E-book series

Microsoft Azure

Analytics Lessons Learned: How Four Companies Drove Business Agility with Analytics

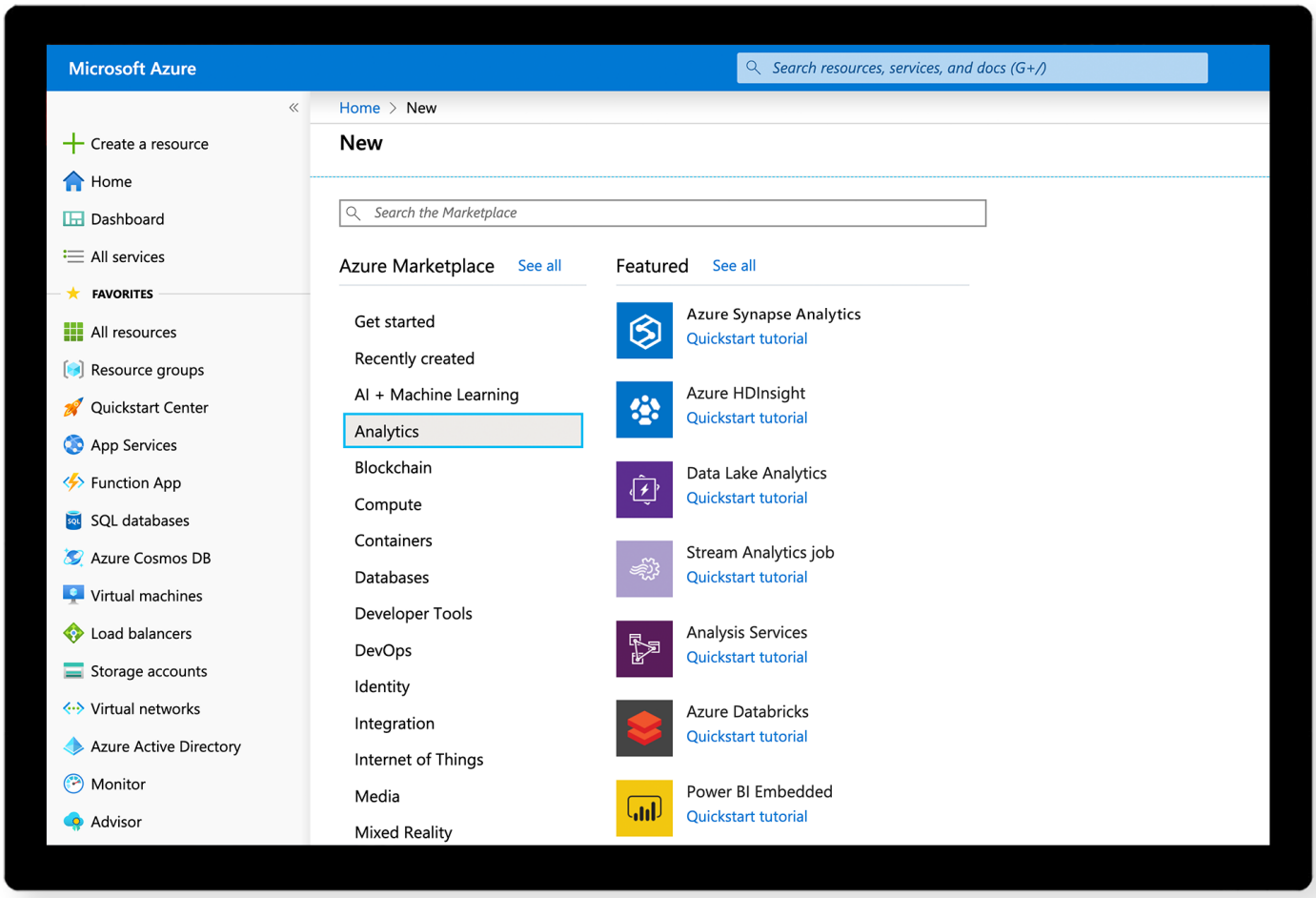
01 / Introduction 3

02 / Use case 1: Just-in-time inventory management 4

03 / Use case 2: Fraud detection 8

04 / Use case 3: Predictive maintenance 12

05 / Use case 4: Marketing analytics (customer 360˚ view) 16



© 2020 Microsoft Corporation. All rights reserved.

