Collaborate Board

Instructions

How to Edit

Click Edit This Slide in the plugin to make changes.
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

DATA TECHNICIAN

CLOUD COMPUTING &
AZURE DATA FUNDAMENTALS (DP900)
-DAY 3-

Just IT

B2W group

Apprenticeships | Training | Recruitment

OBJECTIVES - DAY 3

- **Intent:**

Describe features and capabilities of Azure blob storage, Azure Data Lake Gen2, Azure file storage, and Azure table storage. Describe key features and capabilities of Azure Cosmos DB. Describe common features of large-scale analytical solutions and features of real-time analytics.

- **Implementation:**

Delivered using practical labs. Concepts explained through examples and exercises/quizzes

- **Impact:**

Gain the skills and confidence needed to understand data analysis processes and their types , preparing for more advanced studies or real-world applications. The impact measured through in-class mini-projects, labs & Quiz

Classroom Expectations

BE PREPARED, BE PROFESSIONAL, BE READY TO LEARN!

01 CAMERAS ON

Cameras should be switched on with an appropriate background, and with you visible at your screen. If you have technical/ personal difficulties, inform your tutor.

02 MICS ON MUTE

Place your mics on mute unless asking or answering a question. Use the 'raise hand' function in MS teams to gain attention.

03 APPROPRIATE LANGUAGE USED

Use appropriate and professional language at all times.

04 BE READY

Be focused, ensuring you all have all equipment ready and a drink by your side. Be prepared to learn without interruption.

05 ENGAGE- TAKE NOTES AND ASK QUESTIONS!

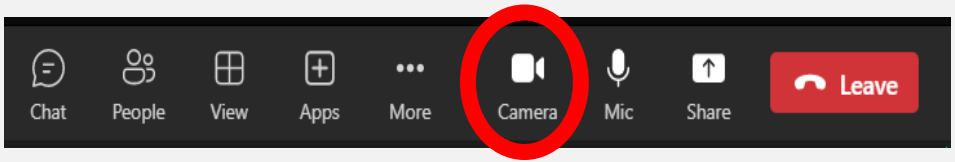
Take notes throughout your session, ask questions and support each other. Help create a positive learning environment for everyone.

CAMERA EXPECTATIONS

Before we start… Please have your camera switched on.

This is a mandatory requirement for the duration of the course.

- We have a duty of care to ensure you are in a safe learning environment.
- Our job is to prepare you for work in an age where in almost any job there is likely to be at least some aspect of remote engagement with your employer.
- Online learning is to be treated the same way as a classroom learning, you must be physically visible to interact with your tutor and other learners in order to create an engaging group dynamic.



Safeguarding Contact: besafe@justit.co.uk

NON-RELATIONAL DATA IN AZURE

Safeguarding Contact: besafe@justit.co.uk

AZURE DATA PROFESSIONAL ROLES

Data professional roles

Database Administrator

- Database provisioning, configuration and management
- Database security and user access
- Database backups and resiliency
- Database performance monitoring and optimization

Data Engineer

- Data integration pipelines and ETL processes
- Data cleansing and transformation
- Analytical data store schemas and data loads

Data Analyst

- Analytical modeling
- Data reporting and summarization
- Data visualization

© Copyright Microsoft Corporation. All rights reserved.

Collaborate Board

Instructions

Job Search Activity: Search for **Data Engineer** and **Data Analyst** roles, then list the main skills required for each on the discussion board using the shared Nearpod link in the chat.

How to Edit

Click Edit This Slide in the plugin to make changes.

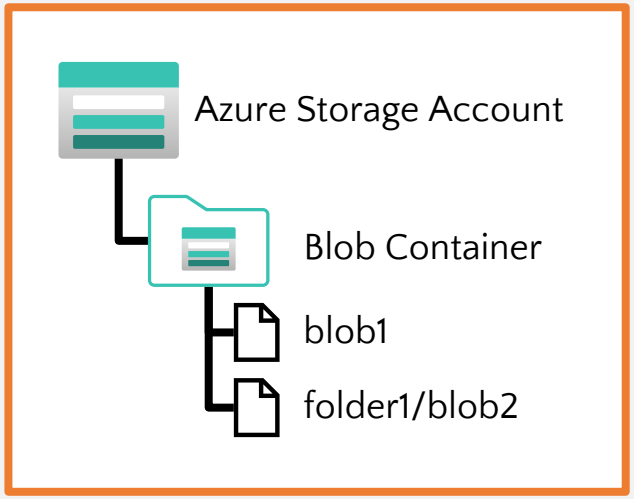
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

AZURE BLOB RECAP

- Storage for data as binary large objects (BLOBs)
  - Block blobs
    - Large, discrete, binary objects that change infrequently
    - Blobs can be up to 4.7 TB, composed of blocks of up to 100 MB
      - A blob can contain up to 50,000 blocks
  - Page blobs
    - Used as virtual disk storage for VMs
    - Blobs can be up to 8 TB, composed of fixed sized-512 byte pages
- Append blobs
  - Block blobs that are used to optimize append operations
  - Maximum size just over 195 GB – each block can be up to 4 MB

