

//ADA STRA

Hadoop to Synapse Migration

Kevin Harmer





Agenda

Adastra Overview

The Synapse Advantage

Migration Approach

Appendix: Synapse Overview

Appendix: Synapse Security

Appendix: Synapse Architecture

Appendix: Ingestion Flexibility



Adastra Overview



Adastra delivers industry leading **Solutions and Services** across the **Data & Digital** spectrum

DATA

Cloud
Data Engineering
CRM & ERP
Architecture
Managed Services

GOVERNANCE

Data Governance
Data Quality
Master Data Mgmt
Reference Data Mgmt
Meta Data Mgmt
Data Lineage

ANALYTICS & AI

Machine Learning
Statistical Analysis
Text Mining
Visual Analytics
RPA & IPA

DIGITAL & DESIGN

Mobile Apps & Web
Development
UX/UI
Personalized Videos
Emerging
Technologies



Adastra Global Footprint

2000+

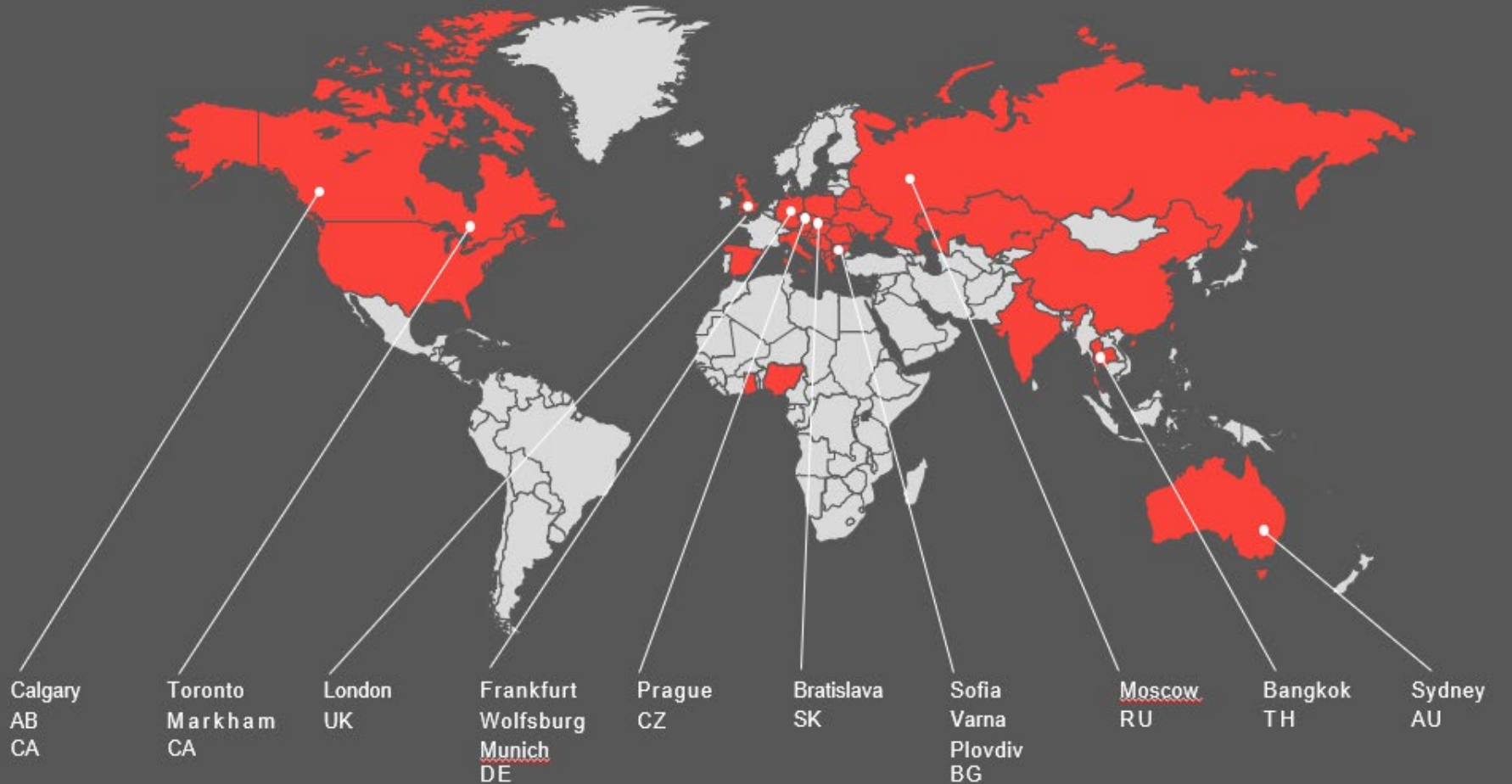
2000+ professionals

1000+

Projects in 46 countries

1

Focus





Adastra Success Stories

Adastra has successfully architected / implemented Azure Modernization solutions for > 65 organizations in the last 24 months. Select clients:



Next Done Now.™





Adastra Difference

Client Focused



Strategy Focused

We develop solutions that meet client's current and future business needs.



