REAL-TIME BIG DATA ANALYTICS IN THE CLOUD 101: EXPERT ADVICE FROM THE ATTUNITY AND AZURE DATA LAKE STORAGE GEN2 TEAMS

Microsoft

ATTUNITY

Today's Speakers

Carole Gunst
Marketing Director
Attunity

Jordan Martz,
Director Technology Solutions,
Attunity

Jeff King,
Senior Program Manager for Azure Data Lake Storage Gen2,
Microsoft

We’ll answer these questions

Why is real-time data important for driving business insights?

What’s a data lake and why would you use one to store your real-time data?

How can you use change data capture (CDC) technology to efficiently transfer data to the cloud?

How can you build sophisticated analytic workflows quickly?

Why is Azure Data Lake Storage Gen2 the best data lake for real-time analytics?

Migrating Data to the Microsoft Data Platform

ATTUNITY

Jordan Martz, Director of Technology Solutions, Attunity

Microsoft + Attunity Strategic Partnership

Cloud and database migrations

to Microsoft Data Platform from any enterprise platform -- on-premises and in Azure

Ingests into Microsoft’s DBs, EDWs and data lakes

from a broad range of data sources: SAP, Oracle, Teradata, mainframes, and more

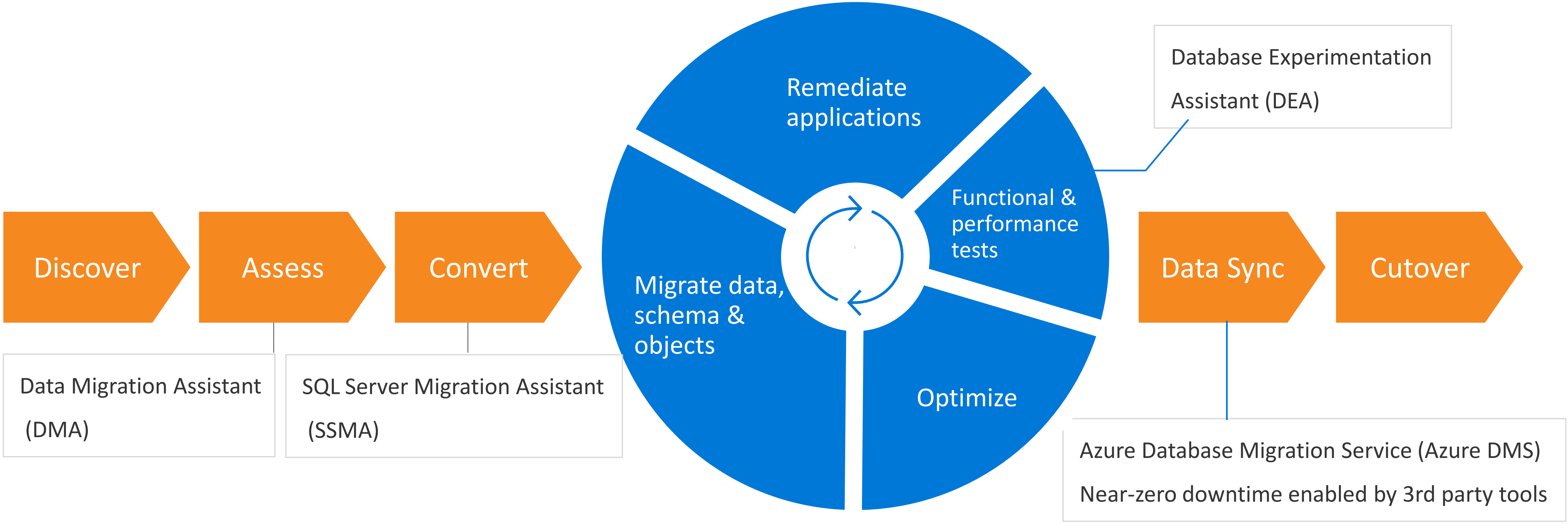
Continuous refresh of data

for zero-downtime migrations, real-time streaming for business intelligence & analytics

#1

DATA MIGRATION & CDC TECHNOLOGY

Data Migration Lifecycle



Discover

Assess

Convert

Remediate applications

Functional & performance tests

Optimize

Migrate data, schema & objects

Data Sync

Cutover

Data Migration Assistant (DMA)

SQL Server Migration Assistant (SSMA)

Database Experimentation Assistant (DEA)

Azure Database Migration Service (Azure DMS)

Near-zero downtime enabled by 3rd party tools

Additional Sources and Microsoft Targets Available

ATTUNITY
REPLICATE

SOURCES

each
with CDC

RDBMS	DW	HADOOP	CLOUD	MAINFRAME	SAP
Oracle **SQL Server** **DB2 iSeries** **DB2 z/OS** **DB2 LUW** **MySQL** **PostgeSQL** **Sybase ASE** **Informix**	**Exadata** **Teradata** **Netezza** **Vertica**	**Hortonworks** **Cloudera** **MapR**	**AWS RDS** **Salesforce**	**DB2 for z/OS** **IMS/DB** **VSAM** **SQL/MP** **Enscribe** **RMS**	**ECC on Oracle** **ECC on SQL** **ECC on DB2**

Microsoft
targets

RDBMS	DW	HADOOP	CLOUD	STREAMING
SQL Server **MySQL** **PostgreSQL**	**Microsoft PDW** **Azure SQL DW**	**Microsoft HDI**	**Azure SQL DW** **Azure SQL DB** **Azure Data Lake**	**Azure Event Hubs** **Kafka**

Attunity Replicate – Optimized and Secure

*“We load **+10 Billion rows/hour doing full load** while replicating from Oracle Exadata with Attunity Replicate.”*

*“Attunity handles **460,000 records/sec doing CDC from large and highly active Oracle** databases into a data warehouse, with peaks of 100Gb per hour”*

