

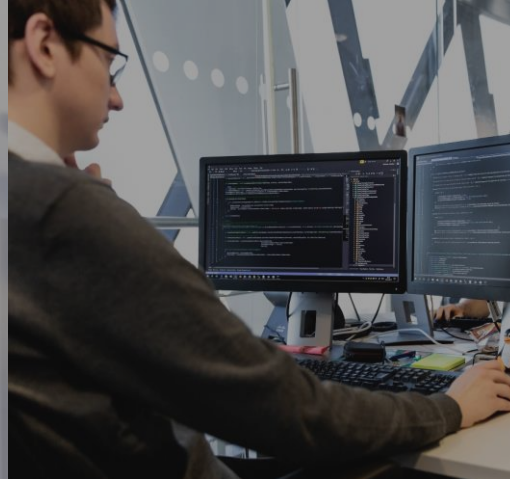


DP-200T01: Working with Data Storage



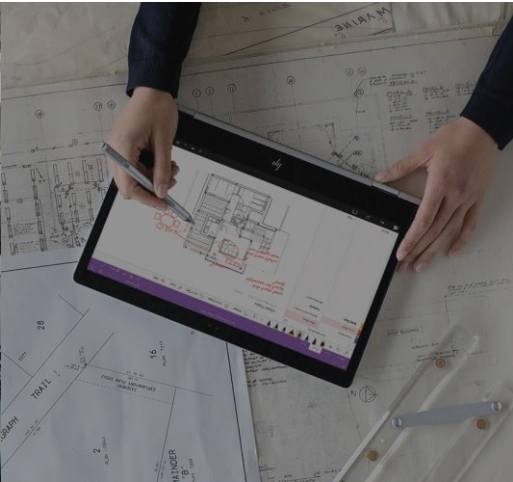
Agenda

- 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



Lesson 01

Choose a Data Storage Approach in Azure



Lesson Objectives

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

Benefits of using Azure to store data

Automated Backup

Global Replication

Encryption Capabilities

Multiple Data Types

Support for Data Analytics

Storage Tiers

Virtual Disks

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

On-premises storage requires up-front expenses. Azure data storage provides a pay-as-you-go pricing model

Reliability

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

Storage types

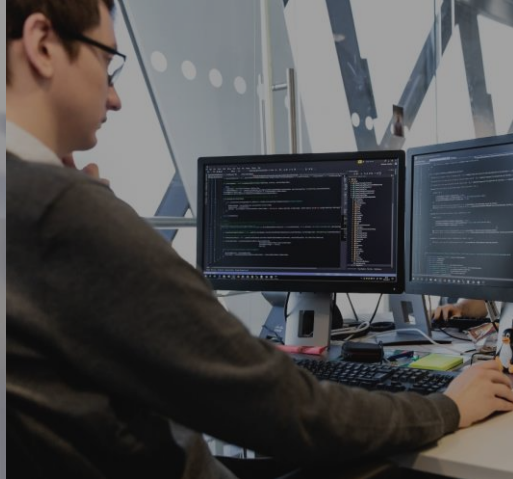
Azure data storage provides a variety of different storage options including distributed access and tiered storage

Agility

Azure data storage gives you the flexibility to create new services in minutes and allows you to change storage back-ends quickly

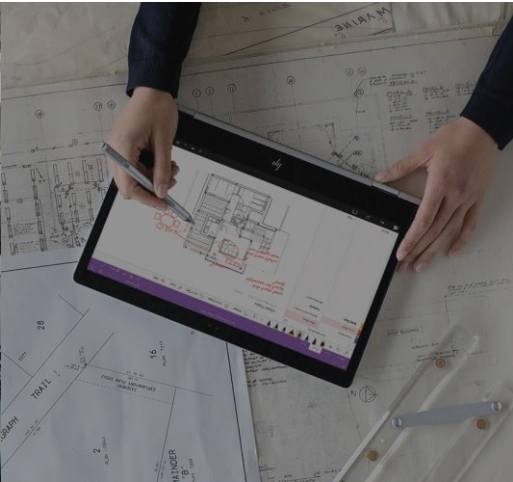
Review Questions

- Q01 – Suppose you work at a startup with limited funding. Why might you prefer Azure data storage over an on-premises solution?
- A01 – The Azure pay-as-you-go billing model lets you avoid buying expensive hardware.
- Q02 - Which of the following situations would yield the most benefits from relocating an on-premises data store to Azure?
- A02 – Unpredictable storage demand that increases and decreases multiple times throughout the year.



Lesson 02

Create an Azure Storage Account



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

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.



Storage Account settings

Create storage account

Basics Networking Advanced Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below.

