



Learn how to build a real-time analytics solution

José Mendes | @ZeMiguelMendes



Senior Data Analytics Consultant

blogs.adatis.co.uk/josemendes

@ZeMiguelMendes



Agenda

Real-Time
Analytics
Solutions

Lambda
Architecture

Demo

Databricks Delta
Architecture

Demo



Agenda

Real-Time
Analytics
Solutions

Lambda
Architecture

Demo

Databricks Delta
Architecture

Demo



Real-Time Analytics Solutions

Over 2.5 quintillion bytes of data are created every single day, and it's only going to grow from there. By 2020, it's estimated that 1.7MB of data will be created every second for every person on earth



Real-Time Analytics Solutions

A growing demand for faster access to data for analysis

Retail giants making real-time price changes

Manufacturing plants using anomaly detection to determine potential problems on the assembly line before they happen

Mine and gas companies using hi-tech drill sensor readings to know precisely what is happening hundreds of feet below the earth's surface as it's happening



Agenda

Real-Time
Analytics
Solutions

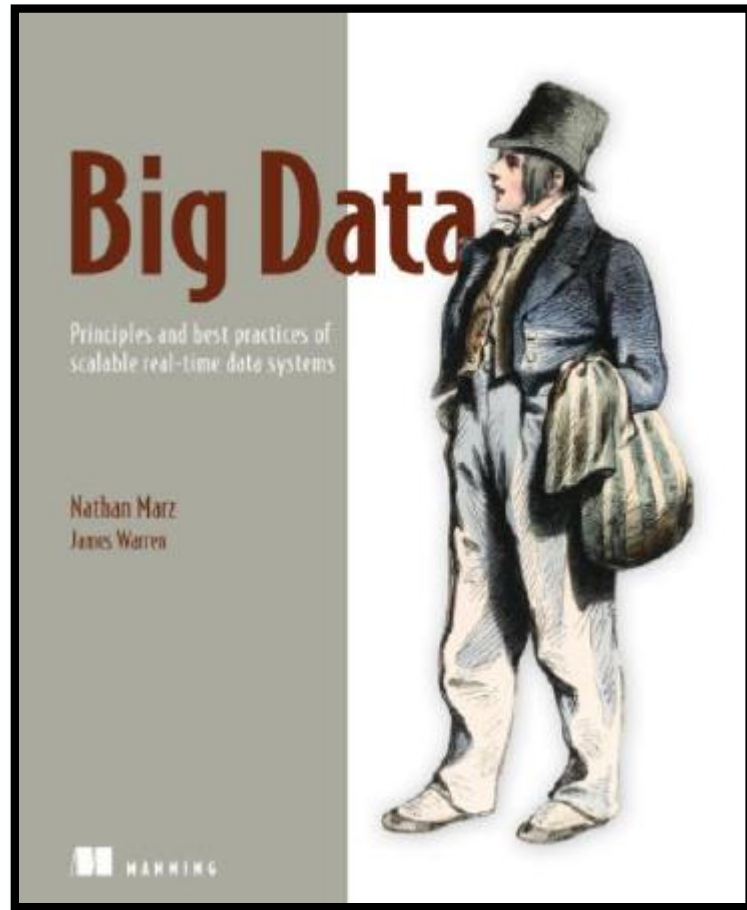
Lambda
Architecture

Demo

Databricks Delta
Architecture

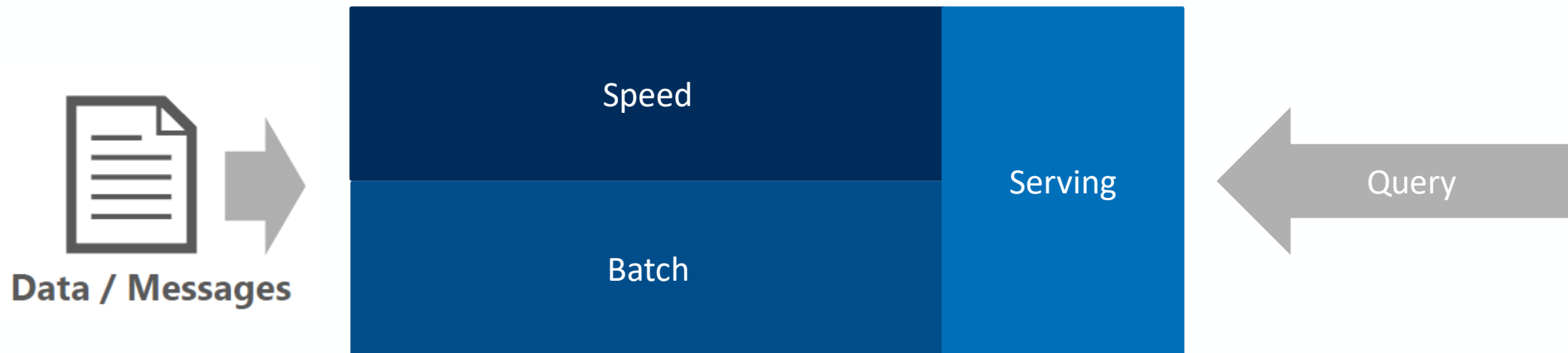
Demo

Lambda Architecture

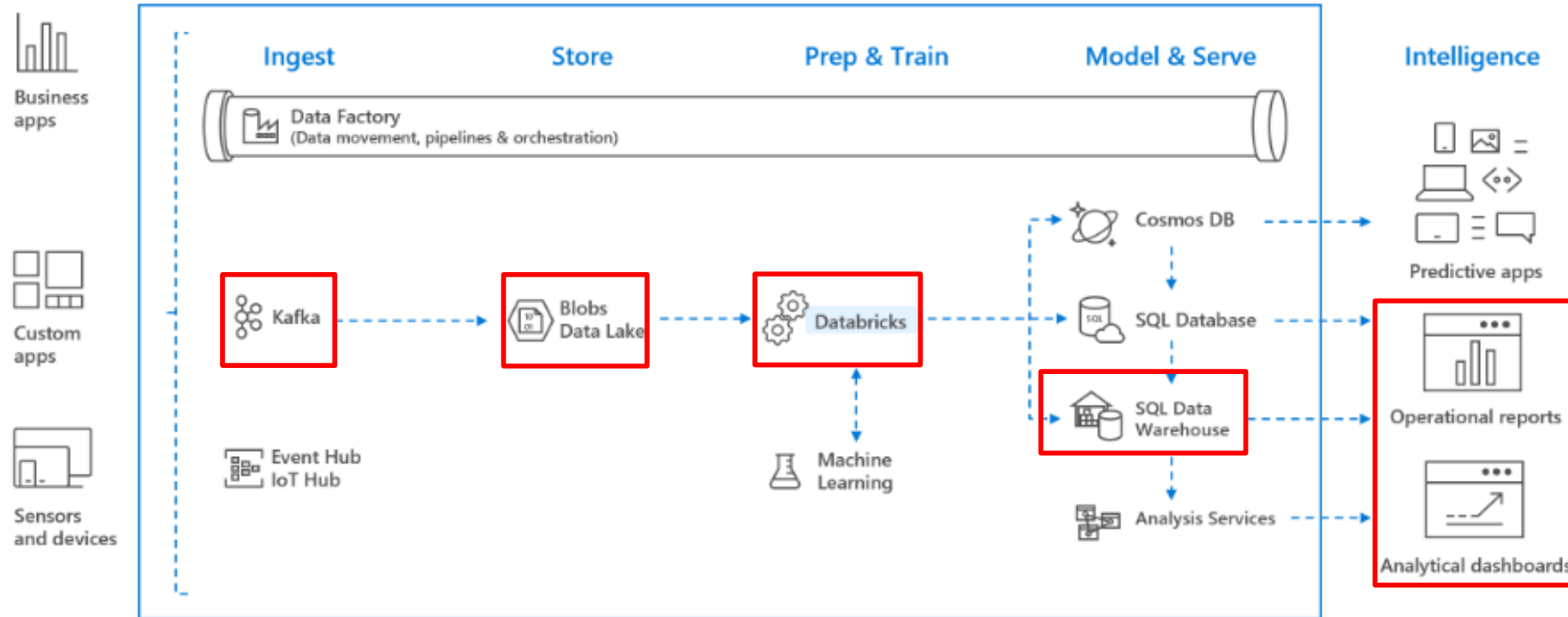


A big data processing architecture that combines both batch and real time processing methods

