

 Microsoft

Microsoft Azure Data Solution



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

2

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
	Data	AZ-500	Microsoft Azure Security Technologies	Security Engineer
		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	Data Scientist
		DP-100	Designing and Implementing a Data Science Solution on Azure	Data Scientist
Advance (Expert)	AI	AI-100	Designing and Implementing an Azure AI Solution	AI Engineer
	Solutions Architect	AZ-400	Designing and Implementing Microsoft DevOps solutions	DevOps Engineer
		AZ-303	Microsoft Azure Architect Technologies	Solutions Architect
Specialty	Data	AZ-304	Microsoft Azure Architect Design	
		DA-100	Analyzing Data with Power BI	Data Analyst
	-	AZ-220	Microsoft Azure IoT Developer	Developer



3

Agenda

- About this course
- Course agenda
- Audience
- Prerequisites

4

About this course

In this course, the students will implement various data platform technologies into solutions that are in line with business and technical requirements including on-premises, cloud, and hybrid data scenarios incorporating both relational and No-SQL data. They will also learn how to process data using a range of technologies and languages for both streaming and batch data.

The students will also explore how to implement data security including authentication, authorization, data policies and standards. They will also define and implement data solution monitoring for both the data storage and data processing activities. Finally, they will manage and troubleshoot Azure data solutions which includes the optimization and disaster recovery of big data, batch processing and streaming data solutions.

5

Course Agenda

- Module 1
 - Azure for the Data Engineer
 - L01 - Explain the evolving world of data
 - L02 - Survey the services in the Azure Data Platform
 - L03 - Identify the tasks that are performed by a Data Engineer
 - L04 - Describe the use cases for the cloud in a case study
- Module 2
 - Working with Data Storage
 - L01 - Choose a data storage approach in Azure
 - L02 - Create an Azure Storage Account
 - L03 - Explain Azure Data Lake Storage
 - L04 - Upload data into Azure Data Lake

6

Course Agenda

- Module 3
- Working with Relational Data Stores in the Cloud
 - L01 - Explain SQL Database
 - L02 - Explain SQL Data Warehouse
 - L03 - Provision and load data in Azure SQL Data Warehouse

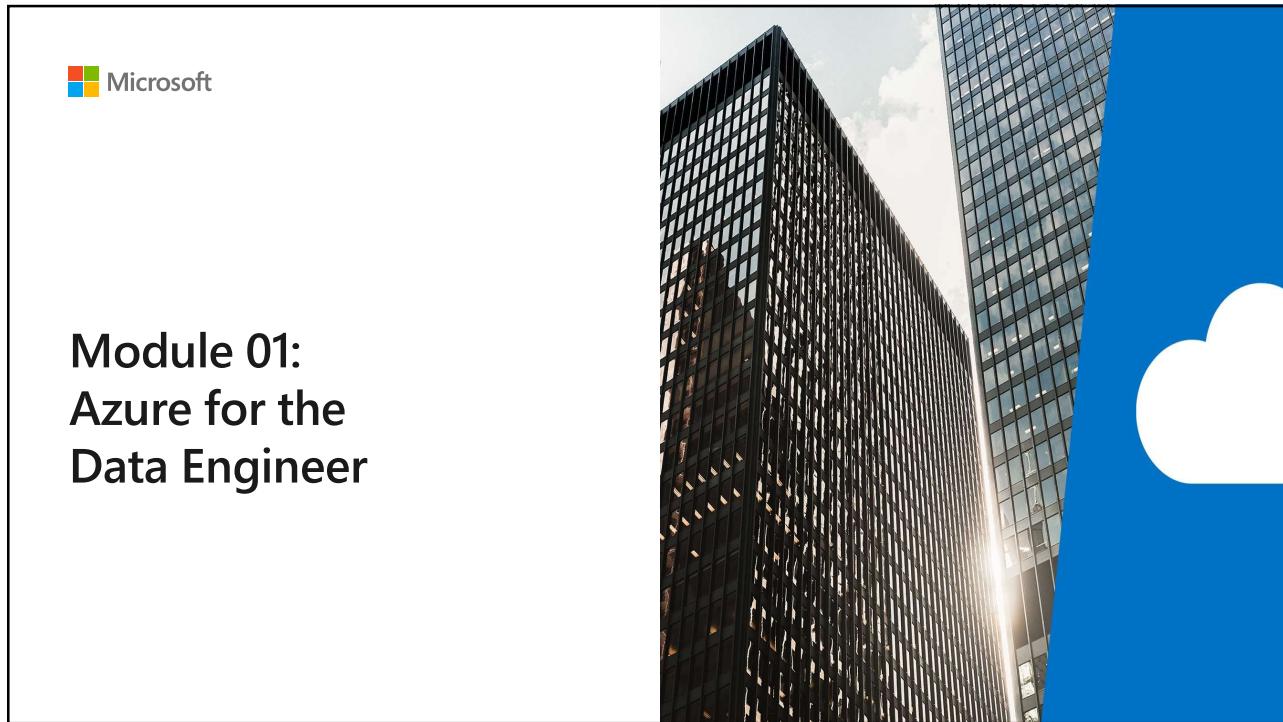
7

Prerequisites

In addition to their professional experience, students who take this training should have technical knowledge equivalent to the following courses:

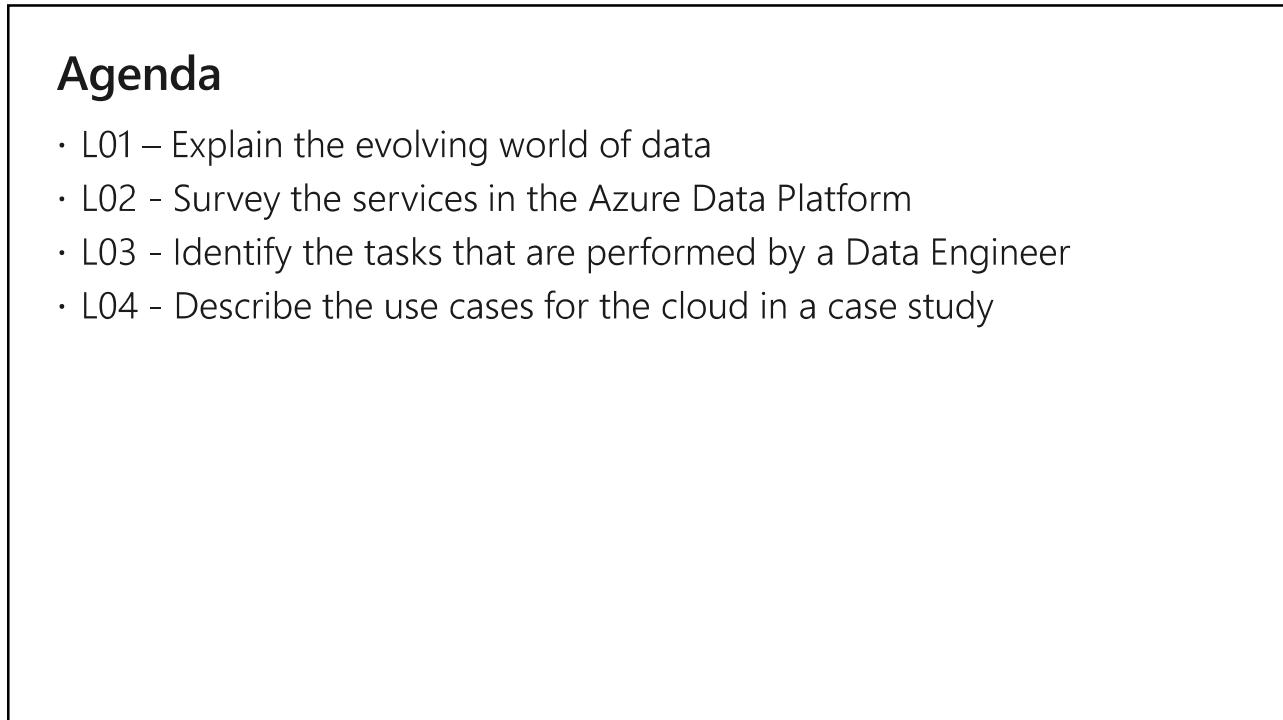
[Azure fundamentals](#)