[Learn more about Azure storage accounts](#)

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](#)

Instance details

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

Storage account name *

Location *

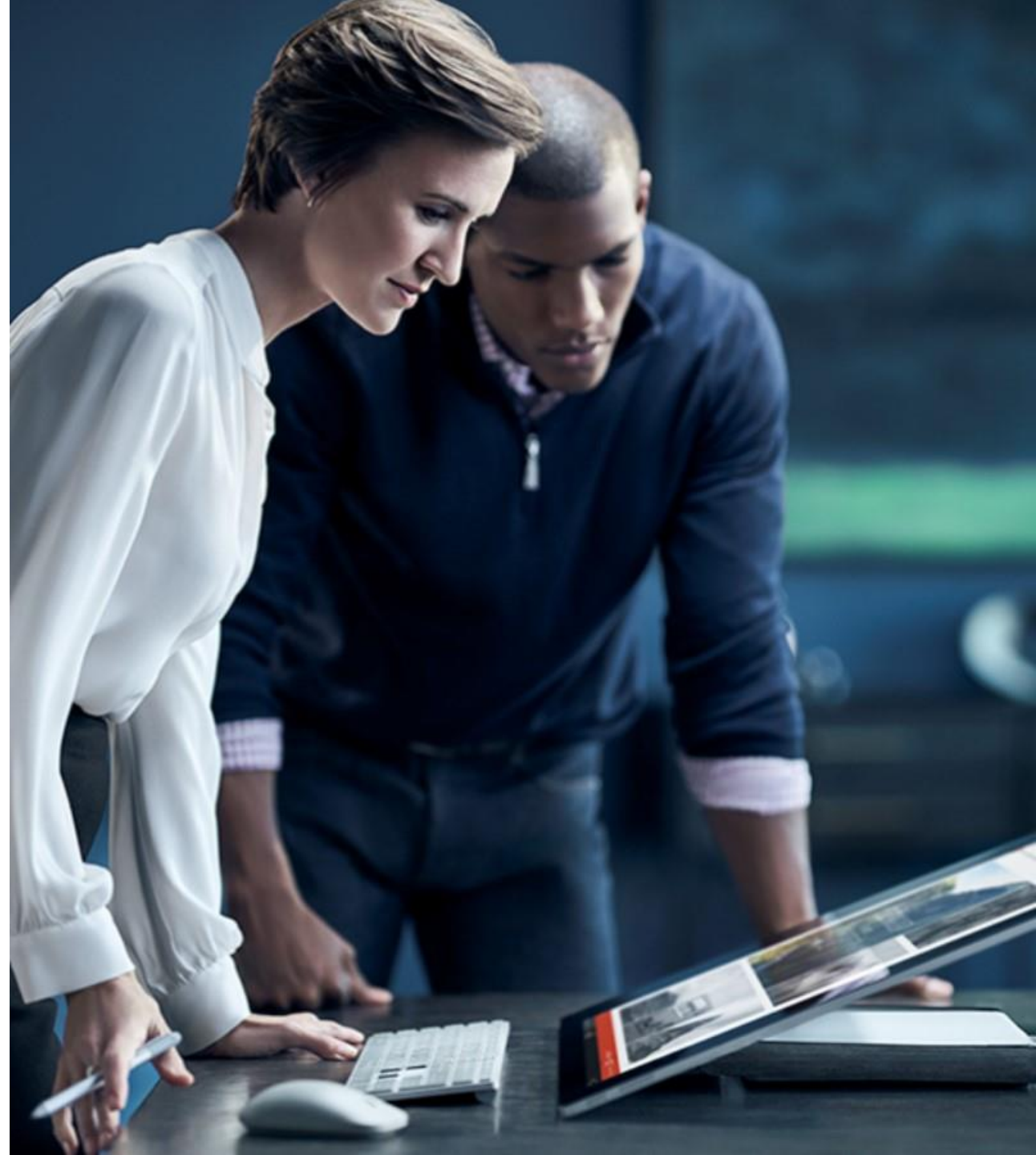
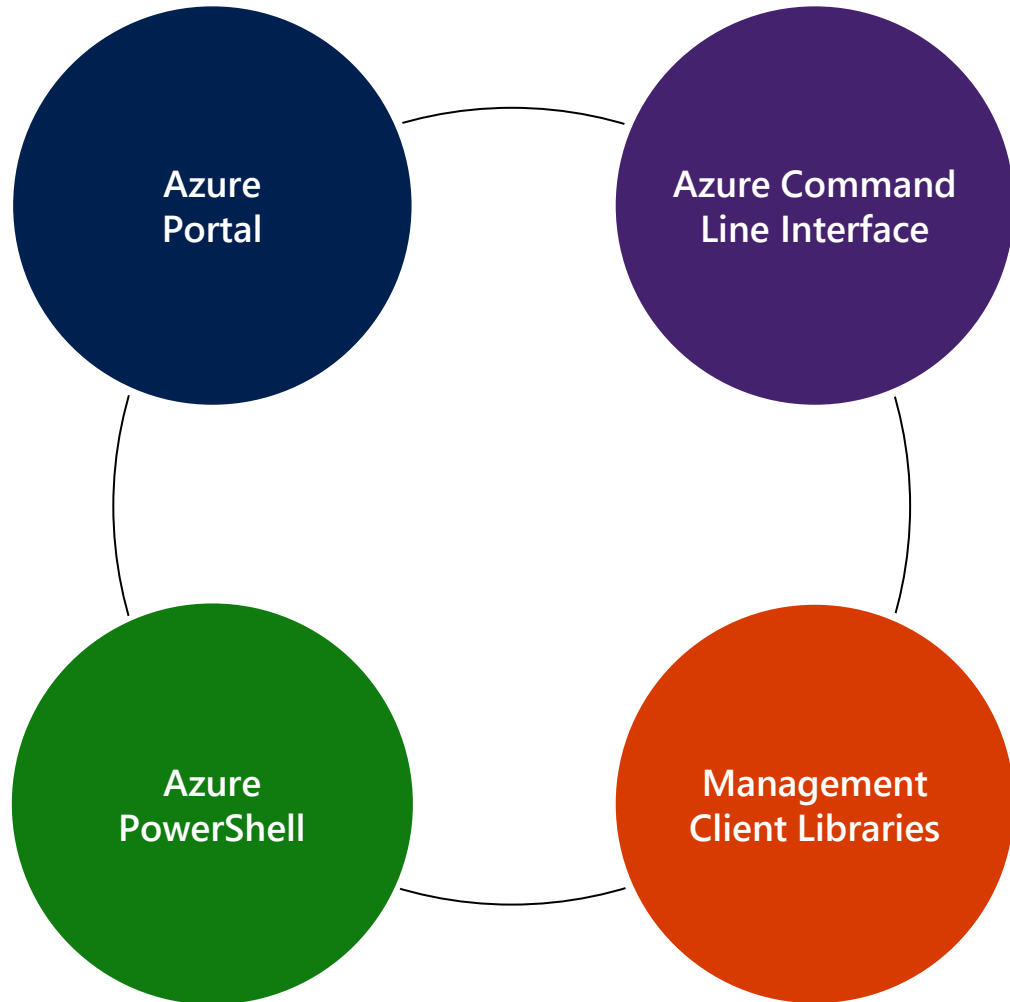
Performance Standard Premium