Lambda Architecture



Real-Time Analytics Architecture





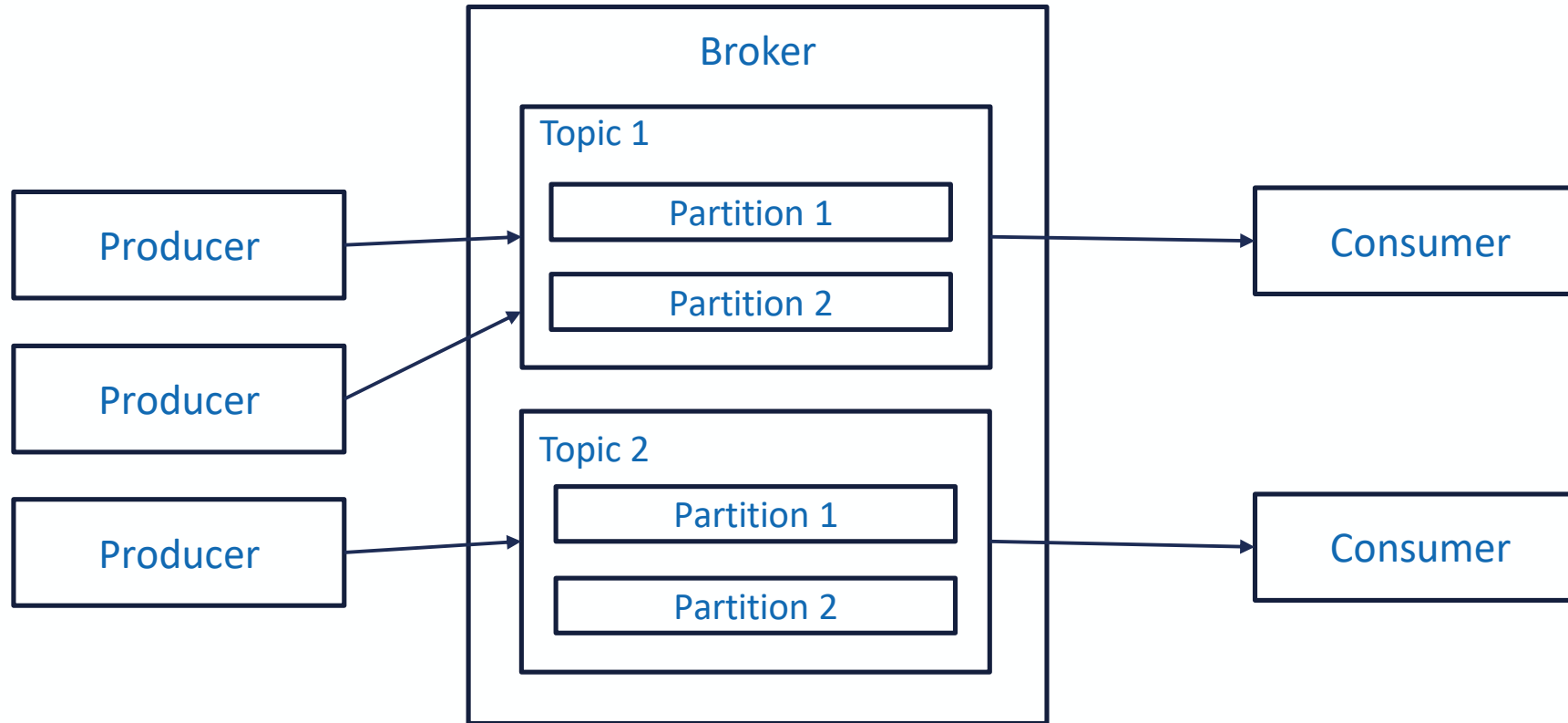
Apache Kafka for HDInsight (Ingest)

Apache Kafka is an open source, distributed, scalable, high-performance, publish-subscribe message broker, originally built at LinkedIn and now part of the Apache Software Foundation