Value Driven

We deliver data-driven value to our customers with measurable ROI.



Nimble

We share knowledge and expertise with nimble work structures that allow us to scale as needed.



Transparent

We work with our clients to manage scope, timelines and level of detail.

Best-In-Class Solutions



End-to-End Data Services

We provide services from inception to deployment and maintenance for effortless client execution.



Excellence of Execution

Nothing matters if it can't be delivered on time, on budget, and within scope.



Top Talent

We hire only the best and the brightest and invest in their training and development.



Best Shoring

Top global experts can be assigned to take advantage of economies of scale and provide 24/7 delivery.

Data

AI & Analytics

Cloud

Governance

Digital



Adastra Microsoft Partnership

Microsoft
Partner



Gold Data Analytics
Gold Data Platform
Gold Cloud Platform
Gold Datacenter
Gold Application Integration

Adastra: Go-To Partner for Data & AI

Microsoft IMPACT Award Recipient:

- 2021 Analytics Impact Award
- 2021 AI Impact Award
- 2021 Data Platform Modernization Award
- 2020 / 2019 Commercial Partner of the Year
- 2019 Manufacturing Innovation Impact Award

Azure Migration Partner

Advanced Specialization for Windows / SQL
Server Migration to Azure

Lead Canadian Partner for Synapse Migration
/ Implementation

Product Team Collaboration for Azure
Synapse / Azure Purview / Azure Databricks



Adastra Azure Specializations

AZURE FOUNDATION

cloud assessment, tenant design, resource naming, service tag approach, network architecture, governance design, tco analysis, hybrid network implementation, devops integration, azure foundation implementation, ...

AZURE APP / MIGRATION

app / data assessment, app / data decisioning (lift / shift vs modernize), app / data architecture, api architecture, microservices architecture, app / data security design, migration roadmap, migration execution, iac pipelines, devops integration, ...

AZURE SECURITY

security assessment, identity strategy, role based access, secrets management, encryption, data loss protection, api management, private zone configuration, siem / soar integration, policy enforcement, security implementation ...

AZURE ANALYTICS

analytics assessment, analytics architecture, analytics roadmap, data zoning, enterprise model design (kimball, inmon, data vault), ETL data pipelines, persona enablement, citizen report development, trusted data as a service, ...

AZURE BIG DATA

big data assessment, data lake design, Hadoop integration, PySpark data engineering, ELT pipelines, spark delta lake, spark streaming, serverless compute, devops integration, ...

AZURE AI / ML

advanced analytics assessment, cognitive service integration, r&d model training / testing, mlops implementation, ai / ml pipelines, data science workbench automation, devops integration, ...

POWER PLATFORM

citizen development assessment, power platform governance, roles / responsibilities, environment strategy, CoE kit, canvas / model apps, power automate flows, power automate rpa, power platform dataverse, power bi datasets / reports, ...

AZURE DATA GOVERNANCE

data governance assessment, data catalog, data classification, data sensitivity, data use governance, data privacy, data lineage, master data management, data quality management, reference data management

AZURE INTEGRATION PAAS

integration paas assessment, api management, logic workflows, service bus management, event grid distribution, peer to peer patterns, pubsub patterns, managed file transfers, iot telemetry streaming, iot edge device management



Adastra Managed Services





The Synapse Advantage



Data Is The Backbone of The Enterprise

Today, all businesses are data businesses

- / Data makes business possible
- / Data is crucial to businesses of all sizes
- / Security data is mandatory

All data businesses need to be analytics businesses

- / Business need full transparency
- / Workers are empowered with intelligent insights
- / Augmented intelligence sees patterns we cannot

