avanade

Do what matters

Considerations and decisions for your Data, Analytics and AI Platform

The information provided in this document is for informational purposes only and does not create a business or professional services relationship between you and Avanade. The information in this document is provided on an AS IS basis, with no warranty; it is not necessarily complete, or up to date, and is subject to change without notice. Any use you make of the information in this document is at your own risk.

“Complexity is your enemy. Any fool can make something complicated. It is hard to keep things simple.”

- Richard Branson

Executive Summary

- The importance of a data, analytics and AI platform is unquestionable in today’s connected and always-on world. A full eco-system of technologies exists to support the capabilities that we find in these platforms, supporting data integration, orchestration, ingestion, storage, processing, analysis, modelling, security and sharing.
- Having greater choice increases your option, when it comes to data platforms this adds complexity that far too often causes organizations many costly problems.
- There are many overheads and risks when designing and delivering a data, analytics and AI platform that is built with many third-party components. The alternative is to build and use an end-to-end unified cloud Platform that does not require non-native components for core needs.
- In this whitepaper we outline appropriate considerations when reviewing data, analytics and AI platform capabilities with Azure Databricks our key reference point.

Background

Many organizations are moving away from legacy data systems to modern data, analytics and AI platforms. Assessing the benefits and capabilities to select a new modern platform architecture is no easy task.

Historically, organizations have combined a myriad of technologies to support specific data and analytics roles. This came with the consequence of high integration costs and the complexity of managing teams of experts across different technologies. Subsequently, engineering overruns, impacted ROI, and in many cases more than one platform was required to work around compatibility issues.

Considered due diligence in assessing your needs carefully will pay dividends. While there are several players in this data and AI landscape, your considerations will need to go beyond basic assessment dimensions Total Cost of Ownership (TCO) and performance.

Avanade recently shared how Azure Synapse analytics and Azure Databricks **are better together in the modern data architecture.**

avanade

What's in a **Data, Analytics & AI platform**?

A data, analytics and AI platform is a technology solution designed for end-to-end data treatment and capitalization of data assets through analysis, analytics and/ or Data Science methods. To achieve this, such platforms must be able be ingest, store, process, analyze and share data and metadata, as well as to connect to upstream and downstream technologies.

Today, the major cloud vendors support varying levels of capability and coverage with their native (out-of-the-box) data, analytics and AI platform offerings. For example, Microsoft offers a unified data and AI platform, the Microsoft Intelligent Data Platform ('MIDP')[1].

With Microsoft Azure, its native technologies are automatically compatible across the entire Azure suite of Microsoft products and services. Services like Azure Data Factory easily connect to Azure Machine Learning, Azure Synapse Analytics, Azure Databricks, etc.

Through Avanade's own accelerators and industry specific data platform blueprints, organizations can quickly deploy their data, analytics and AI (DAAI) platform with all required services and products. We've seen data engineers hands-on in the production environment in less than 12 hours. This is a 100% Azure native data platform.