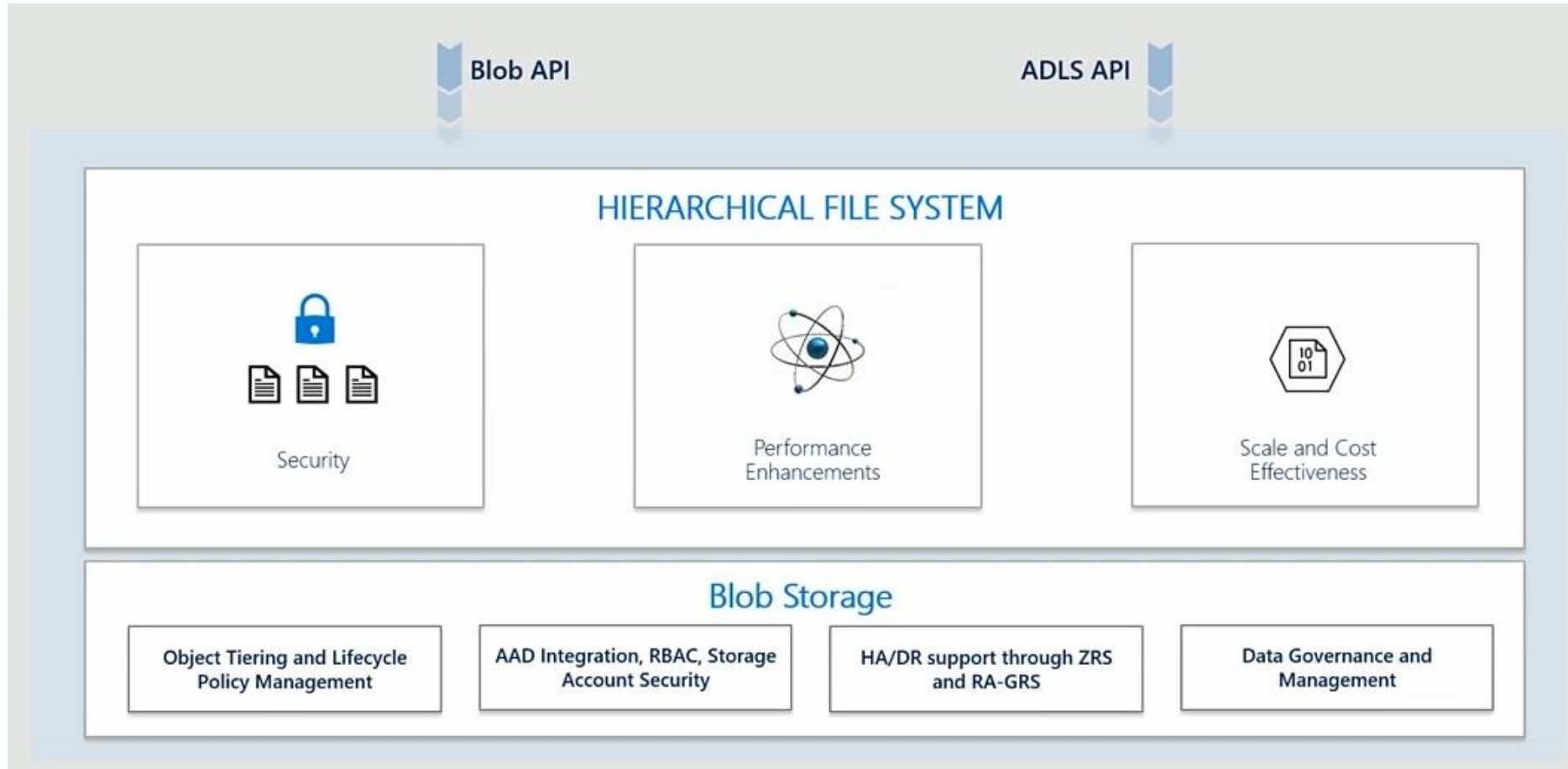
Azure HDInsight is a Hadoop service that enables clusters of managed Hadoop instances

Apache Kafka for HDInsight is a managed service that provides a simplified configuration process