Account kind

Replication

Access tier (default) Cool Hot

Storage Account creation tool

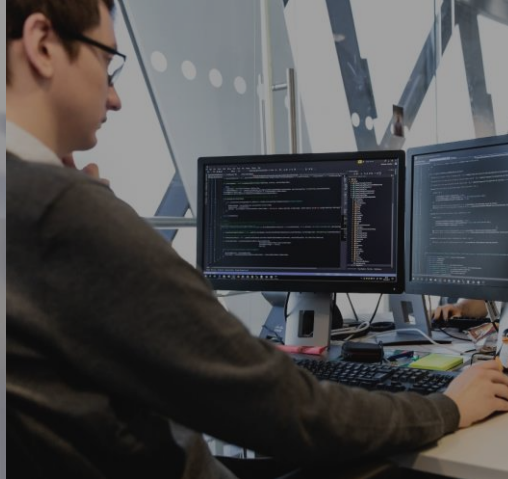


Create a Storage Account



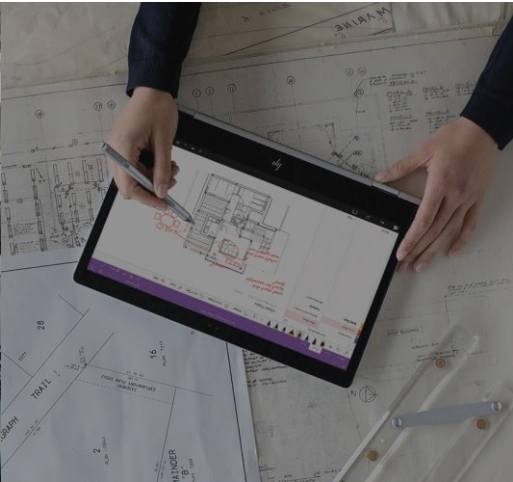
Review Questions

- Q01 - Suppose you have two video files stored as blobs. One of the videos is business-critical and requires a replication policy that creates multiple copies across geographically diverse datacenters. The other video is non-critical, and a local replication policy is sufficient. True or false: To satisfy these constraints, the two blobs will need to be in separate storage accounts?
- A01 – True.
- Q02 - The name of a storage account must be..?
- A02 – Globally unique.
- Q03 - In a typical project, when would you create your storage account(s)?
- A03 – At the beginning, during project setup.



Lesson 03

Azure Data Lake Storage



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

Azure Data Lake Storage – Generation II



**Hadoop
Access**



Security



Performance



Redundancy

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

Home > New > Storage account > Create storage account

Create storage account

Basics Networking **Advanced** Tags Review + create

Security

Secure transfer required ⓘ Disabled Enabled

Azure Files

Large file shares ⓘ Disabled Enabled

i The current combination of storage account kind, performance, replication and location does not support large file shares.

Data protection

Blob soft delete ⓘ Disabled Enabled

i Data protection and hierarchical namespace cannot be enabled simultaneously.

Data Lake Storage Gen2

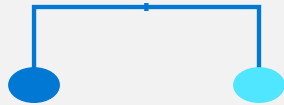
Hierarchical namespace ⓘ Disabled Enabled

NFS v3 ⓘ Disabled Enabled

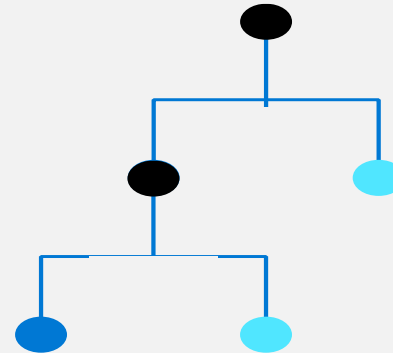
Signup is currently required to utilize the the NFS v3 feature on a per-subscription basis. [Signup for NFS v3](#) ↗

