



Transform your data modernization strategy and watch business value soar

Azure Databricks and Azure Synapse are better together in the modern data architecture

Transform your data modernization strategy
and watch business value soar

Executive summary

Data architectures continue to evolve because organizations need them to meet new needs, greater goals and unexpected challenges.

The data lakehouse, combining the best of the data warehouse and data lake, supports a broad range of patterns and personas from across the organization. The good news in this world of continual change is that adopting the data lakehouse may not require a rip-and-replace of your existing data platform; as a cloud environment, Microsoft Azure provides all the tools and services you need. [Microsoft Intelligent Data Platform](#) (MIDP) is the modular data platform in Azure that provides inherent interoperability and low TCO to help the organization deliver on its strategic needs.

Key analytics components Azure Databricks and Azure Synapse, once neatly categorized as intended for data scientists/engineers and business analysts/developers, respectively, now increasingly offer overlapping capabilities for a common audience. Choosing between them is often a matter of personal preference, training and experience. The use of Power BI as a front end to each reflects further convergence, as well as Microsoft's approach to elevate self-service as an increasingly important part of its strategy. These technologies will play important roles in the continued evolution of data architectures, such as the decentralized data mesh. The good news is that Microsoft and Databricks [continue to strengthen their partnership](#) to help organizations achieve better data and analytics outcomes.



Transform your data modernization strategy
and watch business value soar

As your business evolves, so should your data architecture

You may not know what business disruption or market upheaval is next, but you know that you need to be ready for it. Whatever your strategic direction, it must include new heights of resiliency and agility, preparing you to address the change you can't see – but know is coming.

To achieve this flexibility to face whatever lies ahead, you need to make the best, data-driven decisions, based on the most timely, accurate insights you can gain from your data. If you're not there yet, you're not alone: In a [recent survey](#), just 17% of executives said they felt confident they had the right insights to make informed business decisions during the pandemic. Little wonder. The data landscape continues to grow larger and more complex, accessed by more applications and more types of users. Data architecture has continually evolved to meet these needs and it must continue to do so. The need for a single version of truth among siloed systems and disparate applications led to the data warehouse. But this architecture wasn't designed for the complex, large-scale data and compute requirements of machine learning, which plays an increasingly important role in digitally mature systems.

As business sought to address shortcomings and make better use of unstructured data – such as social media feeds, photos, video and IoT – data lakes were part of the answer. Data lakes succeeded in supporting analysis of unstructured data, in offering lower-cost storage, and in providing somewhat flexible, open, direct access to data.

But getting useful information out of the data lake, with workable quality, governance and performance, has sometimes proved more difficult than getting data *into* the data lake in the first place. The raw data in data lakes exists in a range of formats for data engineering, streaming, data science, machine learning, analytics, business intelligence (BI), artificial intelligence (AI) and more – when unmanaged, is graciously often described by users as resembling a swamp more than a lake. In such data swamp environments, some teams may get little to no value from the data lake.



Transform your data modernization strategy
and watch business value soar

Your unique data estate needs a **customized data modernization strategy**

Today, businesses need a data and analytics strategy that supports a broad range of patterns and personas from across the organization – enabling democratization of data-driven insights in a responsible way. It's a tall order, but with potentially big results and outcomes: First, a more flexible and perceptive approach ensures teams and individuals working with your data can collaborate with others on common use cases, while each works in their preferred environment. Second, it avoids the need for high-cost/low-value customizations, complexity and cost, so you can accelerate innovation and high-value use cases, while keeping data secure and well-governed at all times.

When building your optimal strategy, you need an environment that meets the needs of your data scientists and data engineers, including the ability of the former to better exploit data and of the latter to create reusable data pipelines and products. But you should also ensure that data analysts and business users can work with data with the tools they already use. It requires an interrogation layer on the lake to support batch and streaming, an open format to prevent lock-in, and comprehensive governance, lineage and security. It also requires native support for all data types and modern lower-cost, durable cloud storage.

The data lakehouse is the evolutionary advance in data architecture that can deliver exclusive data and analytics capabilities. It's a single platform that unifies all of an organization's data, analytics and AI workloads by combining the strengths between the data warehouse and data lake.

