



Describing core data concepts

# Revision

# Core data concepts

Who amongst the following would be responsible for tasks such as

1. Setting up a database
2. Granting access to tables for users
3. Setting up back and recovery for the database

Data Engineer

Data Analyst

Machine Learning Engineer

Database Administrator

# Core data concepts

Who amongst the following would be responsible for tasks such as

1. Identifying trends in data
2. Making forecasts based on historical data

Data Engineer

Data Analyst

Machine Learning Engineer

Database Administrator

# Core data concepts

The following belongs to which data type

```
{
  "CustomerID": 1,
  "Title": "Mr.",
  "FirstName": "Orlando",
  "MiddleName": "N.",
  "LastName": "Gee",
  "EmailAddress": "orlando0@adventure-works.com",
  "Phone": "245-555-0173"
},
{
  "CustomerID": 2,
  "Title": "Mr.",
  "FirstName": "Keith",
  "MiddleName": null,
  "LastName": "Harris",
  "EmailAddress": "keith0@adventure-works.com",
  "Phone": "170-555-0127"
},
```

Structured

Relational

Semi-structured

Unstructured

# Core data concepts

The following belongs to which data type

	CustomerID	NameStyle	Title	FirstName	MiddleName	LastName
1	1	0	Mr.	Orlando	N.	Gee
2	2	0	Mr.	Keith	NULL	Harris
3	3	0	Ms.	Donna	F.	Carreras
4	4	0	Ms.	Janet	M.	Gates
5	5	0	Mr.	Lucy	NULL	Harrington
6	6	0	Ms.	Rosmarie	J.	Carroll

Structured

Relational

Semi-structured

Unstructured

# Core data concepts

The following belongs to which data type

```
<?xml version="1.0" encoding="UTF-8" ?>
<root>
  <0>
    <CustomerID>1</CustomerID>
    <Title>Mr.</Title>
    <FirstName>Orlando</FirstName>
    <MiddleName>N.</MiddleName>
    <LastName>Gee</LastName>
    <EmailAddress>orlando0@adventure-works.com</EmailAddress>
    <Phone>245-555-0173</Phone>
  </0>
  <1>
    <CustomerID>2</CustomerID>
    <Title>Mr.</Title>
    <FirstName>Keith</FirstName>
    <MiddleName />
    <LastName>Harris</LastName>
    <EmailAddress>keith0@adventure-works.com</EmailAddress>
    <Phone>170-555-0127</Phone> </1>
</root>
```

Structured

Relational

Semi-structured

Unstructured

# Core data concepts

Which of the following is the correct file format based on the given definition

This is a row-based format. This file format was created by Apache. This is a good format for compressing data.

Avro

ORC

CSV

Parquet



Relational Data

# Revision

# Relational Data

Which of the following is an example of a Data Manipulation Language statement

ALTER

UPDATE

CREATE

DROP

# Relational Data

Which of the following A.C.I.D properties maps to the following description

Either the entire transaction completes or fails.

Consistency

Durability

Atomicity

Isolation

# Relational Data

The Azure SQL Database maps to which Cloud Service Model

1

Infrastructure as  
a service

2

Platform as a  
service

3

Software as a  
service

# Relational Data

When using Platform as a service - Azure SQL database

The entire infrastructure is managed for you

You get built-in high availability.

You can configure scaling for your infrastructure.

You get features such as automated backup.

You have various offerings - Azure SQL Managed Instance.

Easy to get started with.

# Relational Data

1

A relational database is used for storing structured data.

2

We can use the structured query language to query for data in a relational database.

3

In a table we can define primary keys to uniquely identify a row in the table. We can use foreign keys to enforce relationships between tables.

4

In a table you first define the schema before you can start adding rows of information to the table.

# Relational Data

If a company wants to migrate their Microsoft SQL Server instance to Azure and they want to use a service on Azure that provides an almost 100% compatibility with their current setup, which of the following should they consider. They also want to minimize admin tasks.

1

Azure SQL  
Database

2

Azure SQL  
Managed  
Instance

3

Microsoft SQL  
Server on Azure  
virtual machine

# Relational Data

## View

This is a virtual table wherein the data is defined by a query

## Clustered Index

This sorts and stores the data rows in the table based on the key values

## Nonclustered index

Here a separate index or structure is maintained in addition to the data rows.