Analytic businesses need to evolve data science

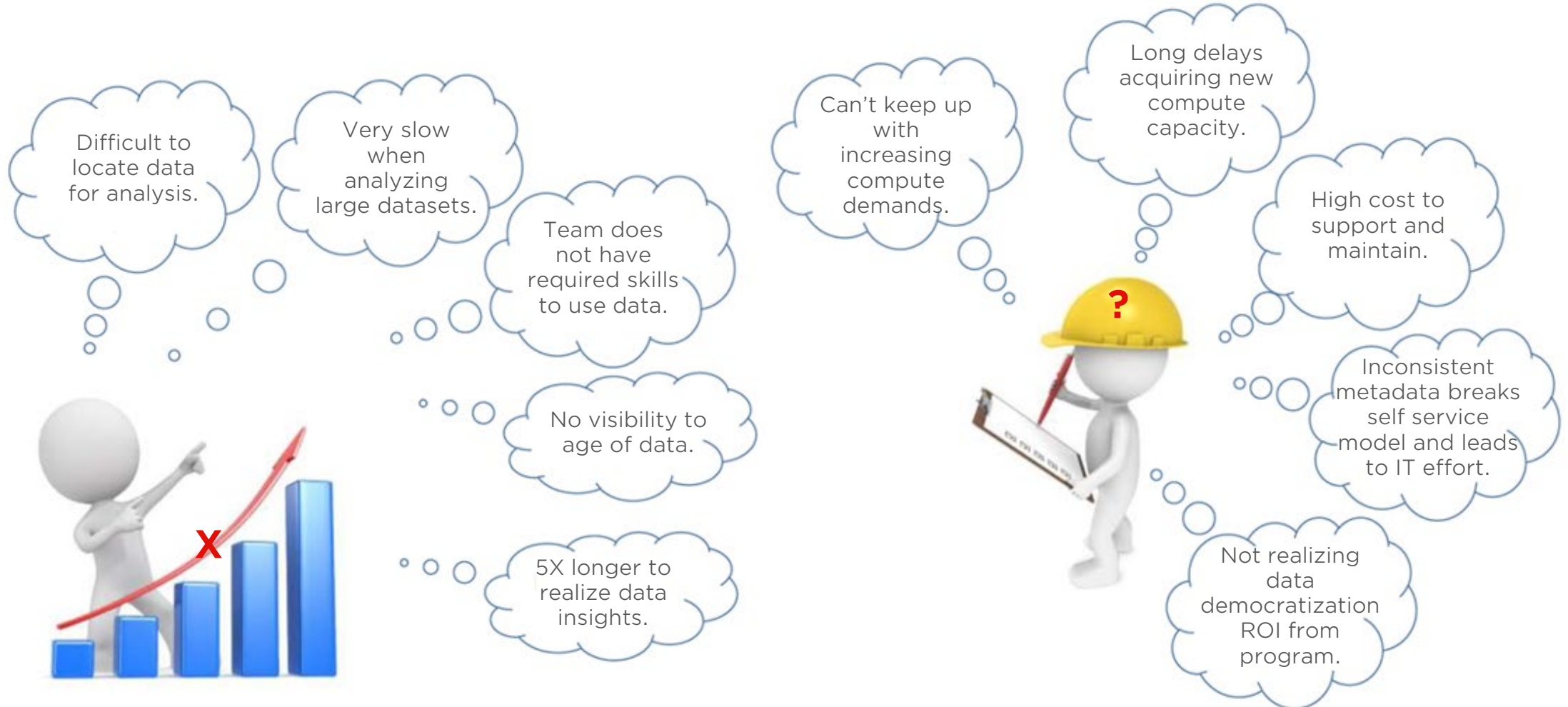
- / Data science empowers us to go farther
- / AI/ML creates smarter organizations
- / Businesses need scalable analytics



Hadoop to Synapse Migration Use Case

Adastra recently helped a large Canadian bank migrate on premise Hadoop to Azure Synapse.

Below are direct quotes regarding their legacy on premise Hadoop environment, that drove the Azure Synapse migration



Business Perspective

IT Perspective



Typical 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



The Azure Synapse Solution

**Implement Services
in Minutes**

Unlimited Scale

Pay for Use Pricing

**Serverless
SQL / Spark Compute**

**Fully Integrated
Platform as a Service**

**Azure Governance
Ecosystem**

**Reduces TCO by up to
50%**

**Realize Data Insights
3X Faster**

**Unleash Citizen
Data Insights / Analytics**



Synapse = One Integrated 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	✓ (OAuth / OpenID)



Hadoop to Synapse Migration Approach



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
- / 3x faster migration via data sync / workload conversion tooling