The emerging data platform: Lakehouse



Transform your data modernization strategy and watch business value soar

Harness the environments you already use

The data lakehouse represents the next generation of data platform, one that takes the right modern approach to the right data for the right need, persona or pattern. And it does so in an environment broad enough to readily provide all the competitive tools and leading capabilities that organizations need, as well as an environment that most of them already use:

[Microsoft Azure](#).

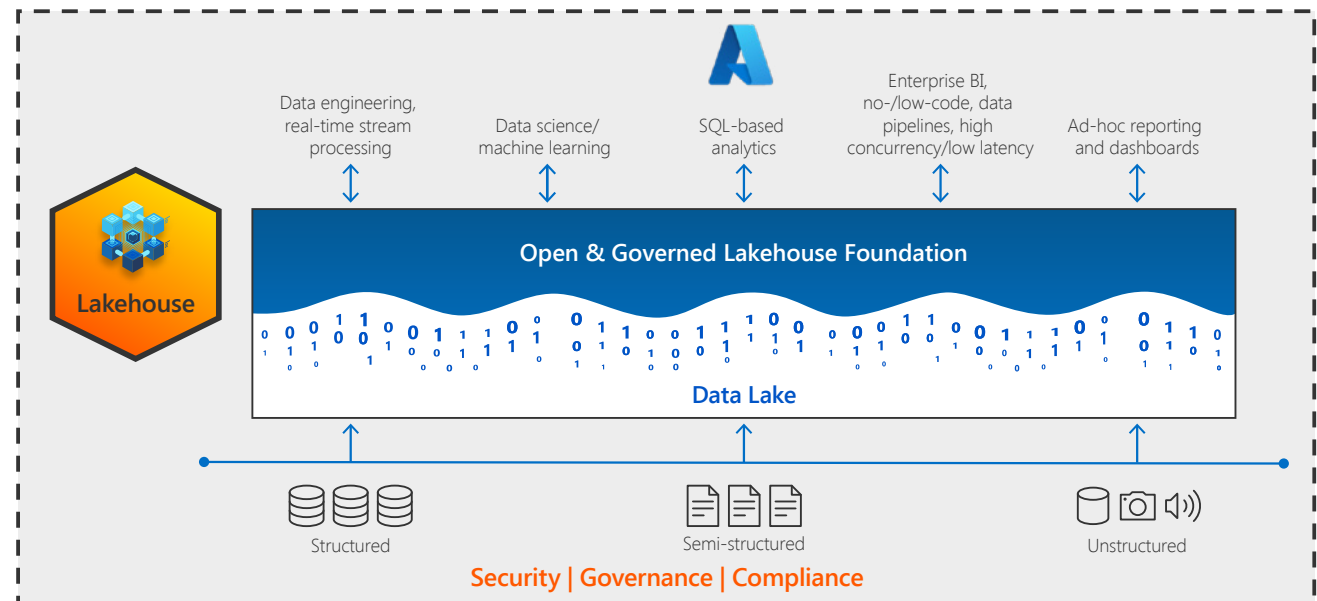
With Azure, enterprises have an integrated solution that includes storage, ETL/ELT, analytics, data governance, and reporting. They gain unmatched total cost of ownership (TCO) value from Azure's single support model, with built-in security and compliance capabilities.

For most modern enterprise organizations, there's simply no need for separate ETL/ELT, analytics or AI solutions that are not only expensive to acquire and maintain but also can introduce security, integration and interoperability challenges. Built on Azure, the Microsoft Intelligent Data Platform includes a set of well-integrated services like Azure Synapse, Azure Databricks, Azure Machine Learning, Power BI and more, while supporting open standards and open libraries.

Azure Synapse and Azure Databricks enable the open and governed lakehouse. With Azure Purview and Unity Catalog, this serves as a foundation for unified and cloud-native data fabric. These services combine harmoniously to execute the most demanding analytics and machine learning applications.

The lakehouse eliminates the need to build, integrate, and operate sub-components for every workload. Through a single-pane experience for data and analytics governance in [Microsoft Purview](#), organizations can be assured of optimal results against their data, analytics, machine learning, and AI objectives.

Bringing the right compute to data with lakehouse



Transform your data modernization strategy and watch business value soar

Databricks, Synapse... or both?