## Stored Procedure

This is a group of SQL statements that can be executed as a unit on the database server.

# Relational Data

Store data  
based on  
dynamic  
schema

Store images  
and videos

Store key-  
value pairs

Store  
structured  
data

Do you use a relational database to ...

# Relational Data

SELECT

Retrieve rows from a database

INSERT

Add rows into a table

UPDATE

Modify column values for a row in a table

WHERE

Specify a search criteria for rows that need to be returned

GROUP BY

Segregates the query results into groups of rows

JOIN

You can retrieve data from multiple tables

CREATE

You can create a new object in the database

# Relational Data

Database Normalization - Helps to organize the data in a database. Here we create multiple tables and establish relationships between tables based on rules.

1

Helps to reduce data redundancy.

2

Helps to improve the overall integrity of data.

3

Here we use JOINS to get results from multiple tables.



# Relational Data

Use Azure Data Studio when you want to work with multiple database engines on different platforms.

As a database administrator, if you want to manage the database like backup and restore operations, preferable to use Microsoft SQL Server Management Studio.



Non-Relational Data

# Revision

# Non-Relational Data

Non-relational data store offers more flexibility.

1

Here there is no need of defining a fixed schema.

2

We don't need to define the schema beforehand.

3

We don't need to maintain relationships between entities.

# Non-Relational Data

1

Azure Storage - Blob - This is an object storage service. Ideal for storage of unstructured and semi-structured data.

2

Azure Storage - File - This is a File share service. Supports mapping and the SMB protocol.

3

Azure Storage - Table - This is used to store key-value pairs. Allows for easy retrieval of values based on keys.

4

Azure Storage - Queue - This is a simple messaging service.

# Non-Relational Data

## Hot tier

Ideal for data that is accessed frequently. Has higher data storage costs but less transactional costs.

## Cool tier

Ideal for data infrequently accessed. The data should be stored for at least 30 days. Less storage costs but higher transactional costs.

## Cold tier

Ideal for data rarely accessed. Data should reside for at least 90 days.

## Archive

Ideal for data that is rarely accessed like backups. Has the most latency, but least storage option.

# Non-Relational Data

## Concepts in Azure Cosmos DB

1

Create an account -  
based on the API.

2

Then create a  
database and a  
container. The  
throughput can  
be set at both  
levels.

3

Partition Key  
used to partition  
data and  
optimize queries  
based on the  
key.

# Non-Relational Data

You want to host a data store on Azure. The data store would be used to store JSON-based documents. Which of the following API would you use in Azure Cosmos DB?

TABLE API

NoSQL API

Cassandra API

Gremlin API

# Non-Relational Data

You want to host a data store on Azure. The data store would be used to store JSON-based documents. Which of the following API would you use in Azure Cosmos DB?

TABLE API

NoSQL API

Cassandra API

Gremlin API

# Non-Relational Data

What about having data in multiple regions.

1

Azure Storage -  
Geo-redundancy.  
Read in many  
regions. But write  
only one region.

2

Azure Cosmos  
DB - Can have  
multiple read  
regions.

3

Azure Cosmos  
DB - Can also  
have multiple  
write regions.



Analytical Workloads

# Revision

# Analytical Workloads

You are planning on making use of Azure Stream Analytics. Which language is used for processing the incoming streams of data?

KQL

JSON

SQL

XML

# Analytical Workloads

You are planning on making use of Azure Synapse Analytics - Data Explorer pool. Which language is used for querying data in the pool?

KQL

JSON

SQL

XML

# Analytical Workloads

Which of the following services has the ability to make use of Apache Spark clusters? Choose 2 answers.

Azure Synapse Analytics

Azure Stream Analytics

Azure SQL Database

Azure Databricks

# Analytical Workloads

1

Data warehousing architecture - Here data from multiple source streams are loaded into a data warehouse.

2

Extract , Transform and Load - Here data is extracted, cleaned, transformed and loaded into the target data store.