8



Module 01: Azure for the Data Engineer

9



10



Lesson 01

The Evolving World of Data

11

Lesson Objectives

- Data abundance
- Differences between on-premises and cloud data technologies
- How the role of the data professional is changing in organizations
- Identify use cases impacted by these changes

12

Data abundance	Processes Businesses are tasked to store, interpret, manage, transform, process, aggregate and report on data
	Consumers There are a wider range of consumers using different types of devices to consume or generate data
	Variety There's a wider variety of data types that need to be processed and stored
	Responsibilities A data engineer's role is responsible for more data types and technologies
	Technologies Microsoft Azure provides a wide set of tools and technologies

13



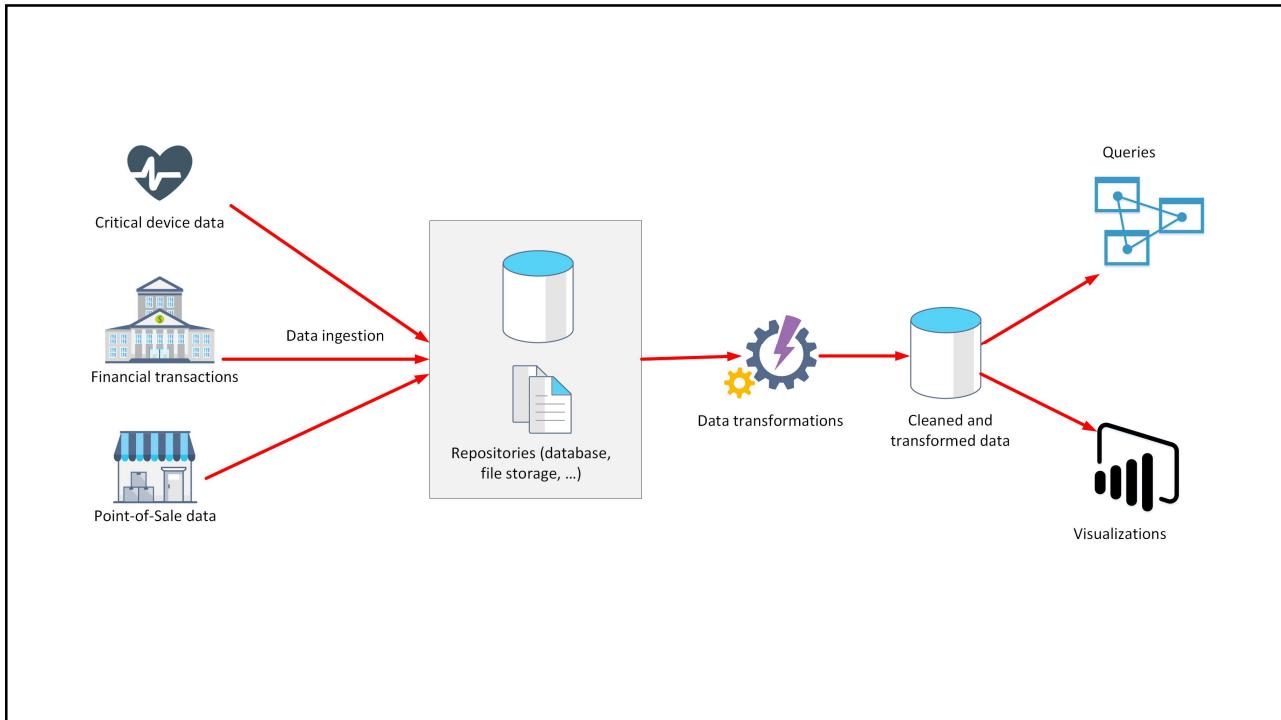
14



Data engineering job responsibilities



15



16

Use cases for the cloud

Here are some examples of industries making use of the cloud

Web retail

Using Azure Cosmos DB's multi-master replication model along with Microsoft's performance commitments, Data Engineers can implement a data architecture to support web and mobile applications that achieve less than a 10-ms response time anywhere in the world

Healthcare

Azure Databricks can be used to accelerate big data analytics and artificial intelligence (AI) solutions. Within the healthcare industry, it can be used to perform genome studies or pharmacy sales forecasting at petabyte scale

IoT scenarios

Hundreds of thousands of devices have been designed and sold to generate sensor data known as Internet of Things (IoT) devices. Using technologies like Azure IoT Hub, Data Engineers can easily design a data solution architecture that captures real-time data

17



Lesson 02

Survey the Services in the Azure Data Platform

18

Lesson Objectives

- The differences between structured and unstructured data
- Azure Storage
- Azure Data Lake Storage
- Azure Databricks
- Azure Cosmos DB
- Azure SQL Database
- Azure SQL Data Warehouse
- Azure Stream Analytics
- Additional Azure Data Platform Services

19

	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.

Data Categories

20

Structured versus unstructured data

There are three broad types of data and Microsoft Azure provides many data platform technologies to meet the needs of the wide varieties of data

Structured	Semi- Structured	Unstructured
<p>Structured data is data that adheres to a schema, so all of the data has the same fields or properties. Structured data can be stored in a database table with rows and columns</p>	<p>Semi-structured data doesn't fit neatly into tables, rows, and columns. Instead, semi-structured data uses _tags_ or _keys_ that organize and provide a hierarchy for the data</p>	<p>Unstructured data encompasses data that has no designated structure to it. Known as No-SQL, there are four types of No-SQL databases:</p> <ul style="list-style-type: none"> • Key Value Store • Document Database • Graph Databases • Column Base

21

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



Non-Relational Database use case

22

```
{"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,"dewPoint":52.42,"humidity":0.8,"pressure":1011.8,"windSpeed":5.08,"windGust":7.73,"windBearing":210,"cloudCover":0.54,"uvIndex":0,"visibility":9.93,"ozOne":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}], "hourly":{ "summary":"Partly cloudy for the day."}}}
```

Non-Relational Database use case

<https://api.darksky.net/>

23

Open API :] สำหรับนักพัฒนา

แสดงที่ปรับปรุงวัน :
[//covid19.th-stat.com/api/open/today](http://covid19.th-stat.com/api/open/today)

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

ข้อมูลเดลต้าเคส :
[//covid19.th-stat.com/api/open/cases](http://covid19.th-stat.com/api/open/cases)

ข้อมูลสรุปจากเคส :
[//covid19.th-stat.com/api/open/cases/sum](http://covid19.th-stat.com/api/open/cases/sum)

แจ้งเตือนพื้นที่ taper คำประกาศ :
[//covid19.th-stat.com/api/open/area](http://covid19.th-stat.com/api/open/area)



Non-Relational Database use case

24

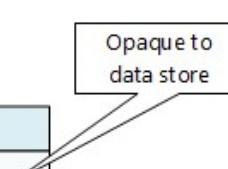
- You might see the term *NoSQL* when reading about non-relational databases.
- NoSQL is a rather loose term that simply means non-relational.
- NoSQL (non-relational) databases generally fall into four categories:
 - key-value stores
 - document databases
 - column family databases
 - graph databases.

What is NoSQL?

25

Key	Value
AAAAA	1101001111010100110101111...
AABAB	1001100001011001101011110...
DFA766	0000000000101010110101010...
FABCC4	1110110110101010100101101...