This document is provided "as is." Information and views expressed in this document, including URL and other internet website references, may change without notice. You bear the risk of using it. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

Introduction

Data analytics is an essential part of any business strategy. Businesses, governments, and individuals worldwide use data to make responsive, informed, and timely decisions using performant data analytics. It's not enough anymore to set a business goal and work toward it; in order to stay relevant in today's market, organizations need to have a finger on the pulse of their business in real time.

For most organizations, this means stitching together a multitude of services to run analytics. This approach results in brittle, high-maintenance architectures that drain resources to manage technology, resources that ought to be spent on responding to new opportunities.

In the world of analytics, speed is paramount, because the value of data depreciates with time. Think of a scenario where you can predict that a part of a manufacturing line is faulty. If you find this data point early enough, you can perform preventative maintenance and avert any potential damage. Besides speed, organizations also need to have agility and scalability to be able to scale their analytics workloads according to their business context and challenges.

This is where cloud analytics platforms, such as Azure Synapse Analytics, become a vital asset to any organization. With Azure Synapse Analytics, companies get a complete, out-of-the-box solution designed to accelerate time-to-insight and increase the agility of their business. Azure Synapse Analytics is the only end-to-end platform that unifies data ingestion, big data analytics, and data warehousing. It also offers turnkey setup and configuration options on a fully managed platform to help businesses get results quickly without getting bogged down. Abstracting all the infrastructure, Azure Synapse Analytics offers a platform as a service (PaaS) and allows customers to choose the best pricing option for each workload, with both serverless and dedicated options.

To help you understand how cloud analytics in general and Azure Synapse Analytics in particular can help you and your business, we will cover three different use cases over the next few pages. These are based on practical, real-world scenarios, where global organizations adopted Azure Synapse Analytics to help their teams uncover new insights and explore data in a scalable, secure, and timely manner. We'll cover key features of Azure Synapse and explore what makes it unique compared to the other cloud analytics platforms on the market.

Photo courtesy of Aggreko

Use case 1: Just-in-time inventory management

Aggreko is a global leader in the supply of temporary power generation, temperature control systems and energy services, providing backup energy and power supply whenever and wherever their customers need it. They need to balance temporary demands alongside permanent business requirements. Their business model requires building custom applications for their customers, who require different specifications and different inventory due to geographical differences, weather, and other factors. Aggreko uses Azure Synapse Analytics to increase operational efficiency with the just-in-time supply of their specialist equipment.

Business challenges and requirements

Aggreko, like many other organizations, had an on-premises data warehouse. The data team pulled data from many source systems, including financial systems, operational stores, business and end user systems, B2B apps, and external datasets. All this data needs to come together for the team to be able to make any meaningful analytics or uncover any possible data trends.

Due to the constraints of on-premises infrastructure, the team used to have to schedule these batch jobs to run every 8 hours, and each time they loaded the data to the warehouse it took about 4 hours to process a new batch. Moreover, the data warehouse had to be rebuilt every day, due to storage limitations. This frustrated the data team and hampered their ability to produce analytics in a timely manner. They were only able to operate on data that was already about 24 hours old, which meant their reporting was not as relevant and they were not able to drive business value as the data arrived.

They wanted to use the data to keep their warehouses well stocked with the correct inventory, by responding quickly to business requirements as the data came in. This was almost impossible with the constraints of the on-premises data warehouse and their batch processing of data.

Not only did they need hours to batch the new data, but they also needed hours to transform and clean the data before they could query it. In fact, the data team estimated that about 30-40% of their time was spent on tackling technology complexities such as infrastructure tuning and data cleaning before they could begin to query and explore data.

Solution review

To improve operational efficiency and introduce near real-time inventory management, Aggreko adopted Azure Synapse Analytics. They started by re-platforming their on-premises data warehouse to Azure Data Lake Storage Gen2, which gave them a single home for data with limitless scale. All the necessary data from internal and external systems came together to this one platform and was ingested, curated, and transformed to meet the needs of Aggreko's varied applications and teams.