Two key data lakehouse services in Azure are Azure Databricks and [Azure Synapse Analytics](#) (Synapse). Each platform service provides data and analytical capabilities that organizations can use to generate insights that create value. That's led some businesses to think they have to decide between either Azure Databricks or Azure Synapse. We've heard unreferenced industry chatter that speculates one technology will supersede the other.

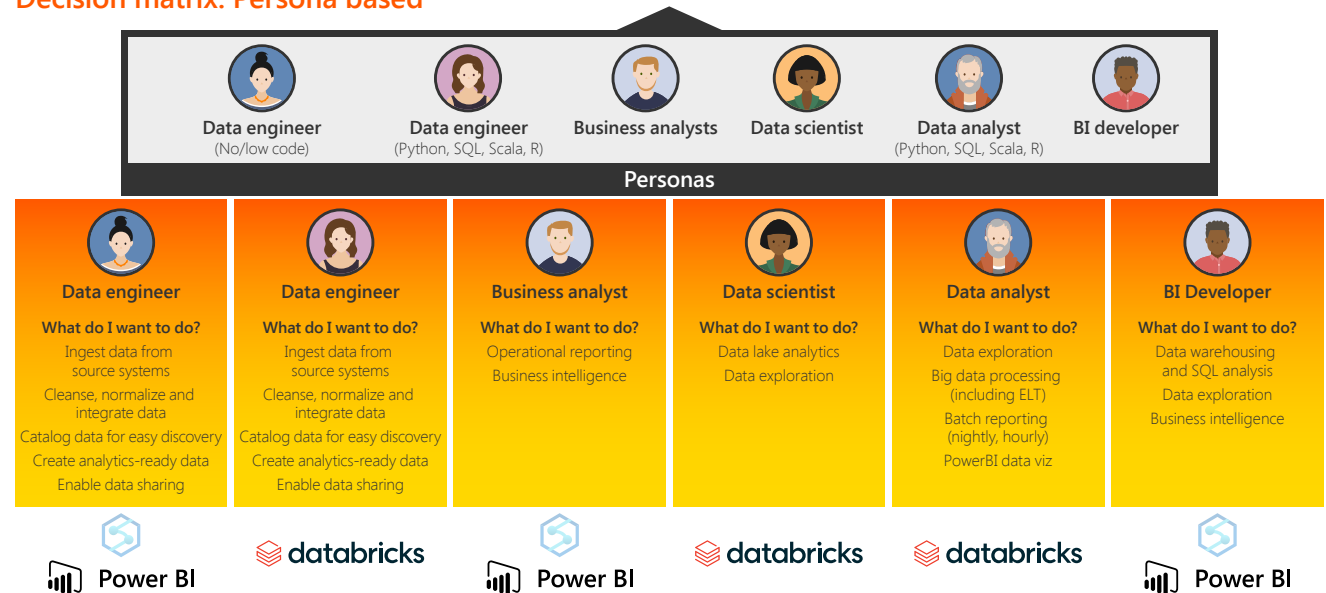
In fact, Azure Databricks and Azure Synapse are complementary, not competitive. Together they enable the broadest and most demanding analytics and AI workloads: across data engineering, streaming, data science and machine learning, SQL analytics, as well as enterprise and local reporting, and dashboarding.

Each of these platform services continues to have its place in the modern data architecture. Sometimes one of the two services will suffice, but using both together allows you to harness maximum value from all your data across Azure. The open and governed lakehouse foundation supports Delta format today but will also support other emerging standards like Iceberg and Hudi, helping to future-proof your business.

Including both Databricks and Synapse in your modern data architecture is prudent because it's as cost efficient as it is effective: Both are first-party Azure Services accessible through any organization's Azure environment, with no additional, third-party license costs for their use. Moreover, because they're based on Azure, rather than an on-premises system or legacy platform, either

Databricks or Synapse – or both – can be easily mobilized and scaled whenever needed and spun down when not in use. That means you can use these technologies with greater agility and cost effectiveness than you can with on-premises or Azure non-first-party products like Snowflake, Cloudera, Netezza, MongoDB or Hive.

Decision matrix: Persona based



Azure Databricks + Azure Synapse = better together: An example

With an early overabundance of options served up by third parties, some organizations saw Databricks and Synapse as individual options for modernizing their data estate. Others recognized that both are necessary to apply the distinct advantages of each as needed for the different scenarios that every organization faces. For example, a manufacturing business may need the business-critical machine learning capabilities of Databricks to serve digital channels, while also needing Azure Synapse to support fast daily decisions on the shop floor.

