E-book series

Microsoft Azure

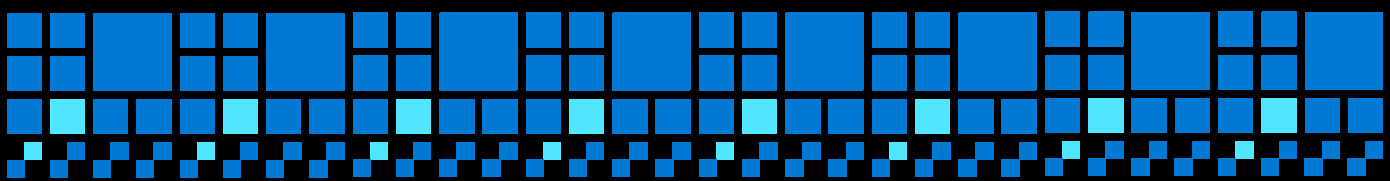
# Building Industry-Aware Analytics Solutions Using Azure Synapse Analytics

# How to use this guide

Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. With the availability of industry-specific database templates in Azure Synapse Analytics, you can shape and integrate data from multiple sources in a common format. This significantly cuts down the time to insight by enabling the rapid building out of analytics-infused industry solutions.

The primary audiences for this paper include architects and technical decision-makers mandated to build industry-aware analytics solutions for their organizations. This may include chief data officers, line-of-business application owners, data scientists, and data engineers. Technically oriented business decision-makers will also benefit from this guide.

Through this guide, you will gain a high-level understanding of Azure Synapse, especially as it relates to building industry-aware analytics solutions relevant to your industry. While the retail industry is used to illustrate the concepts, the approach is applicable to any other industry as well.



# About the authors

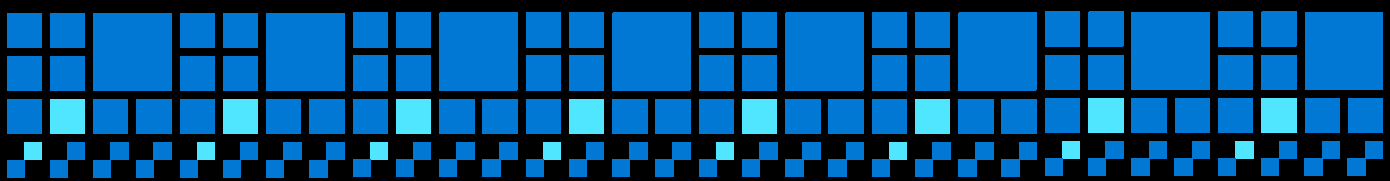
**Gerd Saurer** is a Principal Program Manager within the Microsoft Azure Data & AI group. In the last 10 years with Microsoft, he has had the opportunity to work in different roles spanning presales, partner management, and product management covering several industries, including retail, consumer packaged goods, and financial services. In his current role, he works closely with customers to identify, define, and drive industry-focused analytics solutions.

**Rajeev Jain** is a Senior Product Marketing Manager with Microsoft Azure Data & AI. In his current role, he drives the business growth of the Azure analytics business through the industry and partner lens. In the past, Rajeev has led sales, marketing, and partner businesses at Microsoft, and has worked in the oil and gas and financial services consulting industry outside of Microsoft.

# Editorial credits

Azure Data Engineering: Matt Burdick, Santosh Balasubramanian, Gautam Bharti, Aaron Merrill, Trevor Olson

Azure Marketing: Kaiser Larsen, Daniel Yu



# Building Industry-Aware Analytics Solutions Using Azure Synapse Analytics

# Data and analytics are at the heart of digital transformation

We have been witnessing a second wave of digital transformation over the past year affecting every company and every industry, with many structural changes that will outlast the COVID-19 pandemic.

Consider the retail industry. This industry was already adopting e-commerce rapidly due to the cascading impact of digitization on their supply chains, operations, and customer experiences. The pandemic led to the mainstreaming of many niche trends like contactless shopping and curbside pickup experiences. The increased digitization of processes and experiences led to ever greater data volumes from applications and interactions. This could potentially be combined and analyzed in interesting new ways to generate actionable insights, often in real time, for driving business agility. For example, the ability to capture data around customers' online purchases could be combined with their demographic profile and social media interactions to proactively generate personalized offers or reduce churn.

Similarly, analytics can be applied to all areas of the business to convert data into insights for agile decision-making and execution. In practice, rapidly creating analytics-based applications, like ones that enable personalization, demand forecasting, or supply chain optimization, needs a robust data foundation. This foundation is often a challenge to create due to the multiple internal and external systems not talking to each other, thereby creating data silos. Unlocking and combining data trapped in these systems is hard. A common data blueprint is needed to avoid each scenario being an independent data project that slows down digital transformation and reduces innovation.

Azure Synapse provides the foundational analytics and data shaping capabilities for your business—so that data from different applications and services across different business areas (like customer, product, sales, marketing, and operations) can leverage the same connected dataset and generate a holistic view.

# Overview: Industry-aware analytics with Azure Synapse

As shown in *Figure 1*, Azure Synapse is a limitless analytics service that brings together the foundational capabilities for analytics: data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated resources—at scale. It brings together the worlds of structured and unstructured data with a unified experience to ingest, explore, prepare, manage, and serve data for immediate **business intelligence (BI)** and **machine learning (ML)** needs.

Azure Synapse provides several different components, enabling customers with an integrated and seamless experience to build and operate their data projects in a secure and scalable way:

- An analytics "hub" that provides the limitless capability to analyze any type of data independently of the underlying storage technology. Customers can take advantage of the industry-leading SQL, big data, or data discovery analytics services that—combined with cost-efficient storage capabilities—add semantics and context so that data can become an organized strategic asset.
- The comprehensive selection of prebuilt database templates for customers in different industries. These templates enable you to start your data projects with application-independent data structures. Leveraging proven database templates built with decades of experience across industries reduces project risk and helps accelerate time to value.
- Synapse pipelines give the customer an integrated low-code/no-code experience to move and transform data from silos into a unified structure with hundreds of prebuilt connectors at no additional cost.
- SQL and Spark analytics compute engines for data warehousing and big data analytics.