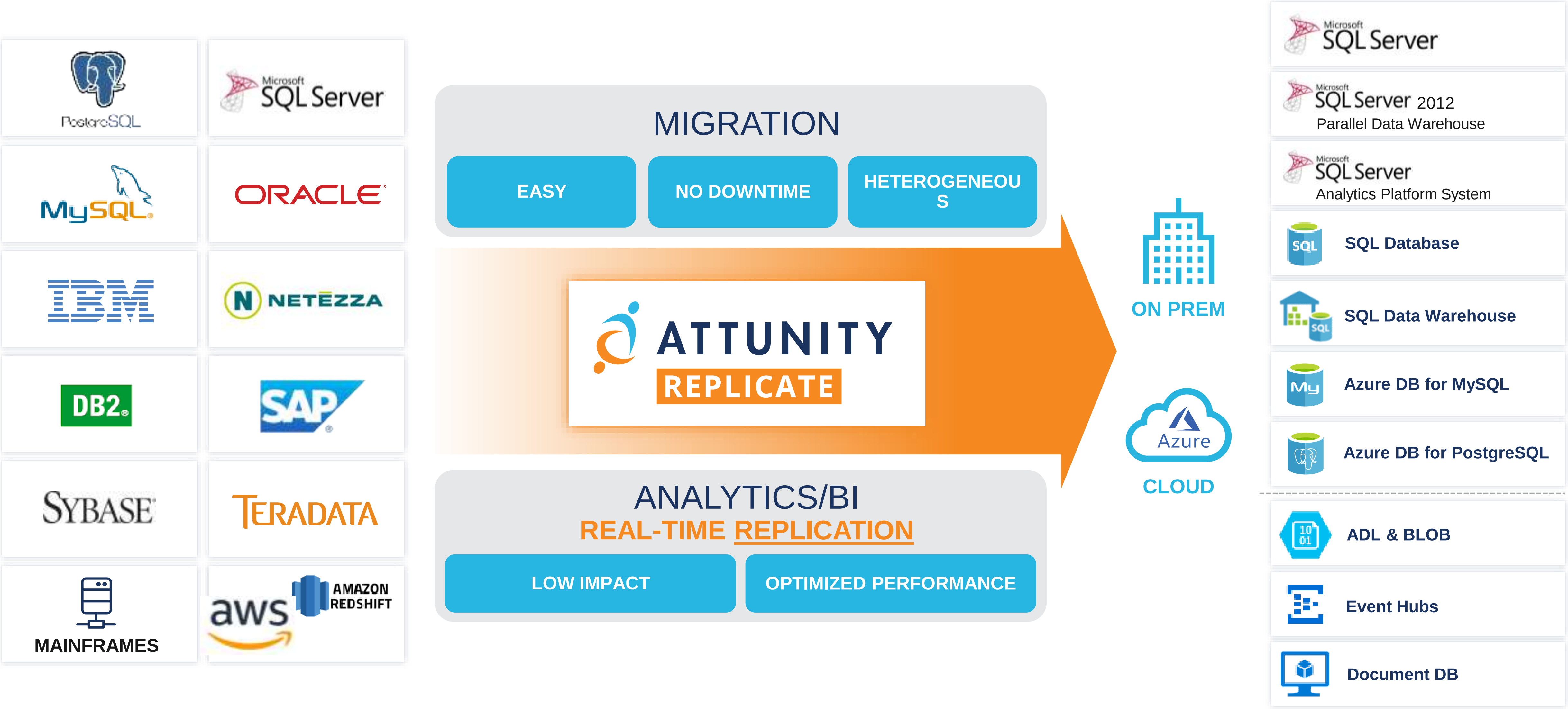
ATTUNITY
REPLICATE

PERFORMANCE-OPTIMIZED DATA TRANSFER

Extraction	**Transfer on-premises**	**Transfer to Azure**	**Ingest**
optimized with CDC and batch bulk-loads	optimized with in-memory streaming	optimized file transfer & compression	optimized for each different target (RDBMS, DW, Hadoop)

SECURE

Universal Solution for the Microsoft Data Platform



PostgreSQL
Microsoft SQL Server
MySQL
ORACLE
IBM
NETEZZA
DB2
SAP
SYBASE
TERADATA
MAINFRAMES
aws AMAZON REDSHIFT

MIGRATION

EASY
NO DOWNTIME
HETEROGENEOUS

ATTUNITY REPLICATE

ANALYTICS/BI

REAL-TIME REPLICATION

LOW IMPACT
OPTIMIZED PERFORMANCE

ON PREM

Azure

CLOUD

Microsoft SQL Server
Microsoft SQL Server 2012 Parallel Data Warehouse
Microsoft SQL Server Analytics Platform System
SQL Database
SQL Data Warehouse
Azure DB for MySQL
Azure DB for PostgreSQL

ADL & BLOB
Event Hubs
Document DB

Microsoft Data Platform Migration

- Microsoft SQL Server
- PostgreSQL
- TERADATA
- MySQL
- ORACLE
- IBM
- NETEZZA
- AMAZON REDSHIFT
- SYBASE An SAP Company

MIGRATION

EASY | **NO DOWNTIME** | **HETEROGENEOUS**

ATTUNITY REPLICATE FOR MICROSOFT MIGRATIONS

ATTUNITY REPLICATE for Microsoft Migrations
Download Attunity Replicate for Microsoft Migrations