AZURE BLOB RECAP

- Per-blob storage tiers:
  - Hot – Highest cost, lowest latency
  - Cool – Lower cost, higher latency
  - Archive – Lowest cost, highest latency



Blobs can be organised in virtual directories, but each path is considered a single blob in a flat namespace – folder level operations are not supported

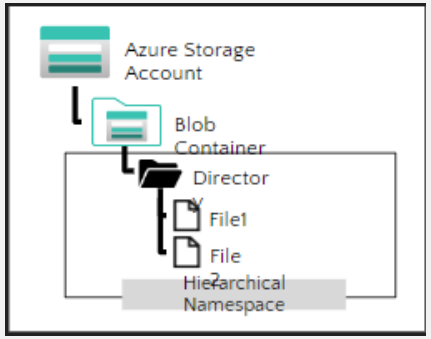
AZURE DATA LAKE STORE GEN 2

Distributed file system built on Blob Storage:

- Combines Azure Data Lake Store Gen 1 with Azure Blob Storage for large-scale file storage and analytics
- Enables file and directory level access control and management
- Compatible with common large scale analytical systems

Enabled in an Azure Storage account through the Hierarchical Namespace option

- Set during account creation
- Upgrade existing storage account
  - One-way upgrade process

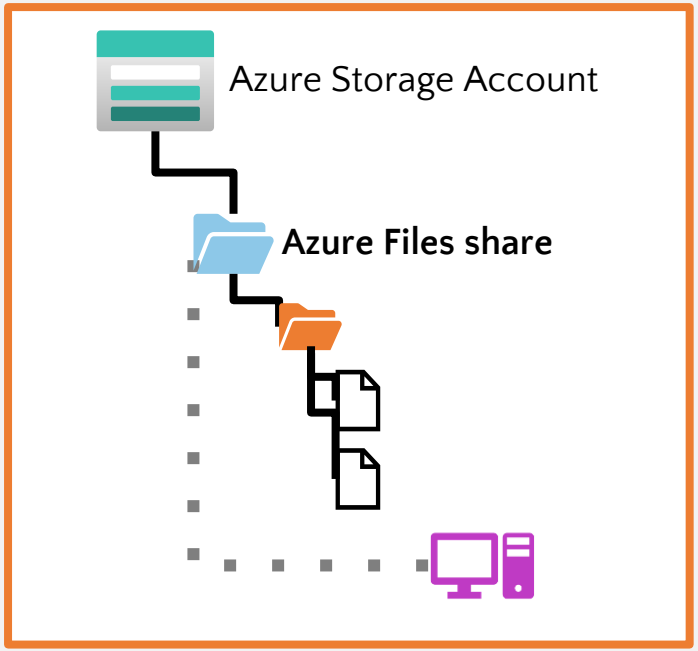


AZURE FILES

File shares in the cloud that can be accessed from anywhere with an internet connection

- Support for common file sharing protocols:
  - Server Message Block (SMB)
  - Network File System (NFS) – requires premium tier

- Data is replicated for redundancy and encrypted at rest

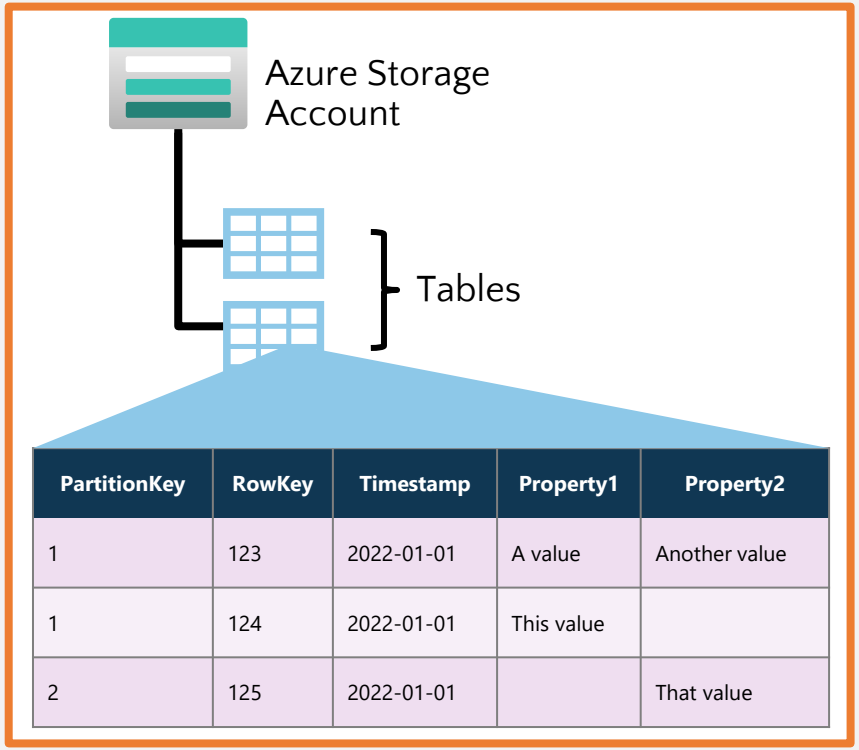


AZURE TABLE STORAGE

Key-Value storage for application data

- Tables consist of key and value columns
  - Partition and row keys
  - Custom property columns for data values
    - A Timestamp column is added automatically to log data changes
- Rows are grouped into partitions to improve performance
- Property columns are assigned a data type, and can contain any value of that type
- Rows do not need to include the same property columns