Analytics in Microsoft Intelligent Data Platform requires a purposeful selection of the right Azure services for a specific job:

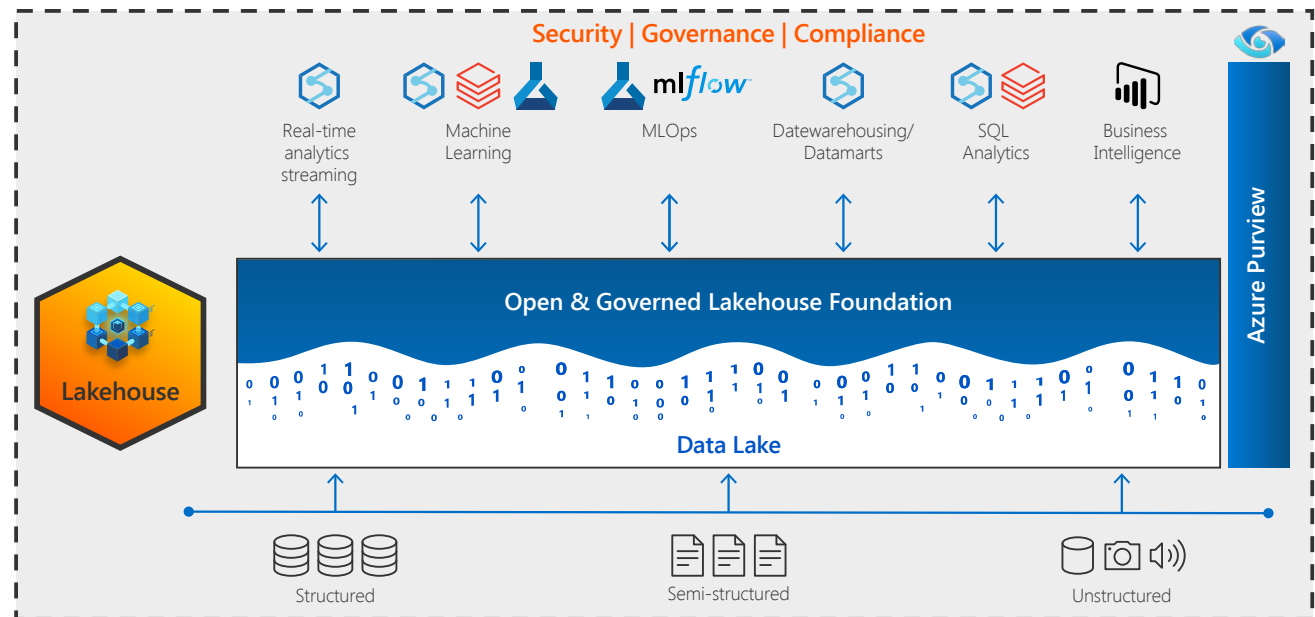
- Azure Data Factory in Azure Synapse Analytics for orchestrating data integration pipelines to integrate data from across hybrid, multicloud, SaaS, and legacy enterprise data source systems
- Azure Data Explorer in Azure Synapse Analytics for real-time streaming analytics
- Azure Databricks for building an open standard data lakehouse using the Delta format
- Azure Machine Learning, Synapse ML, and Azure Databricks for machine learning
- Azure Machine Learning, and Databricks MLflow for MLOps
- Azure Synapse SQL and Databricks SQL for serverless SQL analytics

- Azure Synapse SQL for a Data Warehouse modality for data serving Power BI for Business Intelligence
- Microsoft Purview and Databricks Unity Catalog Federation for Unified Data Governance in Purview spanning Operational, Analytics, and ML/AI data assets on the Microsoft Intelligent Data Platform

This harmonious and complementary approach enables any data platform use case to be designed, deployed, supported, updated, and – when needed – retired in a controlled fashion.

This purposeful strategy is why Avanade has consistently advocated that Azure Synapse and Azure Databricks are better together when architecting an Azure-centric, hybrid, or multi-cloud data platform.

Other perspectives fail to acknowledge that it's possible for an organization to have a high-performance, flexible, open, and cost-efficient organization-wide data platform architecture. Organizations don't have to forego any of these benefits.