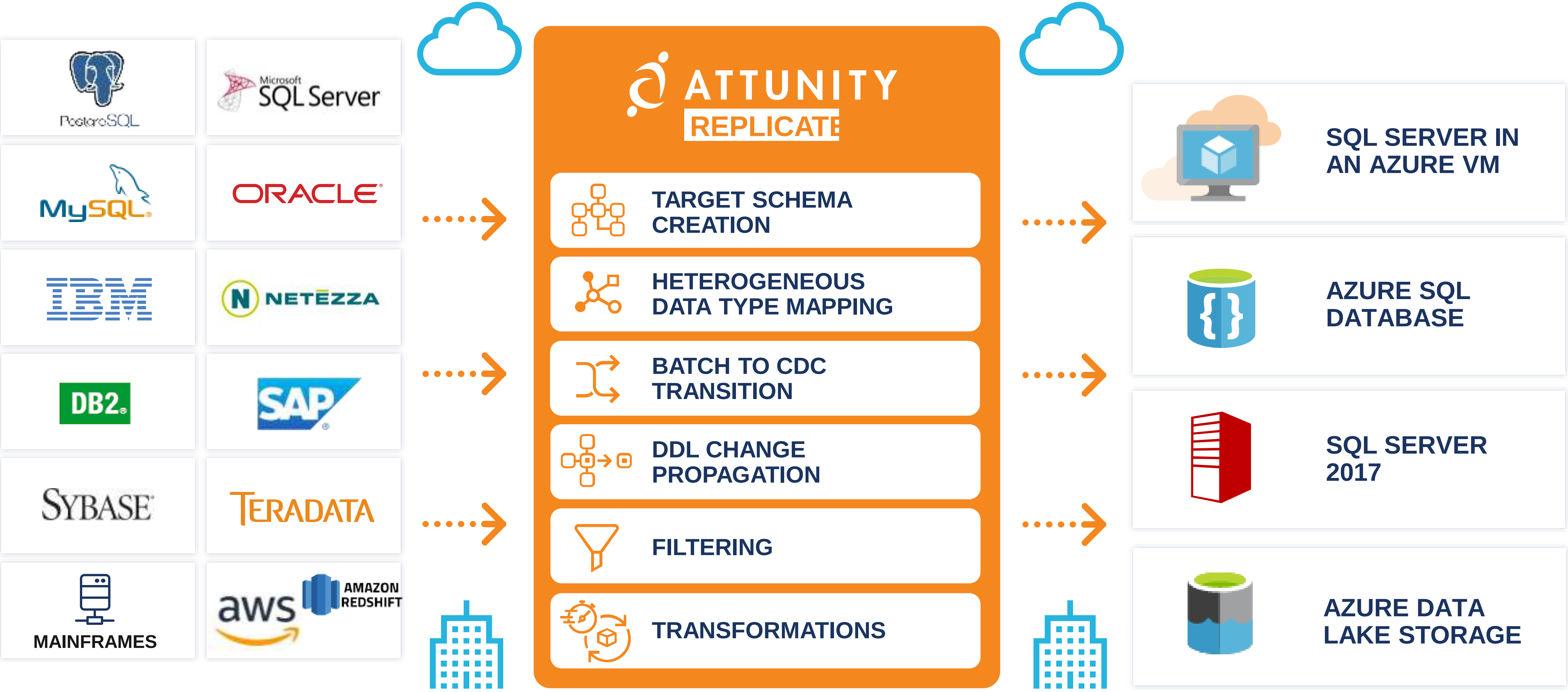
ON PREM

Azure
CLOUD

- Microsoft SQL Server
- Microsoft SQL Server 2012 Parallel Data Warehouse
- Microsoft SQL Server Analytics Platform System
- SQL Database
- SQL Data Warehouse
- Azure DB for MySQL
- Azure DB for PostgreSQL

www.attunity.com/MicrosoftMigrations

Data Migration & On-Going Replication with Attunity


PostgreSQL
Microsoft SQL Server
MySQL
ORACLE
IBM
NETEZZA
DB2
SAP
SYBASE
TERADATA
aws
AMAZON REDSHIFT
ATTUNITY

MAINFRAMES

REPLICATE

- TARGET SCHEMA CREATION
- HETEROGENEOUS DATA TYPE MAPPING
- BATCH TO CDC TRANSITION
- DDL CHANGE PROPAGATION
- FILTERING
- TRANSFORMATIONS

