adatis

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

adatis

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

adatis

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

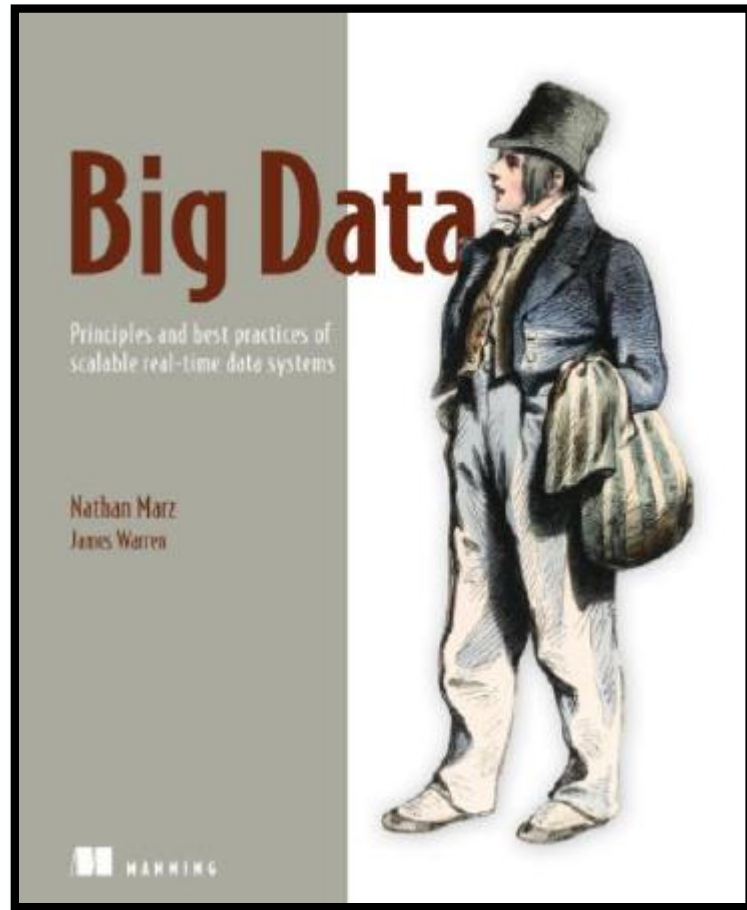# Agenda

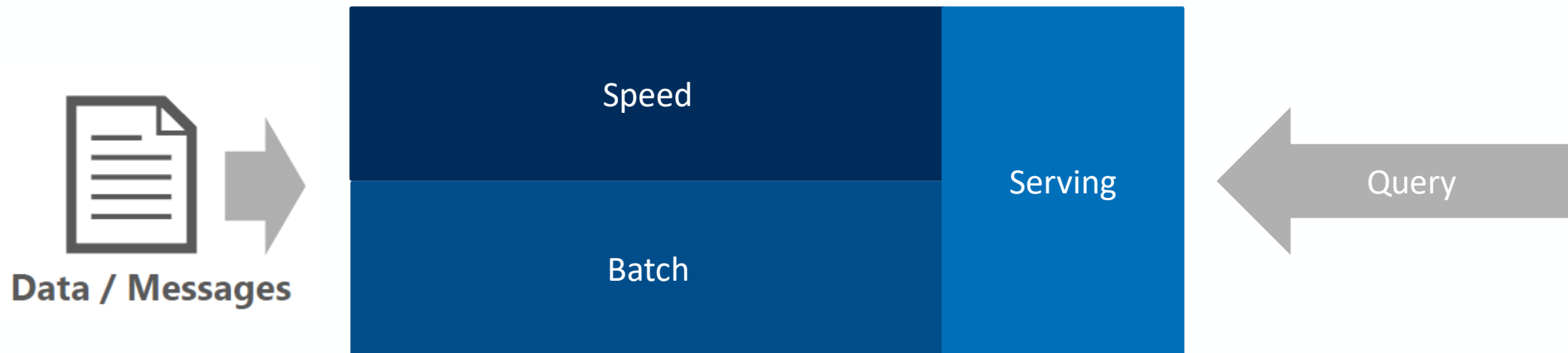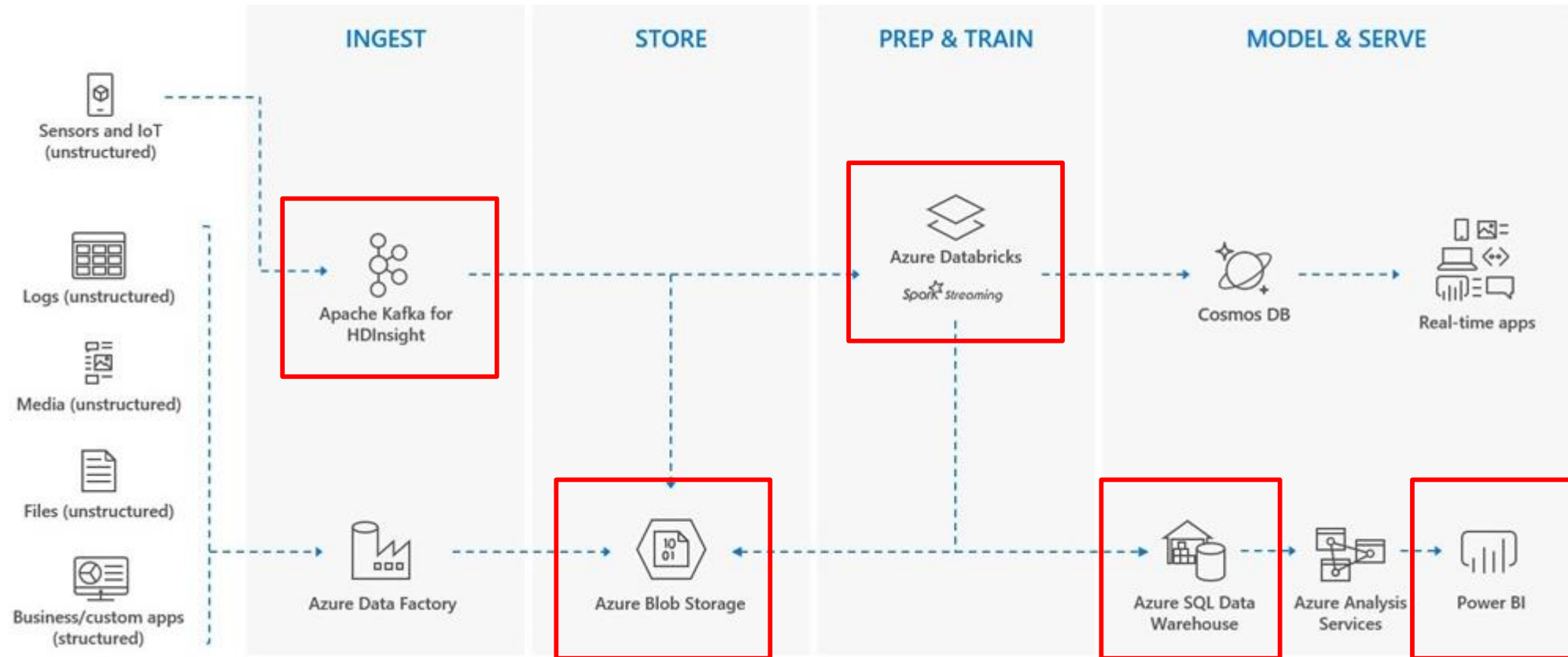| Real-Time Analytics Solutions | Lambda Architecture | Demo | Databricks Delta Architecture | Demo |