At the center of Aggreko's new data analytics architecture is Azure Synapse Analytics, which helps the team by providing both a data warehouse and the unified platform necessary to enable big data analytics in a single service. Azure Synapse pipelines and Azure Event Hubs are used for data ingestion. This helps the data team to significantly reduce data loading job times and makes it possible to query data and uncover trends while the data sits in the data lake. This empowers Aggreko to achieve exponentially faster insights, irrespective of the data volume that they are processing, by querying the data directly and then outputting the resulting dataset into Power BI for interactive reporting and dashboards.

Furthermore, using a Jupyter notebook with a serverless Apache Spark pool in Azure Synapse Analytics, the data team is able to explore data as it arrives in near real-time—every 5 minutes—to highlight any inventory shortages that might cause delays in manufacturing and push the result into a Power BI dashboard. This dashboard then tells planners and customers alike whether Aggreko has the complete bill of materials to build an asset.

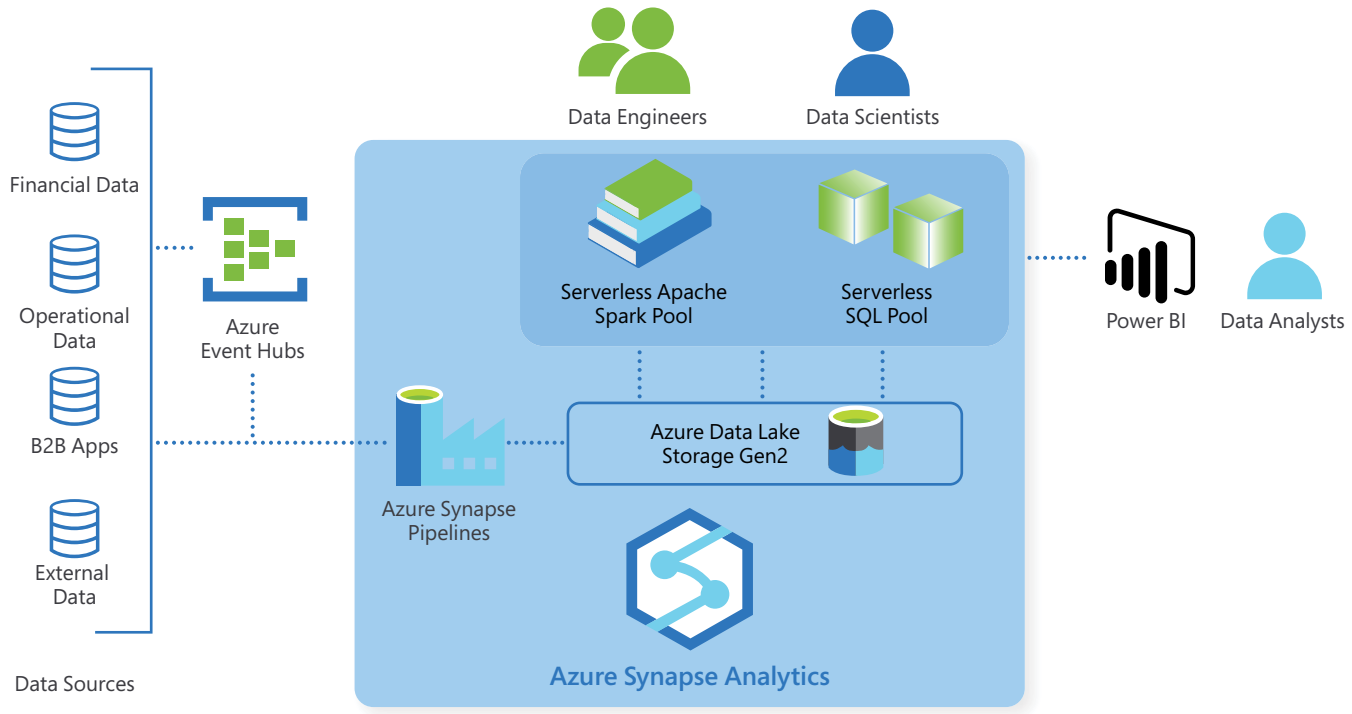


Figure 1: Solution overview for the just-in-time inventory

Outcome

- **The complexity of Aggreko's data ingestion was drastically reduced, and the speed improved.** Ingesting the required data from the source systems used to take 4 hours and could only be loaded every 8 hours. Thanks to the data ingestion pipeline capabilities in Azure Synapse, this now happens as the data arrives, meaning that data can be ingested and processed in near real-time (with up to 5 minutes' delay). This not only improved speed and efficiency but also enabled analytics scenarios that were impossible under the old system, by making the most recent time-series data available for analytics in Azure Synapse.