Opaque to
data store



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

Key-Value Stores

26

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

27

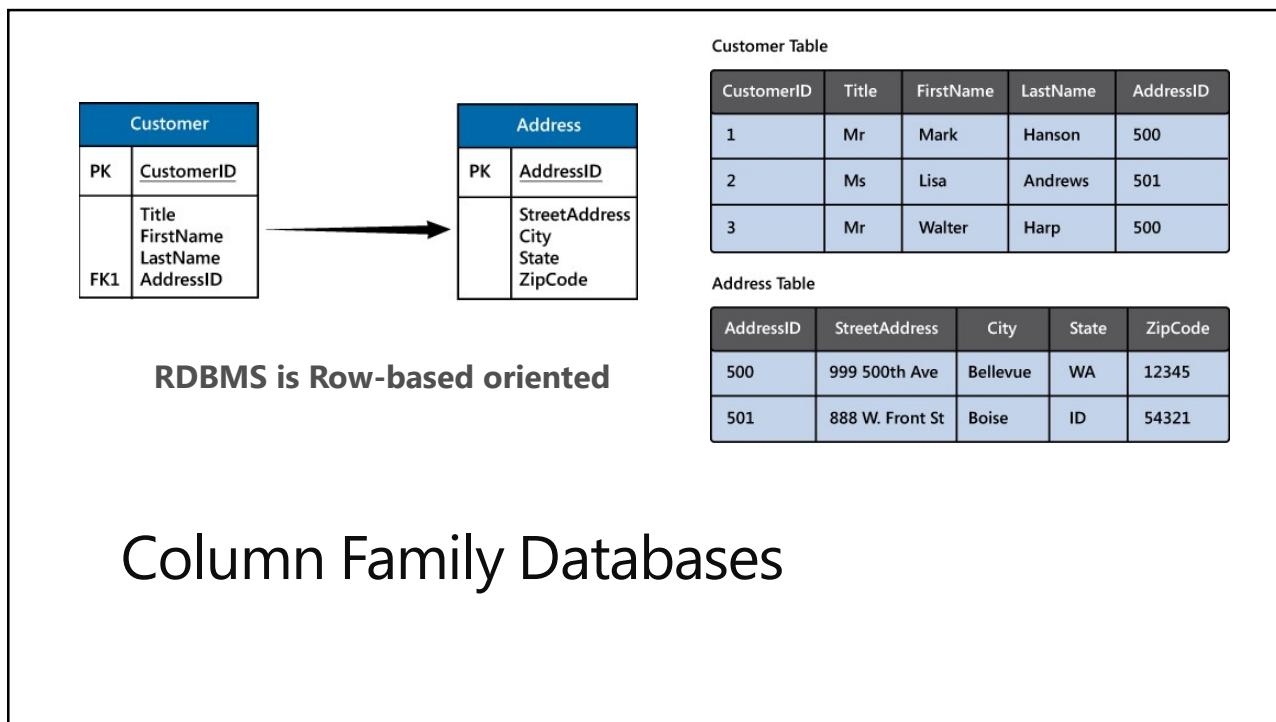
Document Databases

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

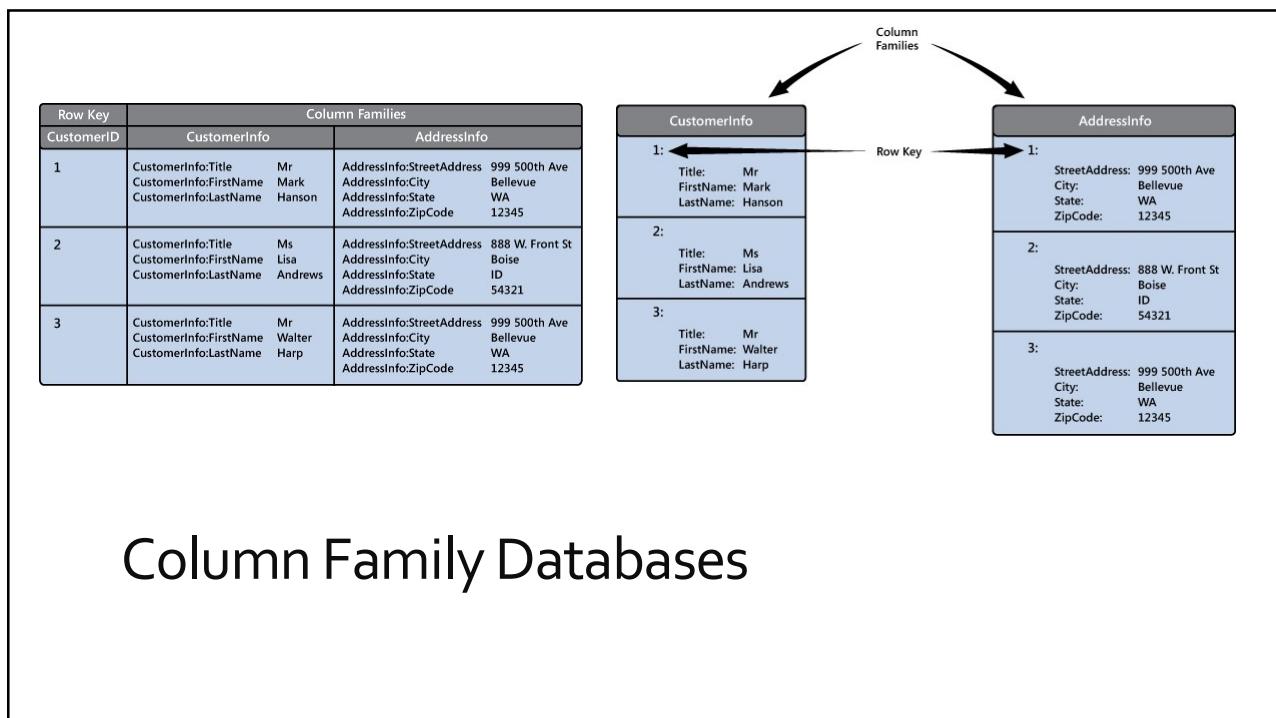


28

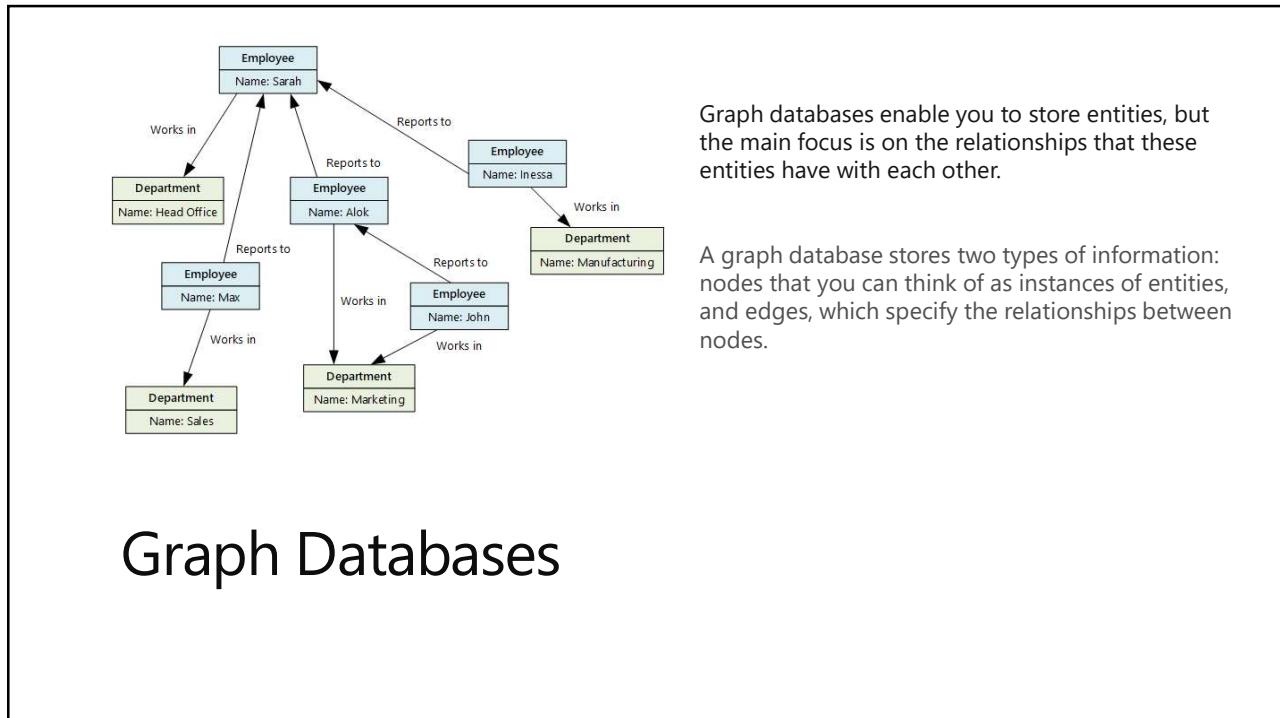


29

Column Family Databases

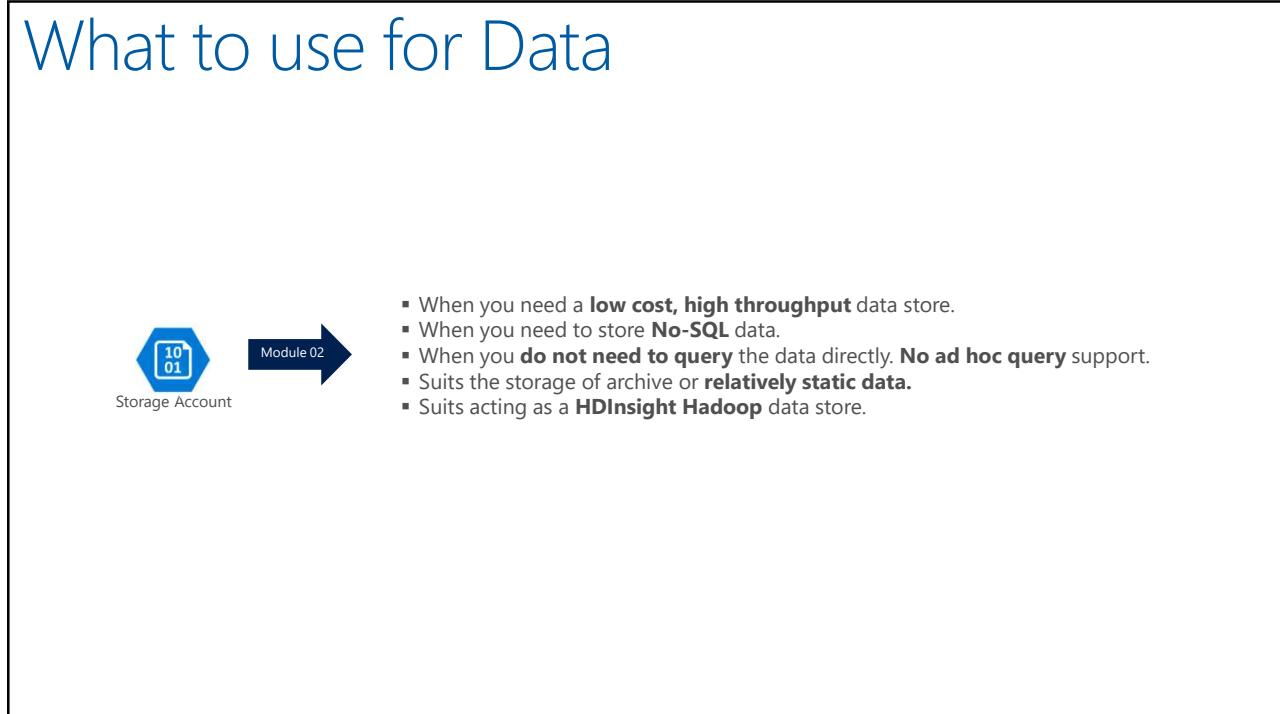


30



Graph Databases

31



32

What to use for Data



Data Lake Store

Module 02 →

- When you need a **low cost, high throughput** data store.
- **Unlimited storage** for **No-SQL** data
- When you **do not need to query** the data directly. **No ad hoc query** support.
- Suits the storage of archive or **relatively static data**.
- Suits acting as a **Databricks**, **HDInsight** and **IoT** data store.

33

What to use for Data



Azure Databricks

Module 03 →

- **Eases the deployment** of a Spark based cluster.
- Enables the **fastest processing** of Machine Learning solutions.
- **Enables collaboration** between data engineers and data scientists.
- Provides **tight enterprise security integration** with Azure Active Directory
- **Integration with other Azure Services** and **Power BI**.

34

What to use for Data



Azure CosmosDB



- Provides **global distribution** for both structured and unstructured data stores.
- **Millisecond query response** time.
- **99.999% availability** of data.
- **Worldwide elastic scale** of both the storage and throughput
- **Multiple consistency levels** to control data integrity with concurrency

35

What to use for Data



Azure SQL Database



- When you require a **relational** data store.
- When you need to manage **transactional workloads**
- When you need to manage a **high volume on inserts and reads**
- When you need a service that **requires high concurrency**
- When you require a solution that can scale **elastically**