Find out more about the specific services at Azure Synapse Analytics:

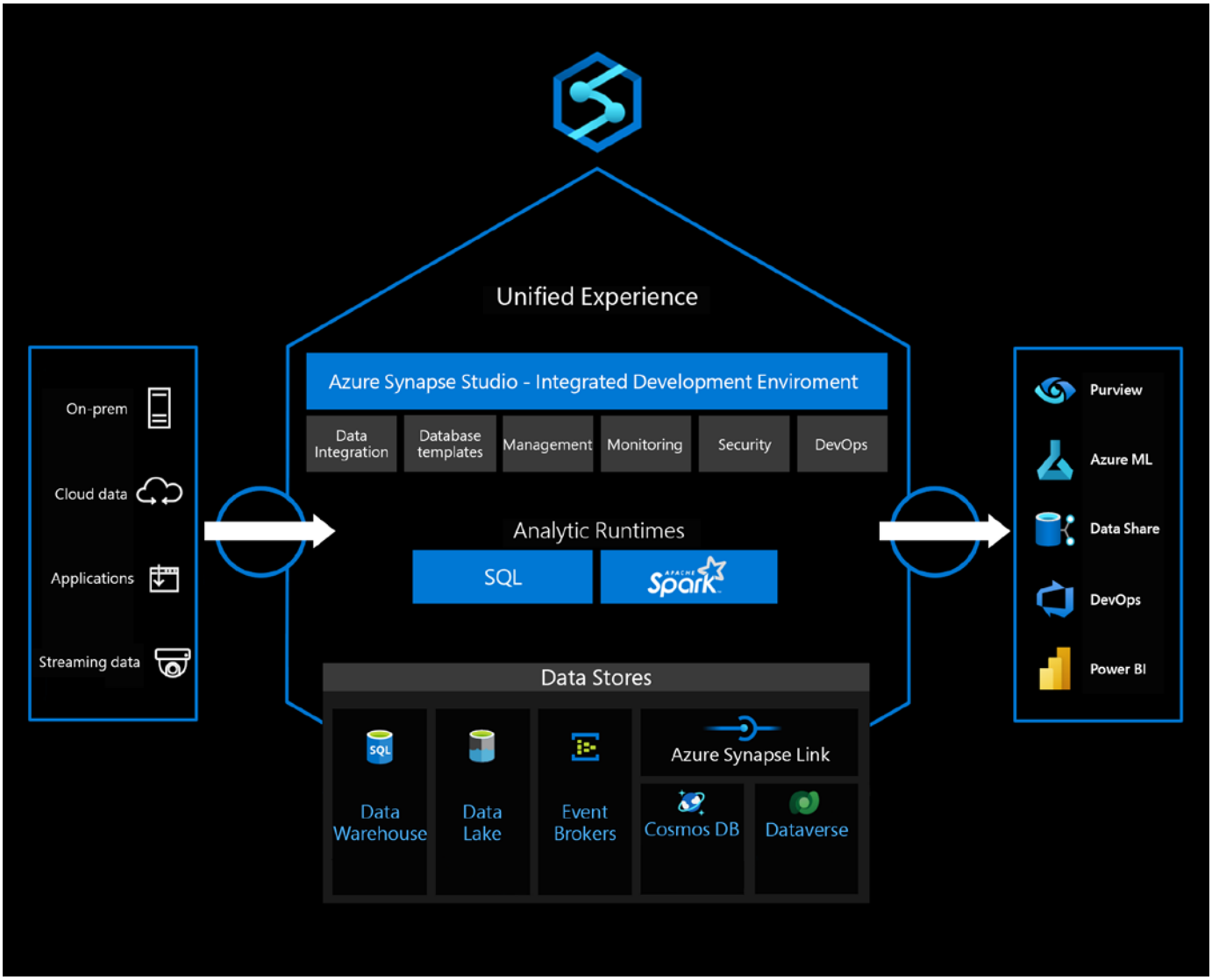



Figure 1: Azure Synapse overview

# What makes Azure Synapse industry–aware?

With the rapid growth in the volume and variety of data, customers have increasingly deployed data lakes for data storage. However, a key challenge that we face is to define how our data should be shaped as it is stored in different applications, forming data silos. These data silos need to be integrated and combined to help you understand your data and make informed decisions.

With the availability of database templates within Azure Synapse, you can shape and describe data in a standardized way, for your specific industry. Database templates also make the integration of different data sources easier. New and existing projects can leverage out-of-the-box database templates that address a wide variety of industry-specific business areas. Finally, the database templates have been built with an ecosystem in mind. You can rapidly build analytics-infused industry use cases by customizing and extending the standard templates using the database editor in Azure Synapse:

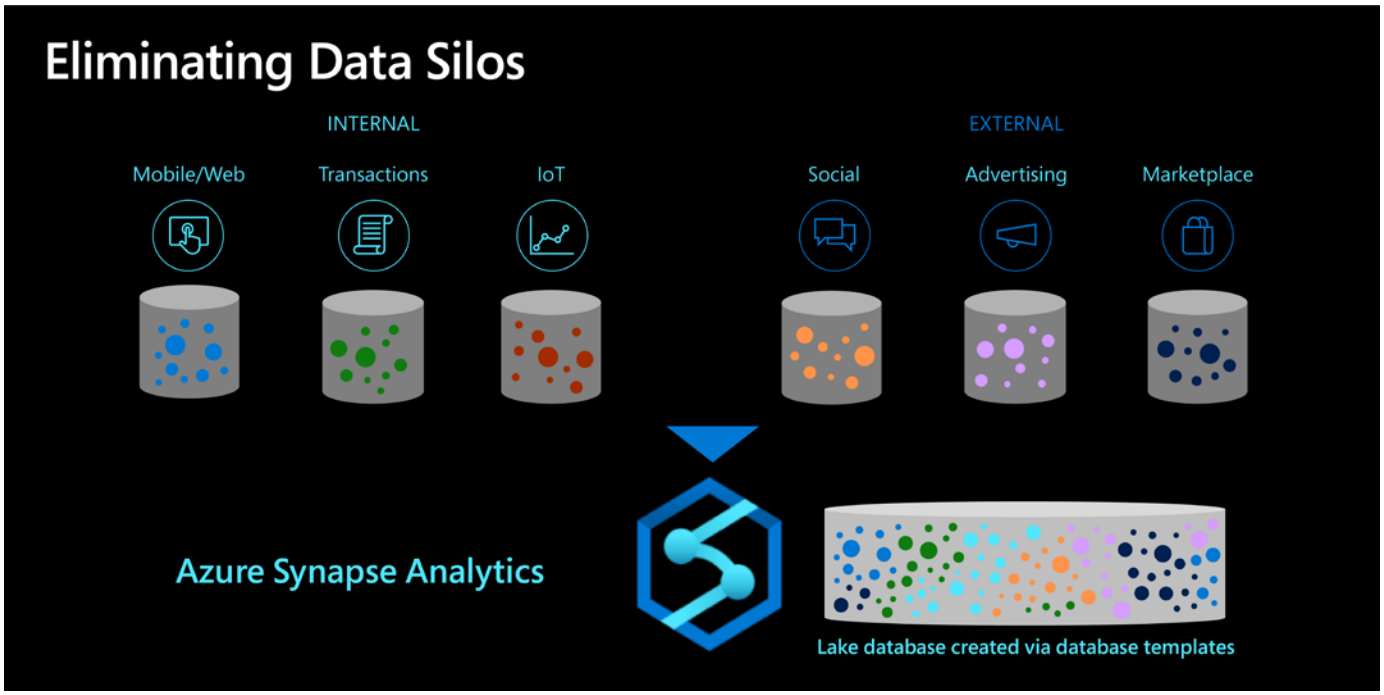



Figure 2: Eliminating data silos

# Understanding database templates in Azure Synapse

Database templates in Azure Synapse are industry-specific, containing a rich set of entities and covering different business areas integrated into a single structure. Each template consists of approximately 20 business areas across 5,000 entities, with clearly defined attributes and relations. These templates that have been created have been used by customers for several decades. *Figure 3* shows the retail database template and some of the different business areas it covers:

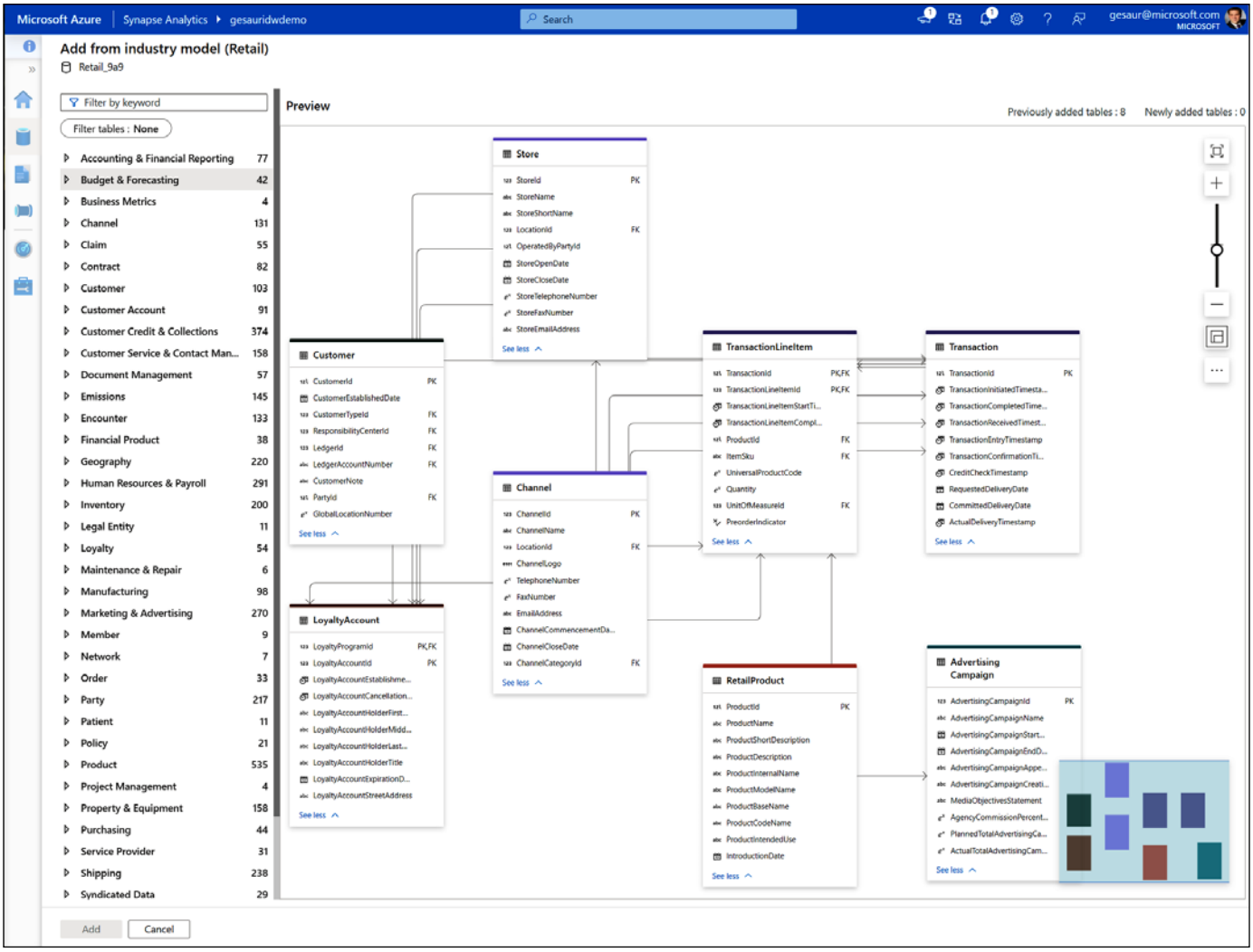
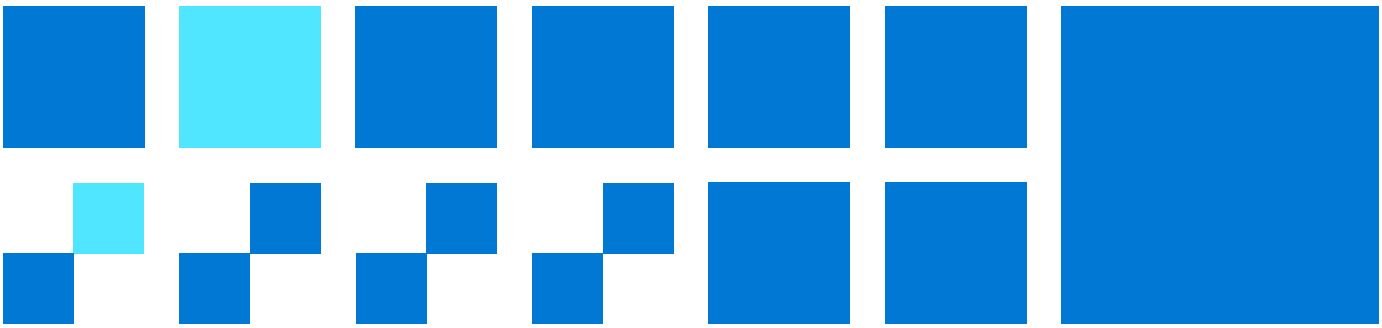



Figure 3: Example—retail database template

With the database design capabilities in Azure Synapse, you can extend and customize database templates to fit your specific scenario. When the database templates are deployed, the metadata (including descriptions, data types, and relationships) is directly attached to the data lake. Once data across heterogeneous sources is ingested and structured in a standardized way, it can be exposed to seamlessly interact with the analytics engines in Azure Synapse (Synapse SQL and Spark), as well as with Power BI. This enables the rapid creation of analytics applications and makes it easy for data analysts, business users, and AI/ML experts to do their job.

Database template-aware components in Azure Synapse, such as pipelines or SQL on-demand, can be used to automate the data lifecycle, including ingestion, exploration, and analysis. This increases productivity and reduces time to insight.



# Industry scenarios enabled by Azure Synapse: Retail example

Let's consider how database templates can help accelerate digital transformation in the retail industry:

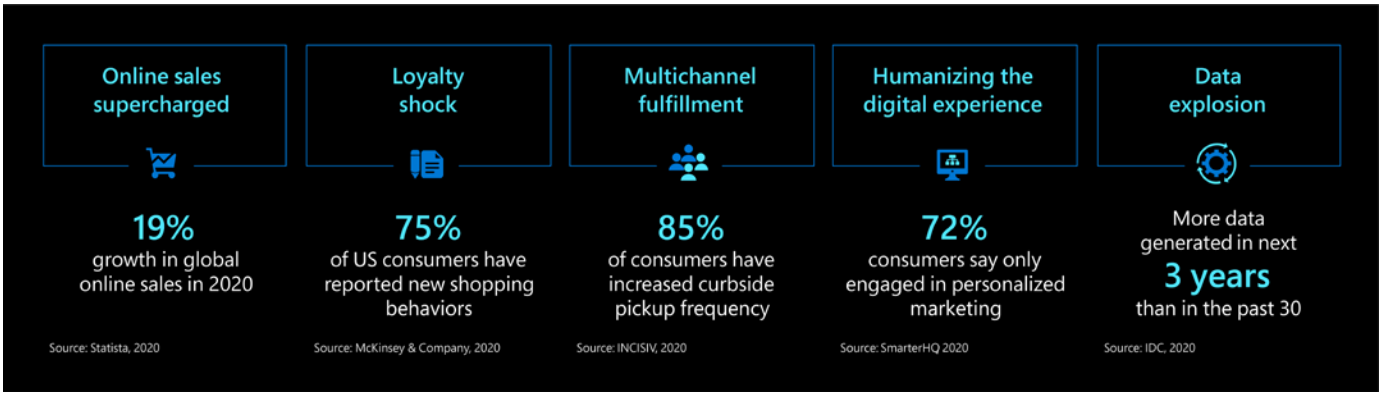



Figure 4: The retail customer journey

In retail, offline merchants are embracing new business models, like contactless shopping and curbside pickup, while e-commerce has been accelerating for several years already. Being able to track and analyze increases in the number of interactions due to e-commerce—before, during, and after purchase—is becoming essential for retailers. Customers not only expect a friction-free experience but also appreciate relevant information sharing in an over-communicated world, versus random messages and offers.