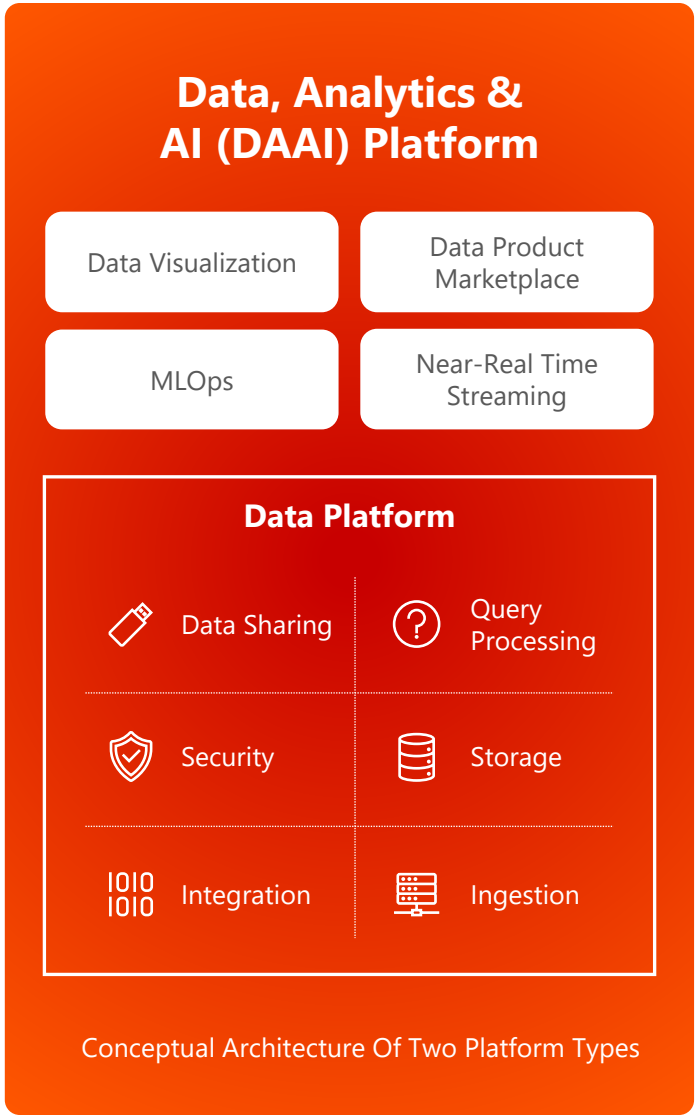
Consideration #1:

How much Is your existing Data and AI non-native platform strategy costing you in effort and associated costs? How many third-party technologies have you had to deploy to satisfy extended integration needs?

The Data Lakehouse

Newer to the scene is Lakehouse architecture which underpins the next generation of data, analytics and AI platforms.

Gartner defines a data lakehouse as "a converged data architecture that combines and unifies the architectures and capabilities of a data warehouse and a data lake deployed on a single platform (usually cloud-based). This setup enables data and analytics leaders to reap the leading benefit of the data lakehouse: the reduction of architectural redundancies"[2]



Conceptual Architecture Of Two Platform Types

Databricks and **Microsoft**

Databricks is one of several players on the Lakehouse scene. Databricks has innovated with a selection of opensource technologies alongside the Apache Spark framework. Delta Lake is the optimized storage layer that provides the foundation for storing data and tables in Databricks Lakehouse Platform[5]. It enables incremental data processing at scale, critical for organizations looking to capitalize on the latest round of AI and generational AI advancements.

Delta Lake is recognized as the most widely used lakehouse formats, with vocal proponents including Apple, Adobe, Amazon, Disney, but to name a few.[7]