Transform your data modernization strategy and watch business value soar

Azure Databricks: Delivering data engineering and data science at scale

Azure Databricks is best in class for data engineering workloads spanning batch and real-time with limitless scale. Businesses can use a range of tools for working with data on a single, unified data platform. It supports these patterns through three primary offerings for developing data-intensive applications:

Databricks Data Science & Engineering is an interactive workspace in which data engineers, data scientists and machine learning engineers can collaborate. Big data – raw or structured – is ingested through Azure Data Factory to a data lakehouse. The data experts can use Apache Spark to turn this information into insights.

Databricks Machine Learning is an integrated, end-to-end machine learning environment incorporating managed services for experiment tracking, model training, and feature development and management.

Databricks SQL makes it easy for analysts who want to run SQL queries on their data lakehouse or create multiple visualization types to explore query results from different perspectives.

Some organizations have used Azure Databricks primarily as an engine for data modernization, but its current capabilities exceed its origins. Businesses now also harness Databricks for data lake analytics, data exploration, big data processing (including ETL), real-time streaming, batch reporting, business critical workloads and cloud-based analytics.

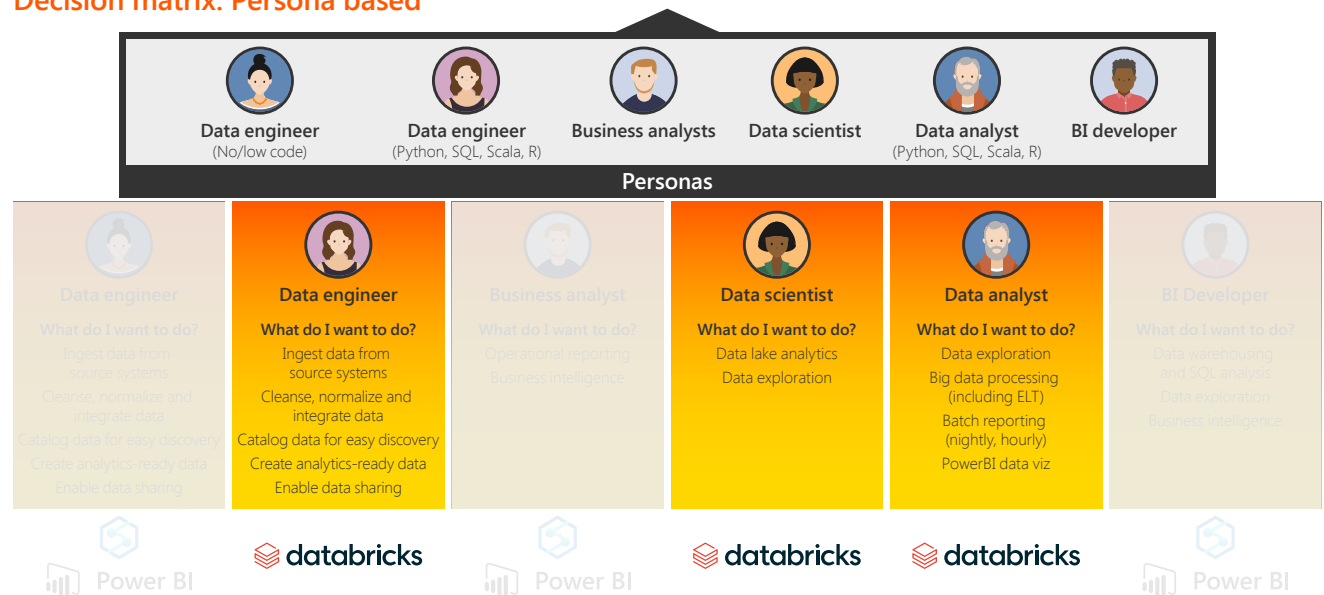
Databricks is likely best known as a tool for data engineers, data scientists and data analysts, who generally come to it from a software engineering background versed in Python, Scala and R. Software developers find Databricks is optimized for their

productivity, with an easier, more efficient and cost-effective environment compared to the on-premises solutions they may know. With the addition of Databricks SQL as well as dashboard capabilities, the platform is also increasingly useful to analysts and teams who come from a data warehouse background.

The Databricks platform is open and standards compliant. It builds upon Databricks' leadership in Apache Spark, Delta Lake and MIFlow with a modern, vectorized execution engine, smart caching and instant compute. The result is a best-in-class data engineering, data science and exploratory data analytics solution.