AZURE TABLE STORAGE



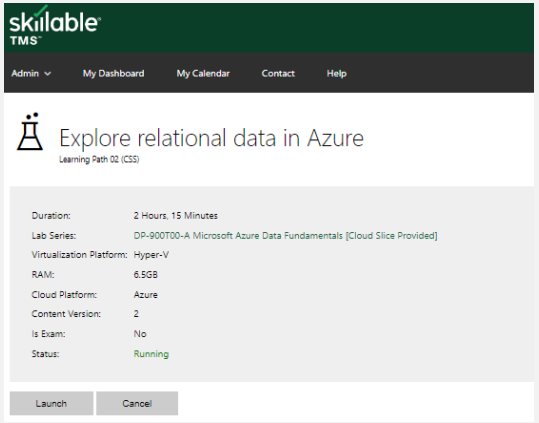
PartitionKey	RowKey	Timestamp	Property1	Property2
1	123	2022-01-01	A value	Another value
1	124	2022-01-01	This value	
2	125	2022-01-01		That value

AZURE VIDEO ACTIVITY

Use the Nearpod link shared in the chat to open this video activity.

While watching the YouTube video **“Azure Storage Account for Beginners,”** you’ll answer a series of questions.

LAB + WORKBOOK ACTIVITY
20 MINUTES



Within your workbook, please complete

Day 3 task 1

If completed early please work your way through MS Learn Azure content

- Complete the 1st lab in Skillable
- Paste a screenshot of the completed lab in the box provided

Safeguarding Contact: besafe@justit.co.uk

Collaborate Board

Instructions

How to Edit

Click Edit This Slide in the plugin to make changes.
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

AZURE COSMOS DB

Safeguarding Contact: besafe@justit.co.uk

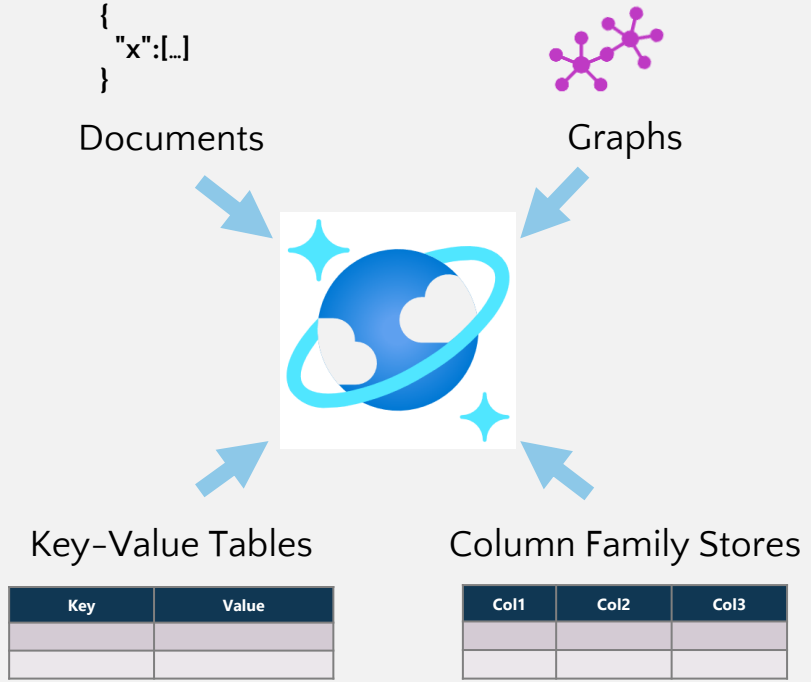
NoSQL Databases

Simply Explained

WHAT IS AZURE COSMOS DB?

A multi-model, global-scale NoSQL database management system:

- Support for multiple storage APIs
- Real time access with fast read and write performance
- Enable multi-region writes to replicate data globally; enabling users in specified regions to work with a local replica



AZURE COSMOS APIs

Azure Cosmos DB for NoSQL

- Native API for Cosmos DB

```
SELECT *
FROM customers c
WHERE c.id =
"joe@litware.com"
```

```
{
   "id": "joe@litware.com",
   "name": "Joe Jones",
   "address": {
       "street": "1 Main St.",
       "city": "Seattle"
   }
}
```

Azure Cosmos DB for MongoDB

- Compatibility with MongoDB

```
db.products.find({
id: 123})
```

```
{
   "id": 123,
   "name": "Hammer",
   "price": 2.99}
}
```

Azure Cosmos DB for PostgreSQL

- Compatibility with PostgreSQL

id	name	dept	manager
1	Sue Smith	Hardware	Joe Jones
2	Ben Chan	Hardware	Sue Smith

