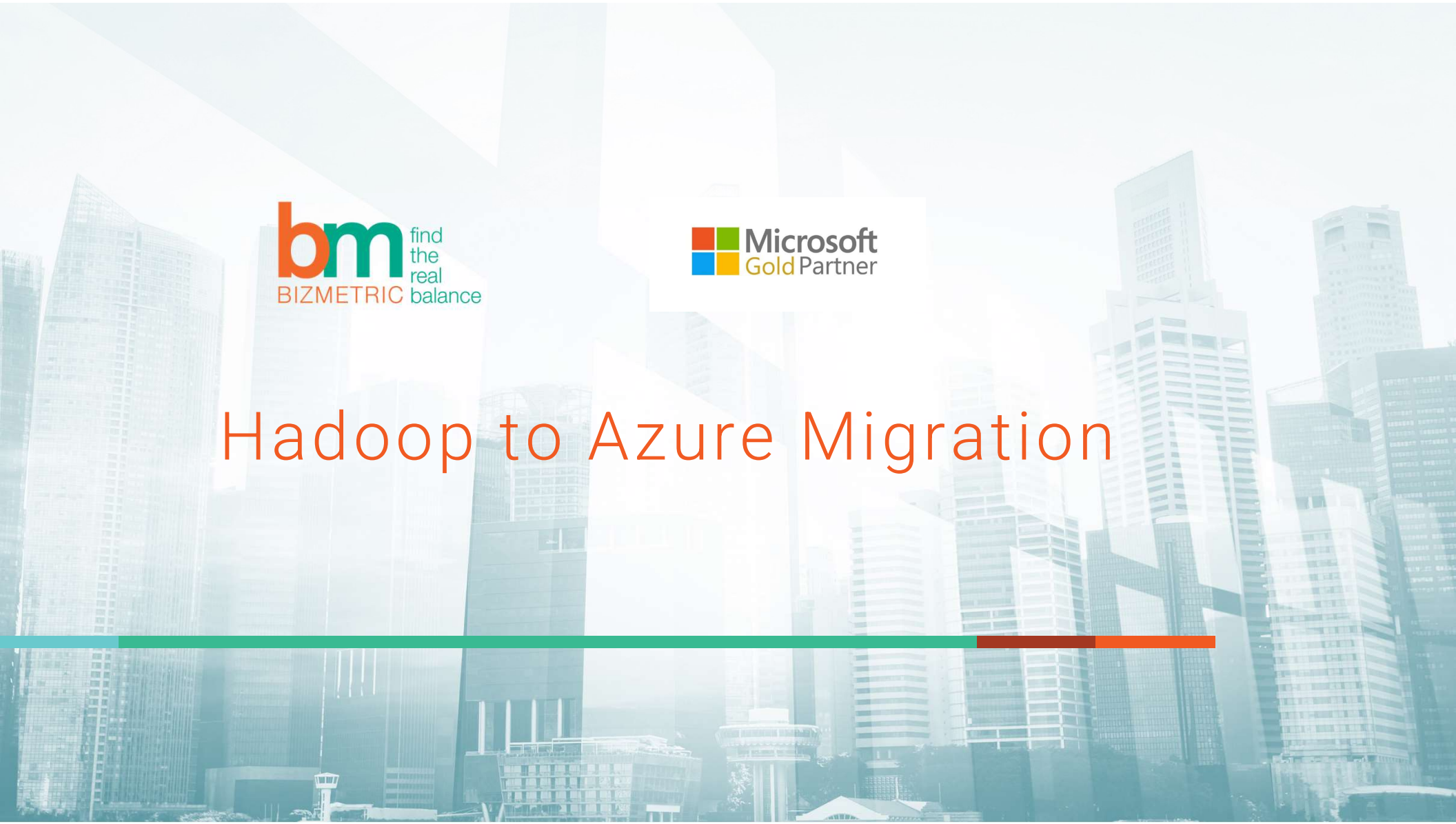

bm
find
the
real
BIZMETRIC balance
Microsoft
Gold Partner
Hadoop to Azure Migration

Hadoop to Synapse Migration Use Case

Bizmetric has helped multiple customers to migrate on premise Hadoop to Azure Synapse.