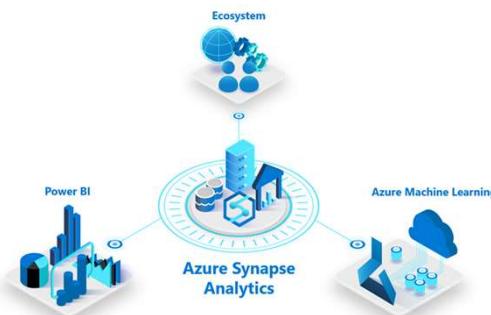
36

What to use for Data

Azure Synapse
Analytics

Module 05 →

- When you require an integrated **relational** and **big data** store.
- When you need to manage **data warehouse** and **analytical workloads**.
- When you need **low cost storage**.
- When you require the ability to **pause and restart the compute**.
- When you require a solution that can scale **elastically**



37

What to use for Data



Azure Stream Analytics

Module 06 →

- When you require a **fully managed event processing** engine.
- When you require **temporal analysis of streaming** data.
- Support for analyzing **IoT streaming** data.
- Support for analyzing application data through **Event Hubs**.
- Ease of use with a **Stream Analytics Query Language**.

38

What to use for Data



Module 07

- When you want to **orchestrate the batch movement** of data.
- When you want to connect to **wide range of data platforms**.
- When you want to **transform or enrich** the data in movement.
- When you want to **integrate with SSIS packages**.
- Enables **verbose logging** of data processing activities.

39

What to use for Data



Azure HDInsight



- When you need a **low cost, high throughput** data store.
- When you need to store **No-SQL** data.
- Provides a Hadoop **Platform as a Service** approach
- Suits acting as a **Hadoop, Hbase, Storm or Kafka** data store.
- **Eases the deployment and management** of clusters.

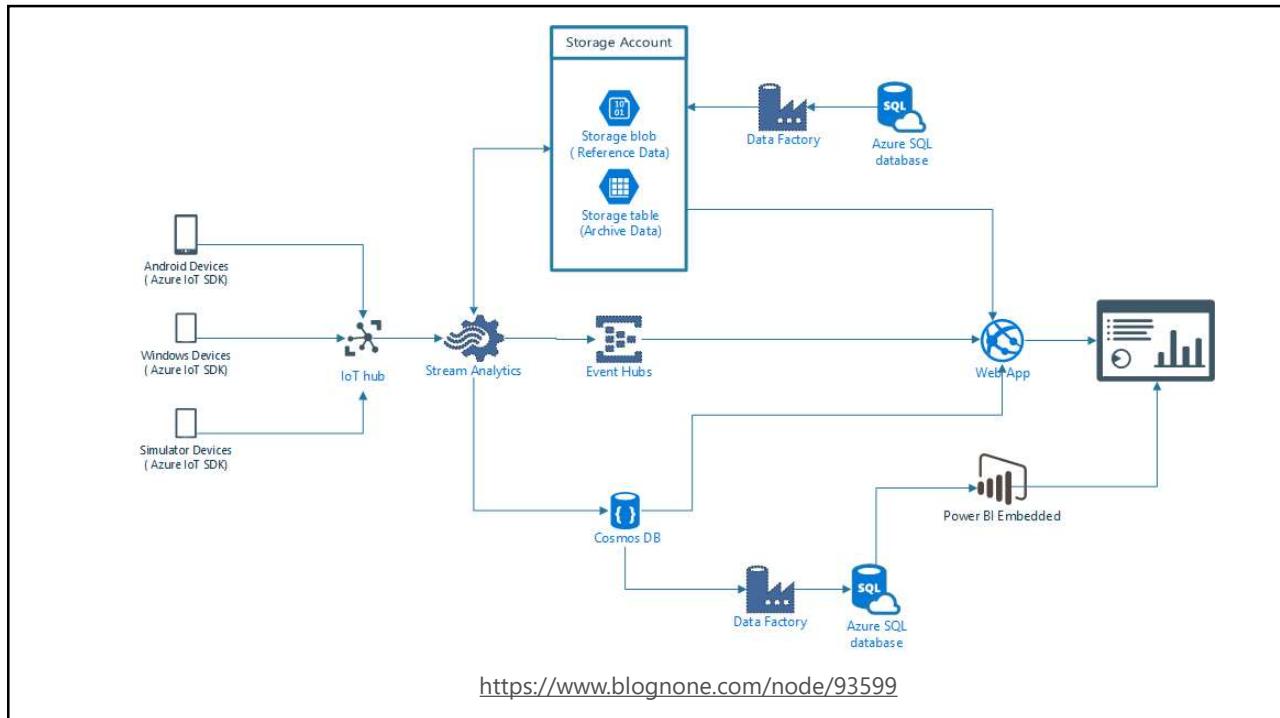


Azure Data Catalog



- When you require **documentation** of your data stores.
- When you require a **multi user** approach to documentation.
- When you need to **annotate data sources** with descriptive metadata.
- A **fully managed cloud service** whose users can discover the data sources.
- When you require a **solution that can help business users** understand their data.