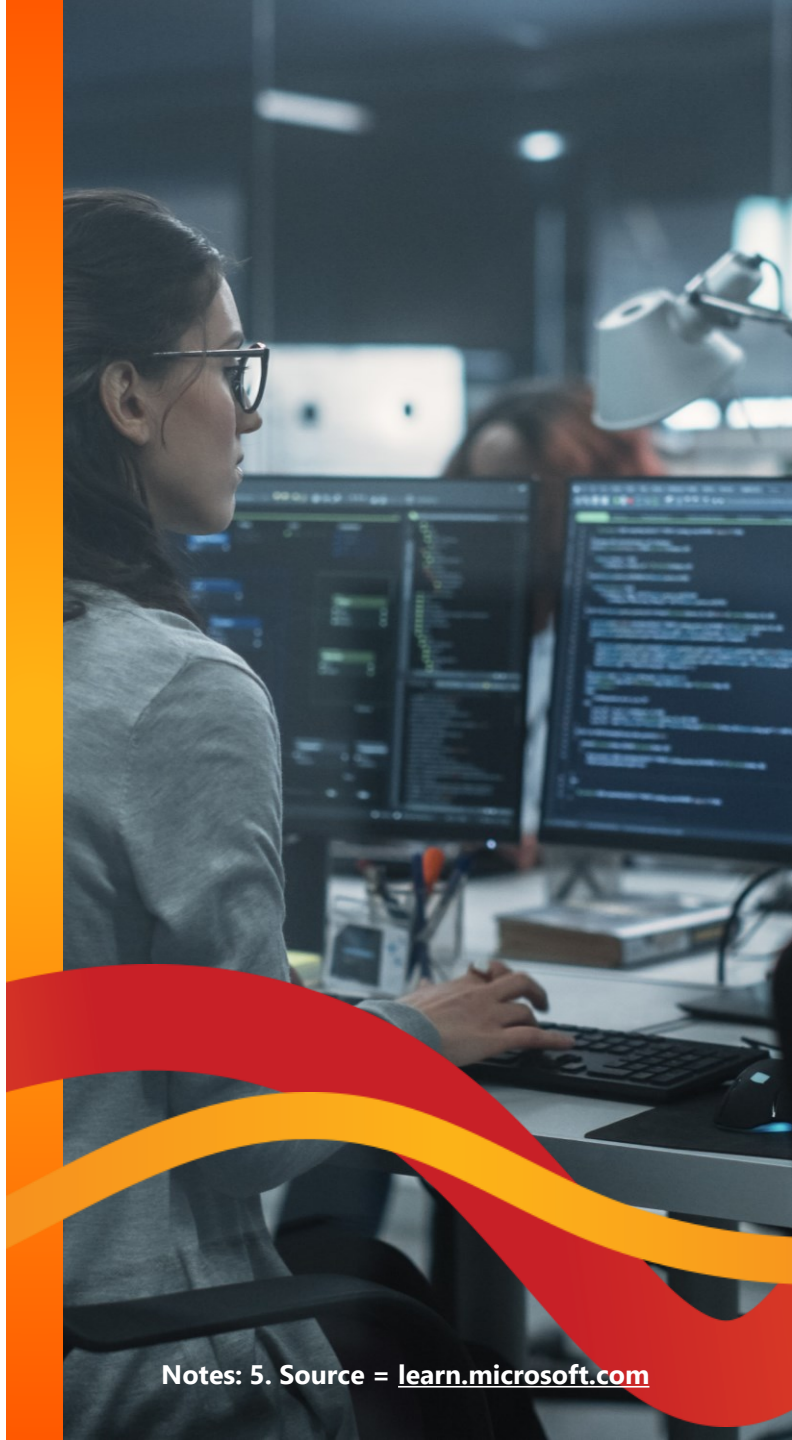
Unlike the previous generation of data lakes, "industrial lakehouse" architectures can index and govern a data lake at scale. Key innovations that have made this possible are Microsoft Purview and Unity Catalog, a Databricks technology, both available across multi-clouds.

Microsoft Purview is focused on managing and governing on-premises, multi-cloud, and software as a service (SaaS) data. It can create a holistic, up-to-date map of a data landscape with automated data discovery, sensitive data classification, and end-to-end data lineage.

Unity Catalog is an oversight and governance service that enables the lakehouse through fully managed and automated capabilities. It facilitates search, discovery, and automated lineage for all workloads in SQL, R, Python, Scala and across all asset types - tables, notebooks, workflows and dashboards.

On MIDP, Microsoft Purview has the potential to be a single connected pane, integrated with Unity Catalog, to govern operational, analytics, and ML/AI data assets, depending on what its users need.

In October 2022, Microsoft announced that the Microsoft Intelligent Data Platform would follow an open and governed Lakehouse foundation, now supporting the Delta format[6]. Thanks to the open nature of MIDP, in practice other emerging standards like Iceberg and Hudi can also be used, helping to future-proof organizations with additional options.



avanade

Notes: 5. Source = learn.microsoft.com

The **Industrial Lakehouse**

Avanade have delivered 500+ customer projects that have leveraged lakehouse architectures and we continue to invest in 1,000+ data engineers and architects who actively work with the lakehouse architecture.

We prefer to use the term "Industrial Lakehouse" to describe a lakehouse that is natively architected, orchestrated and managed. This provides distinction between non-commercial and commercially native solutions which enable a quicker time-to-value and overcome challenges raised by older generation (legacy) capabilities.