Business Perspective

- Difficult to locate data for analysis.
- Very Slow when analyzing Large data sets.
- Teams does not have required skill to use data.
- No visibility to age of data
- 5X longer to realize data insights

IT Perspective

- Can't keep with increasing compute demand
- Long delays acquiring new compute capacity.
- High cost to support and maintain.
- Inconsistent metadata breaks self service model and lead to IT effort.
- Not realizing data democratization ROI from program.

Hadoop Challenges

- High capital infrastructure costs / delays to implement
- Requires infrastructure investment to cover "short duration / peak" compute levels
- Unable to scale to unanticipated compute levels
- Lack of optimized compute leads to slow query performance
- Significant technical debt to maintain / operate environments
- Difficulty achieving operational excellence through automation, governance, and security
- Slow data insights due to missing citizen development / discovery capability
- Delayed platform upgrades to realize new capabilities

BIZMETRIC

Azure Synapse Solution

Implement Service in a Minutes	Unlimited Scale	Pay for Use Pricing
Serverless SQL / Spark Compute	Fully Integrated Platform as a Service	Azure Governance Ecosystem
Reduce TCO by up to 50%	Realize Data Insights 3X Faster	Unleash Data Insights / Analytics

BIZMETRIC

Azure Synapse – One Integration Platform

Hadoop Ecosystem	Synapse	Open-Source DNA ?
HDFS	Synapse Data Lake	✓
Oozie	Synapse Pipelines	✓
Hive	Synapse SQL Pool	✓
Impala	Synapse SQL Pool	✓
Spark	Synapse Spark Pool	✓
Hive / Pig	Synapse Workbooks	✓
Ranger / Sentry	Synapse Security	✓

BIZMETRIC

Migration Approach

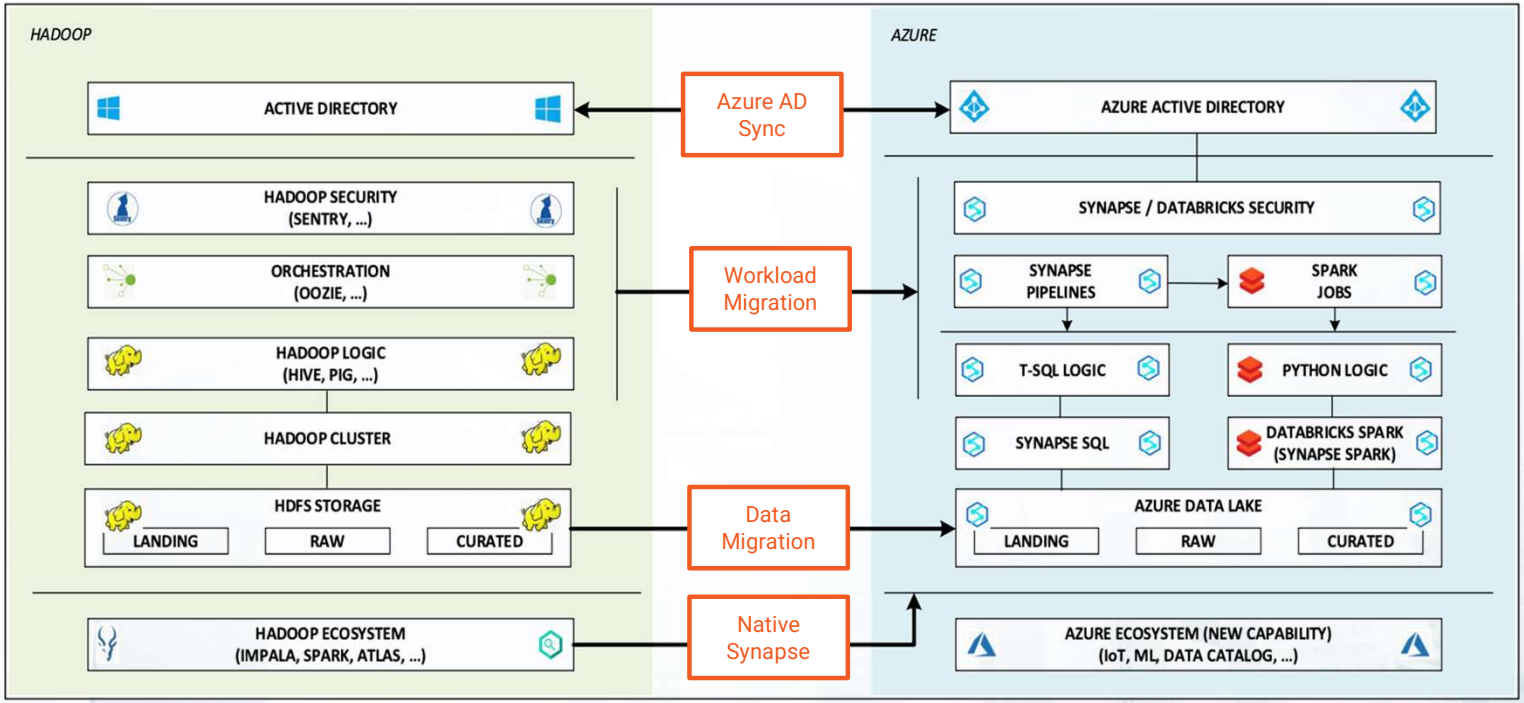
- No service interruption during migration
- Migrate at your pace
  - > Chose which / when Hadoop containers to migrate to Azure
  - > Enables incremental retirement of on premise Hadoop
  - > Enables incremental expansion to Azure Synapse
- Facilitates a parallel run period (for migrated workload validation)
- Ensures no data loss during migration for live content
- Minimal impact to existing HDFS storage, through read once sync