Azure Cosmos DB for Table

- Key-value storage API
- Compatible with Azure Table Storage

PartitionKey	RowKey	Name
1	123	Joe Jones
1	124	Samir Nadoy

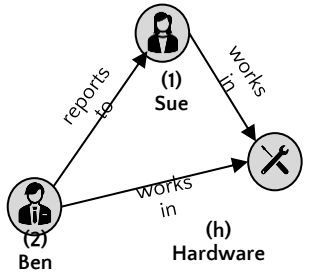
Azure Cosmos DB for Apache Cassandra

- Compatibility with Apache Cassandra

id	name	dept	manager
1	Sue Smith	Hardware	
2	Ben Chan	Hardware	Sue Smith

Azure Cosmos DB for Apache Gremlin

- Used to work with graph data
- vertices are connected via relationships (edges)



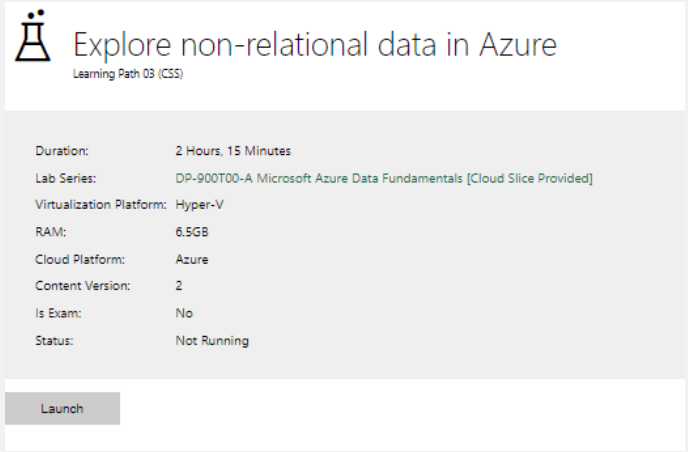
Safeguarding Contact: besafe@justit.co.uk

LAB + WORKBOOK ACTIVITY
40 MINUTES

Within your workbook, please complete

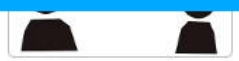
Day 3 task 2

If completed early please work your way through [MS Learn Azure content]



- Complete the 4th lab in Skillable
- Paste the completion in the box provided

Collaborate Board



Instructions

How to Edit

Click Edit This Slide in the plugin to make changes.
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

LARGE-SCALE DATA ANALYTICS

Safeguarding Contact: besafe@justit.co.uk

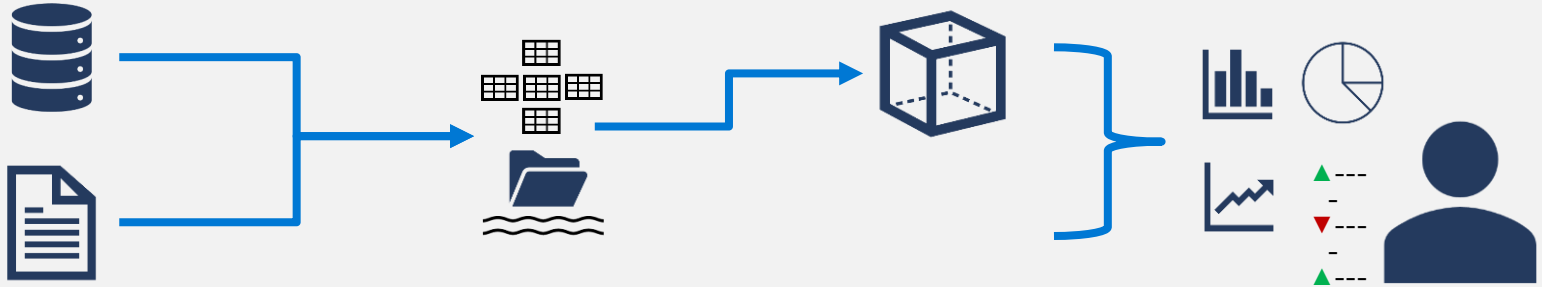
LARGE DATA ANALYTICS IN AZURE VIDEO ACTIVITY

Use the Nearpod link shared in the chat to open this video activity.

While watching the YouTube video **“Big Data & Analytics”** you’ll answer a series of questions.

ELEMENTS OF LARGE-SCALE DATA ANALYTICS

Data ingestion and processing	Analytical data store	Analytical data model	Data visualization



- *Extract, Transform, and Load* (ETL) or *Extract, Load, and Transform* (ELT) orchestration
- Distributed processing to cleanse and restructure data at scale
- Batch and real-time data processing

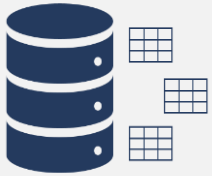
- Flexible, scalable file storage in a *data lake*
- Relational tables in a *data lakehouse* or *data warehouse*

- Semantic models for analytical entities
- Often in the form of aggregated *cubes* that summarize numeric values across one or more *dimensions*

- Reports
- Charts
- Dashboards

Safeguarding Contact: besafe@justit.co.uk

DATA PROCESSING IN LARGE-SCALE DATA ANALYTICS



Relational Database

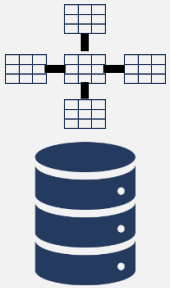
- Well established model for relational data storage and processing
- Comprehensive SQL language support for querying and data manipulation

Apache Spark

- Open-Source platform for scalable, distributed data processing
- Multi-language data processing code (Python, Scala, Java, SQL, …)