Compare Azure Blob Storage and Data Lake Store Gen 2

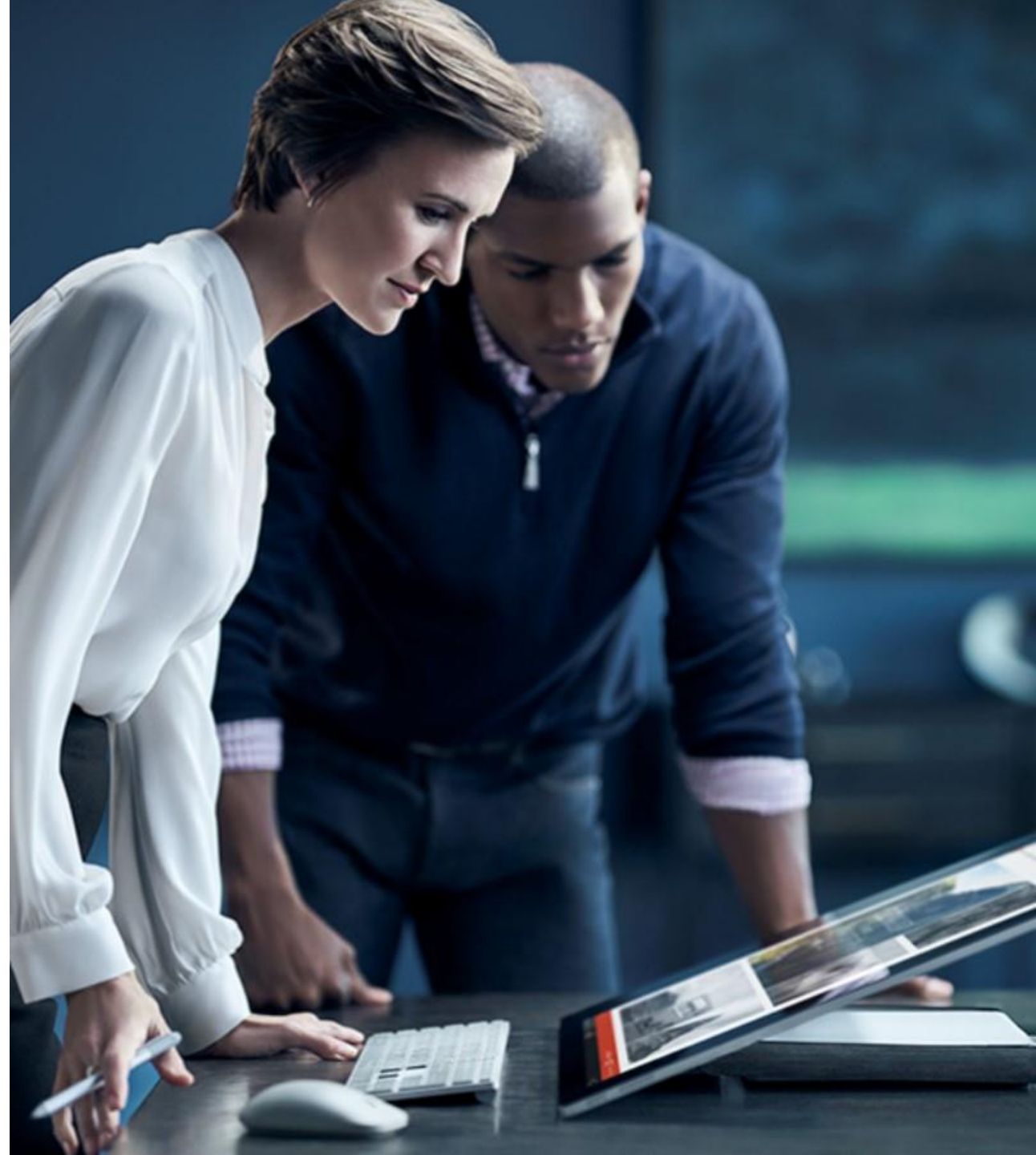
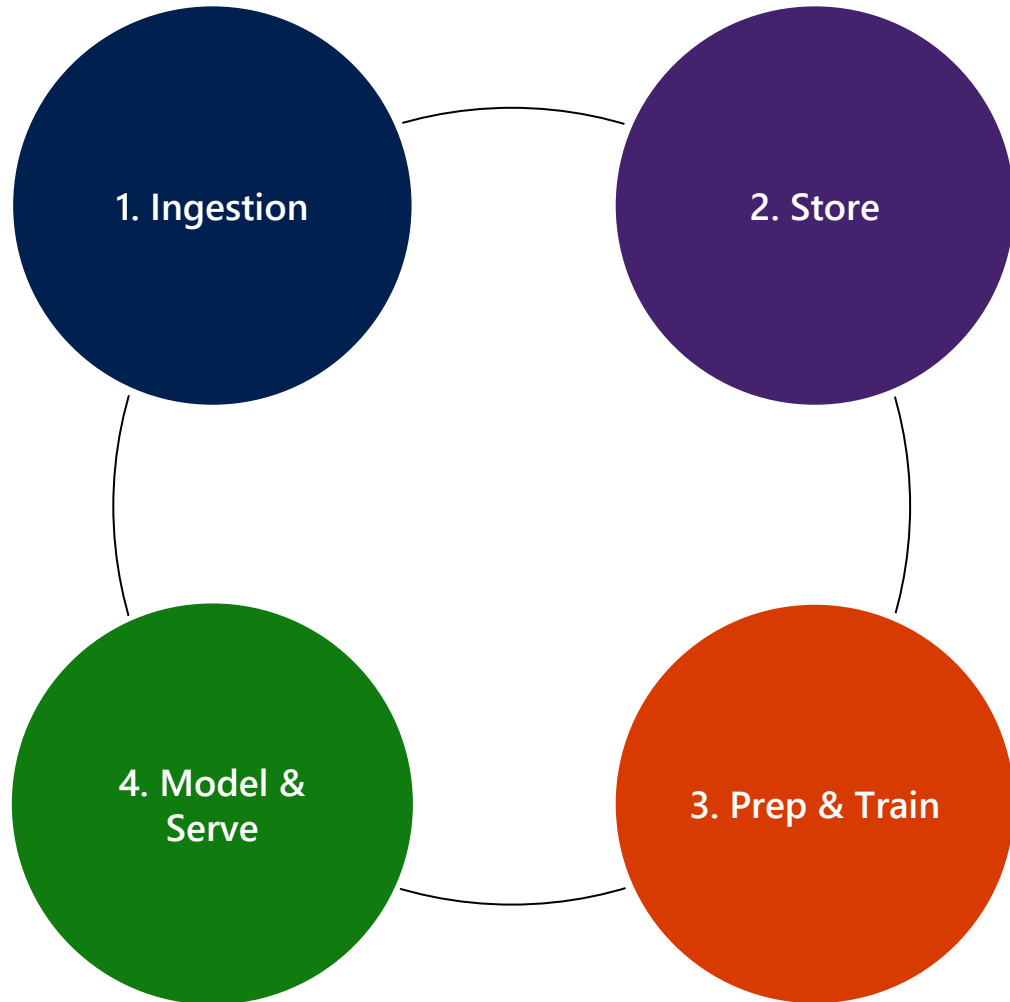
**Azure Blob
Flat Namespace**