BIZMETRIC

Migration Steps

To facilitate a single sign on experience in Synapse, sync internal AD to Azure AD.

- Implement HDFS to Azure Data Lake. Hadoop HDFS storage changes are continuously copied to Azure, to keep Azure data current.
- Workload migration feature, to convert Oozie / Hive / Pig / ... processes to Synapse SQL / Spark processes. Also migrates Sentry security configuration.
- Leverage the Azure ecosystem to easily integrate additional new capabilities with Synapse.



BIZMETRIC

Migration Process

Pre-Migration

Discover

Inventory Database Assets

Application tack

Dependencies

Assess / Design

Profile application Data and Dependencies.

Define Azure Solution design.

Confirm viability for application migration to Azure

Evaluate / Convert

Identify frictions related to Azure modernization.

Convert schema apps to work in Azure if needed.

Migration

Migration Data Address Dependencies

Migrate source schema / data

Implement dependency solution

Perform Unit Testing

Remediate Testing issues

Data Sync

Sync your target data with the source to maintain data currency, during the migration window.

This is relevant for minimal downtime migration.

Post-Migration

Remediate Applications

Iteratively make any necessary changes to your applications.

Optimize

Based on the tests you performed, address any performance issues, and then retest to confirm the performance improvements.

Run Functional & Performance testing

Iteratively run functional and performance tests.

Cutover

Cut over from the source to the target environment.

BIZMETRIC

Azure Case Studies

BIZMETRIC

Microsoft Gold Partner

Case Study Hadoop to Azure Migration

Domain – Oil & Gas Company

Our client is one of the Largest Oil & Gas firm in USA, they were on Hadoop system from last 5 years.

The Problem Statement

- Licensing cost was very high
- Performance issues as different ML Models was running.
- Infrastructure cost was very high.
- Support Cost was very High.
- Project was aimed to migrate all the Hive history tables (8000+)
- 250+ active jobs built in spark, Kafka, Sqoop, MapReduce need to be migrated.

The Business Solution

- Execute end to end assessment of the existing spark and scala code, hive queries and designed a modern data pipeline to spark code seamlessly.
- Migrated Hive HQL to Spark / Python Code based on design pattern Analysis automation.
- Data Validate between the old and new process as performed through automation scripts.
- Metadata validation includes check on the data types of the column and number of columns.
- All the reporting connection were remapped to newer system.

Domain – Global Logistic

Our client is one of the World biggest logistic company, they were on Hadoop (Hortonworks) from last 5+ Years.

The Problem Statement

- Licensing cost and Infrastructure cost was very high.
- Support Cost was very High.
- 9 PB HIVE warehouse and 2 PB Streaming data
- Daily 400+ GB batch load and 200 GB Streaming
- On prem - 64 Hadoop Cluster
- Cluster size was of 3 to 7 nodes

The Business Solution

- Sensor streaming data (location and weather data) in kafka was migrated to Azure event hub and Data Explorer.
- Execute end to end assessment of the existing spark, Kafka, Sqoop code, hive queries and designed a modern data pipeline to spark code seamlessly.
- Migrated Hive HQL and Map Reduce to Spark / Python Code based on design pattern Analysis automation.
- Data Validate between the old and new process as performed through automation scripts.
- Metadata validation includes check on the data types of the column and number of columns.