Kafka Architecture - Components



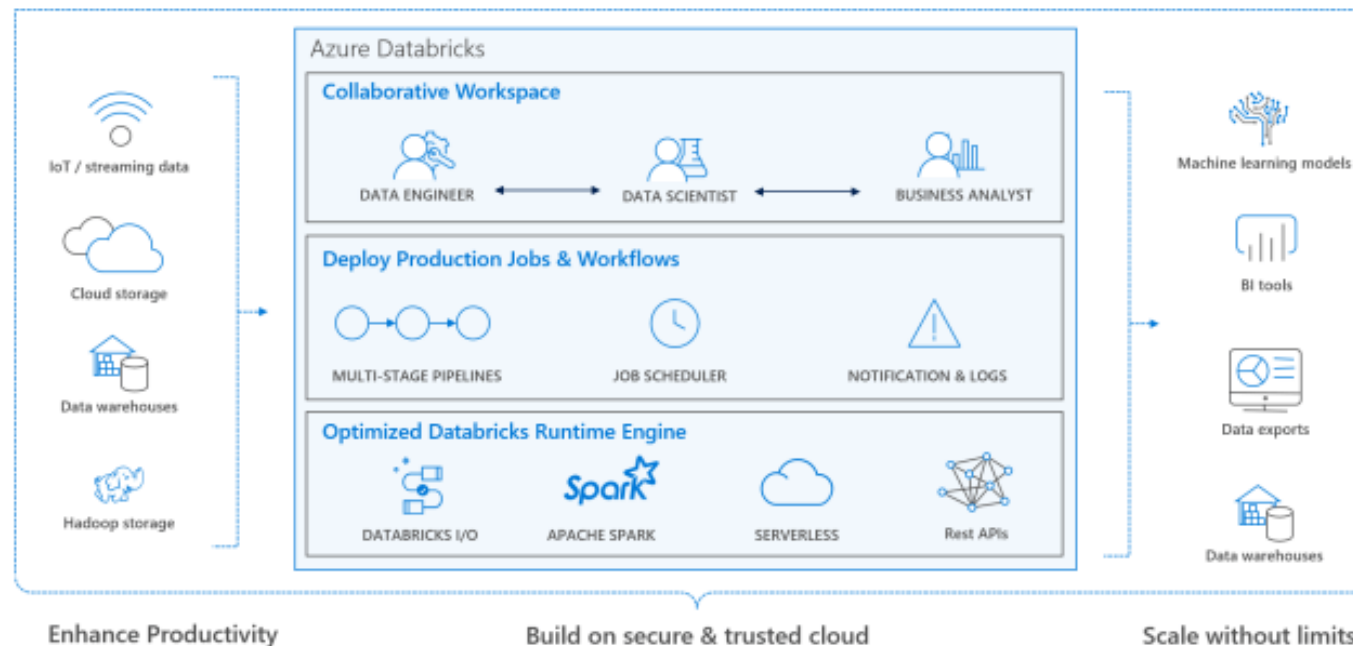
Azure Data Lake Storage Gen 2 (Store)



*credit Microsoft

Azure Databricks (Prep & Train)

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform



*credit Microsoft



Azure SQL Data Warehouse (Model & Serve)

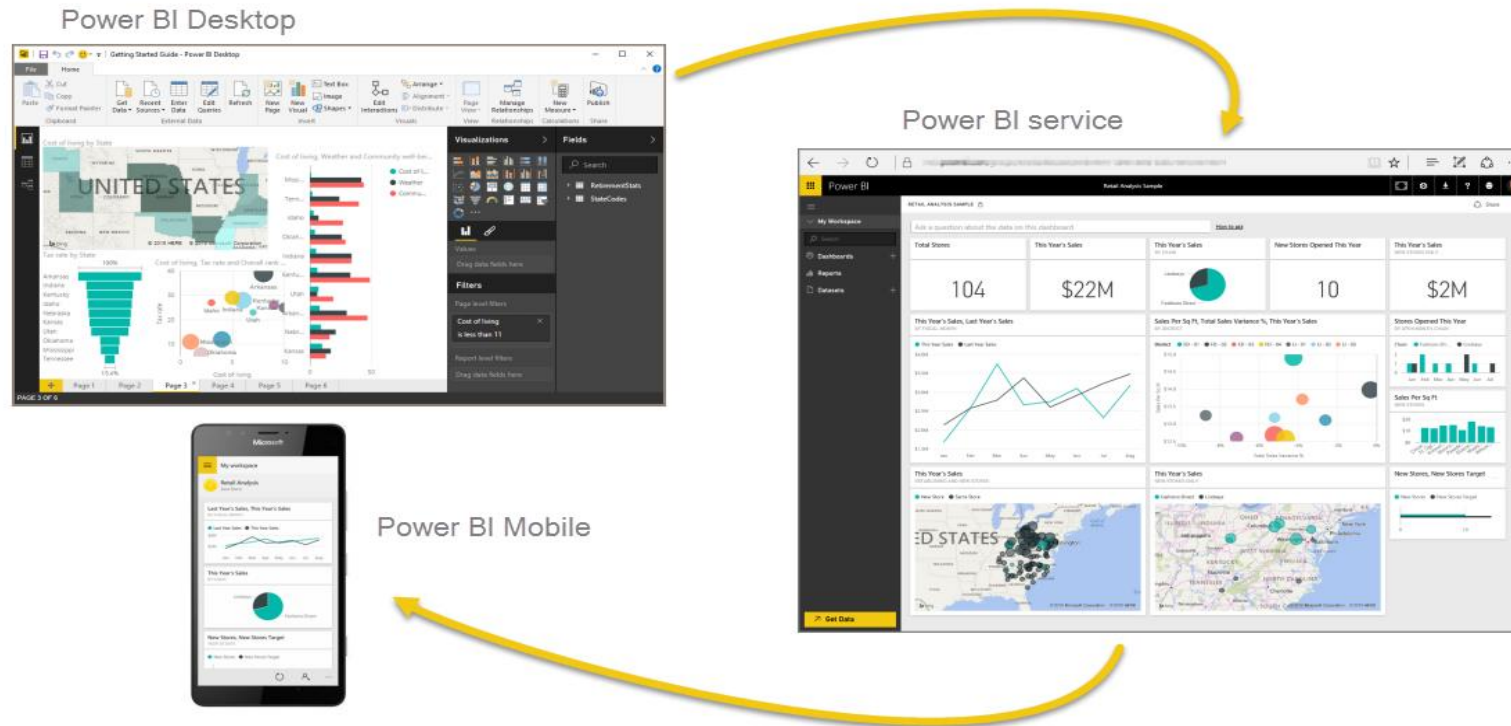
A cloud-based Enterprise Data Warehouse (EDW) that leverages Massively Parallel Processing (MPP) to quickly run complex queries across petabytes of data