*Figure 5* shows a schematic of a personalized offer generation and serving system. The system provides a unified set of capabilities to integrate data from multiple sources (like the customer's prior interactions across physical and virtual channels), develop useful customer analytics, and enable targeted offer generation via a campaign management system.

Azure Synapse is at the heart of such a system, with the retail database templates providing an agile way to build such a capability:

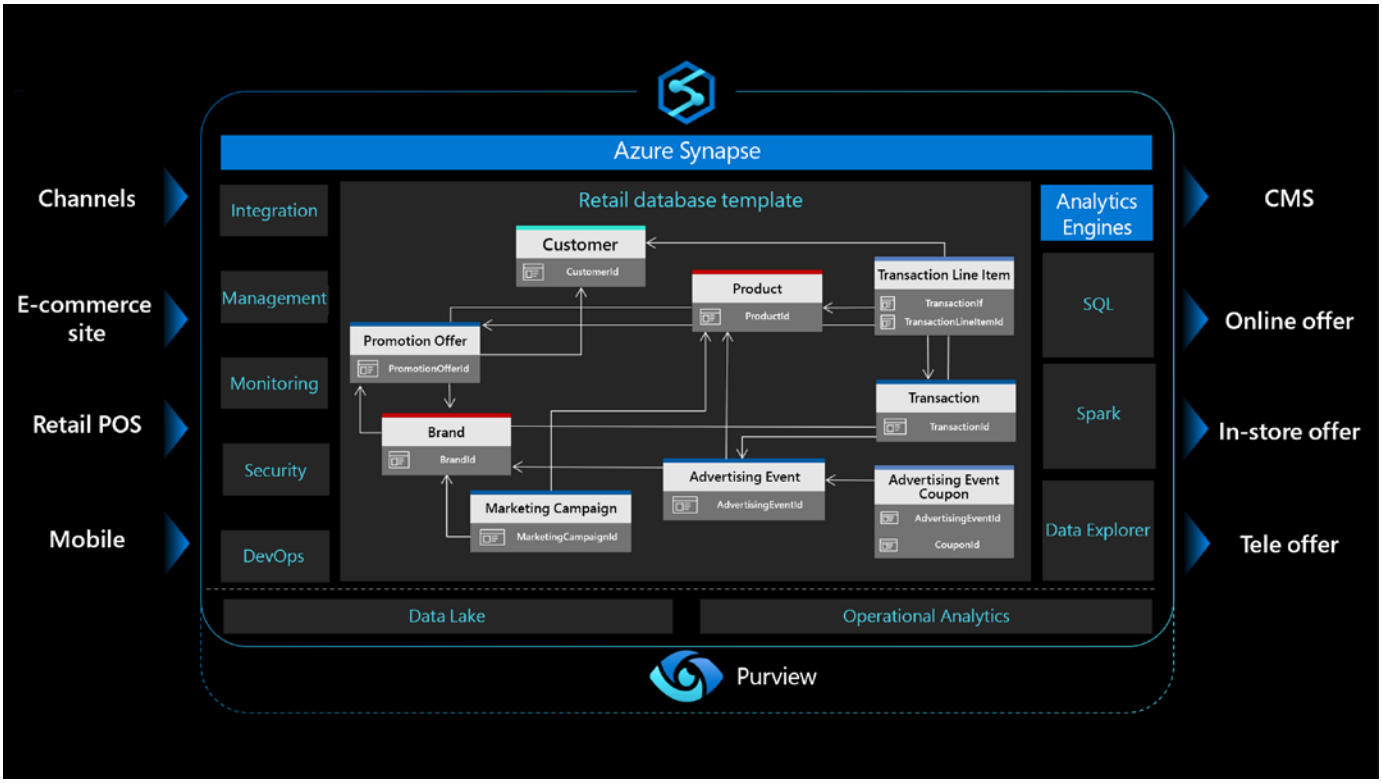
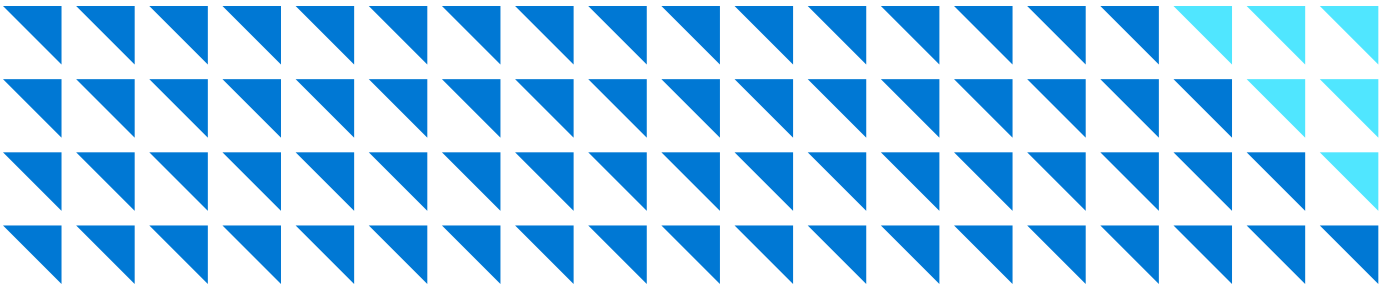



Figure 5: Personalized offer generation system with Azure Synapse database templates