Safeguarding Contact: besafe@justit.co.uk

DATA STORE ARCHITECTURES



Data Warehouse

- Data is stored in a relational database and queried using a SQL query engine
- Tables are *denormalized* for query optimization
  - Typically as a star or snowflake schema of numeric *facts* that can be aggregated by *dimensions*

Data Lakehouse

- Data files are stored in a distributed file system (a *data lake*) and typically processed using Apache Spark
- Metadata is used to define tables that provide a relational SQL interface to the file data
  - Commonly, a *delta lake* format is used to provide transactional database functionality

PAAS DATA ANALYTICS

Azure Synapse Analytics

- Unified solution for relational data warehouse and data lake analytics
- Scalable processing and querying through multiple analytics runtimes
  - Synapse SQL
  - Apache Spark
  - Synapse Data Explorer
- Interactive experience in Azure Synapse Studio
- Built-in pipeline integration for data ingestion and processing

Use for a single, unified large-scale analytical solution on Azure

Azure Databricks

- Azure-based implementation of Databricks cloud analytics platform
- Scalable Spark and SQL querying for data lake analytics
- Interactive experience in Azure Databricks workspace
- Use Azure Data Factory to implement data ingestion and processing pipelines

Use to leverage Databricks skills and for cloud portability

Azure HDInsight

- Azure-based implementation of common Apache "big data" frameworks built on a data lake
  - Hadoop - Query data lake files using Hive tables
  - Spark – Use Spark APIs to query data, and abstract underlying file storage as tables
  - Kafka – Real-time event processing
  - Storm – Stream processing
  - HBase – NoSQL data store

Use when you need to support multiple open-source platforms

SAAS DATA ANALYTICS WITH MS FABRIC

Microsoft Fabric

Data Integration	Data Engineering	Data Warehouse	Data Science	Real-Time Analytics	Business Intelligence	Applied Observability
Data Factory	Synapse	Synapse	Synapse	Synapse	Power BI	Data Activator

Unified data foundation

OneLake

Unified

SaaS product experience	Security and governance	Compute	Storage	Business model

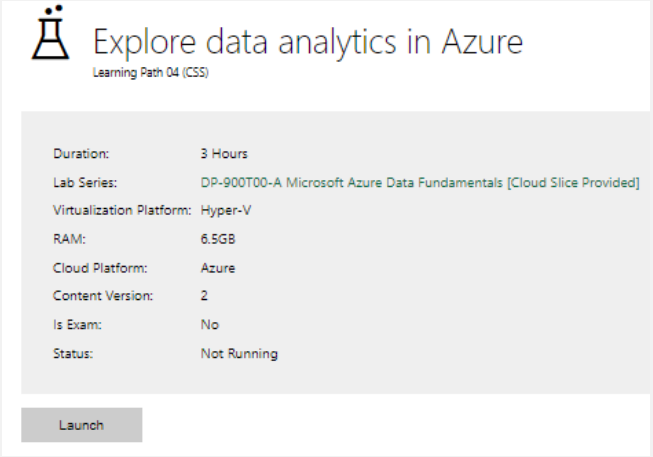
Safeguarding Contact: besafe@justit.co.uk

LAB + WORKBOOK ACTIVITY
40 MINUTES

Within your workbook, please complete

Day 3 task 3 – part 1

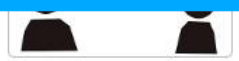
If completed early please work your way through [MS Learn Azure content]



- Complete the 5th lab in Skillable
- Paste the completion in the box provided

Safeguarding Contact: besafe@justit.co.uk

Collaborate Board



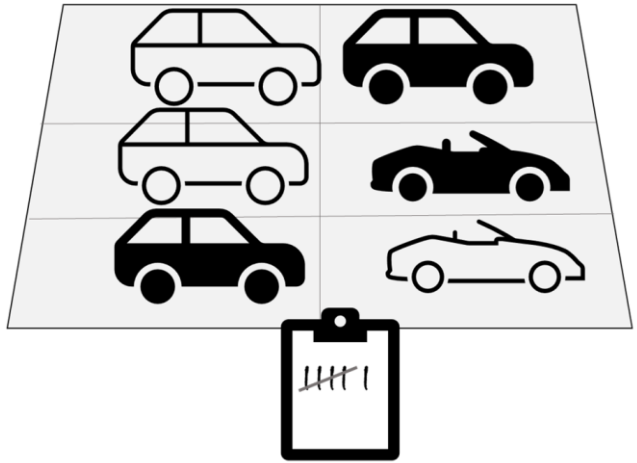
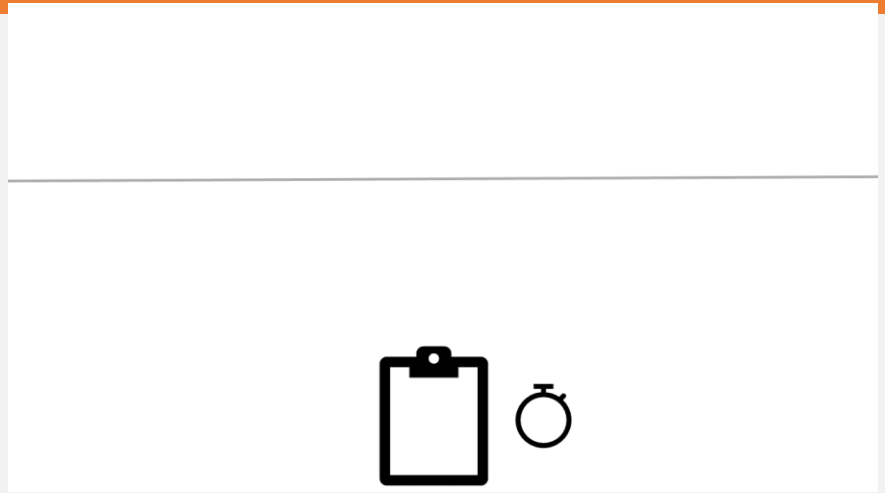
Instructions