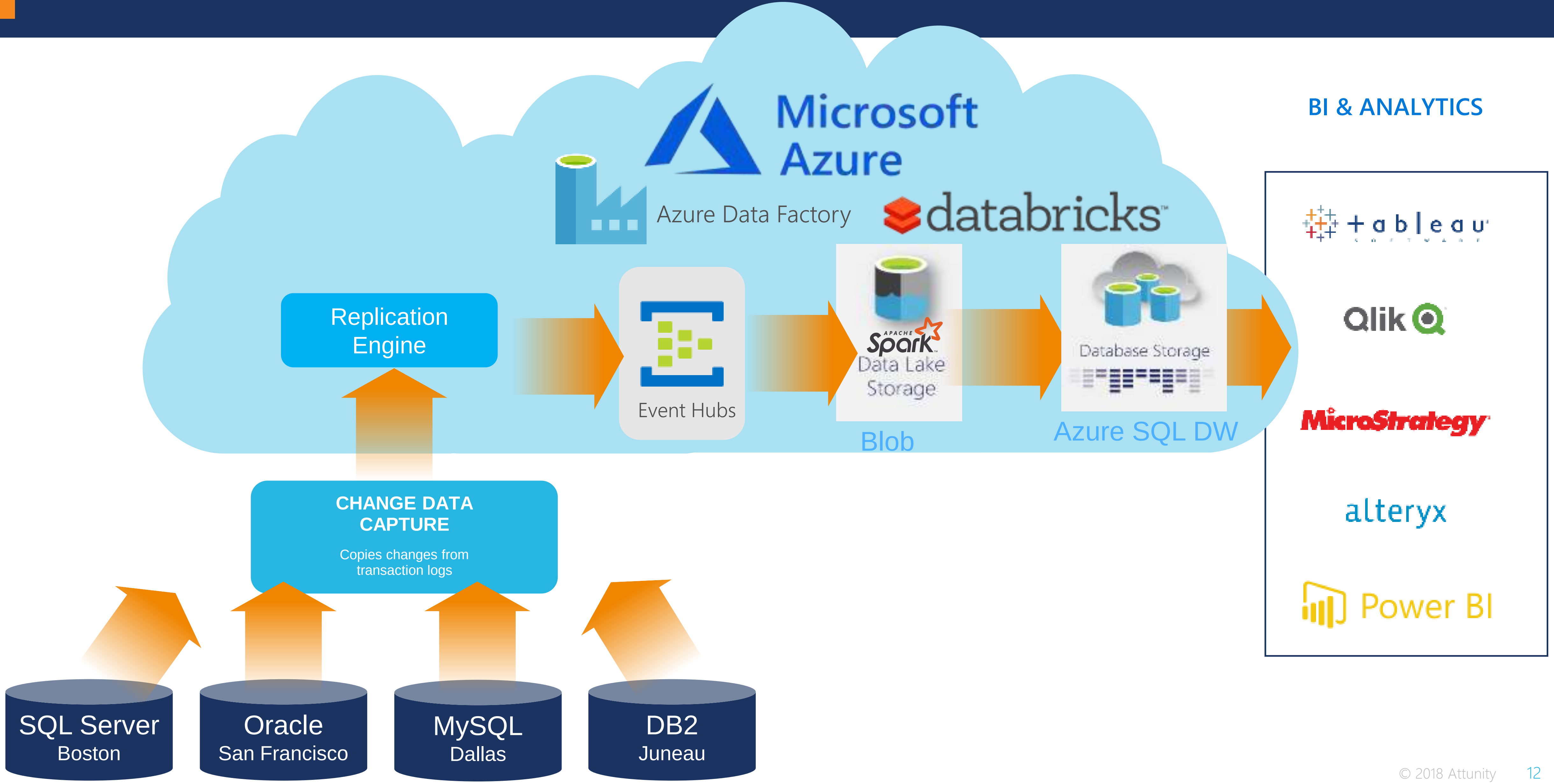
SQL SERVER IN AN AZURE VM

AZURE SQL DATABASE

SQL SERVER 2017

AZURE DATA LAKE STORAGE

Architecture



Microsoft Azure

Azure Data Factory

databricks

Replication Engine

Event Hubs

Spark Data Lake Storage

Blob

Database Storage

Azure SQL DW

CHANGE DATA CAPTURE

Copies changes from transaction logs

SQL Server
Boston

Oracle
San Francisco

MySQL
Dallas

DB2
Juneau

BI & ANALYTICS

- tableau
- Qlik
- MicroStrategy
- alteryx
- Power BI

Attunity Replicate Compliments Azure Data Factory

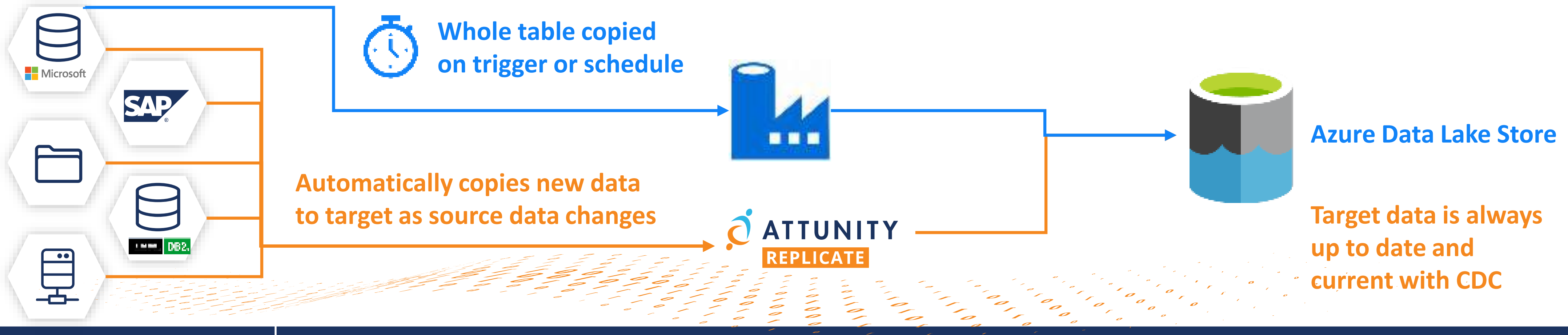
Azure Data Factory

- Data Integration as a Service via fully managed ETL workflow
- Lift-&-Shift compatibility with SSIS integration projects
- Bulk-load connectivity to many source systems
- Easy to use by developers and DBAs

ATTUNITY REPLICATE

- Hi-speed/Hi-volume data ingest for heterogeneous migrations & Big Data analytics
- Automated Change Data Capture (CDC), including replication of changes in metadata (DDL) and data itself
- Support for legacy and enterprise sources to Microsoft Data Services and SQL on-prem
- Easy to use by DBAs – no coding experience

Why Use Attunity Replicate If You Have Azure Data Factory?



Whole table copied on trigger or schedule

Automatically copies new data to target as source data changes

Azure Data Lake Store

Target data is always up to date and current with CDC

Attunity Replicate Features	Attunity Replicate Benefits
Incremental Copy	• Copy just changed rows from any data source in a true CDC manner, not requiring delta table management
Automated Data Transfer	• No TRIGGERS or SCHEDULES – Data moves from source to target as soon as it changes on the source.
High Speed, Large Volume Data Transport	• Optimized for fast data loading into any Azure target(s)*. (*Attunity Replicate can copy source data to multiple targets)
Legacy Systems and Enterprise Applications	• Natively supports legacy, EDW and enterprise systems data sources - Mainframe, iSeries, Exadata, Teradata, Netezza, Vertica, Grenplum/Pivotal. DB2 for z/OS, IMS/DB, VSAM, SAP, SAP Hana, SQL/MP, Oracle, Microsoft SQL Server, DB2 iSeries, DB2 z/OS, DB2 LUW, MySQL, PostgeSQL, Sybase ASE, Informix, Apache Kafka, and many more.
Example Use Case (above)	• A retailer copies their ERP data from IBM DB2 to Azure Data Lake Store using Azure Data Factory. • However their Point of Sales (POS) tables are 200-300GB and transfer is slow with ADF ODBC. • **Attunity Replicate loads the data incrementally (as it changes in real time)**

Microsoft Use Cases Enabled by Attunity Solutions

Microsoft SQL Server

SQL Server Modernization: Upgrade from SQL 2005, 2008, etc in one step

ORACLE
SYBASE
DB2

Competitive Database Migrations: Move from Oracle, Sybase, DB2, etc.