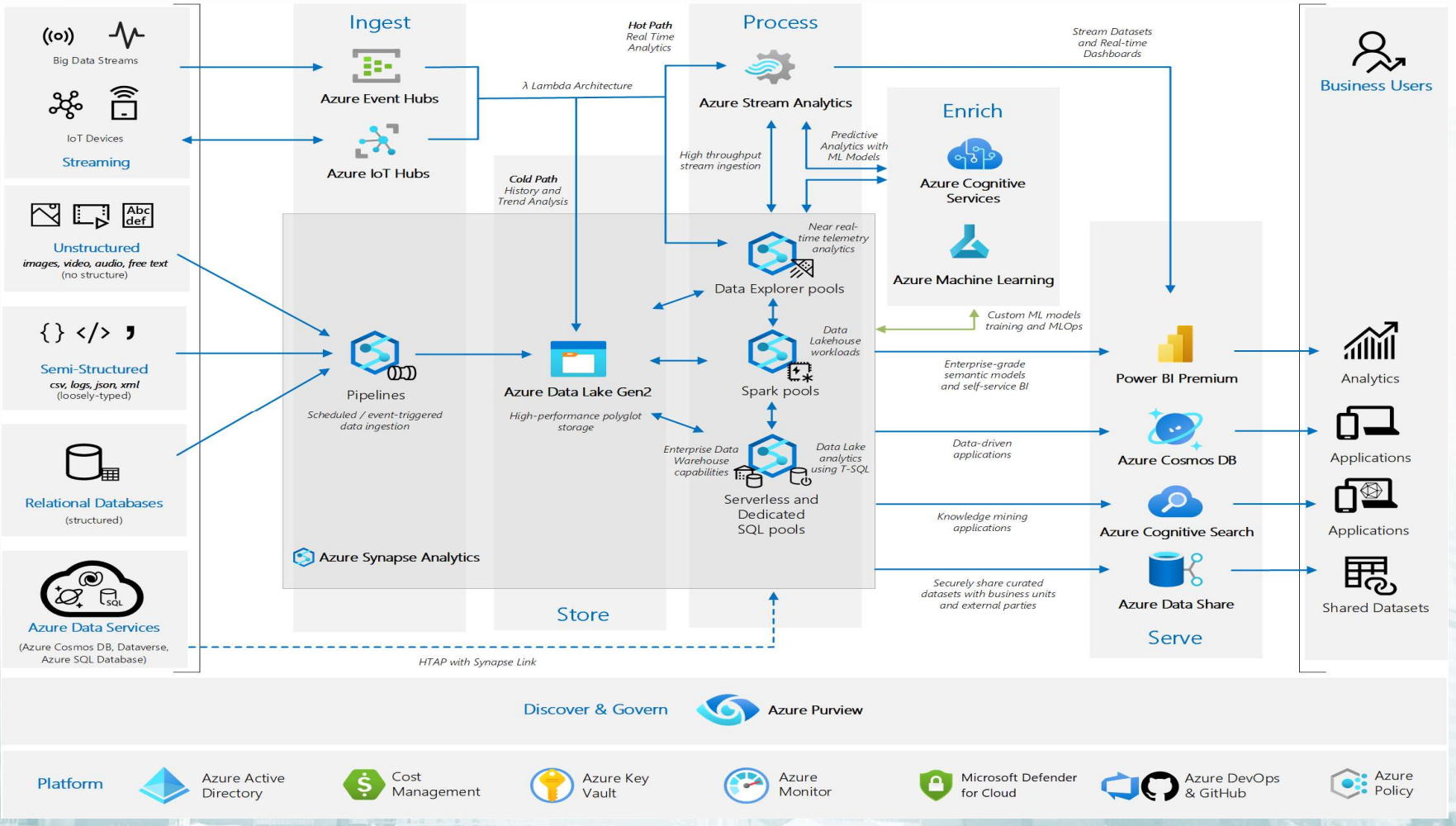
Key Results

- Hadoop Services was retired
- Saving the Annual infrastructure and Support Cost
- Retired Hadoop Nodes

BIZMETRIC

Microsoft Gold Partner

High Level Architecture



BIZMETRIC