- **An accurate, near real-time report on available materials became available.** The resulting dataset was then pushed to Power BI as a Materials Status report to be used by planners, customers, and analysts to help them understand whether they have the full bill of materials to build an asset. The report shows users a summary of the status of the parts. For each project, the report shows the materials needed with a color-coded shortages graph (red, orange, and white).

- **Purchase orders were made easier for planners to place.** Taking advantage of the Power Platform, the data team built an app that was embedded into the Materials Status report to enable planners to place purchase orders for items that had shortages. This made it very easy for planners to take action quickly and easily fill any such shortages.

- **Time was freed up that used to be spent solving technology problems.** Shifting the data analytics to Azure Synapse and simplifying ingestion pipelines meant the data team gained around 30-40% of their time back that was previously spent tackling technology problems and working around infrastructure and storage limitations. This time can now be used to focus on solving business problems and uncovering new insights.

- **A unified platform has created a unified team.** Lastly, the Aggreko data team now has a unified data analytics platform that brings together data scientists, data engineers, analysts, planners, and business users to collaborate and innovate faster and easier than ever before.

This use case is based on a real-world scenario where Aggreko adopted Azure Synapse Analytics as their analytics platform. To learn more about this customer story, you can watch **this interview with Aggreko's Director of Data Insights**.

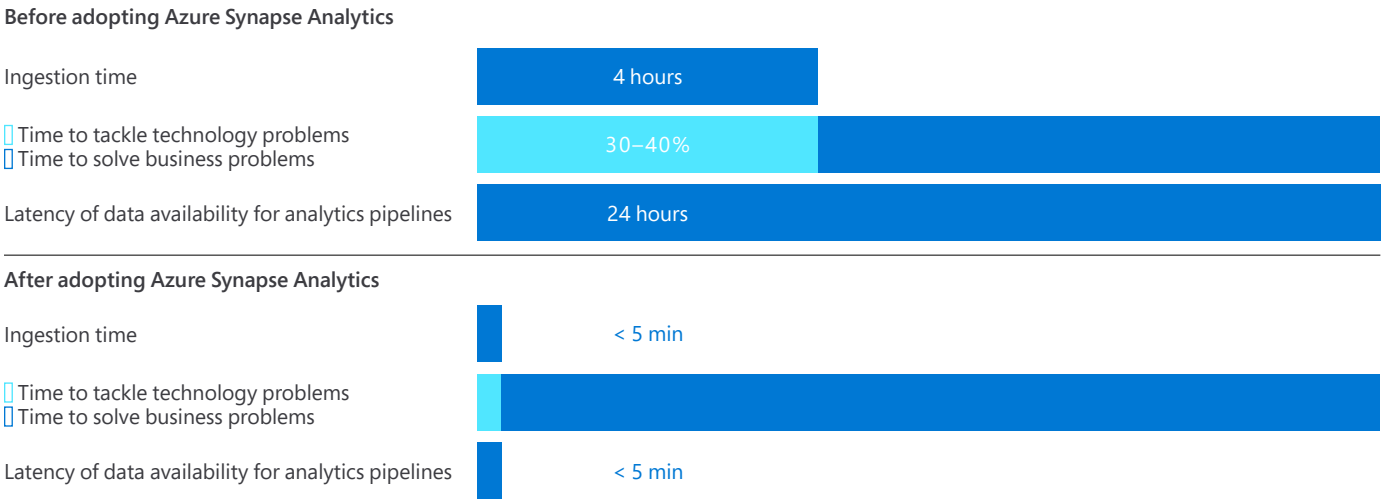


Figure 2: Measuring the outcome of adopting Azure Synapse for just-in-time inventory

Photo courtesy of Clearsale

Use case 2: Fraud detection

The size and number of online transactions are exploding, and as a result, the level of fraud is rising and the techniques that are used to commit it are becoming increasingly more sophisticated.