Legacy data and AI architectures were once focused on supporting data warehousing use cases. Today, most organizations are looking to leverage AI and machine learning capabilities, for front-office business and operational needs.

If you're considering a Lakehouse or industrial Lakehouse here are some quick answers to some burning questions that we've heard from clients.

Business Questions Answered:

Q1 Does it matter if a DAAI platform is closed or proprietary for business need considerations?

Answer: Some closed platform technologies hold data in a proprietary format which can require third-party components to deliver an end-to-end data, analytics and AI platform (e.g. ETL, BI). Open platforms can be easier to triage and fix issues and can result in less data conversion complexities when working with other business systems.

Q2 How granularly do you I need to manage and understand vendor pricing?

Why it's important: With open platforms like Microsoft Azure and Databricks everything is componentized for transparency. This level of granularity is necessary to understand as data workloads and use cases grow.

Q3 Do multiple separate bills make it hard to know the real cost of a Lakehouse-centric platform?

Answer: With platforms like Azure Databricks, which is a first-party service, all billing is covered through an Azure subscription with detailed breakdown of costs.

avanade

The **Devil** is in the **Detail**

Technology and Solution Questions:

Technology strategists and solution architects have an evolving set of questions and tests they need to answer before carefully considering their initial procurement plans and potential vendors.

Q1 Does a Lakehouse add complexity of managing separate services in a decoupled architecture?
Answer: We have found that non-native Lakehouse based platforms often require multiple services. Native based services such as Azure Databricks do not.

Q2 Does a Lakehouse add complexity when managing opensource services?
Answer: With native data platform technologies there is little if any complexity to managing opensource services.

Q3 Do I need to understand our volumes of data egress?
Answer: In some cases, platform technologies charge for data egress. Most open DAAI platforms do not.

Things to know about Azure Databricks:

Databricks is an end-to-end DAAI cloud-native, scalable platform that supports a broad set of needs from data engineer needs to data scientists to data analysts. This enables your data team to work on the data results you need to interpret your data for greater business impact.

Databricks' end-to-end coverage as a major positive as it enables greater synergy and collaboration.

Databricks is an open data, analytics and AI platform which means they do not have proprietary data formatting.

Databricks have two types of licensing, you can get started with pay-as-you-go and upgrade to an enterprise agreement when ready. If you are running non-business critical workloads, pay-as-you-go should suffice. If you using Databricks for business critical or enterprise dependence workloads, then an enterprise agreement is a 'must have'. Granular pricing enables flexibility to fine tune performance to maximize ROI.

avanade

Consideration **Matrix – 1/4**

This four-page consideration matrix supports a high-level procurement exercise when comparing two Cloud Data Platform services against your business requirements. For each service, score your business importance and then score your vendors. You may wish to weight the scoring based on your particular scenario or use case needs. We can provide advice and guidance for any/ all aspects of such a comparison should you need help.

	Importance to our business requirements	Service Vendor – E.g. Databricks	Service to be compared to	Service to be compared to	Notes
Azure Native (1st Party) Service	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Data integration/ orchestration	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Data Pipelines (Prepare/Transform)	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Compute Performance	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Cost transparency/ granularity	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Egress costs	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Monitoring of overall platform	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	

avanade

Consideration **Matrix – 2/4**

	Importance to our business requirements	Service to be compared to	Service to be compared to	Service to be compared to	Notes
Unstructured data capabilities	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Stand-alone data platform capabilities	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Level of vendor 'lock-in'	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Support of Data Mesh framework	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Openness of platform	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Integration with other platform components	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	

avanade

Consideration **Matrix – 3/4**

	Importance to our business requirements	Service to be compared to	Service to be compared to	Service to be compared to	Notes
Open source contribution	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Development frameworks	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Native Business Systems integration	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Machine Learning & Data Science	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Native Data Fabric	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Security features	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
InfoSec suitability	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	

avanade

Consideration **Matrix – 4/4**

	Importance to our business requirements	Service to be compared to	Service to be compared to	Service to be compared to	Notes
Near real-time streaming	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Innovation in Open Source	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
External data onboarding	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Developer experience	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Sustainability credentials	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	
Generative AI credentials	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	

This list of comparative capabilities is not-exhaustive and care should be given based on your organization's needs and wants. The purpose of this consideration matrix is to highlight many common features that should be assessed as a minimum, subject to market and environmental changes.

avanade

How to think about **Data Mesh**

At Avanade we're seeing enterprise clients moving away from centralized platforms and embracing data streaming over traditional batch processes, sometimes known as Data Mesh.