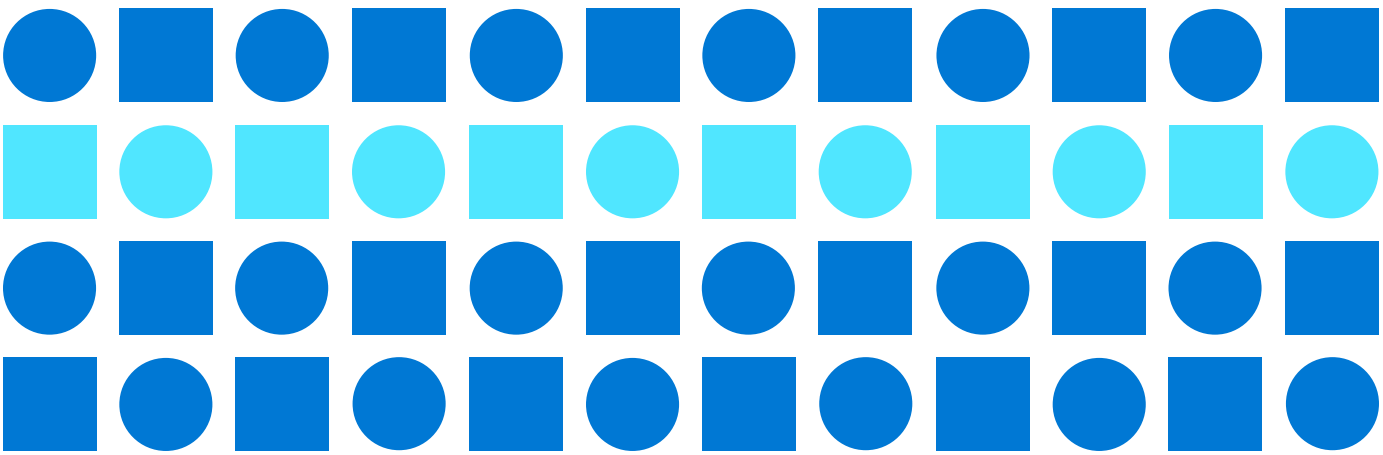
The template being used in *Figure 5* is just a small selection of customizable tables that are available and needed for the personalized offer generation use case.



# Solution architecture

To build an industry-aware solution (like the retail example in *Figure 5*), there are several questions that need to be answered before you can start with the implementation. The answer to those questions also helps to outline the solution architecture that is used for the specific project:

- **What is the end goal?** In our example, we want to generate relevant offers so that customers feel that we know them intimately. These offers should be served across the channels (for example, in-store, online, or telesales) and be based on the buying and browsing history of the customer.
- **What data is needed?** The next logical step is to identify the different datasets that you need to build such a solution. Based on what needs to be achieved, there are different data sources that need to be tapped and relevant data combined. In our case, data from the e-commerce site (transactions and browsing history), the POS system, general customer information, and product information need to be used to generate an appropriate solution.
- **What tools are needed?** Finally, we need to decide what tools should be used to bring the data together, analyze data, and feed the relevant information or insights to different upstream and downstream systems.



Azure Synapse provides an integrated environment that provides data integration, data shaping, and analytics engines to rapidly build an industry-relevant solution:

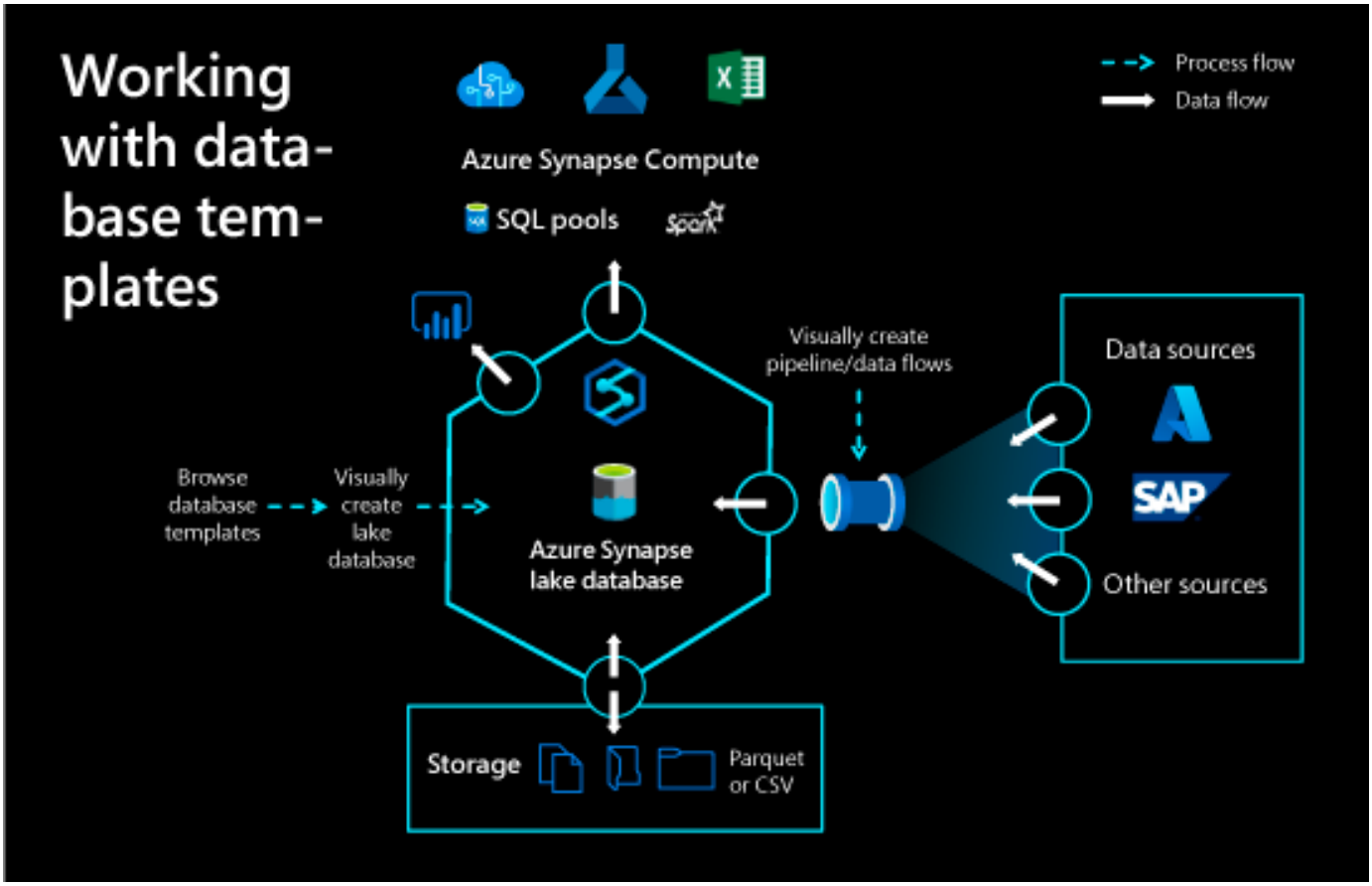



Figure 6: Working with database templates

For our solution, Azure Synapse pipelines are used to connect to the source systems and a lake database is used to bring all the information together in a single location. Database templates build the foundation for the database where all the different data sources are combined.

A Spark notebook is used to analyze data and make recommendations for customer offers that are stored in the lake database where different systems can pick it up. All this tooling is integrated within Azure Synapse Studio, a unified environment to develop, test, and get the solution into production.

Thanks to deep integration with Power BI and the fact that data sits with all the relationships and the associated metadata in the lake database, data analysts can themselves access, analyze, and derive insights.

# Implementation

In this section, we will demonstrate how convenient it is to build an industry-aware analytics solution with the capabilities of Azure Synapse. We will start by creating a lake database.