Clearsale, a leading fraud detection company based in Brazil, used Azure Synapse Analytics to modernize their operational analytics data platform. Clearsale helps customers verify an average of half a million transactions daily using big data analytics to detect fraud across the world; they collect information about purchases, as well as client behavior and other data points. The client behavior data includes many observations about the client, such as the number of credit cards they used, the size of the transaction, and their approximate location. Clearsale runs this data through machine learning models, as well as business rules and historical data from different industries. All this must run in seconds, and based on the results, Clearsale provides a prediction on whether this transaction was fraudulent or not.

Business challenges and requirements

The Clearsale dataset doubles in size every 2 years. They had several SQL servers on-premises that allowed them to ingest data, and they also ran their analytics workloads on these servers. Due to the limitations of their infrastructure, on occasion, they needed to create a new SQL server with replicated data to be able to process their analytics pipelines when there was high demand. Clearly, this was inefficient, and it caused delays in fraud detection.

Fraud detection must take place in seconds. For 99% of transactions, the machine learning models can detect if there's any suspicious activity. For the remaining 1%, an investigation team takes over to process the remaining transactions.

Fraud is also very dynamic. A data leak in one industry could quickly impact others and create many fraudulent transactions, so being able to correlate Clearsale data with data from other data leaks in a timely manner is critical. The data team wanted to be able to ingest external datasets, for instance, data from major data breaches or industry-specific datasets, quickly and easily. This was almost impossible because their systems were limited by the pre-existing on-premises computing and storage capacity.

Solution review

To address their operational data storage needs, Clearsale adopted an Azure SQL Database Hyperscale to help them scale their operational computing. This data store is used to address operational needs, such as serving apps and feeding machine learning models.

For data warehousing requirements, Clearsale now uses Azure Synapse Analytics to bring all their operational and historical data together. They use Azure Data Lake Storage Gen2 as the backing data store. By adopting Azure Synapse Analytics, Clearsale is now able to query and explore their data in near real time. This significantly speeds up the time-to-insight and makes it easy to perform advanced data exploration without impacting the operational data store.

Clearsale also uses Azure Synapse Analytics to periodically train and enhance their machine learning models, training them against the full dataset in the data warehouse, before pushing the results to the Azure SQL Database Hyperscale to improve their fraud detection capability.

Figure 3 depicts the high-level architecture of Clearsale's specific data analytics solution. It includes an Extract-Transform-Load (ETL) pipeline to load the data from the on-premises SQL servers to Azure (SQL Hyperscale and Azure Synapse):

