family Databases

52



Graph databases enable you to store entities, but the main focus is on the relationships that these entities have with each other.

A graph database stores two types of information: nodes that you can think of as instances of entities, and edges, which specify the relationships between nodes.

Graph Databases

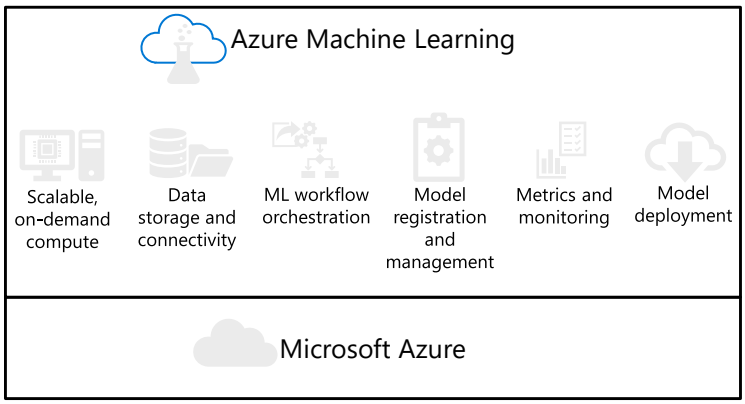
53

Azure Machine Learning : No-Code with Designer

54

What is Azure Machine Learning?

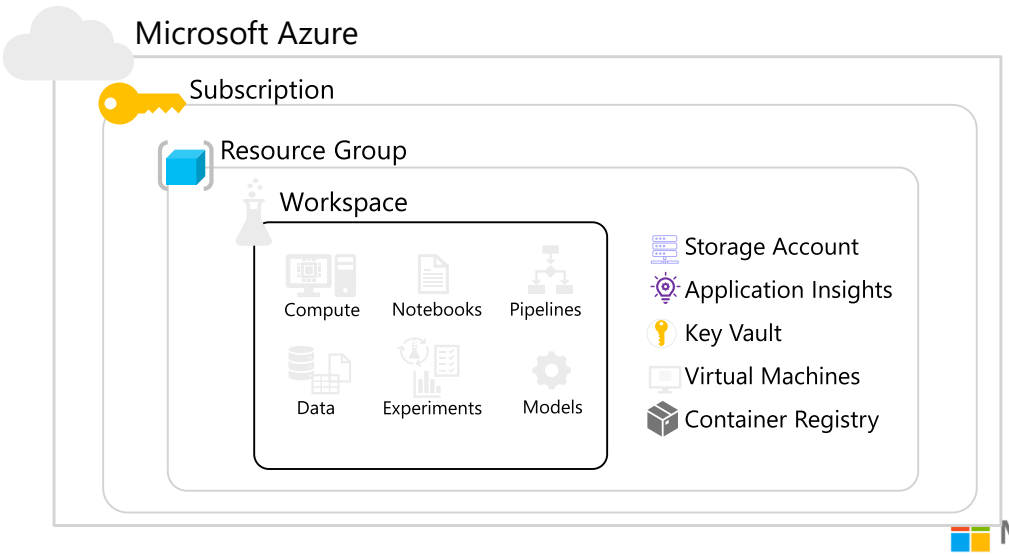
A platform for operating machine learning workloads in the cloud



Microsoft

55

Azure Machine Learning Workspaces



Microsoft

56

Considerations for Creating a Workspace

Region

Check Azure Resource availability

For example, NC-Series Virtual Machines for GPU processing

Edition

Enterprise

- All features

Basic

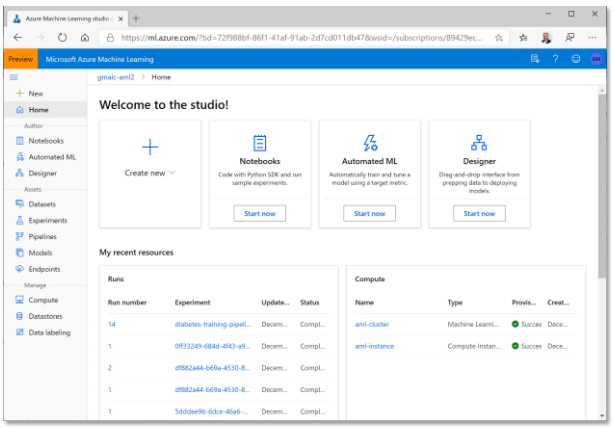
- No Visual Designer
- No Automated ML user interface
- No Data Drift user interface