How to Edit

Click Edit This Slide in the plugin to make changes.
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

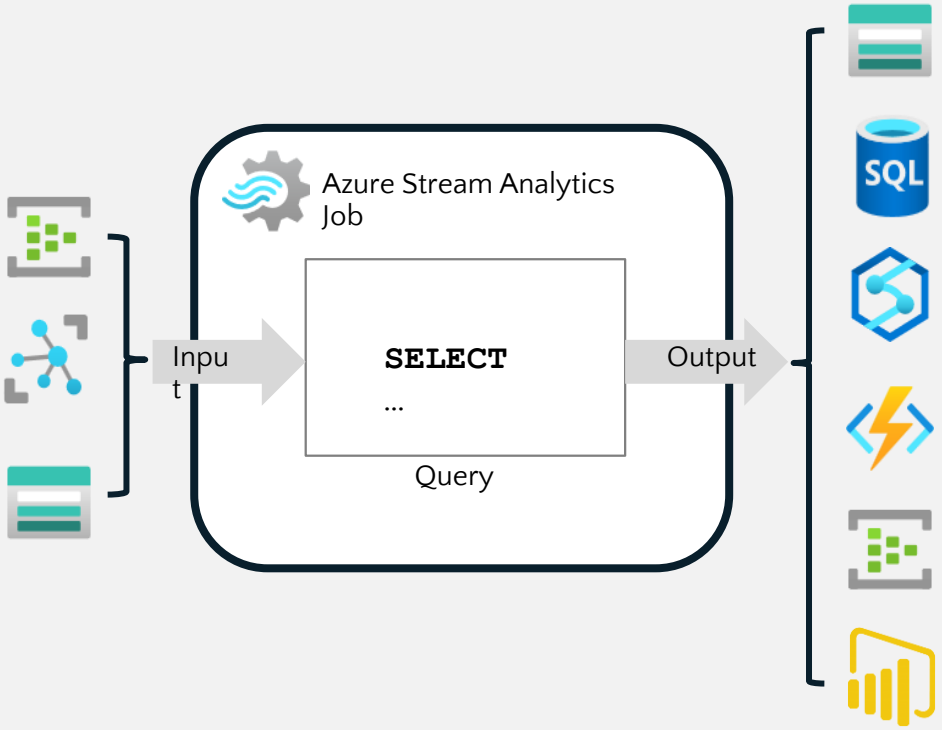
STREAMING & REAL-TIME ANALYTICS

BATCH VS STREAM PROCESSING

Batch processing	Stream processing
	
Data is collected and processed at regular intervals	Data is processed in (near) real-time as it arrives