## Creating a lake database

The first step toward implementation is to configure and deploy the database template. The **Gallery** pane in an Azure Synapse workspace provides data engineers with the capability to browse and select an industry database template that can be used as a starting point to build the underlying database as shown in *Figure 7*:

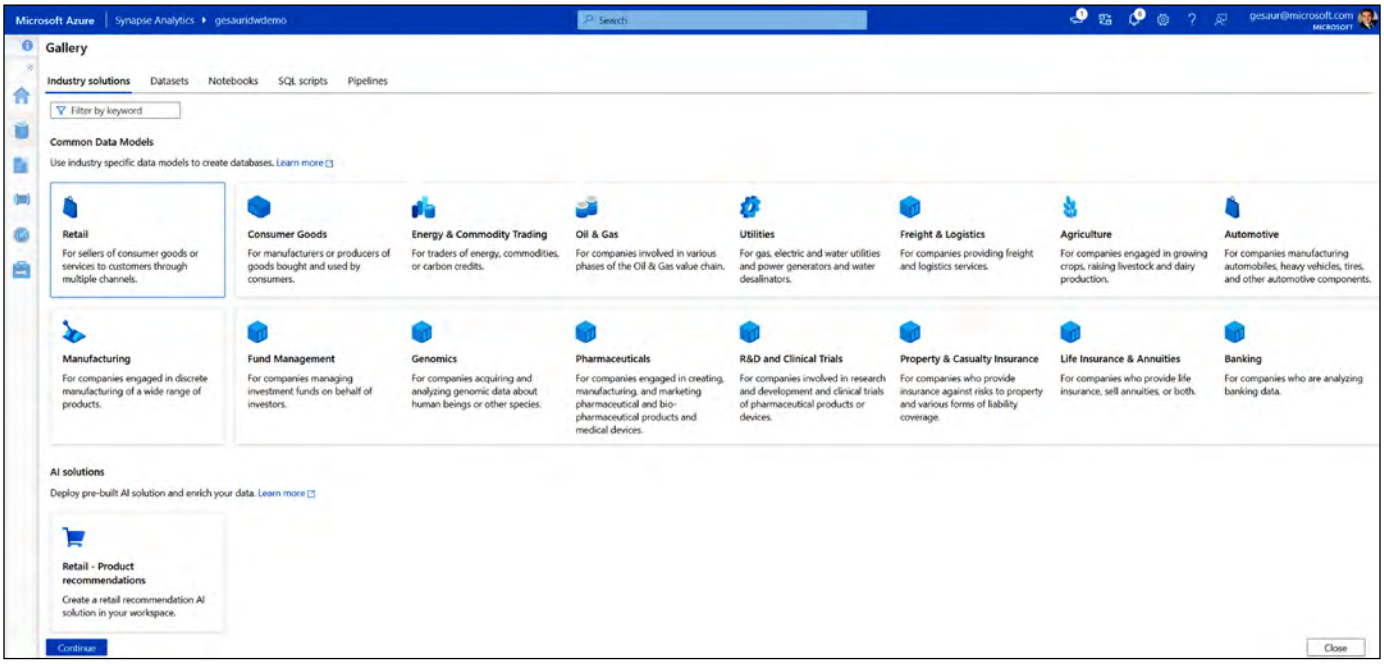



Figure 7: Database templates in Azure Synapse Studio

In our case, we will use the retail template and start selecting the appropriate tables from the business areas that are relevant to the desired retail industry solution (for example, a personalized offer solution).

Different business areas, like customer, sales, finance, and marketing, can be found on the left panel. These business areas have hundreds of prepopulated business entities (tables) that can be used as basic building blocks to customize and create a lake database layout that can easily be extended in the future:

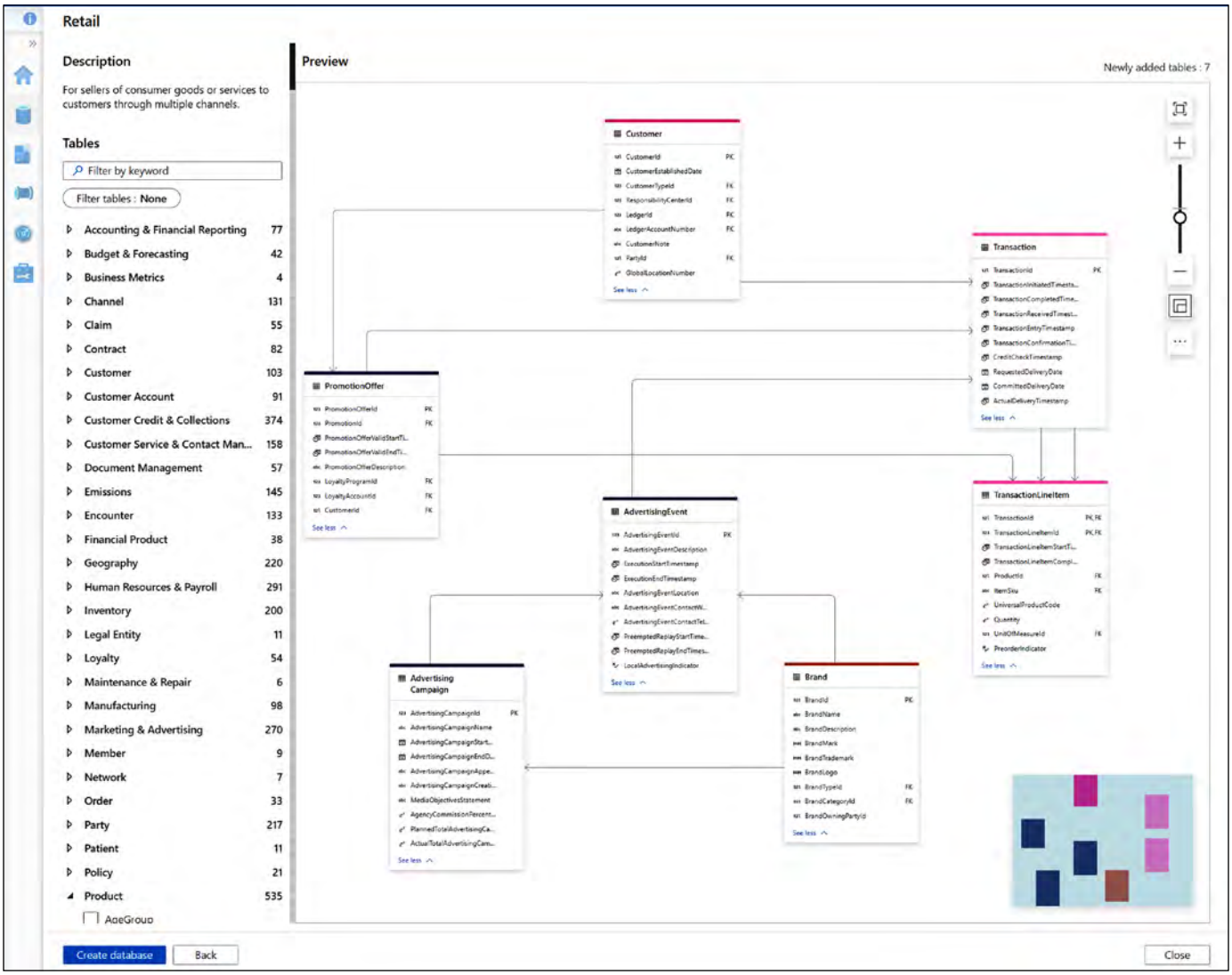



Figure 8: Database template browsing experience in Azure Synapse

As soon as you have selected the entities you are interested in to start with, the **Create database** action will bring you back into the Azure Synapse workspace, where a new lake database is created that can be fine-tuned to your specific need.