What is a Data mesh?

According to Databricks, Data mesh is a paradigm that describes a set of principles and logical for scaling data analytics platforms. The purpose is to derive more value from data as an asset at scale.

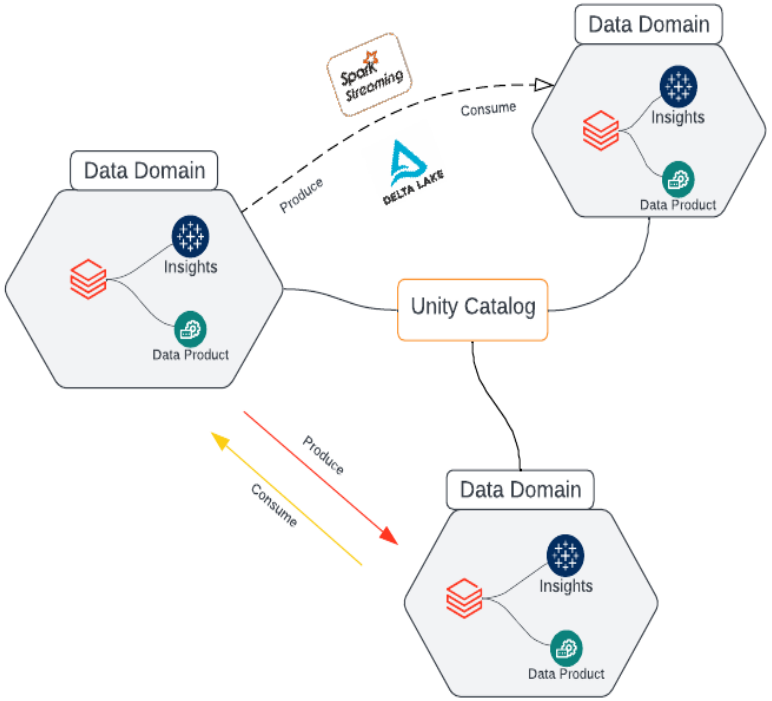
This enables insights to happen more quickly and against vast amounts of data - from days to minutes and seconds.

How does Data mesh play into a Lakehouse strategy?

Scaling your analytics platform enables you to produce quicker insights against vast amounts of data at scale. Streaming data with a data mesh on open standards allows organizations to reduce data silos.

Data Mesh is more not a solution per se, but an architectural and organizational paradigm, usually comprised of the following components.

Source Data	Self-service compute resources and orchestration	Domain-oriented Data products served to other teams and domains	Insights ready for consumption by business users	Adherence to federated computational governance policies



Value you can **Count On**

You know your business. Avanade knows data.

Your people are busy running your business. Bring in the experts who can partner with you to help improve productivity and deliver your vision.

In return, you'll gain confidence and trust as we help you unlock the value of your data, realize the promise of AI and make your company processes and experiences more intelligent.

Data platform optimization

Our people can help you unpick complex challenges that can unlock and accelerate the value of technology. This expertise is packaged into variety of tools that enable clients accelerate their transformation.

Solutioning expertise

Our people are deeply skilled – whether your challenge is rooted in the complexity of data strategy, technology adoption, security, architecture, or governance, we can help you to define direction, build comprehensive plans, and start to execute at pace.

Local and global delivery

We understand that keeping the lights on whilst building and executing on forward thinking strategies can be demanding – we provide our experts into our clients' business to accelerate key initiatives and solve complex problems quickly.