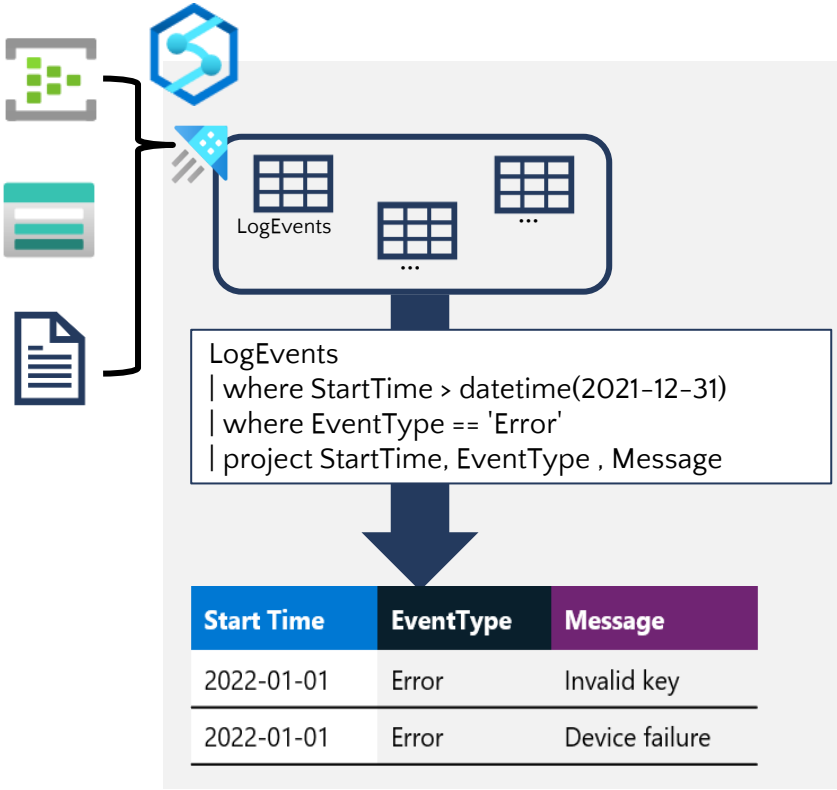
REAL-TIME DATA PROCESSING

- Create an individual Azure Stream Analytics *job* or an Azure Stream Analytics *cluster*
- Ingest data from an *input*, such as:
  - Azure Event Hubs
  - Azure IoT Hub
  - Azure Blob Storage
- Process data with a perpetual *query*
- Send results to an *output*, such as:
  - Azure Blob Storage
  - Azure SQL Database
  - Azure Synapse Analytics
  - Azure Function
  - Azure Event Hubs
  - Power BI



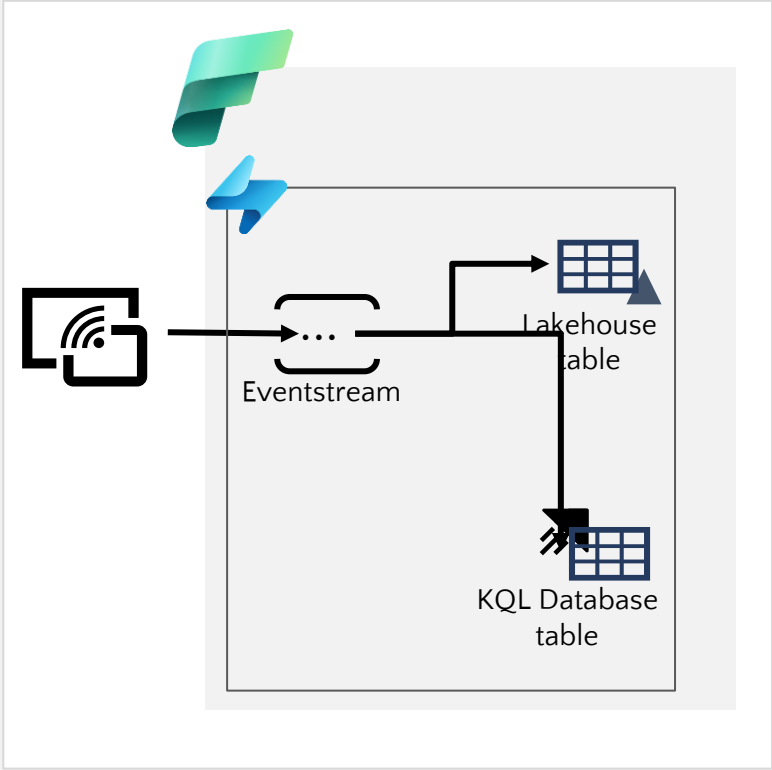
REAL-TIME LOG & TELEMETRY ANALYSIS THROUGH DATA EXPLORER

- High throughput, scalable service for batch and streaming data
  - **Azure Data Explorer** dedicated service
  - **Azure Synapse Data Explorer** runtime in Azure Synapse Analytics
- Data is ingested from streaming and batch sources into tables in a database
- Tables can be queried using *Kusto Query Language* (KQL):
  - Intuitive syntax for read-only queries
  - Optimized for raw telemetry and time-series data



REAL-TIME ANALYTICS IN FABRIC

- Support for continuous data ingestion from multiple sources
- Capture streaming data in an **eventstream**
- Write real-time data to a table in a Lakehouse or a KQL database
- Query real-time data using SQL or KQL
- Build real-time visualisations

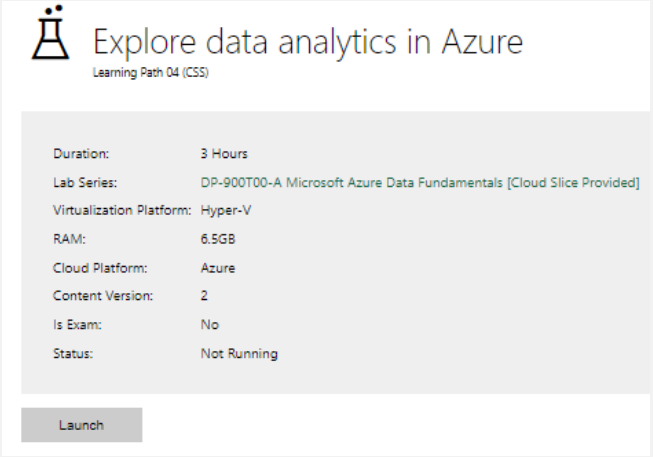


LAB + WORKBOOK ACTIVITY
40 MINUTES

Within your workbook, please complete

Day 3 task 3 – part 2

If completed early please work your way through [MS Learn Azure content]