Microsoft

57

Azure Machine Learning studio

Manage compute and data
Run experiments
View metrics
Manage and deploy models
Manage endpoints
Use graphical modeling tools:
Designer - "no-code" model development
Automated Machine Learning - find the best model for your data



Microsoft

58

The Azure Machine Learning SDK for Python

Code-based configuration for machine learning assets:
Automate repeatable asset creation
Ensure consistency across development, test, and production environments
Incorporate machine learning asset configuration into DevOps

```
pip install azureml-sdk
```

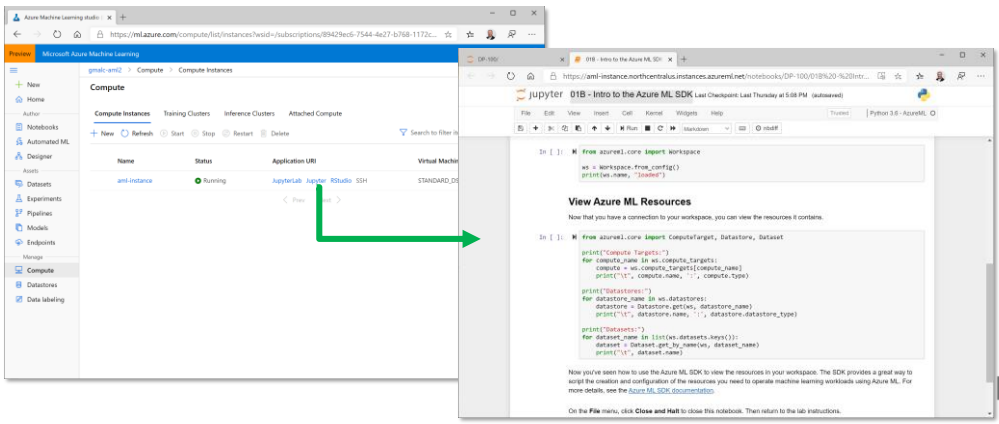
```
from azureml.core import Workspace

ws = Workspace.from_config()
for compute_name in ws.compute_targets:
    compute = ws.compute_targets[compute_name]
    print(compute.name, ":", compute.type)
```

Microsoft

59

Compute Instances

Jupyter Notebook and JupyterLab servers in your workspace
Choose the compute specifications you need



Microsoft

60

Walkthrough :

Creating an Azure Machine Learning Workspace

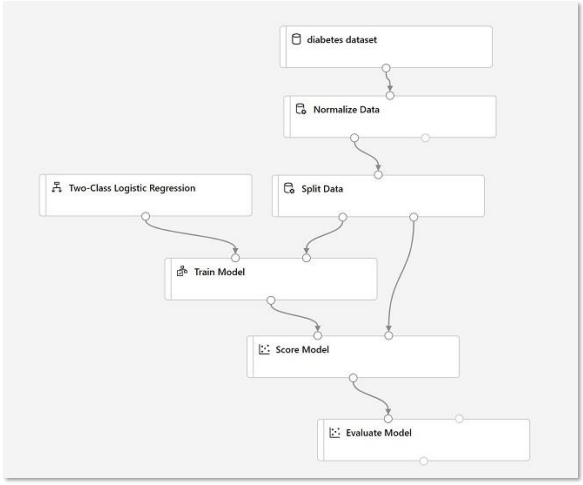
Microsoft

61

What is Azure Machine Learning Designer?

Drag-and-Drop Interface for:

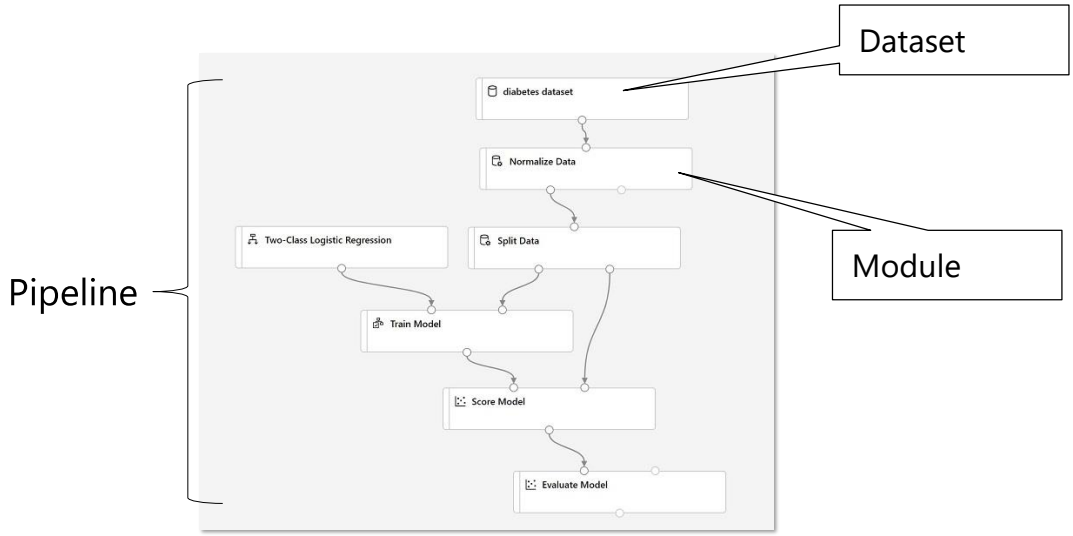
Preparing data and training models
Publishing models as services



Microsoft

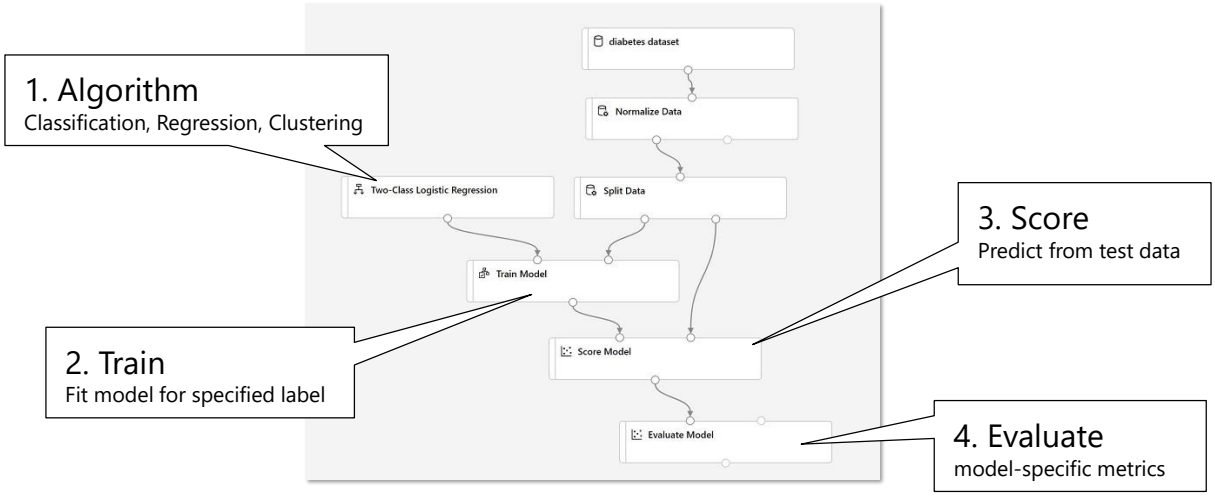
62

Designer Pipelines and Modules



63

Training, Scoring, and Evaluating Models



64

Custom Code Modules

Apply SQL Transformation	Use a SQL statement to process up to three input tables
Execute Python Script	Implement a custom Python function to process up to two dataframes
Create Python Model	Implement a custom Python model in place of a built-in algorithm
Execute R Script	Implement a custom R function to process up to two dataframes

Microsoft

65

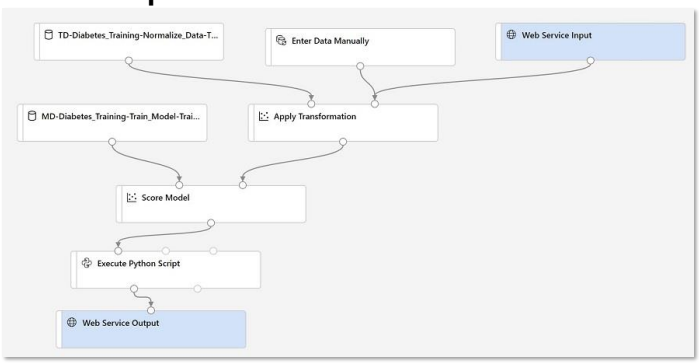
Walkthrough :

Creating a Training Pipeline with the Azure ML Designer

Microsoft

66

What is an Inference Pipeline?