**Data Lake (Gen II)
Hierarchical Namespace**



Processing Big Data with Azure Data Lake Store



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.

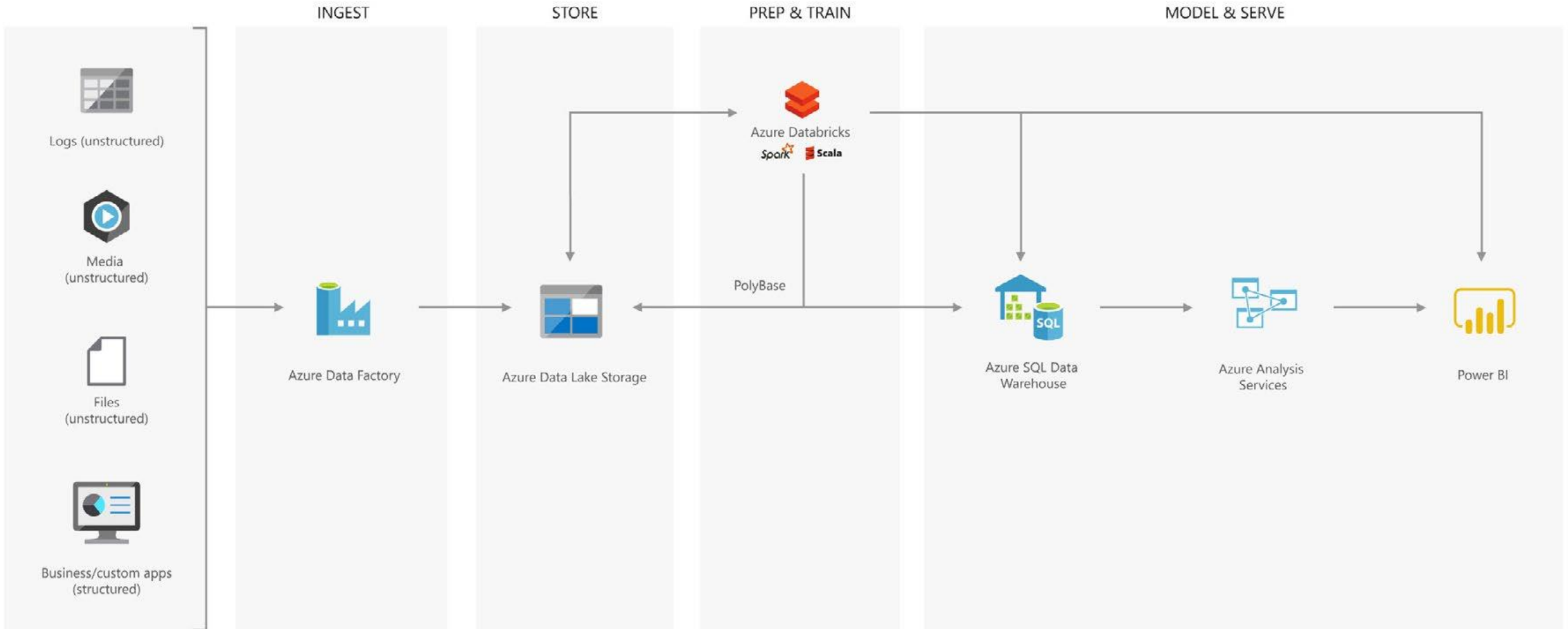
Advanced Analytics

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.