40



41



42

Lesson Objectives

- List the new roles of modern data projects
- Outline data engineering practices
- Explore the high-level process for architecting a data engineering project

43

The slide features a dark blue background on the left side containing a diagram and a photograph on the right.

Diagram: A circular diagram illustrating the interconnected roles in data projects. Five blue circles are arranged in a cluster, connected by thin white lines. The circles are labeled: "Data Engineer" at the top, "Data Management" on the left, "AI Engineer" at the bottom-left, "Data Analyst" at the top-right, and "Data Scientist" at the bottom-right.

Photograph: On the right side, a woman with blonde hair, wearing a blue cardigan over a dark top, is seated at a light-colored wooden conference table, looking down at a laptop. In the background, there is a large window showing a cityscape with buildings. Two other people, a man and a woman, are standing near the window; the man is holding a white coffee cup.

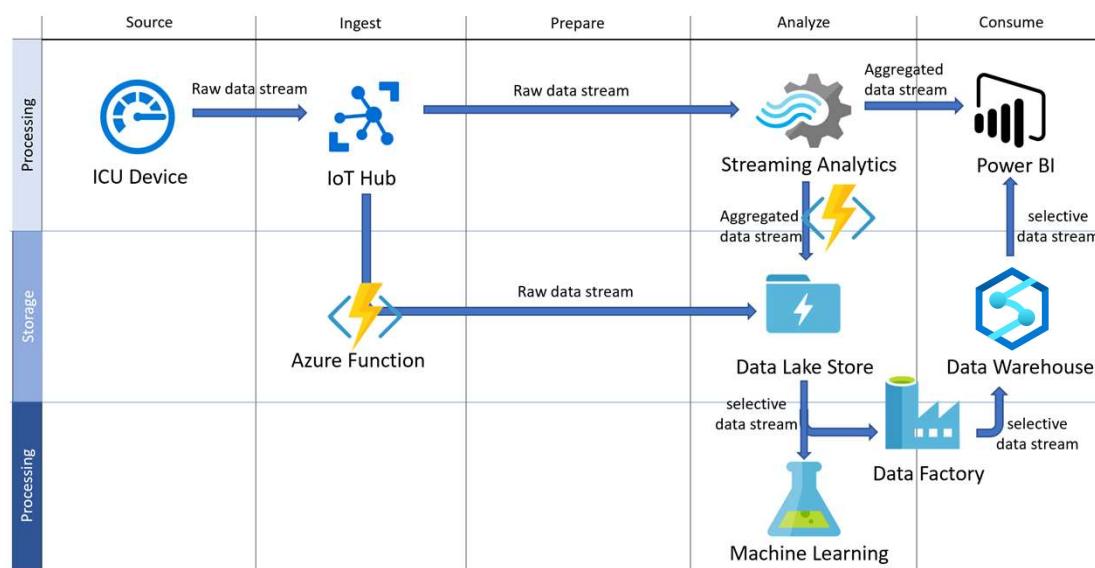
44

Data Engineering Practices



45

Architecting Projects – an example



46

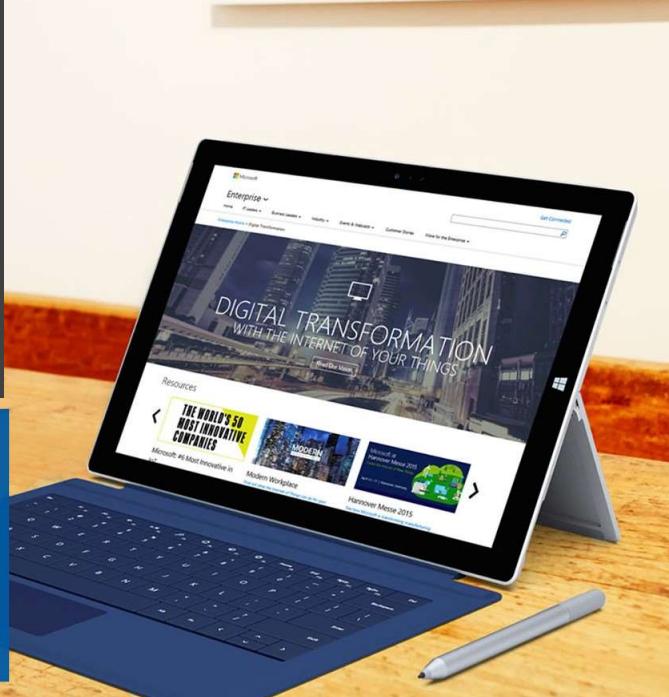
Module Summary ➔

In this module, you have learned about:

- The evolving world of data.
- The services in the Azure Data Platform.
- The tasks that are performed by a Data Engineer.
- A fictitious Case Study for use in labs.

Next steps ➔

After the course, consider visiting [\[the Microsoft Customer Case Study site\]](#). Use the search bar to search by an industry such as healthcare or retail, or by a technology such as Azure Cosmos DB or Stream Analytics. Read through some of the customers stories.



47

 Microsoft

Module 02: Working with Data Storage



48

Agenda

- L01 – Choose a data storage approach in Azure
- L02 - Create an Azure Storage Account
- L03 - Explain Azure Data Lake Storage
- L04 - Upload data into Azure Data Lake Store

49



50

Lesson Objectives

- The Benefits of using Azure to store data
- Compare Azure data storage with on-premises storage

51

Benefits of using Azure to store data



52

Comparing Azure to on-premises storage

The term "on-premises" refers to the storage and maintenance of data on local hardware and servers.

Cost effectiveness	Reliability	Storage types	Agility
On-premises storage requires up-front expenses. Azure data storage provides a pay-as-you-go pricing model	Azure data storage provides backup, load balancing, disaster recovery, and data replication to ensure safety and high availability. This capability requires significant investment with on-premises solutions	Azure data storage provides a variety of different storage options including distributed access and tiered storage	Azure data storage gives you the flexibility to create new services in minutes and allows you to change storage backends quickly

53



54

Lesson Objectives

- Describe storage accounts
- Determine the appropriate settings for each storage account
- Choose an account creation tool
- Create a storage account using the Azure portal

55

Storage accounts

What is a Storage Account

It is a container that groups a set of Azure Storage services. Only data services can be included in a storage account such as *Azure Blobs*, *Azure Files*, *Azure Queues*, and *Azure Tables*.

How many do you need?

The number of storage accounts you need is typically determined by your data diversity, cost sensitivity, and tolerance for management overhead.

The number of storage accounts you need is based on:

Data Diversity

Organizations often generate data that differs in where it is consumed and how sensitive it is.

Cost Sensitivity

The settings you choose for the account do influence the cost of services, and the number of accounts you create



Management Overhead

Each storage account requires some time and attention from an administrator to create and maintain.