adatis

Lambda Architecture

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

Lambda Architecture

Real-Time Analytics Architecture
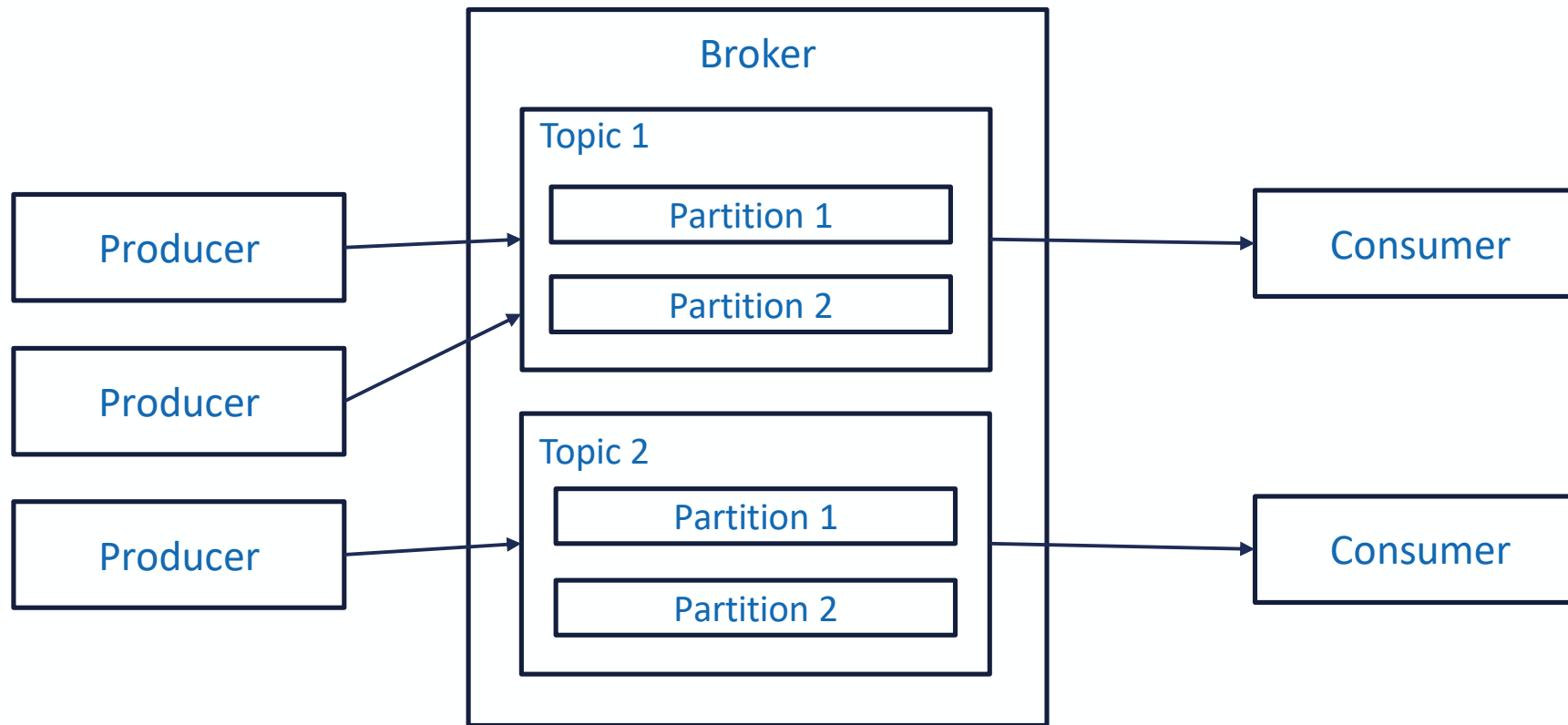

*credit Microsoft