TERADATA
NETEZZA
AMAZON REDSHIFT

Competitive Data Warehouse Migrations: from Teradata, Netezza, AWS Redshift

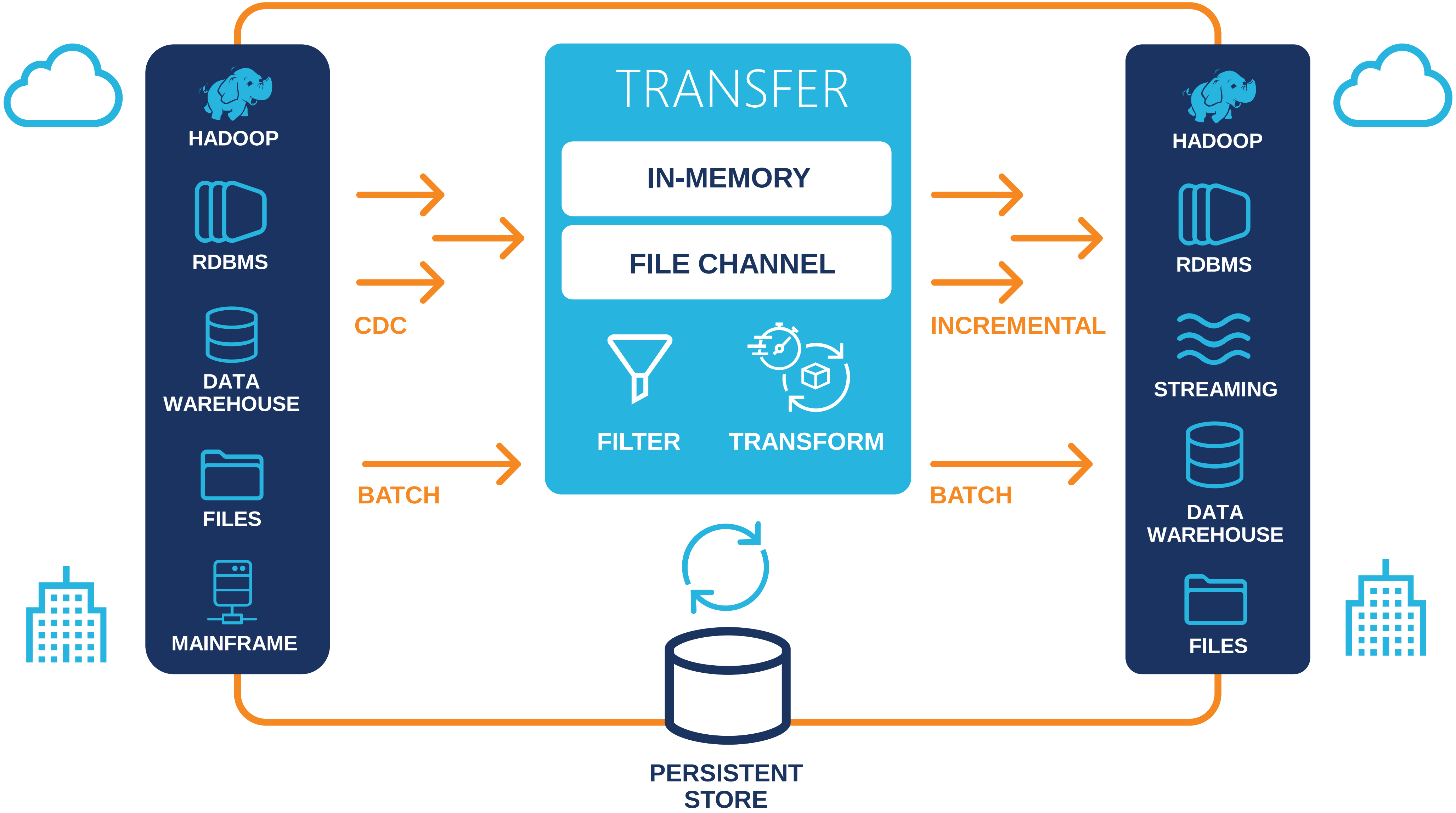
Azure

Azure Data Lake Store Gen 2 for BI / Analytics: Ingest data from Operational Data Stores

Azure

Real-time streaming into Azure Event Hubs and HDInsight for Stream Analytics

Attunity Replicate Architecture


HADOOP
RDBMS
DATA WAREHOUSE
FILES
MAINFRAME
CDC
BATCH
TRANSFER
IN-MEMORY
FILE CHANNEL
FILTER
TRANSFORM
PERSISTENT STORE
INCREMENTAL
BATCH
HADOOP
RDBMS
STREAMING
DATA WAREHOUSE
FILES

Azure Data Lake Storage Gen 2

MICROSOFT

Jeff King, Senior Program Manager for Azure Data Lake Storage Gen2, Microsoft

What Makes a Great Data Lake?

Massive scale	Secure	Optimized for maximum performance	Integration Friendly	Cost Effectiveness
PB-scale, data accessible everywhere, growth on demand	Granular security and protection against accidental data loss	Lightning-quick job execution	Supports multiple methods of data ingress, processing, egress and visualization	Cloud economic model with ability to intelligently manage costs

Rich data management and governance

Azure Data Lake Storage Gen2

A **“no-compromises”** data lake: secure, performant, massively-scalable data lake storage that brings the cost and scale profile of object storage together with the performance and analytics feature set of data lake storage

SECURE

- ✓ Support for fine-grained ACLs, protecting data at the file and folder level
- ✓ Multi-layered protection via at-rest Storage Service encryption and Azure Active Directory integration

MANAGEABLE

- ✓ Automated Lifecycle Policy Management
- ✓ Object Level tiering

FAST

- ✓ Atomic file operations means jobs complete faster
- ✓ High throughput

SCALABLE

- ✓ No limits on data store size
- ✓ Global footprint (50 regions)

COST EFFECTIVE

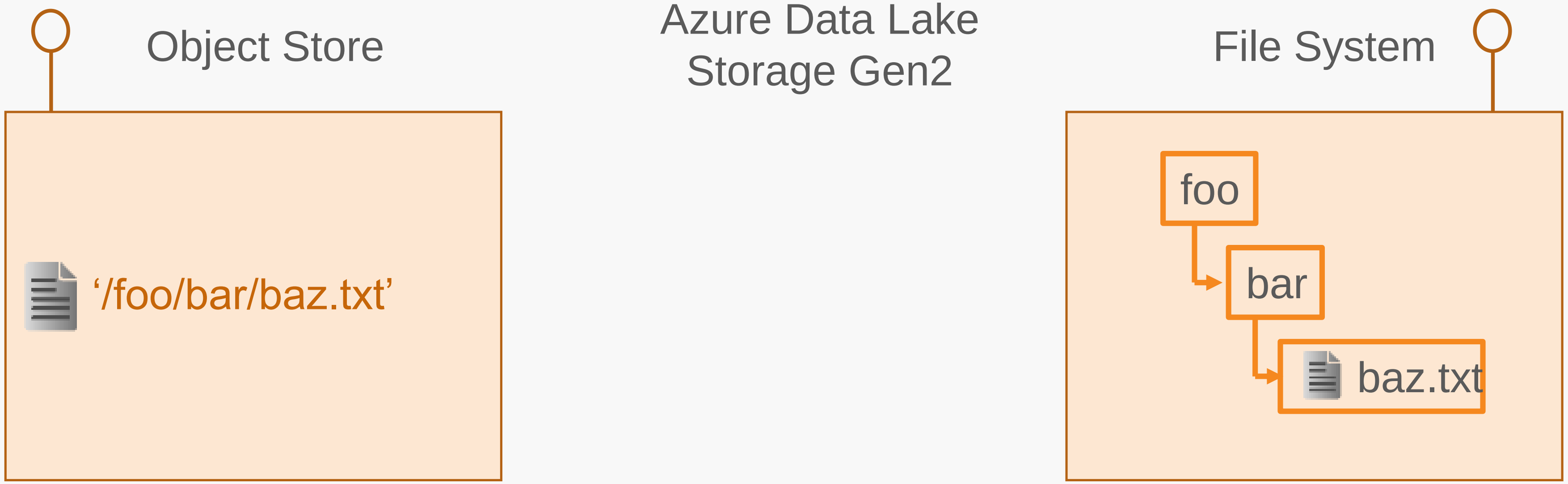
- ✓ Object store pricing levels
- ✓ File system operations minimize transactions required for job completion