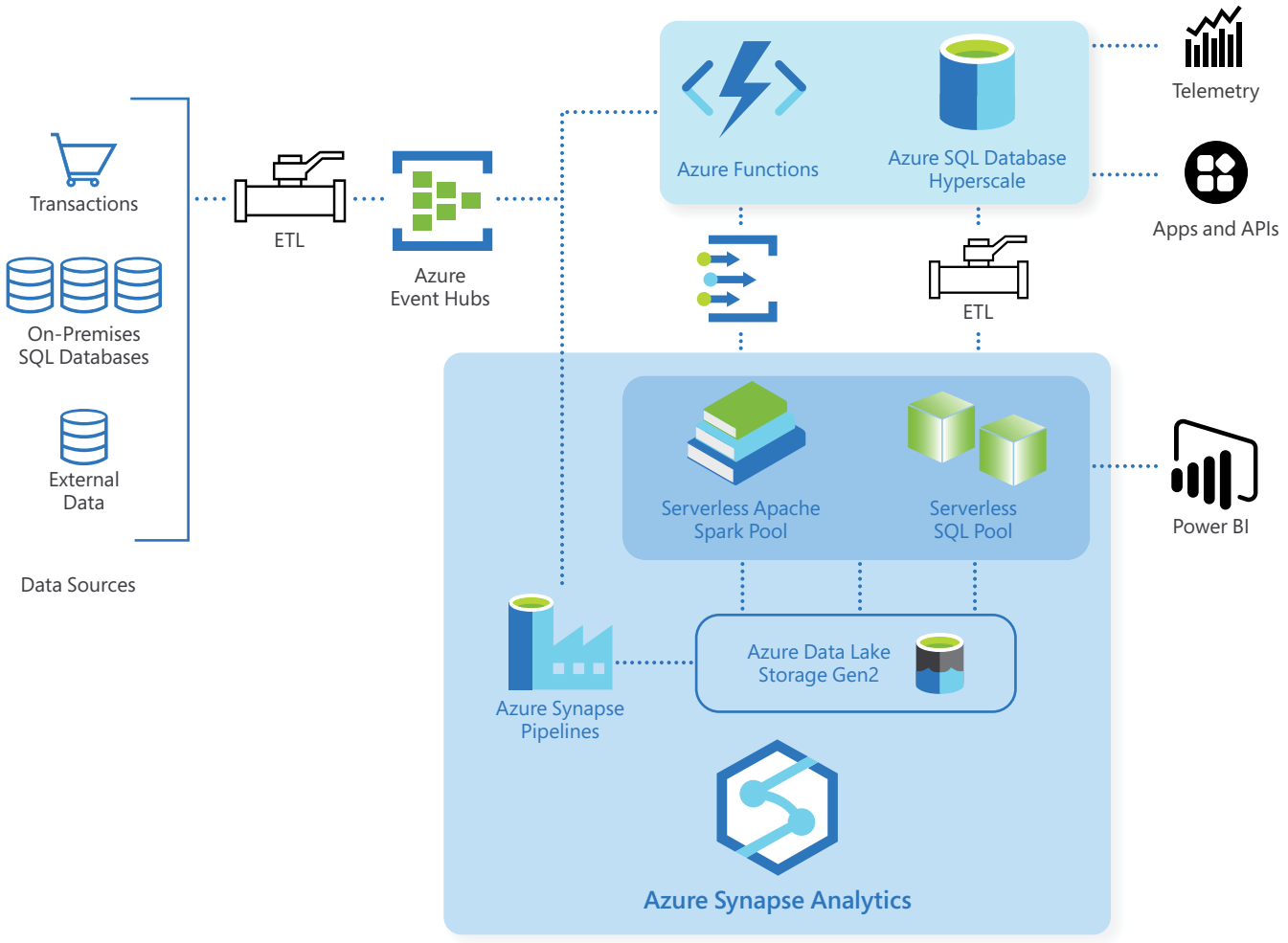


Figure 3: Solution overview for Clearsale's fraud detection on Azure

Outcome

Clearsale is now taking advantage of Azure Synapse Analytics' limitless scale and out-of-the-box features to modernize their data estate. Overall, the outcome of this early adoption of Azure Synapse Analytics is greatly positive. The benefits that Clearsale has realized by moving to Azure Synapse Analytics can be summarized as follows:

- **The Clearsale data team has significantly reduced the time it takes to train new models to improve their fraud detection capability.** Using their previous on-premises platform, it used to take close to a week to ingest, prepare, and train the machine learning models. Using Azure Synapse Analytics, this has now been slashed to under 6 hours. This is a massive improvement that has enhanced their capability, improved efficiency, and reduced operational overhead.

- **The team can now ingest large datasets easily and quickly, by using Azure Synapse as their data warehouse with Azure Data Lake Gen2 as a data store.** This includes not only their operational data and their historical data but other datasets that are relevant to their fraud detection capabilities.

- **Knowledge and experience on other systems transfer to Azure Synapse.** The Clearsale data team can carry over all their knowledge and experience of working with SQL, Spark, and Azure Synapse Studio as they shift to Azure Synapse Analytics. This makes it easy to adopt Azure Synapse for teams that have made investments in their skillset and their tools. Azure Synapse Analytics is a modern platform with a new architecture, but it builds on existing industry standards, open source technologies and familiar languages (such as SQL). This makes transitioning to Azure Synapse Analytics very easy.