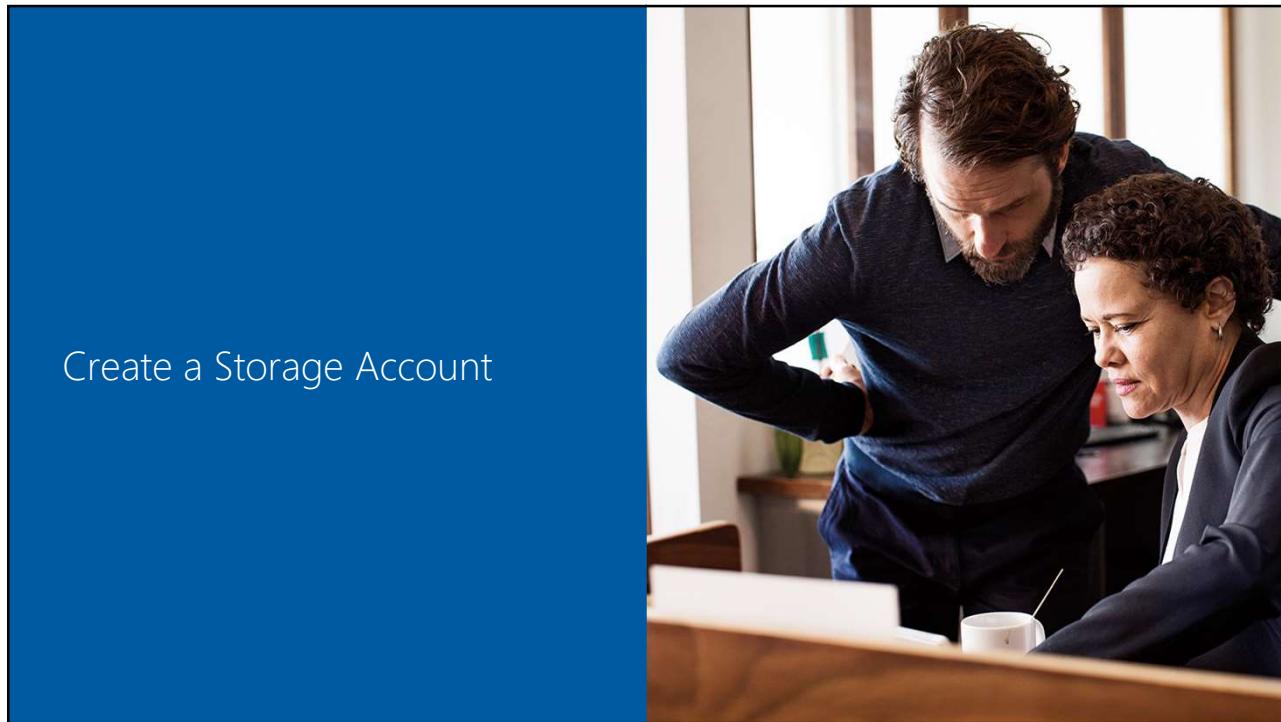
56

The screenshot shows the 'Create storage account' page in the Azure portal. At the top, there's a breadcrumb navigation: Home > New > Storage account > Create storage account. Below it, a title 'Create storage account' and a tab bar with 'Basics' (which is selected), 'Networking', 'Advanced', 'Tags', and 'Review + create'. A descriptive text explains that Azure Storage is a Microsoft-managed service providing cloud storage. The 'Project details' section asks for a subscription (selected: 'chtestao') and a resource group ('Select existing...' or 'Create new'). The 'Instance details' section includes fields for 'Storage account name' (empty), 'Location' (selected: '(US) South Central US'), 'Performance' (selected: 'Standard'), 'Account kind' (selected: 'StorageV2 (general purpose v2)'), 'Replication' (selected: 'Read-access geo-redundant storage (RA-GRS)'), and 'Access tier (default)' (selected: 'Hot').

57

The diagram on the left illustrates the 'Storage Account creation tool' with four interconnected components: 'Azure Portal' (dark blue circle), 'Azure Command Line Interface' (purple circle), 'Azure PowerShell' (green circle), and 'Management Client Libraries' (orange circle). Arrows show bidirectional connections between all four components. To the right of the diagram is a photograph of a man and a woman standing at a desk, looking down at a laptop screen together.

58



59



60

Lesson Objectives

- Explain Azure Data Lake Storage
- Create an Azure Data Lake Store Gen 2 using the portal
- Compare Azure Blob Storage and Data Lake Store Gen 2
- Explore the stages for processing Big Data Using Azure Data Lake Store
- Describe the use cases for Data lake Storage