Entities and attributes can be added, changed, or removed to reflect the changes in your business processes, while the service will keep the lineage back to the original database template in case you need to add information in the future or want to use a newer version as the basis for your model:

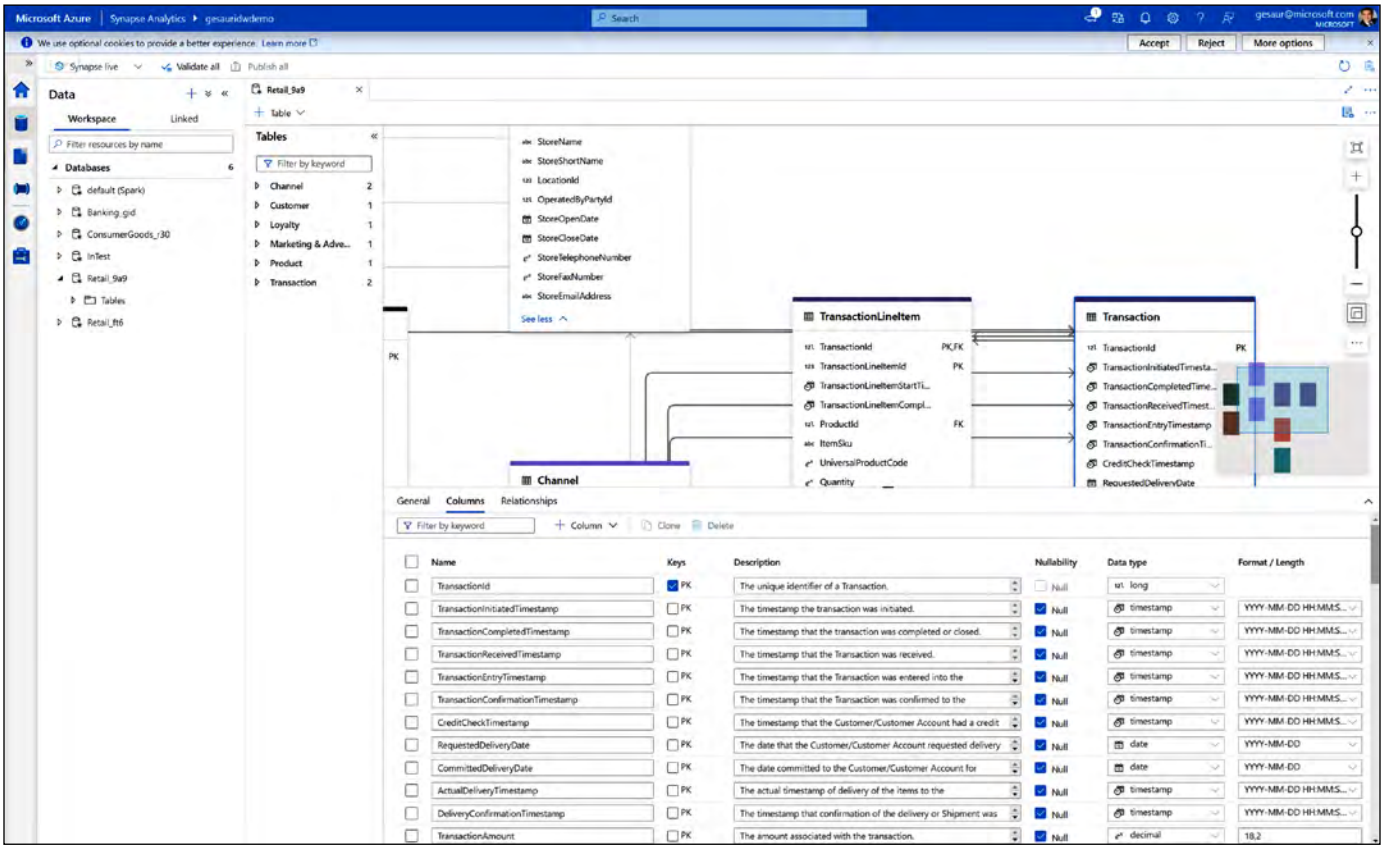



Figure 9: Database editing experience in Azure Synapse

For each table, the Azure Synapse database designer allows you to change the name, attributes, and relationships as shown in *Figure 9*. However, for the most impactful AI and ML integration, we recommend minimizing changes and staying as close to the original database template as possible.

# Bringing data into Azure Synapse

After the database is created and deployed, other services can access the newly created lake database within Azure Synapse. The first step is to align data from different sources into the newly created database. To do this, Azure Synapse provides low-code/no-code data pipelines that are directly integrated with the lake database as shown in *Figure 11*.

Data processing pipelines provide connectors to different sources, which makes it easy to connect to different source systems as shown in *Figure 10*. There are hundreds of different connectors available that can be used out of the box to get data into the new lake database easily (for example, SAP ECC, Dynamics CRM, Magento, and others):

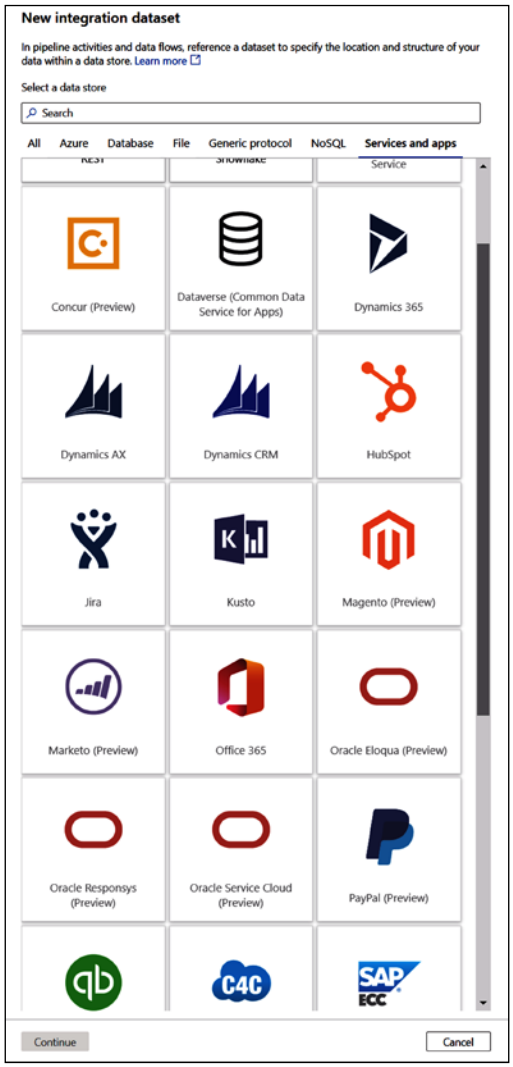



Figure 10: Different Synapse connectors available

Additional information about available connectors can be found at [Connector overview – Azure Data Factory & Azure Synapse | Microsoft Docs](Connector overview – Azure Data Factory & Azure Synapse | Microsoft Docs).