In all, Databricks helps organizations ingest vast data stores and apply the machine learning and data science that helps accelerate both innovation and time-to-value.

Decision matrix: Persona based



Transform your data modernization strategy and watch business value soar

Azure Synapse: For business users and beyond

Azure Synapse is another crucial component of the modern data architecture. It's optimized for data warehousing and SQL analytics, BI, data exploration, non-batch reporting and a range of business workloads. It's the preferred platform for business analysts and BI developers – users who need to access and explore data relatively quickly and easily. For those who prefer a no-code/low-code environment, Synapse offers a familiar way to access, analyze and output data.

Organizations value Azure Synapse for its high-quality user experience, single suite of accessible services and powerful, dedicated pools. Offering fast and flexible data access, Synapse helps you scale, compute and store data elastically and independently, with a massively parallel processing architecture.

Azure Synapse is widely known for its underpinning data reporting capabilities as a result of its tight integration with Microsoft Power BI (see below). Synapse SQL enables batch, streaming and interactive processing with T-SQL. With Apache Spark in Synapse, the platform provides big data processing with Python, Scala, R and .NET.

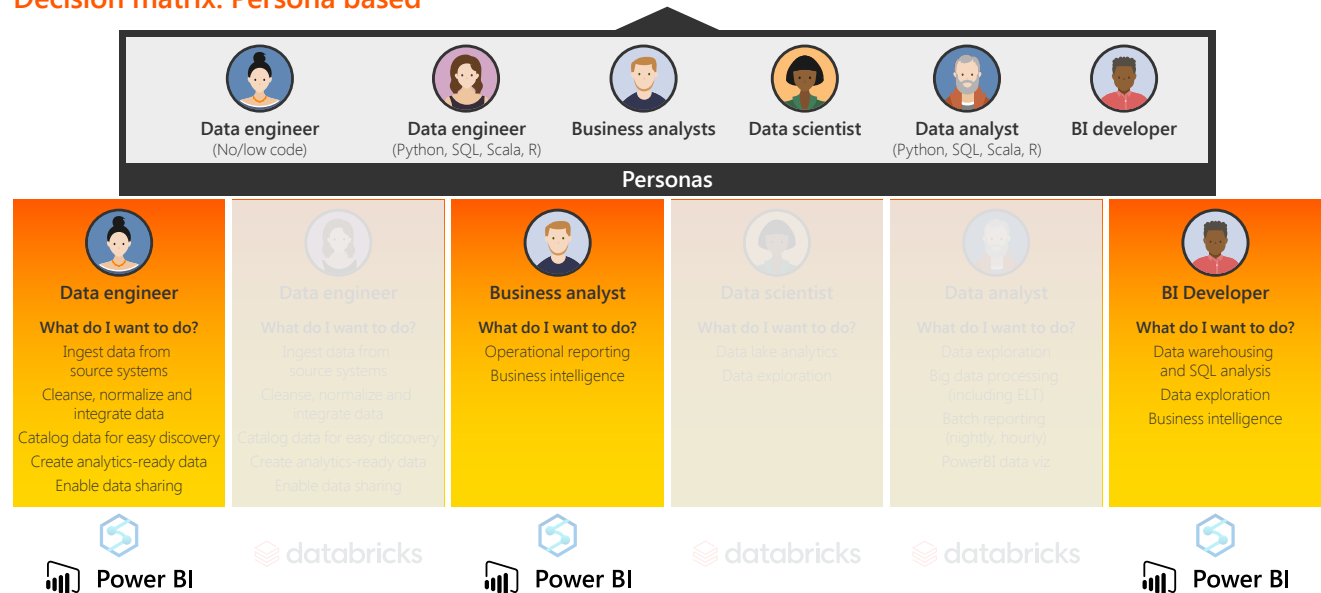
The Azure Synapse platform continues to expand. Additional capabilities include:

Azure Synapse Studio is a one-stop shop for data engineering and analytics development, from data exploration to data integration and large-scale data analysis. It provides a unified workspace for data prep, data management, data warehousing, big data and AI tasks.

Azure Synapse Link is an extension of Azure Synapse that broadens the operational data sources – via a connection to cloud-native hybrid transactional/analytical processing (HTAP) – available to users while minimizing the amount of data engineering required.