Modern data warehouse



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.

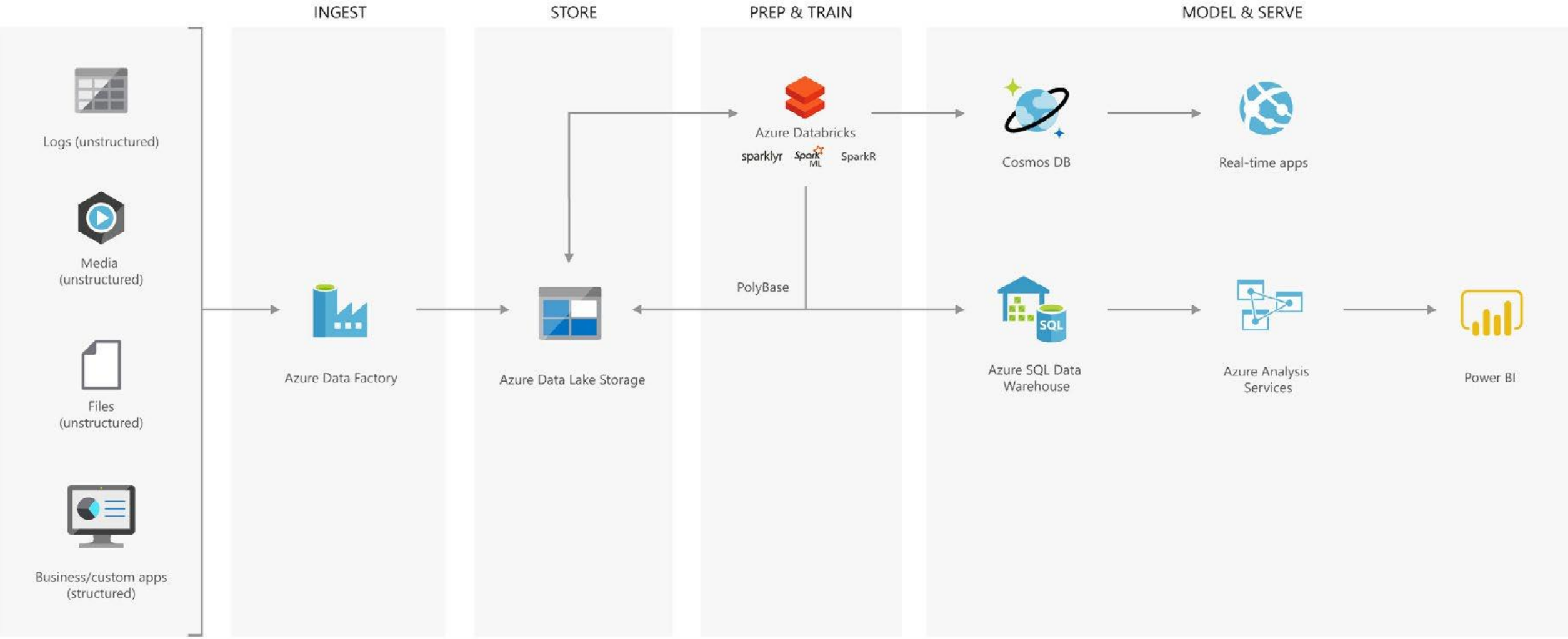
Advanced Analytics

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.

Advanced analytics on big data



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.