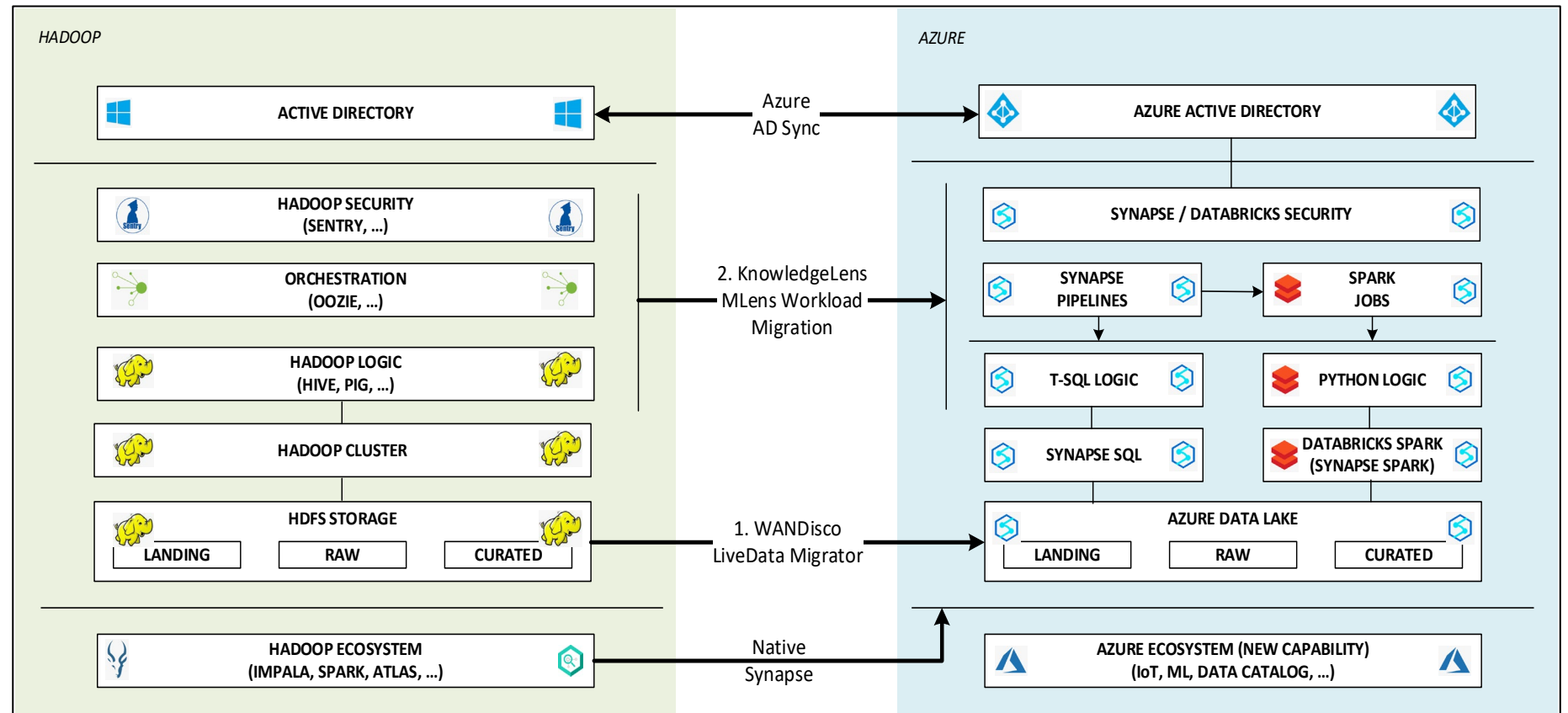


Migration Steps

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

1. Implement HDFS to Azure Data Lake live sync via 3rd party tool LiveData Migrator. Hadoop HDFS storage changes are continuously copied to Azure, to keep Azure data current.
2. Use 3rd party tool KnowledgeLens MLens's 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.





WANDISCO LiveData Migrator



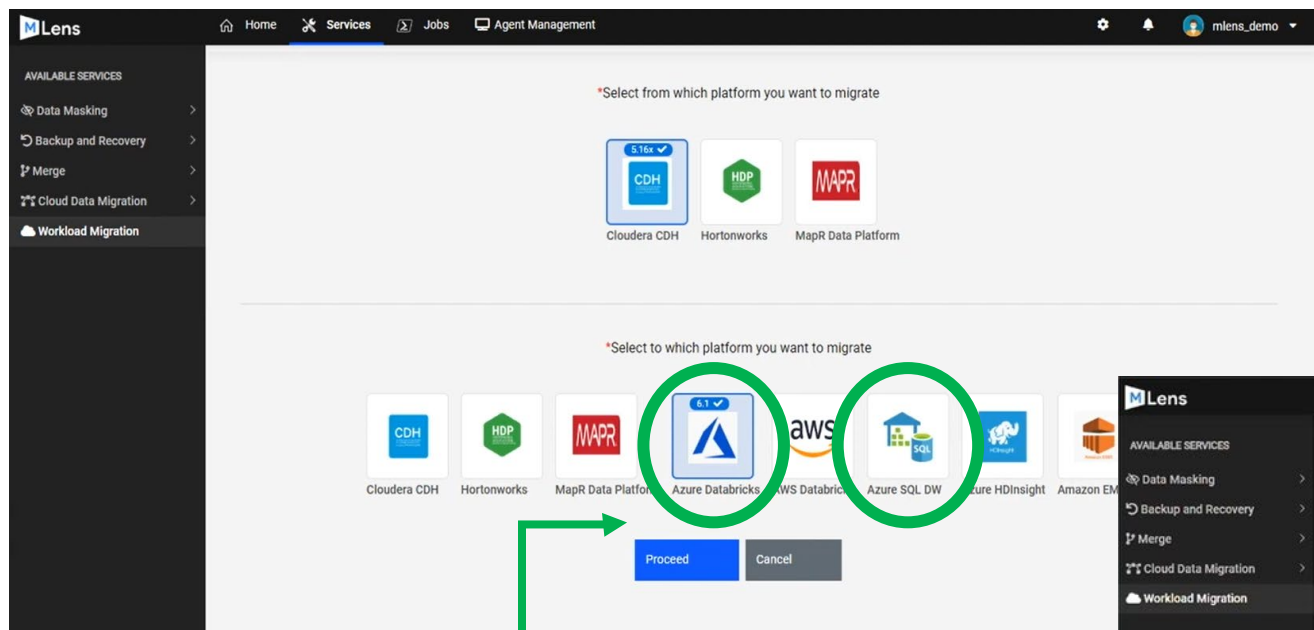
Live Data Migrator eliminates risks for big data migration to cloud
with its ONE Scan, ONE Consensus, and ONE Touch Replication capabilities