61

Azure Data Lake Storage – Generation II

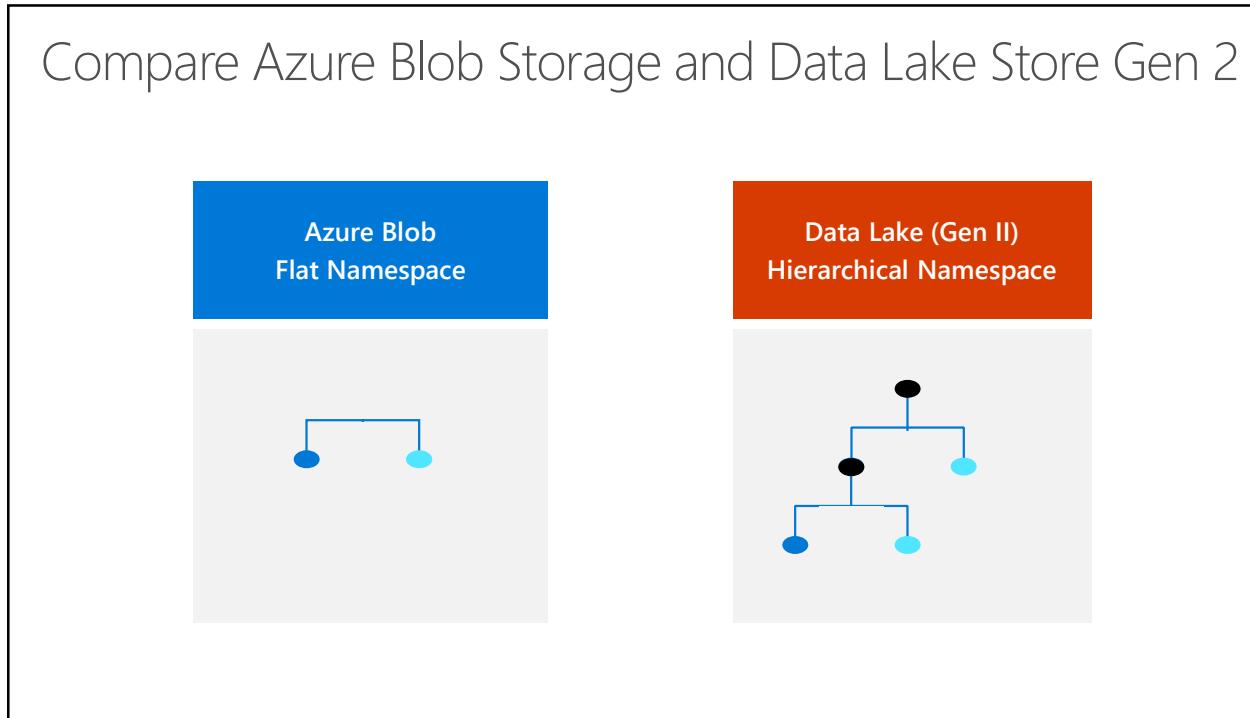


62

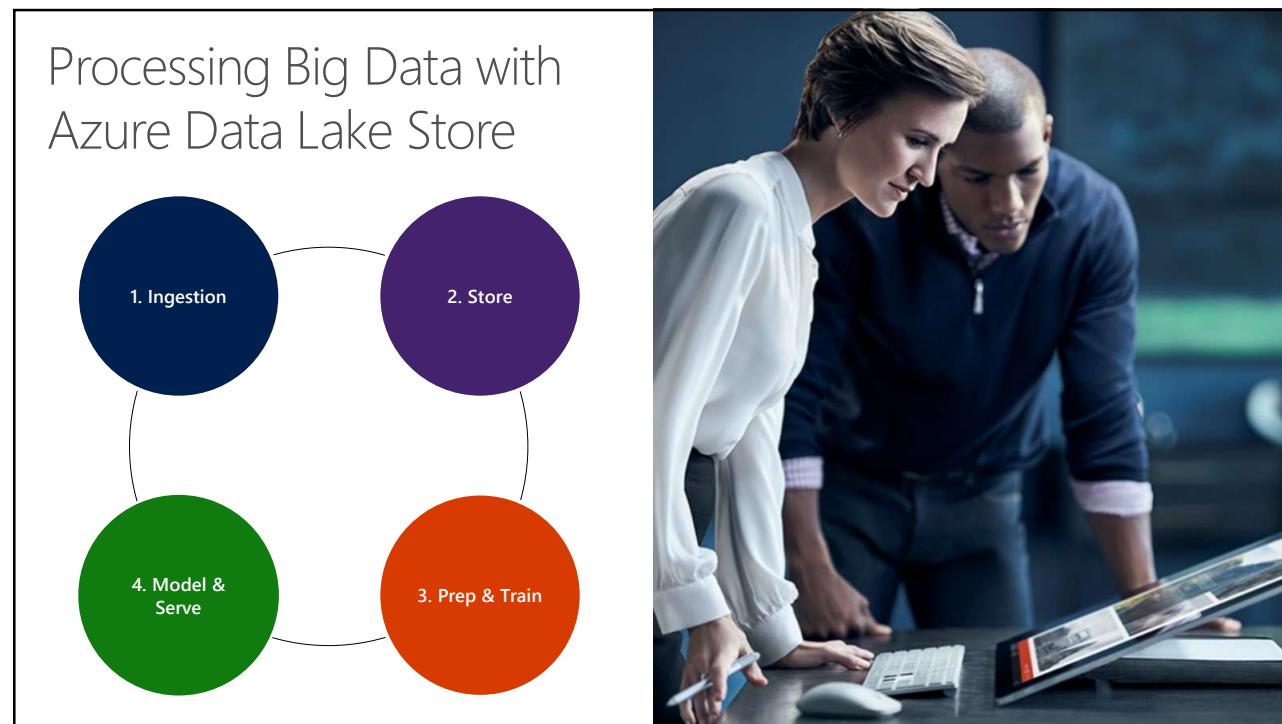
Create a Azure Data Lake Store (Gen II) using the Portal.

The screenshot shows the 'Create storage account' page in the Azure portal. The 'Advanced' tab is selected. Under 'Data protection', the 'Hierarchical namespace' section is highlighted with a red box. Inside this box, the 'Enabled' radio button is selected. Other sections like 'Security' and 'Azure Files' also have their own configuration options.

63



64



65

Big Data Use Cases

Let's examine three use cases for leveraging an Azure Data Lake Store

Modern Data Warehouse	This architecture sees Azure Data Lake Storage at the heart of the solution for a modern data warehouse. Using Azure Data Factory to ingest data into the Data Lake from a business application, and predictive models built in Azure Databricks, using Azure Synapse Analytics as a serving layer.
Advanced Analytics	In this solution, Azure Data factory is transferring terabytes of web logs from a web server to the Data Lake on an hourly basis. This data is provided as features to the predictive model in Azure Databricks, which is then trained and scored. The result of the model is then distributed globally using Azure Cosmos DB, that an application uses.
Real Time Analytics	In this architecture, there are two ingestion streams. Azure Data Factory is used to ingest the summary files that are generated when the HGV engine is turned off. Apache Kafka provides the real-time ingestion engine for the telemetry data. Both data streams are stored in Data Lake store for use in the future.

66



Lesson 04

Upload Data into Azure Data Lake Store

67

Lesson Objectives

- Create an Azure Data Lake Gen2 Store using PowerShell
- Upload data into the Data Lake Storage Gen2 using Azure Storage Explorer
- Copy data from an Azure Data Lake Store Gen1 to an Azure Data Lake Store Gen2

68

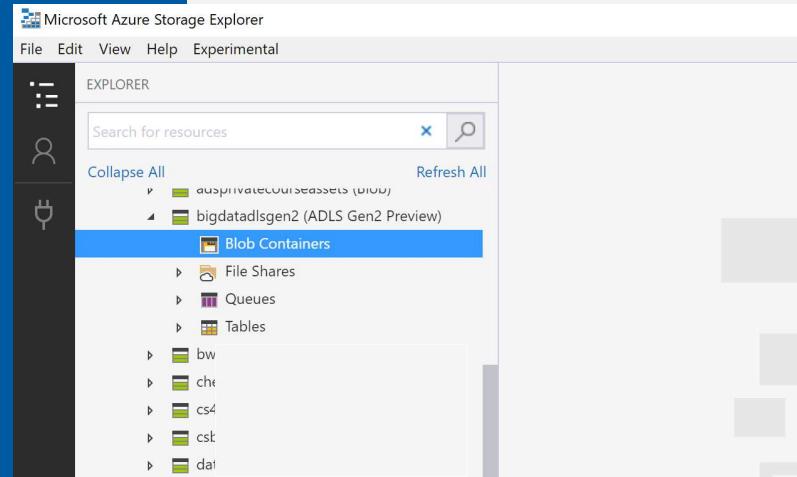
Create a Azure Data Lake Store (Gen II) using PowerShell.