A data flow defining a web service for using the trained model

A **Web Service Input** defines the input data schema
Transformations based on training data are encapsulated in datasets
The trained model is encapsulated in a dataset
A **Web Service Output** defines the output data schema
You may want to modify the pipeline before deploying its as a web service

Microsoft

67

Publishing a Service Endpoint

Deploy a Real-Time Pipeline:

Requires Azure Kubernetes Services Inference Compute
Submit new data to HTTP endpoint for immediate results

Publish a Batch Pipeline

Requires Azure Machine Learning Training Compute
Initiate pipeline experiment run through HTTP endpoint
Results saved in run output

Microsoft

68

Consuming a Service Endpoint

View endpoints in Azure Machine Learning studio
Use starter code to build client applications

```
data = {"Inputs": {"input0":[{'feature1': "123", 'feature2': "99",},],},
        "GlobalParameters":  {}}
body = str.encode(json.dumps(data))

url = 'http://10.0.0.1:80/api/v1/service/diabetes_predictor/score'
api_key = 'a1234567890x'
headers = {'Content-Type':'application/json',
           'Authorization':('Bearer '+ api_key)}

req = urllib.request.Request(url, body, headers)
response = urllib.request.urlopen(req)
result = response.read()
```

Microsoft

69

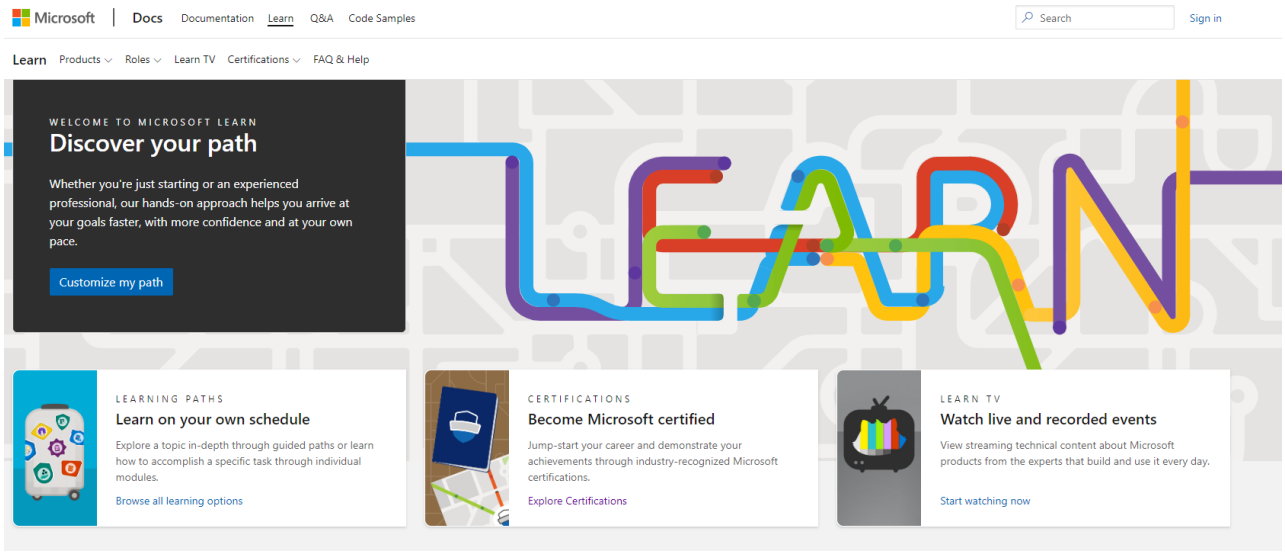
Walkthrough :

Deploying a Service with the Azure ML Designer

Microsoft

70

Reference : Microsoft Learn



71

Complete the Course

Microsoft

72