The pipelines also provide an easy way to join, filter, select, and modify data that is coming in, as you can see in *Figure 11*. The graphical user interface makes it easy for everyone to understand what is happening with the initial data. A data lake sink understands the format of the new lake database using rich metadata from the template that has been used. No additional configuration is needed, as the service already knows how and where to store data from the lake database that was created:

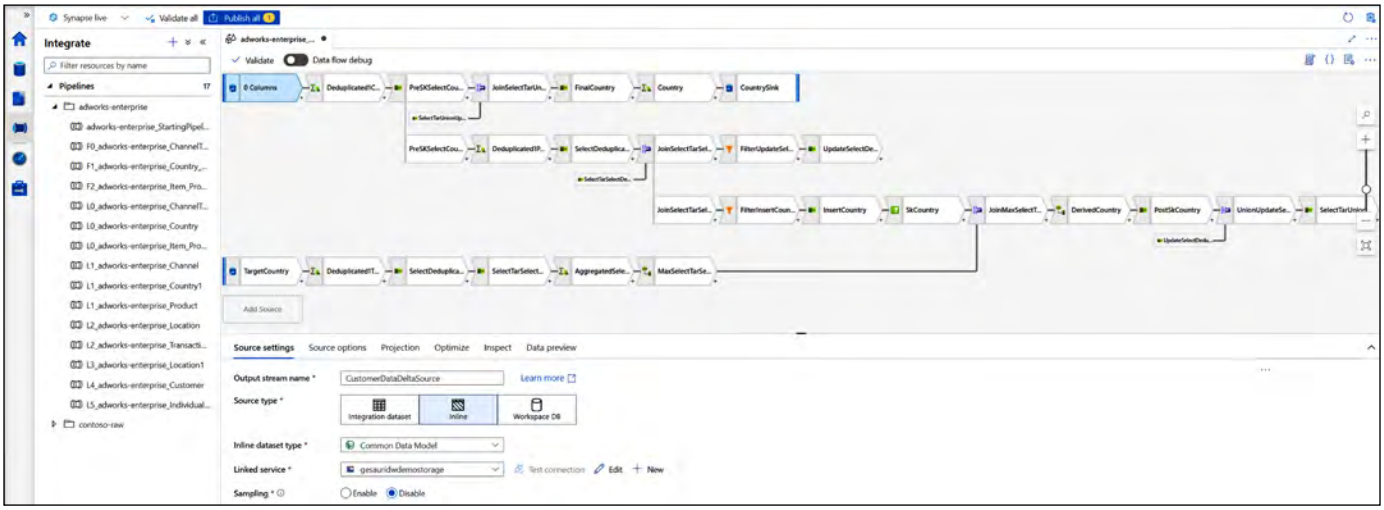



Figure 11: Pipeline example to map data

Once you have landed the data in the common format, you can integrate visualization or ML models that work directly out of the box with the format of your data as shown in *Figure 12*. They can leverage the data stored in the lake database and provide data back, as well. Leveraging database templates as the blueprint for the lake database helps to integrate these components easily as the solutions already know the format of the data.

In the retail example, the ML solution will create personalized promotional offers and advertising events that are stored back in the lake database where they can be picked up by other solutions:

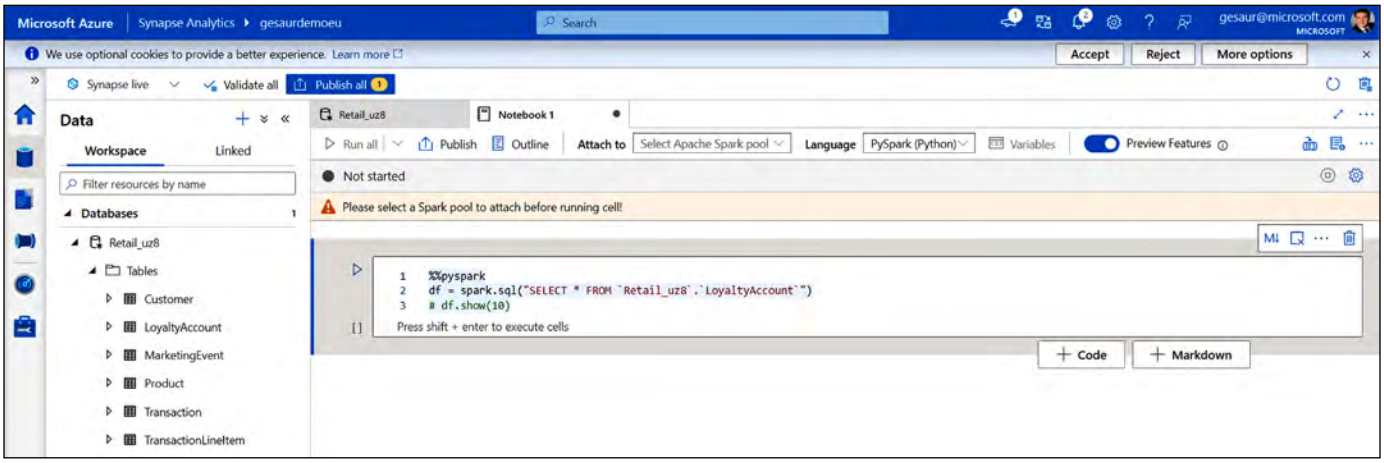



Figure 12: Example of how to access the data using Spark

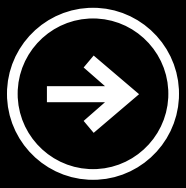
Now that the data is available in a standardized format and enriched with metainformation, it can be exposed and viewed by different services. Data processing pipelines can be used to publish the newly generated insights into other services. The connectors mentioned earlier can be used for this as well (for example, a web shop, the POS system, or social media channels, as depicted in *Figure 5*). In addition, the data can also be exposed over a SQL interface provided by Azure Synapse out of the box on top of the created lake database. SQL on-demand can be used on top of the lake database and make use of the rich metadata stored in the model. This enables every solution that can connect to SQL to connect and access the data in the lake.

The data expressed in the lake database in Synapse can be used in Power BI and other Power Query-enabled apps (for example, Power Apps and Customer Insights) as well as using the Synapse Power Query connector. With the ease of data import, along with the automatic import of relationships and the additional metadata in the lake database (names, descriptions, and other metadata), business users can use Power Query's no-code transformation capability to get faster and better results for their ad hoc business needs. It is easier than ever for business users to select the right data across tables, by leveraging relationships, and create queries with joins and filters without needing to write code or understand relational database concepts.

# Summary and conclusion

In this guide, you got a sense of how you can accelerate the digital transformation of your business with Azure Synapse Analytics. We highlighted the power of database templates and integration into the lake database to create an industry-aware data foundation for your specific industry solution. We used a retail example to highlight how the database templates are structured and work. Azure Synapse comes prepackaged with database templates supporting several industries, including retail, consumer packaged goods, banking, and funds management.

For more information and implementation guidance, check out the Azure Synapse database templates documentation.



To start with Azure Synapse Analytics today, you can:

- Create an Azure account to create a new Azure Synapse instance and explore your data instantly.
- Visit the Azure Synapse Analytics documentation webpage to see how cloud analytics can help your business.
- Request a call from an Azure Synapse Analytics sales specialist if you have any questions or need help.