Industry Accelerators

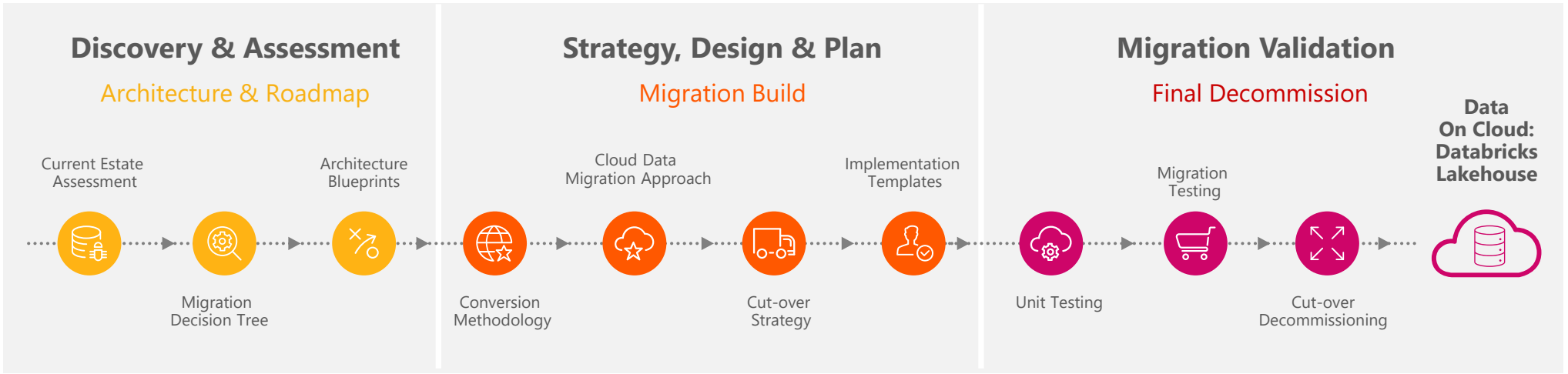
Our work across multiple industries and technologies allows us to provide valuable insights – whether this be industry, technology, geography specific, we can deploy assets and our accelerators to suit your needs and deliver business value for today – whilst getting you ready for tomorrow.

avanade

A proven path to **Azure Databricks**

We have data migration pathways to help you take advantage of Avanade's dedicated data migration factory to accelerate any data migration.

- Snowflake
- Cloudera
- Oracle®
- Tableau
- MicroStrategy®
- SAP® solutions
- + many more



50+ Migration Templates | **30+ Tools & Accelerators** | **Migration at speed**

Continuous Learning & Innovation

avanade

Get started today

An Avanade Data Strategy Workshop is a great way to initiate or further your data journey, helping to ensure that your investment in data modernization with Azure delivers optimal business value and benefit. To learn more or to engage with an Avanade solution architect, contact us today

Special Thanks

Avanade would like to thank the team of data experts for their effort, support and contribution on this point of view, in particular our Microsoft MVPs and Databricks Technical Champions. We would also like to thank all non-Avanade contributors, who helped provide the supporting research we have used and referenced to bring this point of view to life, in particular the Barcelona Supercomputing Center, TCP and GigaOM.

About Avanade
Avanade is the leading provider of innovative digital and cloud services, business solutions and design-led experiences on the Microsoft ecosystem. Every day, our professionals bring bold, fresh thinking combined with technology, business and industry expertise to help make a genuine human impact on our clients, their customers and their employees. We are the power behind the Accenture Microsoft Business Group, helping companies to engage customers, empower employees, optimize operations and transform products, leveraging the Microsoft platform. Avanade has 60,000 professionals in 26 countries, bringing clients our best thinking through a collaborative culture that honors diversity and reflects the communities in which we operate. Majority owned by Accenture, Avanade was founded in 2000 by Accenture LLP and Microsoft Corporation. Learn more at www.avanade.com.

North America
Seattle
America@avanade.com

South America
Sao Paulo
AvanadeBrasil@avanade.com

Asia-Pacific
Australia
AsiaPac@avanade.com

Europe
London
Europe@avanade.com

avanade

Do what matters