Synapse makes it easy to harness the full Azure ecosystem, where you can connect and use other Azure services, including: Azure Data Factory for fast and flexible data access, Microsoft Purview for data governance and Azure Machine Learning for model-building and AI services.

Decision matrix: Persona based



Transform your data modernization strategy
and watch business value soar

Power BI: The insight platform for all modern data users

[Microsoft Power BI](#) is a key example of the greater value businesses gain by building their modern data architectures on Azure. Power BI turns data into action by putting data-driven insights into the hands of decision-makers. As a front end to Synapse, it democratizes data, giving business leaders and other less-technical users an easy and effective way to access, explore and analyze data. Moreover, as part of the Power Platform, which includes Power Automate and Power Apps, it spans the gap between insight and action so organizations can act more effectively on their new insights.

While closely integrated with Synapse, Power BI is also now increasingly used as a front end to Azure Databricks. That evolution gives business users unprecedented access to lakehouse data in near-real time, with no-code reporting and dashboards to help drive better-informed decisions. From engineers to the C-suite, all users can collaborate on a common interface and more easily share insights.

Power BI is a key to Microsoft's strategic direction to infuse the entire data platform with greater self-service. All users can get the data and insights they need, when they need it, without continually putting the burden on data engineers or IT staff. As a bonus, IT doesn't have to acquire and maintain separate, third-party reporting and analysis tools, reducing costs and boosting IT productivity.

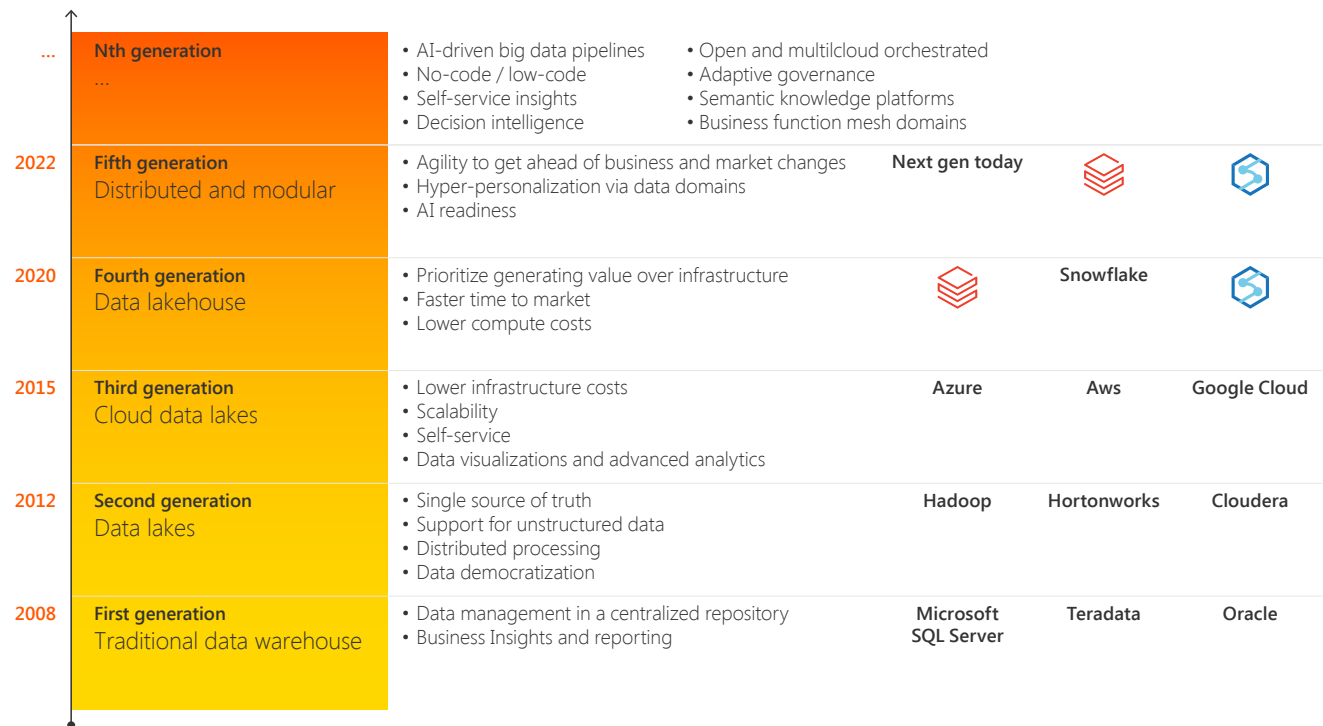


Look ahead to **data mesh**

From data warehouse to data lake and then to data lakehouse, data architectures keep evolving. There's no reason to think that the current iteration of data modernization will be the last. In fact, the next generation in data architecture is already being adopted at scale: data mesh.

