

The background is a complex digital illustration. It features a perspective view of a tunnel or corridor that recedes into the distance. The walls and floor of the tunnel are composed of a dense, mosaic-like pattern of small, glowing blue and orange squares. The perspective lines of the tunnel converge towards a bright, glowing orange and yellow light source at the far end, creating a sense of depth and movement. The overall color palette is dominated by deep blues, bright oranges, and yellows, giving it a high-tech, futuristic feel.

# Data Lake

Course overview  
Data Lake & Mesh  
Concepts  
Hands on





## **Intro to BigData**

Overview of BigData  
BigData Ecosystems  
Setting up BigData  
env.

## **BigData with Kafka**

Real-time data  
processing  
Key components of  
Kafka  
Hands on

## **Apache Spark**

Architecture  
Spark streaming  
and jobs  
Hands on

## **Data Lake**

Design &  
Architecture  
Pipeline  
Hands on

What is a Data Lake?

# What is a Data Lake?



Data Ingestion Layer  
Storage Layer  
Processing layer  
Cataloging and metadata  
Access layer

# Data lake vs Data warehouse

Feature	Data Lake	Data Warehouse
Purpose	Big Data, ML/AI, analytics	BI, structured analytics
Data Structure	Raw, varied data (structured/unstructured)	Structured data only
Scalability	Highly scalable, low cost	Scalable but higher cost
Performance	Slower for querying	Optimized for querying
Schema	Schema-on-Read	Schema-on-Write
Cost	Lower storage costs	Higher storage costs
Tools	Hadoop, Spark, S3	Snowflake, Redshift



# Set up Google Cloud Command line

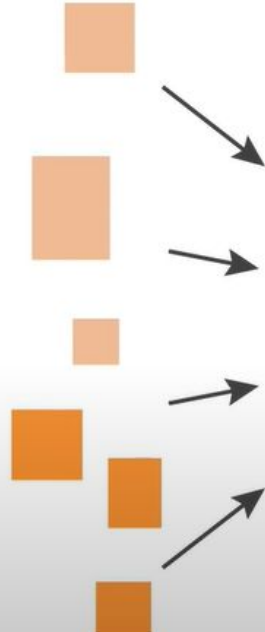
Set Up a Data Lake  
Ingest Data  
Query and Analyze Data

Field Name	Type	Mode
flight_id	INT64	NULLABLE
flight_date	DATE	NULLABLE
airline	STRING	NULLABLE
origin	STRING	NULLABLE
destination	STRING	NULLABLE
duration	INT64	NULLABLE
delay	INT64	NULLABLE
on_time	BOOL	NULLABLE

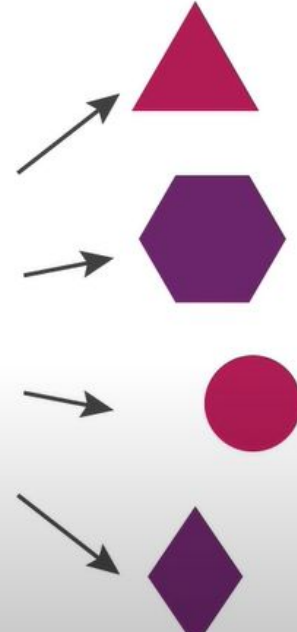
Is BigQuery Data Lake?

# Centralized | Monolithic

Ubiquitous Data



Innovation Agenda





# Data Mesh

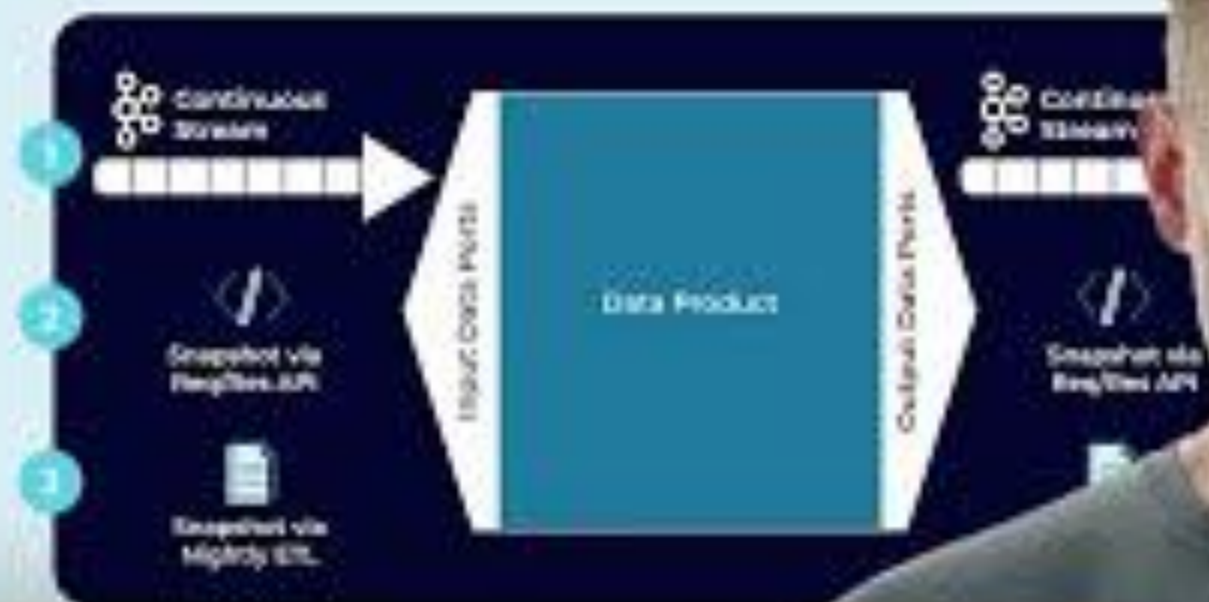


# WHAT IS **DATA MESH?**



Feature	Data Lake	Data Mesh
Architecture	Centralized repository for all raw and processed data.	Decentralized, domain-driven data ownership and architecture.
Ownership	Managed by a centralized data engineering team.	Owned by individual domain teams (e.g., marketing, sales).
Governance	Centralized governance for data quality, access, and policies.	Federated governance across domains with global standards.
Scalability	Can face bottlenecks as data volume grows.	Scales across domains as each team manages its own data.
Focus	Focuses on storage and access.	Focuses on creating data as a product for usability.

# Data as a Product



## DATA MESH 101



Confluent



# Data Mesh Demo Setup

# Implementation



# DATA MESH 101



# Exploring Data Mesh

Create a Data Product

Next steps.....



AI-Native Data Architectures  
Data Spaces and Ecosystems  
Decentralized Data Architectures

Have Questions?

**Working with DATA  
is an ART**