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

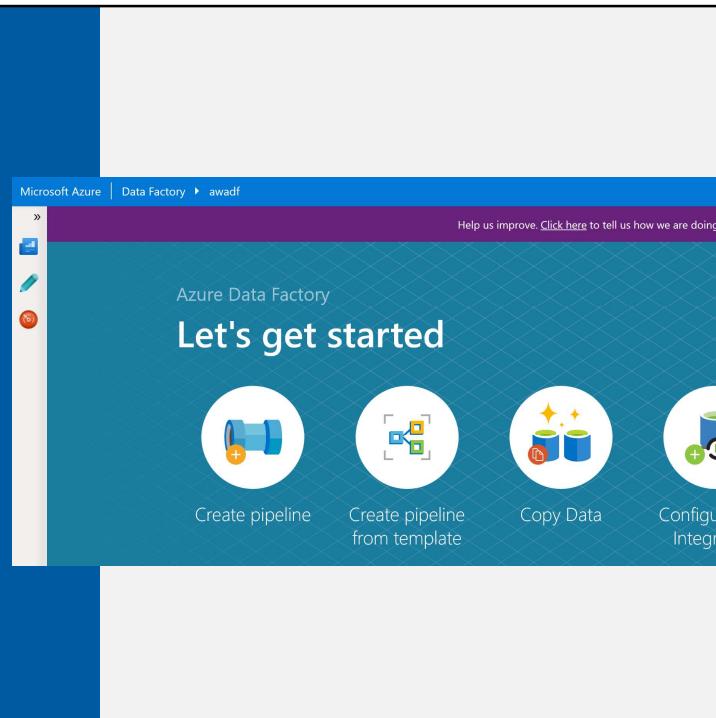
PS C:\Users> $location = "westus2"
>>
>> New-AzStorageAccount -ResourceGroupName $resourceGroup `
>>   -Name "storagequickstart" `
>>   -Location $location `
>>   -SkuName Standard_LRS `
>>   -Kind StorageV2 `
>>   -EnableHierarchicalNamespace $True
```

69

Uploading data with Azure Storage Explorer



70



Copy data from an Azure Data Lake Store Gen1 to an Azure Data Lake Store Gen2

Microsoft Azure | Data Factory > awadf

Help us improve. [Click here](#) to tell us how we are doing

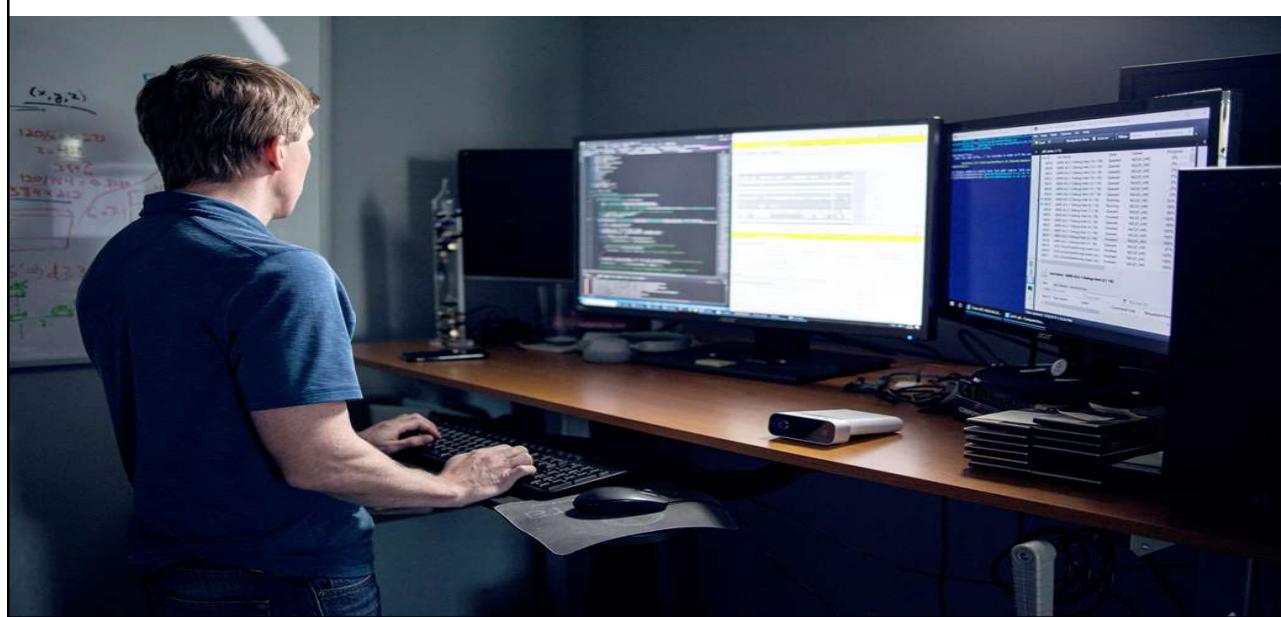
Azure Data Factory

Let's get started

Create pipeline Create pipeline from template Copy Data Config Integr

71

Lab: Working with Data Storage



72



Module 03: Working with Relational Data Stores in the Cloud



73

Agenda

- L01 - Work with Azure SQL Database
- L02 - Work with Azure Synapse Analytics
- L03 - Provision and query data in Azure Synapse Analytics
- L04 - Import data into Azure Synapse Analytics using PolyBase

74



75

Lesson 01

Azure SQL Database

Lesson Objectives

- Why Azure SQL Database is a good choice for running your relational database
- What configuration and pricing options are available for your Azure SQL database
- How to create an Azure SQL database from the portal
- How to use Azure Cloud Shell to connect to your Azure SQL database, add a table, and work with data

76

Why Azure SQL Database is a good choice

Convenience

Cost

Scale

Security

77

Azure SQL Database configuration options

When you create your first Azure SQL database, you also create an *Azure SQL logical server*. Think of a logical server as an administrative container for your databases.

DTUs	vCores	SQL elastic pools	SQL Managed Instances
DTU stands for Database Transaction Unit and is a combined measure of compute, storage, and IO resources. Think of the DTU model as a simple, preconfigured purchase option	vCore gives you greater control over what compute and storage resources you create and pay for. vCore model enables you to configure resources independently	SQL elastic pools relate to eDTUs. They enable you to buy a set of compute and storage resources that are shared among all the databases in the pool. Each database can use the resources they need	The SQL managed instance creates a database with near 100% compatibility with the latest SQL Server on-premises Enterprise Edition database engine, useful for SQL Server customers who would like to migrate on-premises servers instance in a "lift and shift" manner

78

Create an Azure SQL Database.

Create SQL Database

Home > New > SQL Database > Create SQL Database

Basics • Networking Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Resource group *

[Create new](#)

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *

Server * [Create new](#)

The value must not be empty.

Want to use SQL elastic pool? * Yes No

Compute + storage *

79

sqlcmd -S tcp:contoso-1.database.windows.net,1433
-d Logistics -U martina -P "password1234\$" -N -l 30

CREATE TABLE Drivers (DriverID int, LastName varchar(255), FirstName varchar(255), OriginCity varchar(255)); GO

SELECT name FROM sys.tables; GO

INSERT INTO Drivers (DriverID, LastName, FirstName, OriginCity) VALUES (123, 'Zirne', 'Laura', 'Springfield'); GO

Use Azure Cloud Shell to connect to your Azure SQL database

80



Lesson 02

Azure Synapse Analytics

81

Lesson Objectives

- Explain Azure Synapse Analytics
- Explain Azure Synapse Analytics features
- Types of solution workloads
- Explain Massively Parallel Processing concepts
- Compare table geometries

82

Azure Synapse Analytics

What is Azure Synapse Analytics

A unified environment by combining the enterprise data warehouse of SQL, the Big Data analytics capabilities of Spark, and data integration technologies to ease the movement of data between both, and from external data sources.

Data warehouse capabilities.

SQL Analytics

A centralized data warehouse store that provides a relational analytics and decision support services across the whole enterprise

SQL Pools

CPU, memory, and IO are bundled into units of compute scale called SQL, determined by Data Warehousing Units (DWU)

Future features

Will include a Spark engine, a data integration and Azure Synapse Analytics Studio

83

Azure Synapse Analytics features

Workload Management	Result-Set Cache	Materialized Views	SSDT CI/CD support
<p>This capability is used to prioritize the query workloads that take place on the server using Workload Management. This involves three components:</p> <ul style="list-style-type: none"> • Workload Groups • Workload Classification • Workload Importance 	<p>Result-set caching can be used to improve the performance of the queries that retrieve these results. When result-set caching is enabled, the results of the query are cached in the SQL pool storage.</p>	<p>A materialized view pre-computes, stores, and maintains its data like a table. They are automatically updated when data in underlying tables are changed.</p>	<p>Database project support in SQL Server Data Tools (SSDT) allows teams of developers to collaborate over a version-controlled Azure Synapse Analytics, and track, deploy and test schema changes</p>

84

Types of solution workloads

The modern data warehouse extends the scope of the data warehouse to serve Big Data that's prepared with techniques beyond relational ETL



Modern data warehousing

"We want to integrate all our data—including Big Data—with our data warehouse"



Advanced analytics

"We're trying to predict when our customers churn"

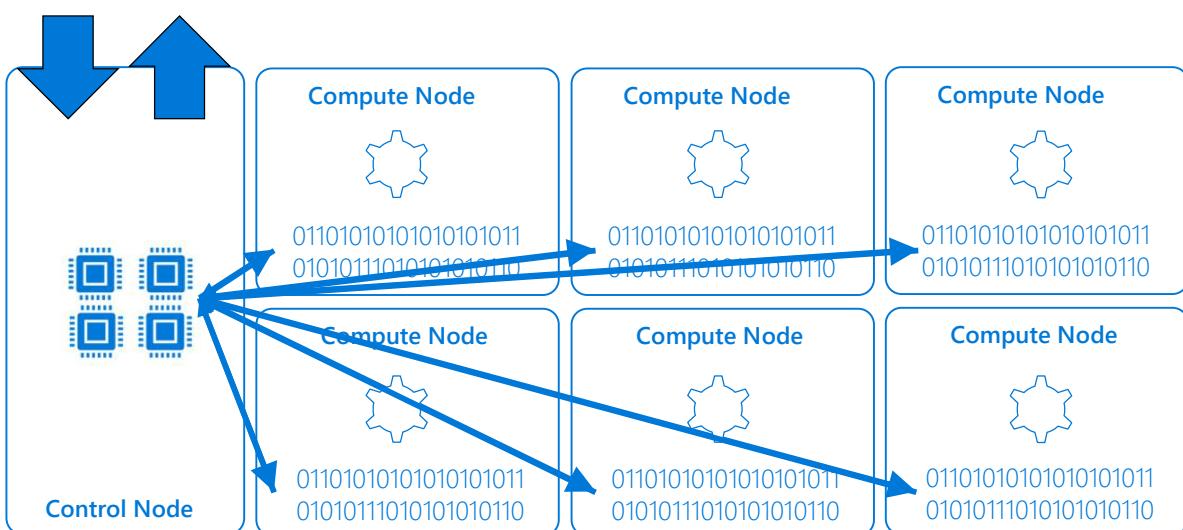


Real-time analytics

"We're trying to get insights from our devices in real-time"

85

Massively Parallel Processing (MPP) concepts



86



The End