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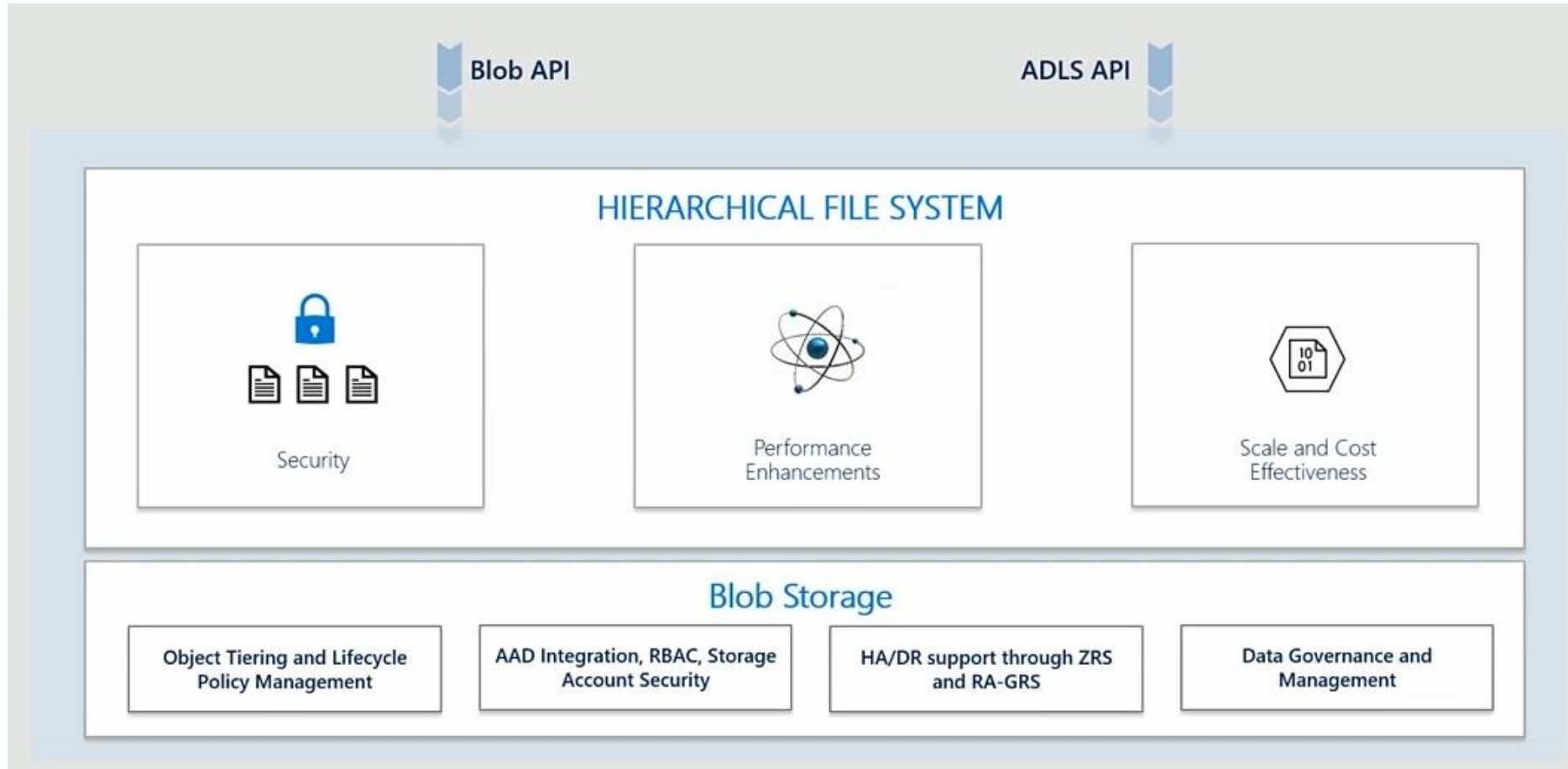
Azure HDInsight is a Hadoop service that enables clusters of managed Hadoop instances