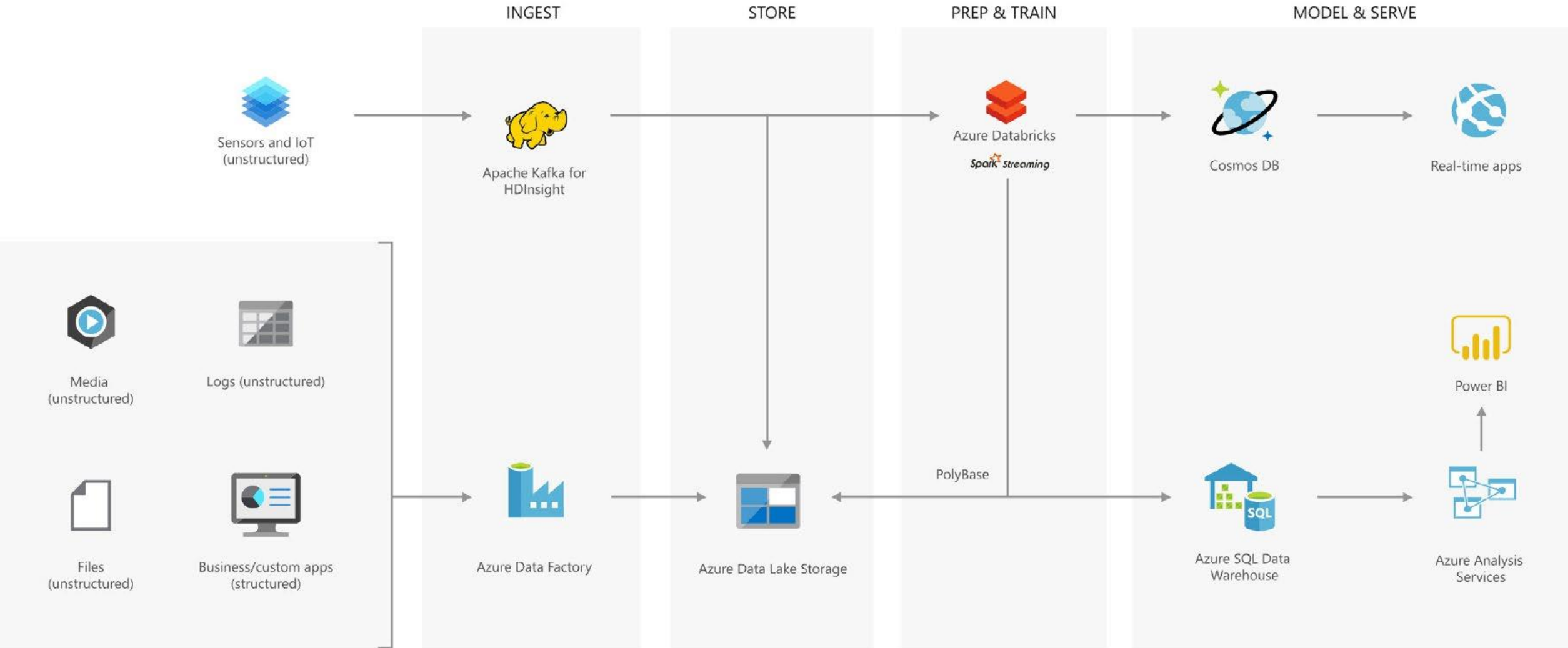
Advanced Analytics

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

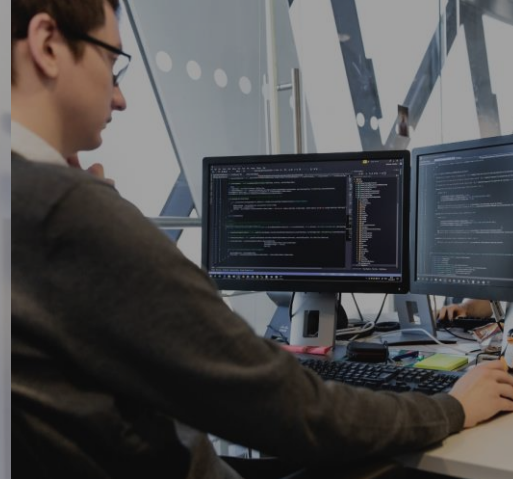
In this architecture, there are two ingestion streams. Azure Data Factory is used to ingest the summary files that are generated when the HGV engine is turned off. Apache Kafka provides the real-time ingestion engine for the telemetry data. Both data streams are stored in Data Lake store for use in the future.

Real-time analytics



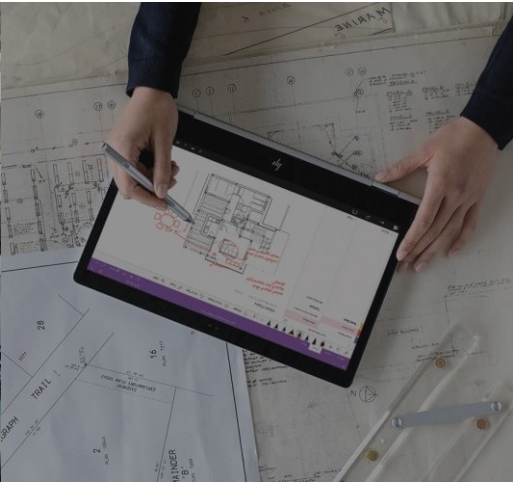
Review Questions

- Q01 – Mike is creating an Azure Data Lake Storage Gen 2 account. He must configure this account to be able to processes analytical data workloads for best performance. Which option should he configure when creating the storage account?
- A01 – On the Advanced tab, set the Hierarchical Namespace to enabled.
- Q02 - In which phase of big data processing is Azure Data Lake Storage located?
- A02 – Store.