INTEGRATION READY

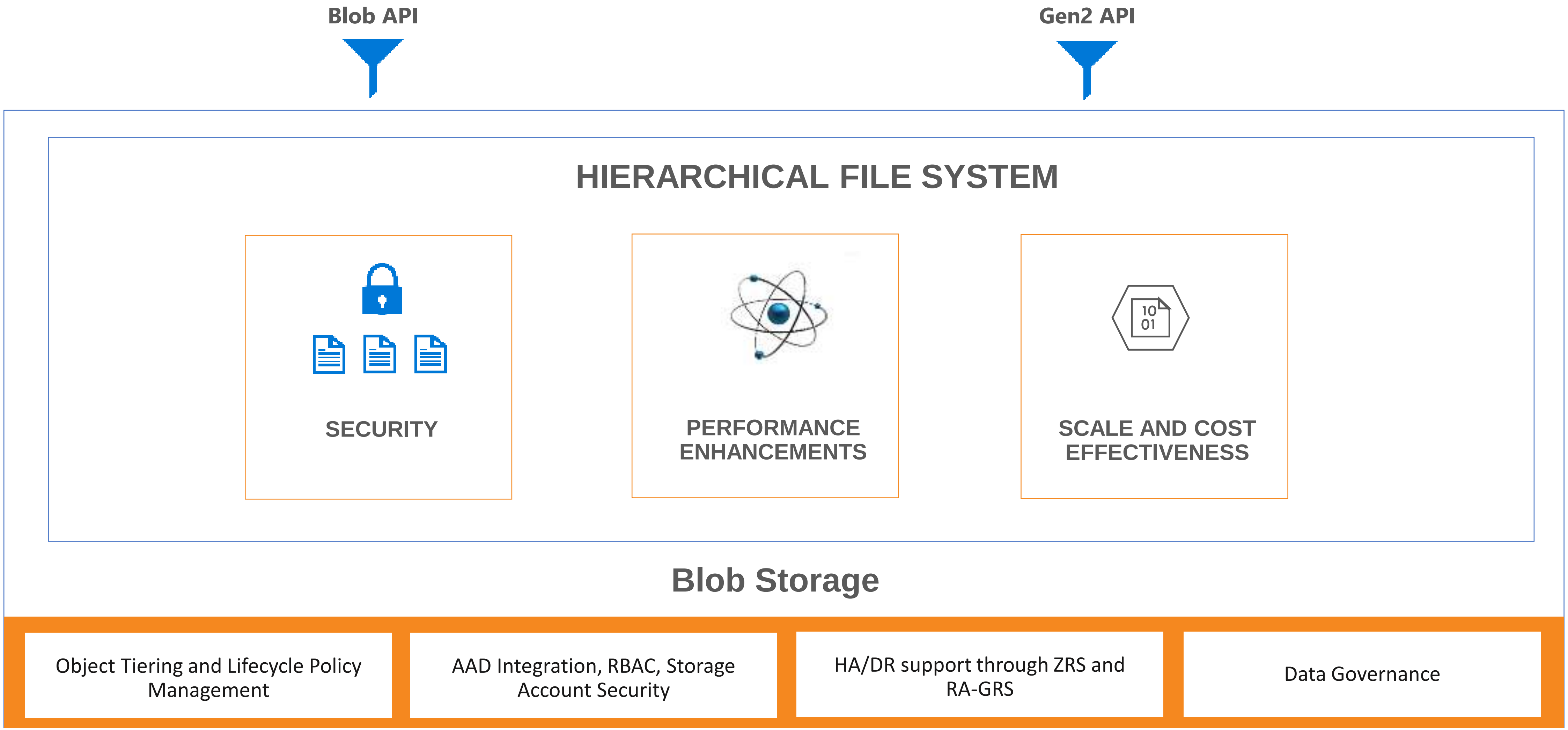
- ✓ Optimized for Spark and Hadoop Analytic Engines
- ✓ Tightly integrated with Azure end to end analytics solutions

A Good Data Lake Should be Multi-modal…

Object Store and File System access over the same data at the same time


Object Store
'/foo/bar/baz.txt'
Azure Data Lake
Storage Gen2
File System
foo
bar
baz.txt