As an emerging data framework, data mesh envisions a decentralized data architecture that relieves the bottlenecks that occur in centralized data platforms. It's a new approach to creating a fully open analytical data estate, fueling the next generation of domain-driven data products. Under the right circumstances, a data mesh approach can further reduce data management and operational costs while accelerating higher-quality data insights. More broadly, it could also foster a more productive and collaborative data culture.

Just as data lakehouses are a newer approach to data architecture that are interoperable with existing Azure technologies such as Azure Databricks and Azure Synapse without requiring a rip-and-replace, so too Azure supports data mesh without requiring massive new technology investments. Azure Databricks and Azure Synapse will continue to play crucial roles in the data mesh architecture alongside Microsoft Purview to construct the modular and distributed data product blueprint.



Transform your data modernization strategy
and watch business value soar

Avanade can help you customize **your data modernization journey**

As a joint venture of Microsoft and Accenture, Avanade has privileged access to Microsoft technology and Accenture industry expertise, giving us an ideal foundation from which to apply Azure technologies to the data modernization needs of businesses like yours.

Our credentials include:



More certified
professionals for
Microsoft Azure than
any other Microsoft
partner



2022 Databricks Global
Partner of the Year
(fourth year in a row)
and seven-time recipient
of regional Databricks
Partner of the Year



Deep partnership with
Databricks, working
together to build
end-to-end enterprise
data strategy at clients
around the world



IDC Marketscape
leader for Microsoft
Implementation
Services



2022 Microsoft Global
Alliance SI Partner
of the Year for the
17th time

“What we really like about Avanade and what it’s really bringing to the table, is the fuel for innovation, it’s the passion for innovation that is fundamental if you want to embrace digital transformation. But technology is not enough. You need people. And you need the right combination of people and technology to make a human impact.” — [Junior Achievement Europe](#)

“To be able to extract value from data, you need to understand data. And in order to understand data, you have to have deep domain experience. It’s not only a business problem to reduce costs. It’s also to see the opportunities for data.” — [DNV](#)

Transform your data modernization strategy
and watch business value soar

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 about our Data Strategy Workshops or to engage with an Avanade solution architect, [contact us](#) today.

Special thanks

Avanade would like to thank Microsoft and Databricks for their effort, support and contribution on this point of view. We'd also like to thank Avanade contributors, including our Microsoft MVPs and Databricks Technical Champions, who helped bring this point of view to life.

North America

Seattle
Phone +1 206 239 5600
America@avanade.com

South America

Sao Paulo
AvanadeBrasil@avanade.com

Asia-Pacific

Australia
Phone +61 2 9005 5900
AsiaPac@avanade.com

Europe

London
Phone +44 0 20 7025 1000
Europe@avanade.com

About Avanade

Avanade is the leading provider of innovative digital, cloud and advisory services, industry solutions and design-led experiences across the Microsoft ecosystem. Every day, our 60,000 professionals in 26 countries make a genuine human impact for our clients, their employees and their customers.

We have been recognized, together with our parent Accenture, as Microsoft's Global SI Partner of the Year more than any other company. With the most Microsoft certifications (60,000+) and 18 (out of 18) Gold-level Microsoft competencies, we are uniquely positioned to help businesses grow and solve their toughest challenges.

We are a people first company, committed to providing an inclusive workplace where employees feel comfortable being their authentic selves. As a responsible business, we are building a sustainable world and helping young people from underrepresented communities fulfill their potential.

Majority owned by Accenture, Avanade was founded in 2000 by Accenture LLP and Microsoft Corporation. Learn more at www.avanade.com.

©2022 Avanade Inc. All rights reserved. The Avanade name and logo are registered trademarks in the U.S. and other countries. Other brand and product names are trademarks of their respective owners.