Lesson 04

Upload Data into Azure Data Lake Store



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

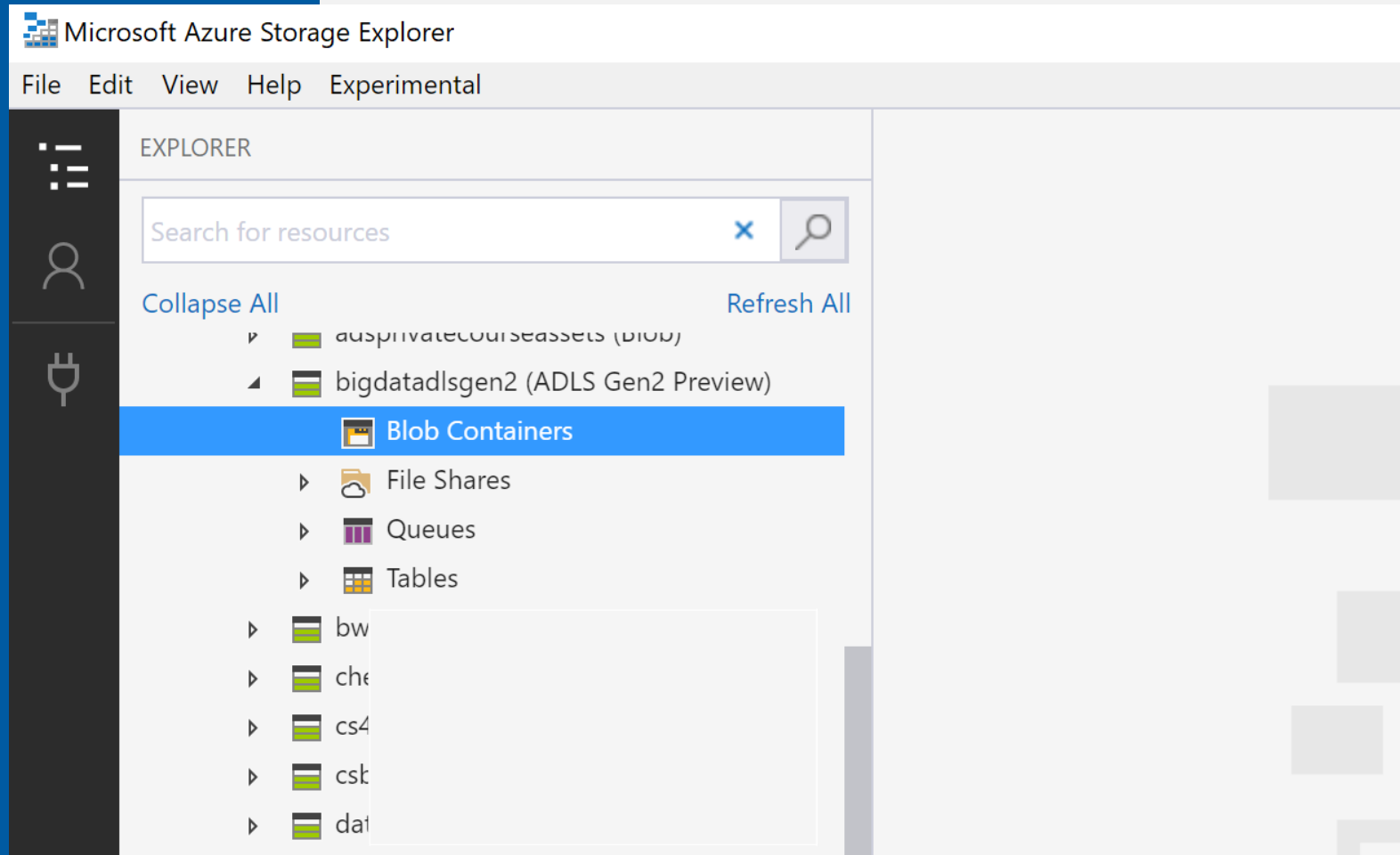
Create a Azure
Data Lake Store (Gen II)
using PowerShell.

Windows 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
```

Uploading data with Azure Storage Explorer



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
- Configure Integrations

Review Questions

- Q01 – Contoso has a Data Lake Storage Account (Gen 2). Which tool would be the most appropriate tool to perform a one time without the installation or configuration of a tool to upload of a single file?
 - A01 – Azure Portal.

- Q02 - Contoso has a Data Lake Storage Account (Gen 2). Which tool would be the most appropriate tool to perform a movement of hundreds of files from Amazon S3 to Azure Data Lake Storage ?
 - A02 – Azure Data Factory.

Lab: Working with Data Storage



Lab overview

In this lab, the students will be able to determine the appropriate storage type to implement against a given set of business and technical requirements. They will be able to create Azure storage accounts and Data Lake Storage account and explain the difference between Data Lake Storage generation 1 and generation 2. They will also be able to demonstrate how to perform data loads into the data storage of choice.

Lab objectives

After completing this lab, you will be able to:

1. Choose a data storage approach in Azure
2. Create an Azure Storage Account
3. Explain Azure Data Lake Storage
4. Upload data into Azure Data Lake

Lab scenario

You have been hired as a Senior Data Engineer to implement a technology solution that is part of a digital transformation project. The organization is migrating an Internet Information Services (IIS) that hosts the company website to Azure. The developers are in the process of transferring the web application and its logic to Azure Web Apps and they have asked you to prepare a data store for them that can be used to host the static images that are used on the website.

In addition, the information services department have informed you that their team is expanding and that they will soon be joined by data scientists that will start the process of building a predictive analytics solution. You have been asked to set up a solution that will be used to host the production environment of their work. In the first instance, you will assess what is the appropriate storage tier to create for the solution.

At the end of this work, you will have:

1. Choose a data storage approach in Azure
2. Create an Azure Storage Account
3. Explain Azure Data Lake Storage
4. Upload data into Azure Data Lake

Lab review

- Exercise 1 – Why did you select the option to store the image files?
- Exercise 2 – Apart from the Azure Portal, are there other methods to automate the deployment of storage accounts?
- Exercise 3 – How is the storage structure of Data Lake Storage Gen II different to Storage Accounts?
- Exercise 4 – Where in the Azure Portal would you find Microsoft Azure Storage Explorer?

Module Summary >

In this module, you have learned about:

- A data storage approach in Azure
- How to create an Azure Storage Account
- How Azure Data Lake Storage Gen2 works
- How to upload data into Azure Data Lake

Next steps >

After the course, consider watching [this video with James Baker and Lara Rubbelke that provides an overview of the enhancements of Azure Data Lake Storage Gen2](#)