Azure Data Lake Storage Gen2 - Architecture



Blob API

Gen2 API

HIERARCHICAL FILE SYSTEM

SECURITY

PERFORMANCE ENHANCEMENTS

SCALE AND COST EFFECTIVENESS

Blob Storage

Object Tiering and Lifecycle Policy Management

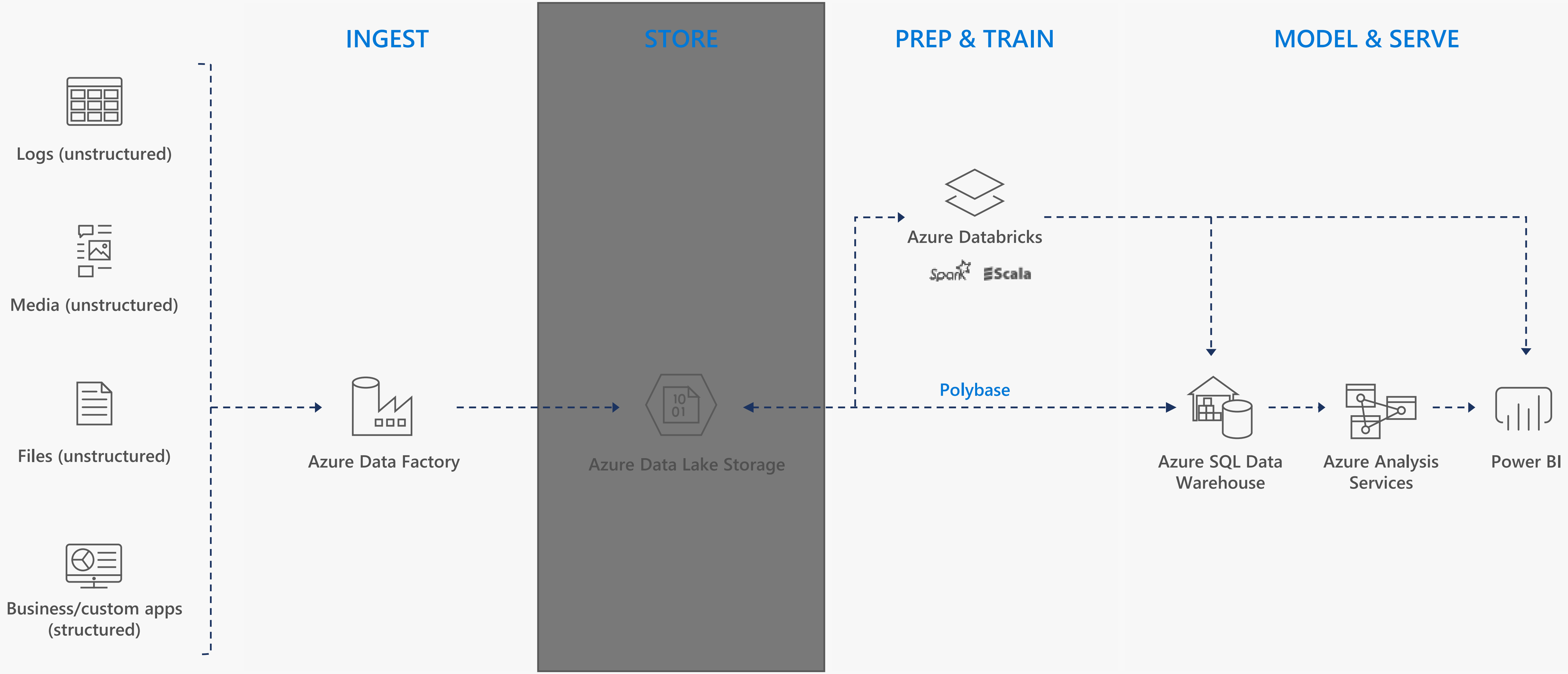
AAD Integration, RBAC, Storage Account Security

HA/DR support through ZRS and RA-GRS

Data Governance

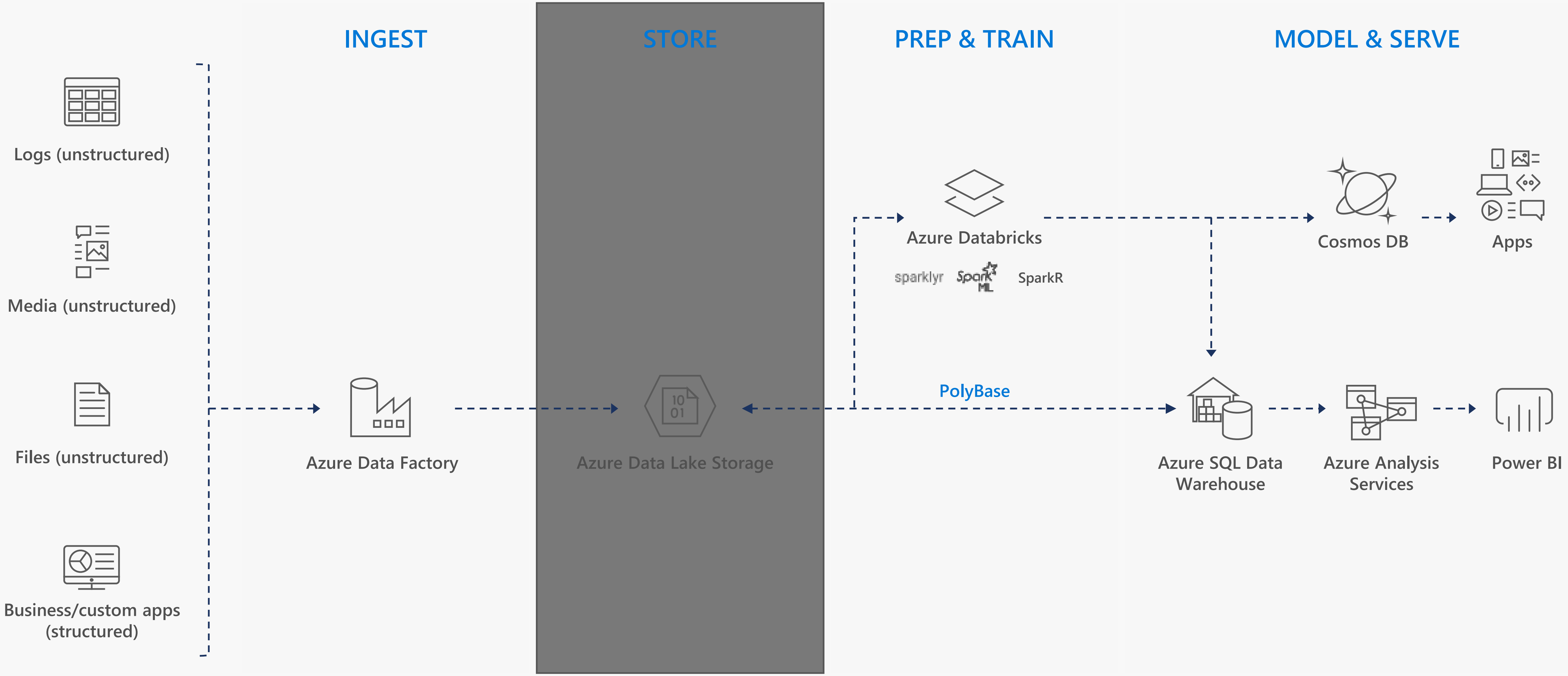
End-to-End Analytics

Modern Data Warehouse


INGEST
STORE
PREP & TRAIN
MODEL & SERVE
Logs (unstructured)
Media (unstructured)
Files (unstructured)
Business/custom apps (structured)
Azure Data Factory
Azure Data Lake Storage
Azure Databricks
Spark
Scala
Polybase
Azure SQL Data Warehouse
Azure Analysis Services
Power BI