Once data is stored in SQL Data Warehouse, it can run analytics at massive scale

Uses PolyBase to query the big data stores

Power BI (Intelligence)

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights



*credit Microsoft



Agenda

Real-time
analytics solutions

Lambda
Architecture

Demo

Databricks Delta
Architecture

Demo



Agenda

Real-time
analytics solutions

Lambda
Architecture

Demo

Databricks Delta
Architecture

Demo



Databricks Delta Architecture

Databricks Delta is a transactional storage layer designed specifically to harness the power of Apache Spark and Databricks DBFS

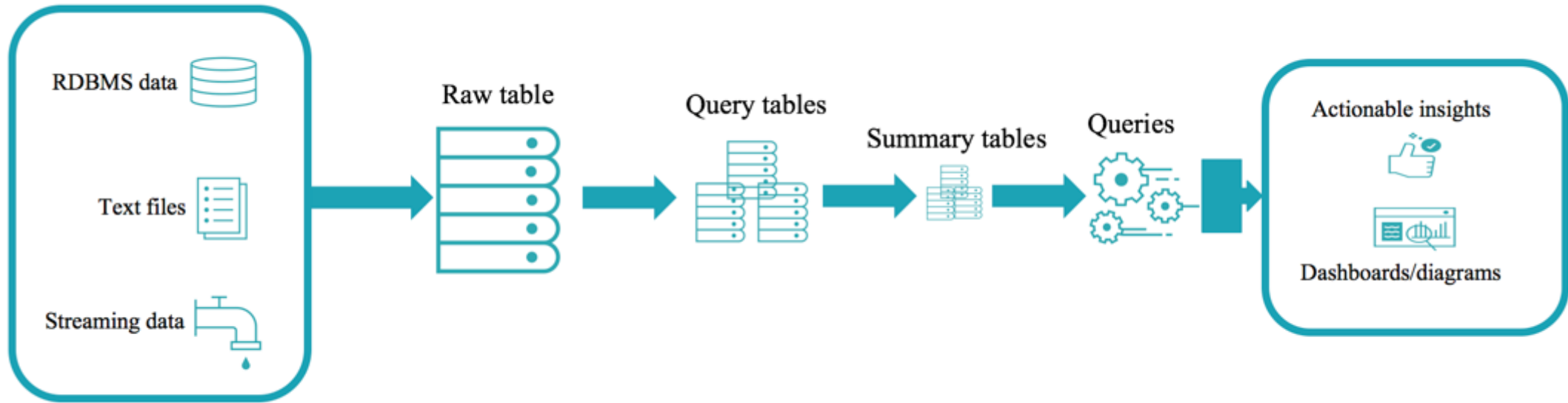
ACID transactions

Deletes/Updates/Upserts

Automatic file management

Statistics and data skipping

Databricks Delta Architecture



*credit Databricks



Agenda

Real-time
analytics solutions

Lambda
Architecture

Demo

Databricks Delta
Architecture

Demo

PRESENTATION CONTENT

<https://github.com/adatisjose/realtimeanalytics>



London:

8 Boundary Row, London SE1 8HP

Farnham:

Broadmede House, Farnham Business Park, Weydon Lane , Farnham, Surrey, GU9 8QT

Bulgaria:

Betahaus, Shipka 6 Street, Floor 3, 1504 Sofia

adatis.co.uk