- Complete the 5th lab in Skillable
- Paste the completion in the box provided

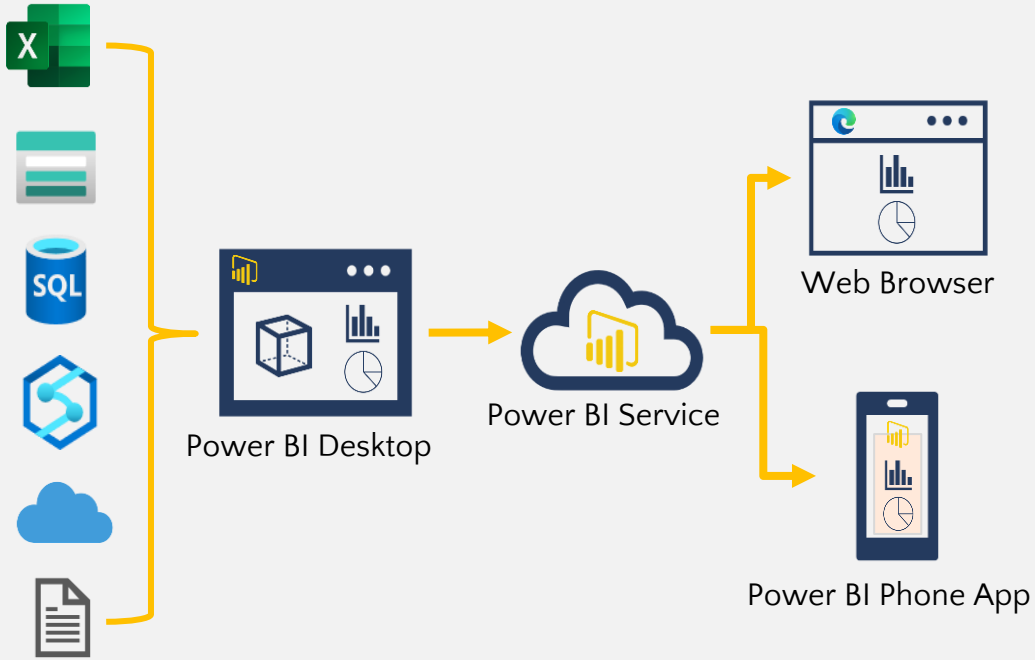
Safeguarding Contact: besafe@justit.co.uk

DATA VISUALISATION

Safeguarding Contact: besafe@justit.co.uk

DATA VISUALISATION WITH POWER BI

- Start with Power BI Desktop
  - Import data from one or more sources
  - Define a data model
  - Create visualizations in a report
- Publish to Power BI Service
  - Schedule data refresh
  - Create dashboards and apps
  - Share with other users
- Interact with published reports
  - Web browser
  - Power BI phone app



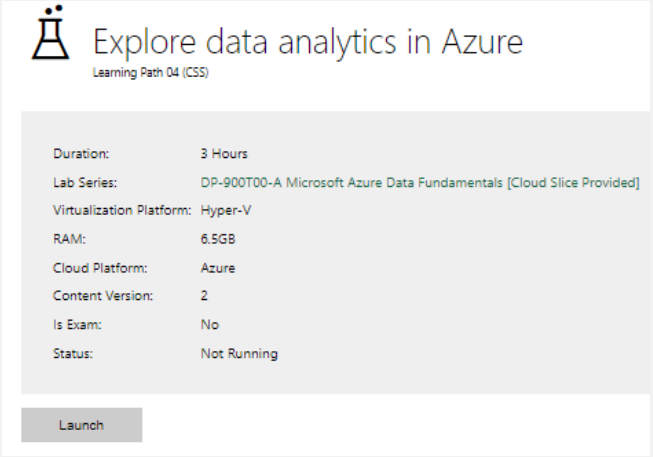

WHAT IS
MICROSOFT FABRIC?

LAB + WORKBOOK ACTIVITY
40 MINUTES

Within your workbook, please complete

Day 3 task 3 – part 3

If completed early please work your way through [MS Learn Azure content]



- Complete the 5th lab in Skillable
- Paste the completion in the box provided

Safeguarding Contact: besafe@justit.co.uk

Open Ended Question

Ready? Enter your answer here.

How to Edit

Click Edit This Slide in the plugin to make changes.

Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

OBJECTIVES - DAY 3 RECAP

- **Intent:**

Describe features and capabilities of Azure blob storage, Azure Data Lake Gen2, Azure file storage, and Azure table storage. Describe key features and capabilities of Azure Cosmos DB. Describe common features of large-scale analytical solutions and features of real-time analytics.

- **Implementation:**

Delivered using practical labs. Concepts explained through examples and exercises/quizzes

- **Impact:**

Gain the skills and confidence needed to understand data analysis processes and their types , preparing for more advanced studies or real-world applications. The impact measured through in-class mini-projects, labs & Quiz

Collaborate Board

Instructions

How to Edit

Click Edit This Slide in the plugin to make changes.
Don't have the Nearpod add-on? Open the "Add-ons" menu in Google Slides to install.

OBJECTIVES - DAY 4

- **Intent:**

To further extend your knowledge on Azure, complete an Azure practice assessment and complete a business proposal activity. You'll be able to identify the key questions of the exam and be able to sell the services of Azure to a 'client'.

- **Implementation:**

Delivered using practical labs. Concepts explained through examples and exercises/quizzes

- **Impact:**

Gain the skills and confidence needed to understand data analysis processes and their types , preparing for more advanced studies or real-world applications. The impact measured through in-class mini-projects, labs & Quiz