End-to-End Analytics

Advanced Analytics



INGEST

STORE

PREP & TRAIN

MODEL & SERVE

Logs (unstructured)

Media (unstructured)

Files (unstructured)

Business/custom apps (structured)

Azure Data Factory

Azure Data Lake Storage

Azure Databricks

sparklyr Spark ML SparkR

PolyBase

Cosmos DB

Apps

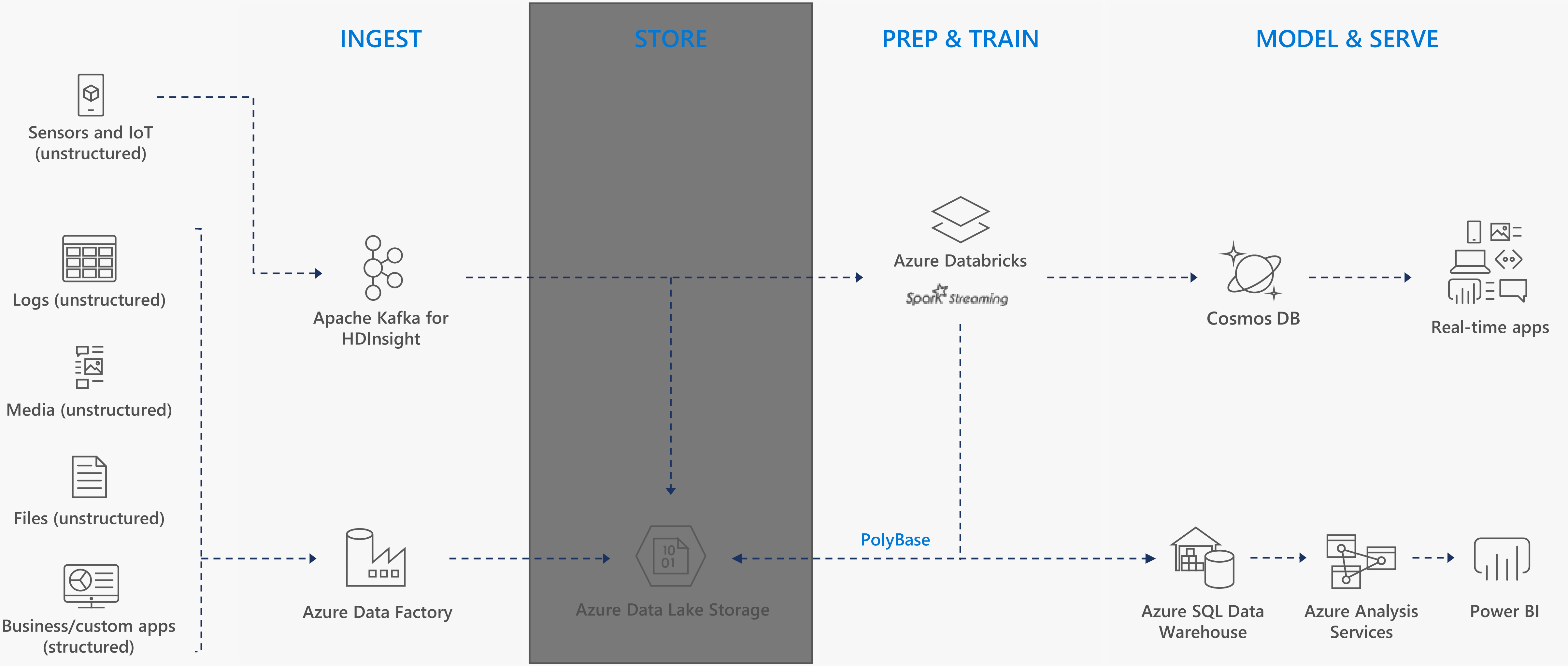
Azure SQL Data Warehouse

Azure Analysis Services

Power BI

End-to-End Analytics

Realtime Analytics



INGEST

STORE

PREP & TRAIN

MODEL & SERVE

Sensors and IoT (unstructured)

Logs (unstructured)

Media (unstructured)

Files (unstructured)

Business/custom apps (structured)

Apache Kafka for HDInsight

Azure Data Factory

Azure Data Lake Storage

Azure Databricks

Spark Streaming

PolyBase

Cosmos DB

Real-time apps

Azure SQL Data Warehouse

Azure Analysis Services

Power BI

Partner Ecosystem – Industry Verticals

- **Financial Services & Insurance**
  - Compliance Archiving Insurance Risk Modeling
  - Electronic Document Storage Financial Risk Modeling
- **Healthcare & Life Sciences**
  - Genomics Medical Imaging (PACS, VNA)
- **Media & Entertainment**
  - Media Asset Management Post Production
  - Media Storage and Archiving Distribution
- **Public Safety & National Security**
  - National Security (Body Cam) Enterprise Video Surveillance
- **Automotive**
  - Autonomous Driving/Assisted Driving Connected Cars
- **Oil and Gas AEC Govt Retail**

Partner Ecosystem - Horizontals

Data Center Backup and Recovery

- COMMVAULT
- veeam
- VERITAS
- druva
- actifio Radically Simple
- Quest
- Acronis
- MICRO FOCUS
- COHESITY
- rubrik
- CARBONITE
- FalconStor
- IBM Spectrum Protect
- UNITRENDS
- DELL EMC

Distributed File Systems and Object Storage

- VERITAS
- NetApp
- caringo
- CLOUDIAN
- DELL EMC
- COHESITY
- SCALITY
- Igneous

Cloud Storage Gateway

- NASUNI
- panzura
- TALON
- NetApp
- ctera
- PEER

Data Integration

- ATTUNITY

Attunity Replicate for Microsoft Migrations

ATTUNITY

ATTUNITY REPLICATE FOR MICROSOFT MIGRATIONS

DOWNLOAD NOW

FREE DOWNLOAD

https://aka.ms/attunity-replicate

ATTUNITY REPLICATE for Microsoft Migrations

Download Attunity Replicate for Microsoft Migrations

JOIN THE COMMUNITY

http://attunity.com/MicrosoftMigrations

ASK QUESTIONS

attunity-replicate@microsoft.com

Thank you

Microsoft

ATTUNITY