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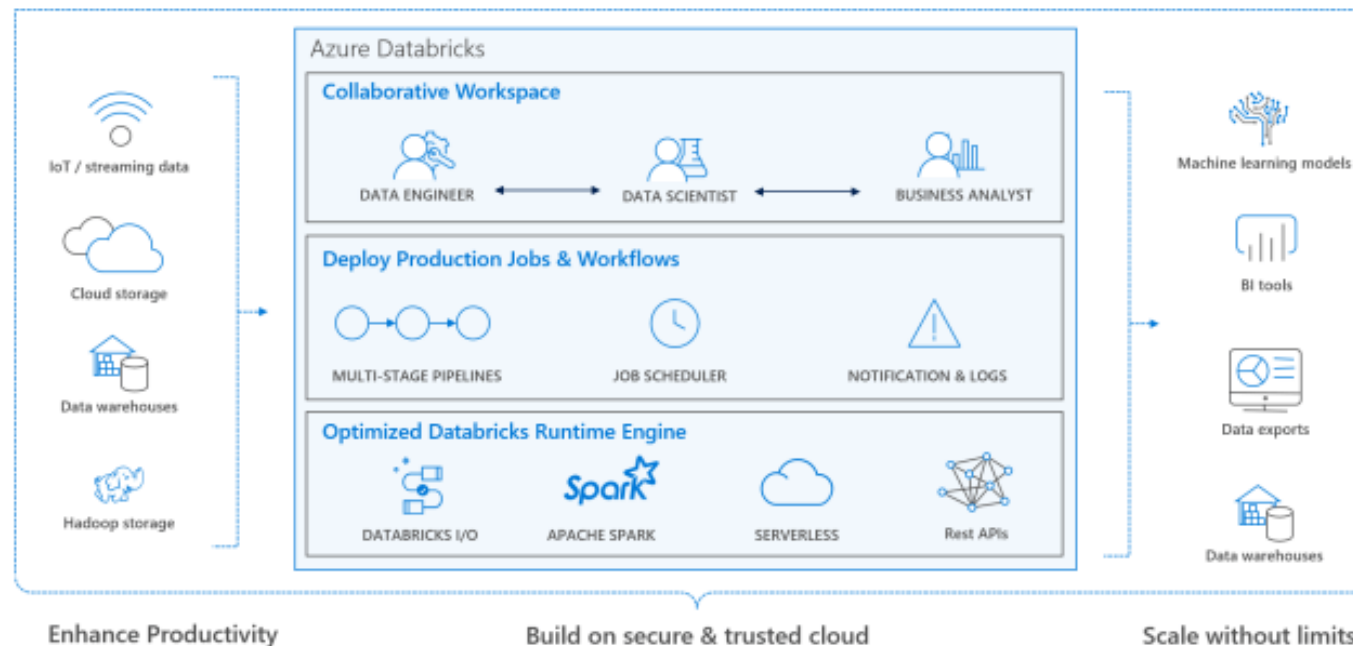
adatis

Kafka Architecture - Components

Azure Data Lake Storage Gen 2 (Store)

Blob API   **ADLS API**

HIERARCHICAL FILE SYSTEM

Security

Performance Enhancements

Scale and Cost Effectiveness

Blob Storage

Object Tiering and Lifecycle Policy Management

AAD Integration, RBAC, Storage Account Security

HA/DR support through ZRS and RA-GRS

Data Governance and Management

*credit Microsoft

adatis

Azure Databricks (Prep & Train)