3

Extract, Load and Transform - Here the data is loaded directly into the target store. The data is transformed at the destination.

4

SQL data warehouse - Your data would be structured and in the denormalized form.

# Analytical Workloads

You need to analyze historical log data based over the past 3 months. This is an example of which type of workload?

1

Batch

2

Stream

3

OLTP

# Analytical Workloads

You need to analyze real-time telemetry data? This is an example of which type of workload?

1

Batch

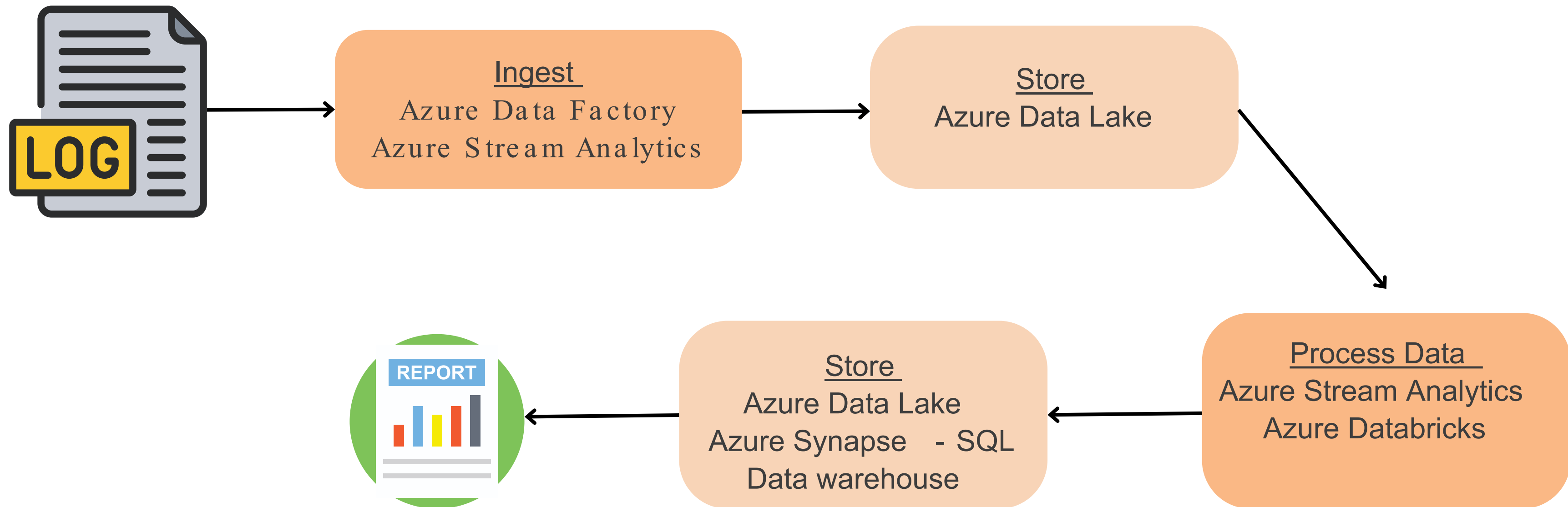
2

Stream

3

OLTP

# Analytical Workloads



# Analytical Workloads - Power BI

## Power BI Dashboards

This is a single page in which you can create visualizations

## Reports

In the dashboard, you can attach multiple reports with each report connecting to a different model.

## Power BI service

Dashboards are only available in the Power BI service and not in the desktop.

## Tiles

Text boxes, images, videos, streaming data and web content.

# Analytical Workloads - Data Factory

Pipelines

This is a collection of Activities.

Activity

This represents something that needs to be done.

Data set

This is the data structure in the data stores.

Linked Service

This is used to connect to the data stores.

Integration Runtime

This is the compute environment.

Trigger

How do you trigger the pipeline.

Control flow

Orchestration of Activities - For each activity.

# Analytical Workloads

Which of the following visualization can be used to display the major contributors to a major result or value.

Key influencers chart

Scatter chart

TreeMaps

Waterfall chart

# Analytical Workloads

Which of the following visualization can be used to display points at the intersection of x and y values.

Key influencers chart

Scatter chart

TreeMaps

Waterfall chart