LiveData Migrator for Azure continuously replicates changes to provide
LIVE DATA
that is always available, accurate, and protected



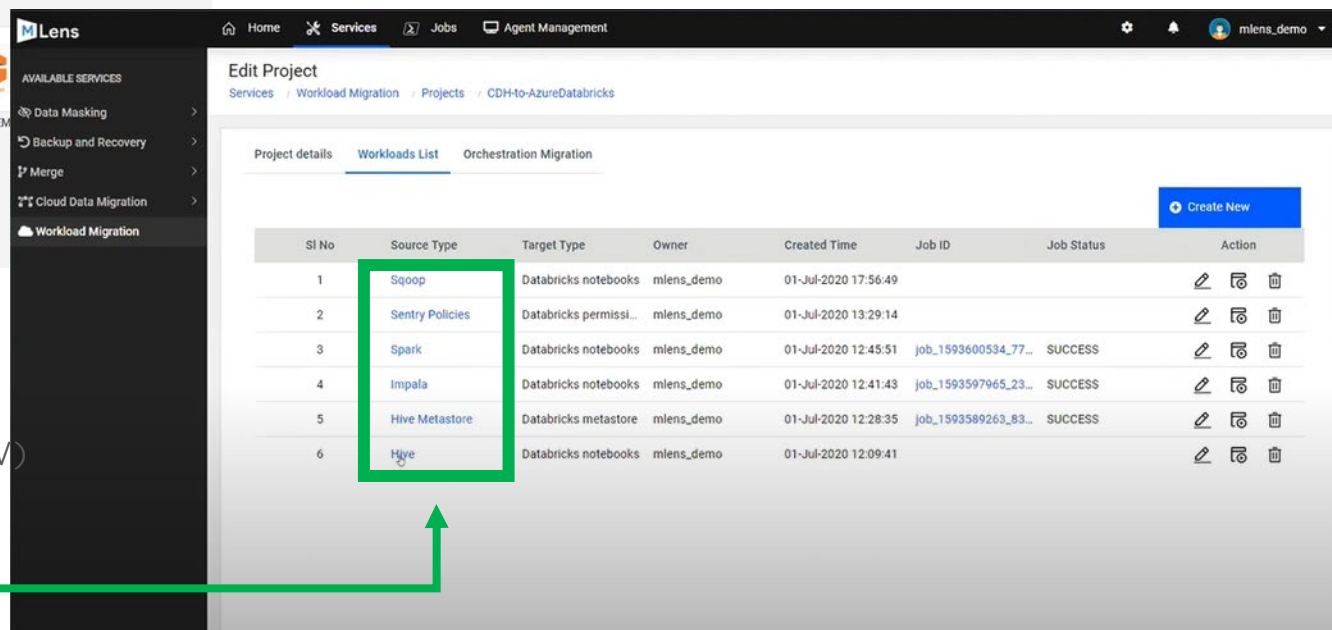
KnowledgeLens MLens Workload Migration

MLens Workload Migration transforms Hadoop workloads to Azure Synapse workloads.



Migrate Hadoop workloads to one of:
Azure Databricks Spark
Azure Synapse Spark (thru Databricks)
Azure Synapse SQL (formerly Azure SQL DW)

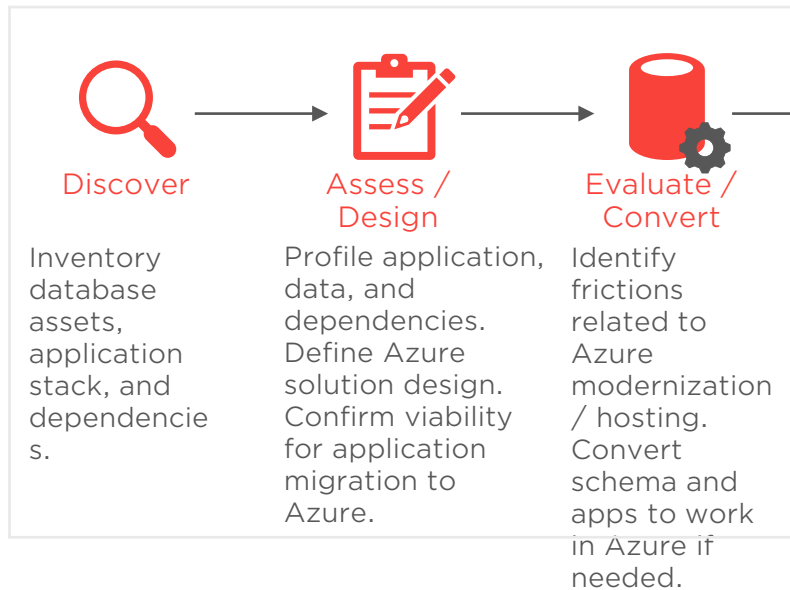
Various Hadoop workload types supported.