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

*credit Microsoft
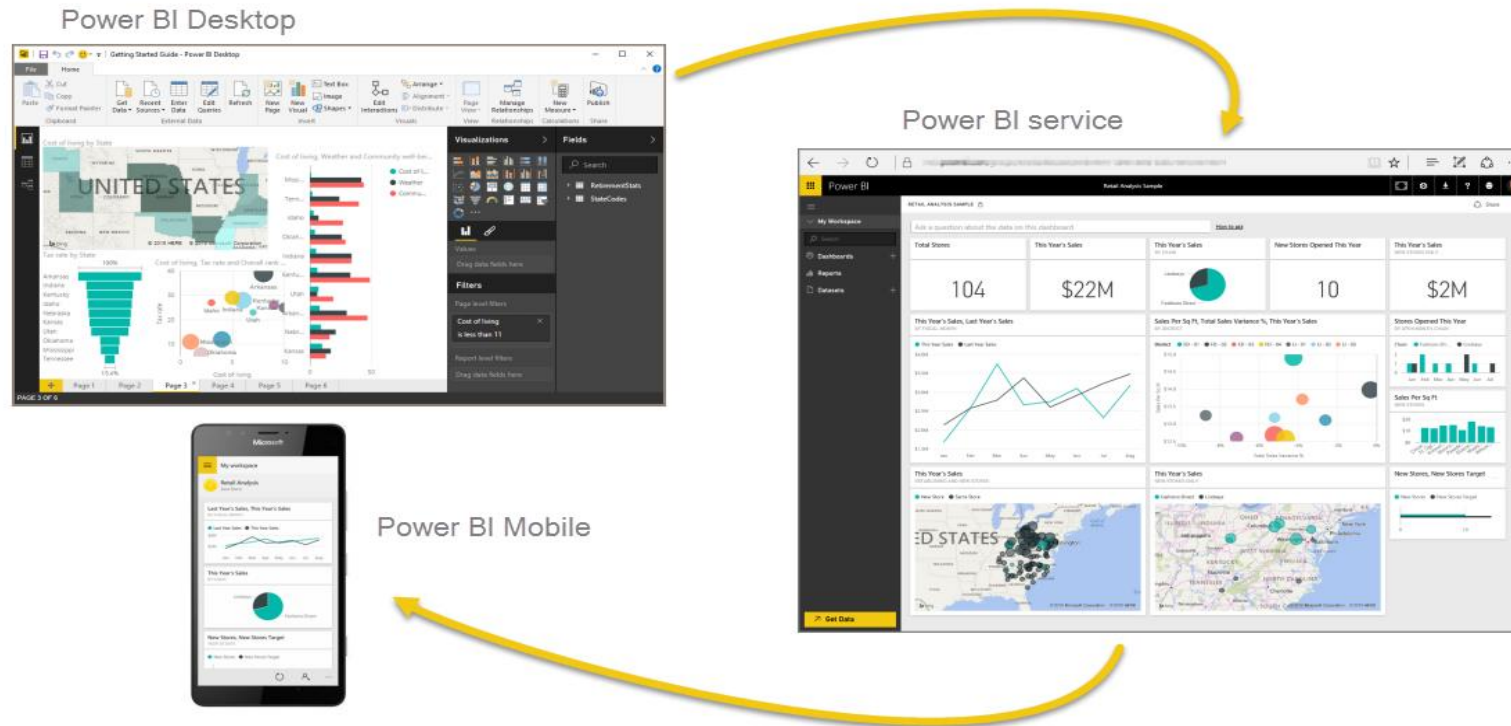
Azure SQL Data Warehouse (Model & Serve)

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 (Model & Serve)

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

adatis

Databricks Delta Architecture

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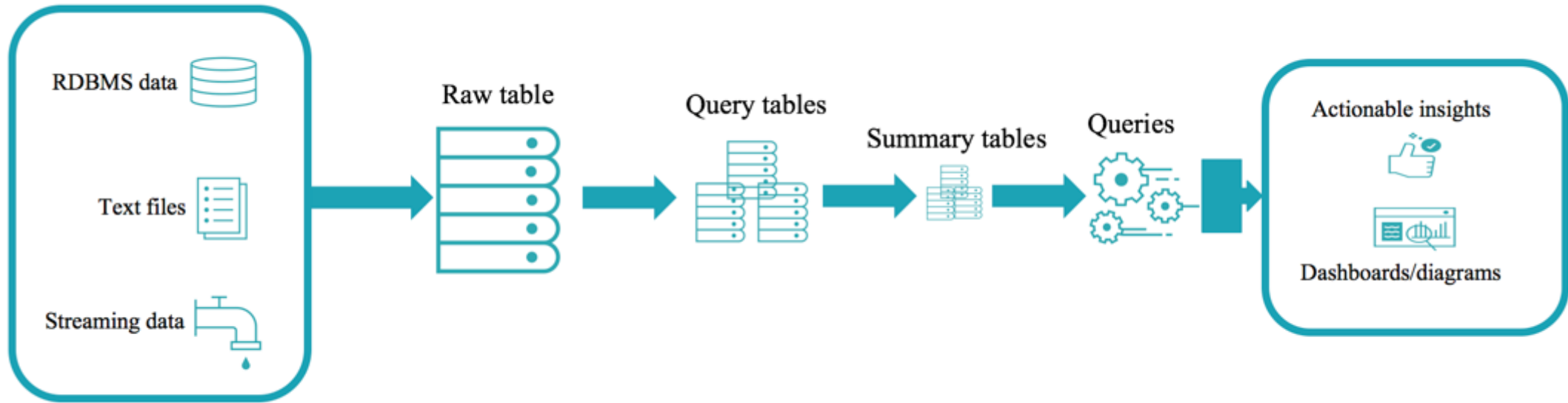
ACID transactions

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 |

adatis

PRESENTATION CONTENT

https://github.com/adatisjose/realtimeanalytics