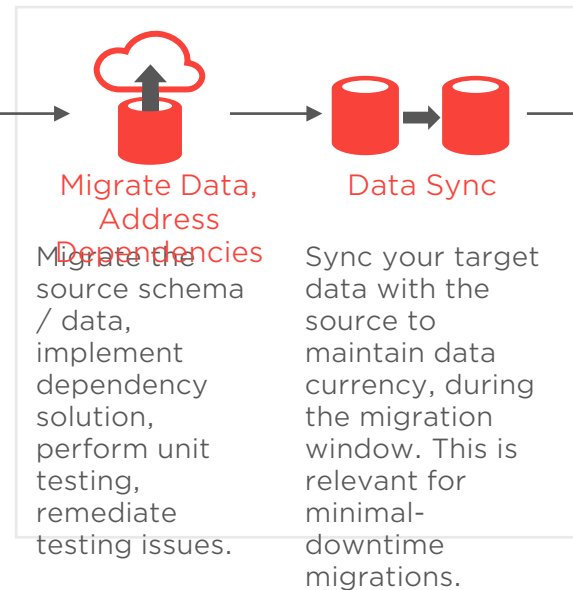


Migration Process

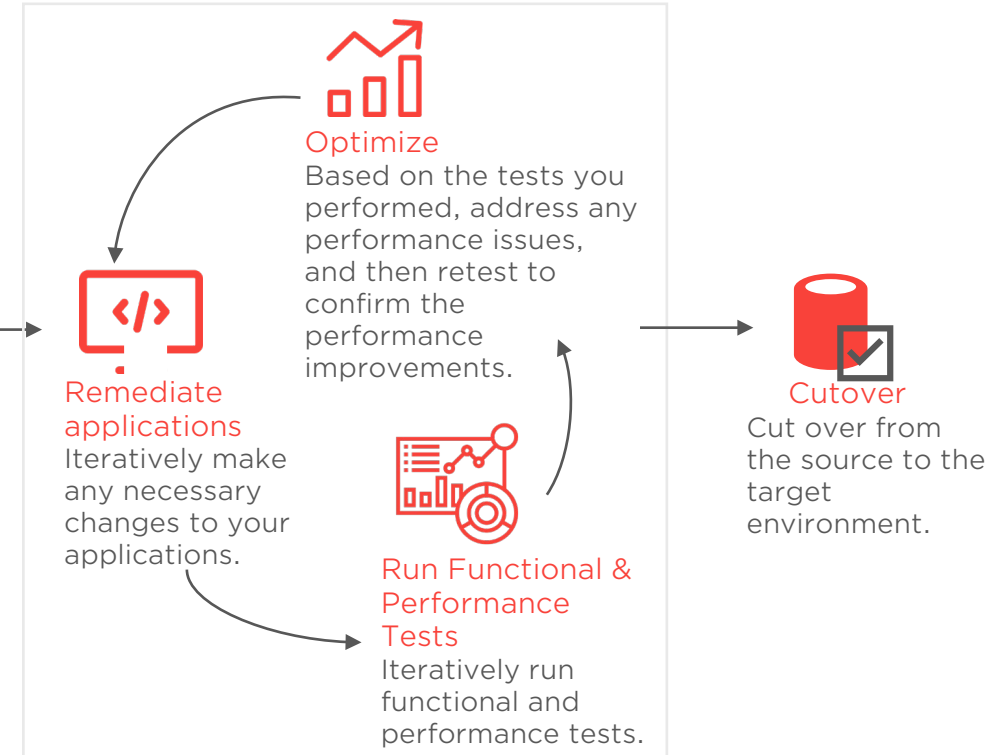
Pre-migration



Migration

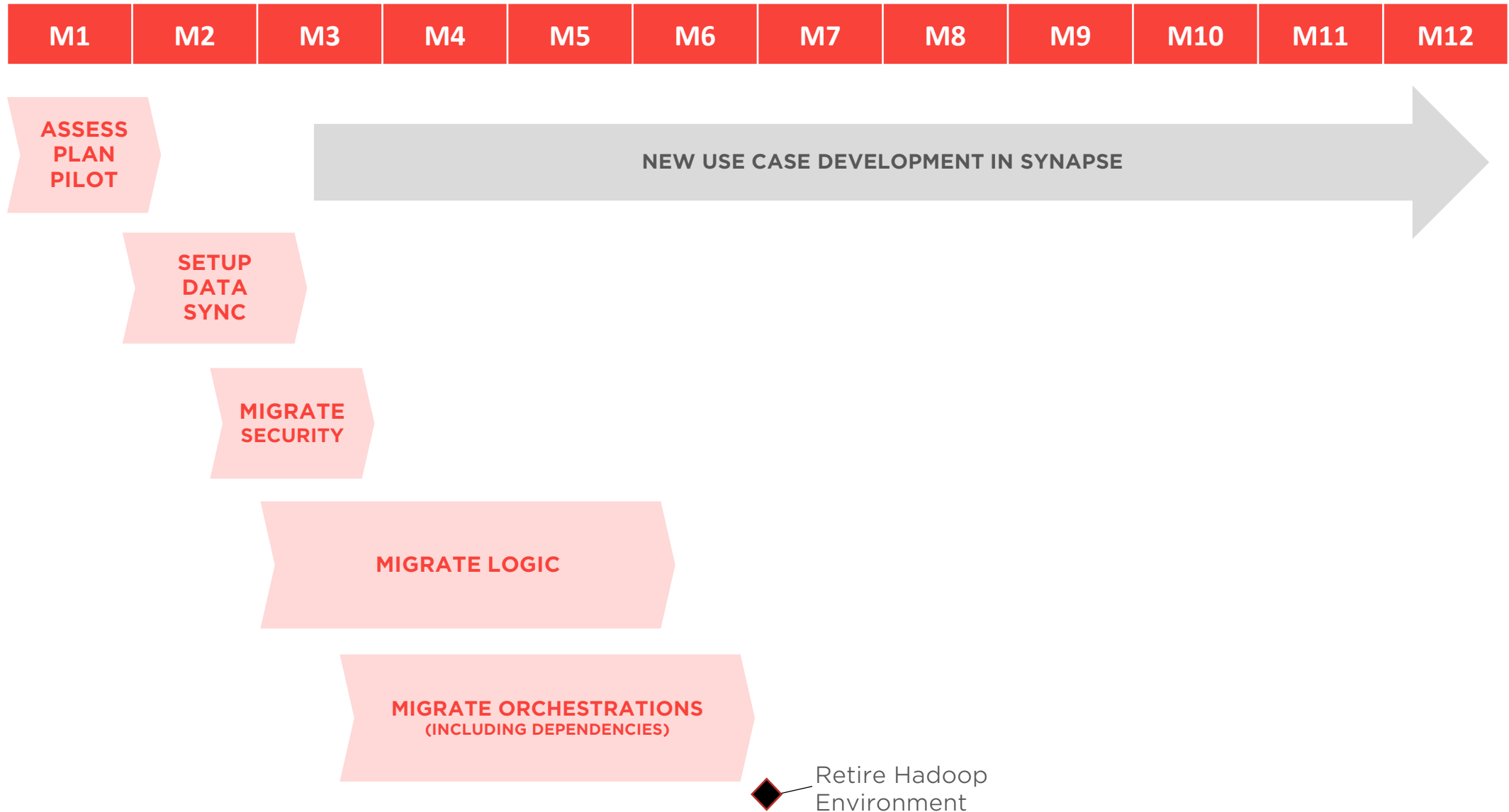


Post-migration





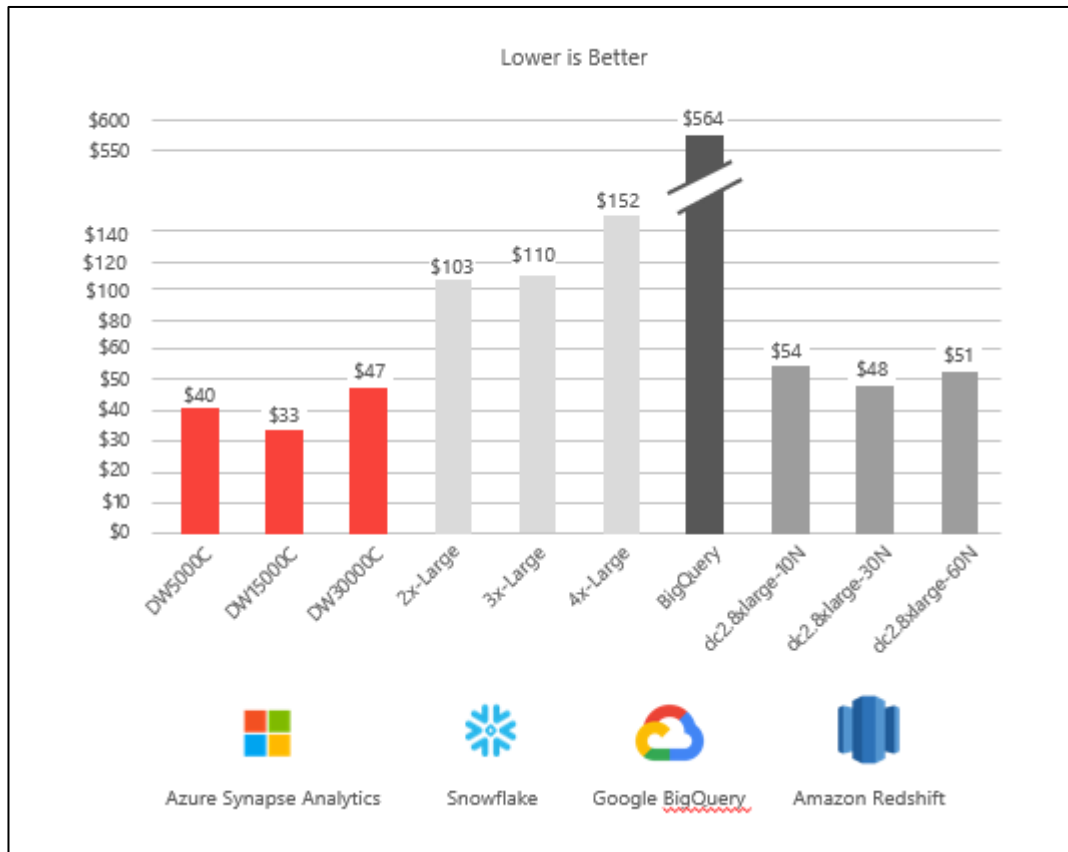
Hadoop to Synapse: Typical Roadmap



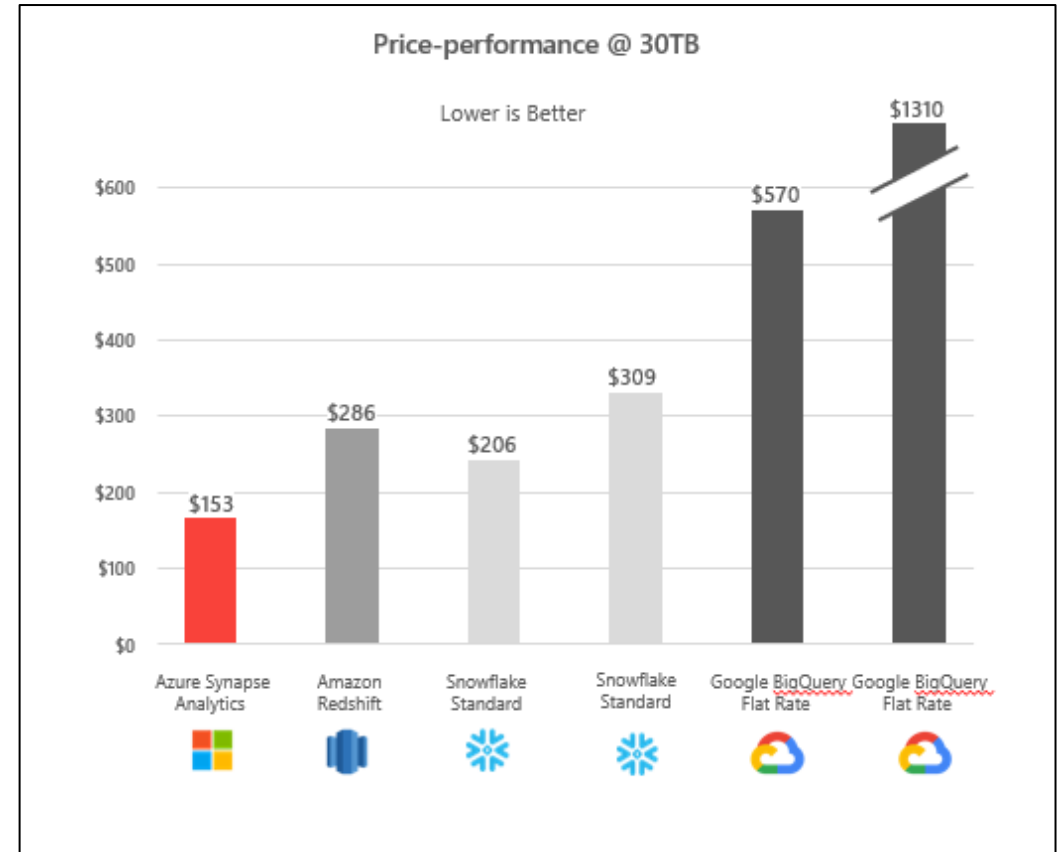


Synapse = Lower TCO Cloud Computing

Cost per TPC-H Small Query vs Competitor Similar Clusters



Cost per TPC-H Large Query vs Competitors



(costing via independent 3rd party GigaOm report)



Adastra Offer: Free Migration Assessment



Roadmap your migration from Hadoop to Synapse with Adastra's free 30-day migration assessment.

Our no-commitment offer will accelerate your Synapse adoption and address current and future analytics goals, delivering tangible business value.

Through our assessment and roadmap, Adastra will:

- Assess current Hadoop Environment
- Validate logic for conversion to Synapse
- Design / size / cost Synapse service design and network design
- Define security design for Synapse environment
- Define technical governance for Synapse environment
- Build migration roadmap / plan
- Pilot migration to validate approach
- Free use of WANdisco (up to 25TB) for pilot

Value: \$50k



//ADA STRA

For Questions contact:

Kevin Harmer

Microsoft Practice Lead

kevin.harmer@adastragr.com

647-990-2101