- **New security features are available to the business.** Since Clearsale deals with financial transactions, they have always focused on security in all aspects of their business. By adopting Azure Synapse Analytics, they can now use Azure Synapse's security features, such as Dynamic Data Masking and row- and column-level encryption.

- **Clearsale can now manage resources and control workload priorities on Azure Synapse in a simpler, more powerful way.** The data team created different workload management groups, such as groups of BI analysts, data engineers, and data scientists. Using these management groups, they can assign precise priorities to submitted workloads. This delivers a more optimal use of resources, better performance, and reduced costs.

This use case was based on a real-world scenario, where Clearsale modernized their data estate using Azure Synapse Analytics. To learn more about this customer story, you can watch **this interview with Clearsale's Data Engineer on Microsoft Mechanics**.

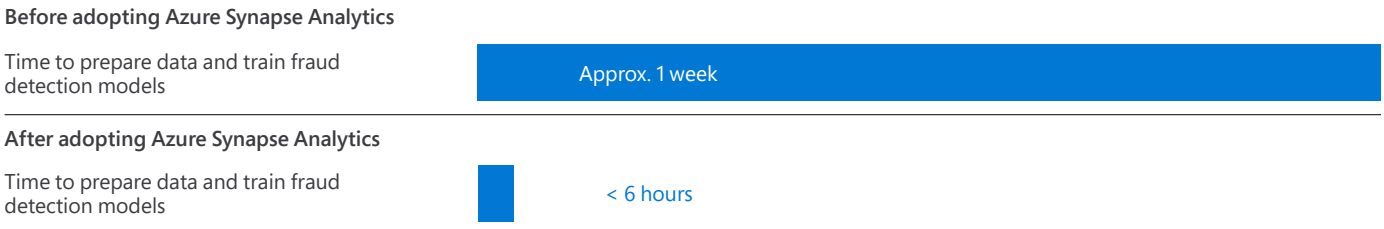


Figure 4: Measuring the outcome of adopting Azure Synapse for fraud detection

Photo courtesy of Microsoft

Use case 3: Predictive maintenance

GE Aviation's Digital Group is a world leader in manufacturing airplane engines and developing aviation software. They produce a wide range of products, from specialized military aircraft engines to the most powerful civil airplane engines. On top of manufacturing, GE also provides advanced data analytics to many airlines around the world, with a focus on safety analytics.

GE, like most organizations, is looking to modernize their data estate. They implemented Azure Synapse Analytics as part of their next-generation data platform for global safety analytics.

Business challenges and requirements

GE ingests massive amounts of data from flights worldwide. For each flight, they ingest the flight's time-series data for the entire flight, which includes as many as 350,000 data points. This data comes combined with data from other sources, such as aircraft type, flight plans, runway, and airport data. GE also ingests data from external sources, such as forecasted and actual weather data. All this generates a massive amount of data that needs to be ingested, curated, and transformed to be ready for consumption.

Over the last 20 years, GE has built their data platform using on-premises technologies. They built a data enabling layer on top of the flight data processing system, which is called EMS. EMS is used by many large airlines around the world to decode and transform flight data recorded by sensors on planes. GE has also built a computation layer on-premises, called the Distributed Processing System. This system runs descriptive analytics, using a library of over 10,000 predefined aviation analytics, to detect abnormal events through every phase of a flight.

GE offers its safety analytics as a service to customer airlines. It provides remote desktop access to its customer airlines so that customers can log in and use GE's safety analytics platform. GE also provides its customer airlines with the ability to access its safety analytics service via APIs.

GE has a massive and very mature data platform. However, data ingestion, movement, and exploration were not easy tasks with the platform as it was. Scalability was also another area where GE wanted to improve, since their platform analytics were bound by the limitations of their infrastructure. Figure 5 shows the high-level architecture of GE's data analytics platform, pre-Azure Synapse.

One key scenario that the data team at GE wanted to explore with Azure Synapse Analytics was the ability to train and produce predictive maintenance machine learning models against a sample set of flight records. These models are developed to produce condition indicators for an aircraft's pneumatic system.

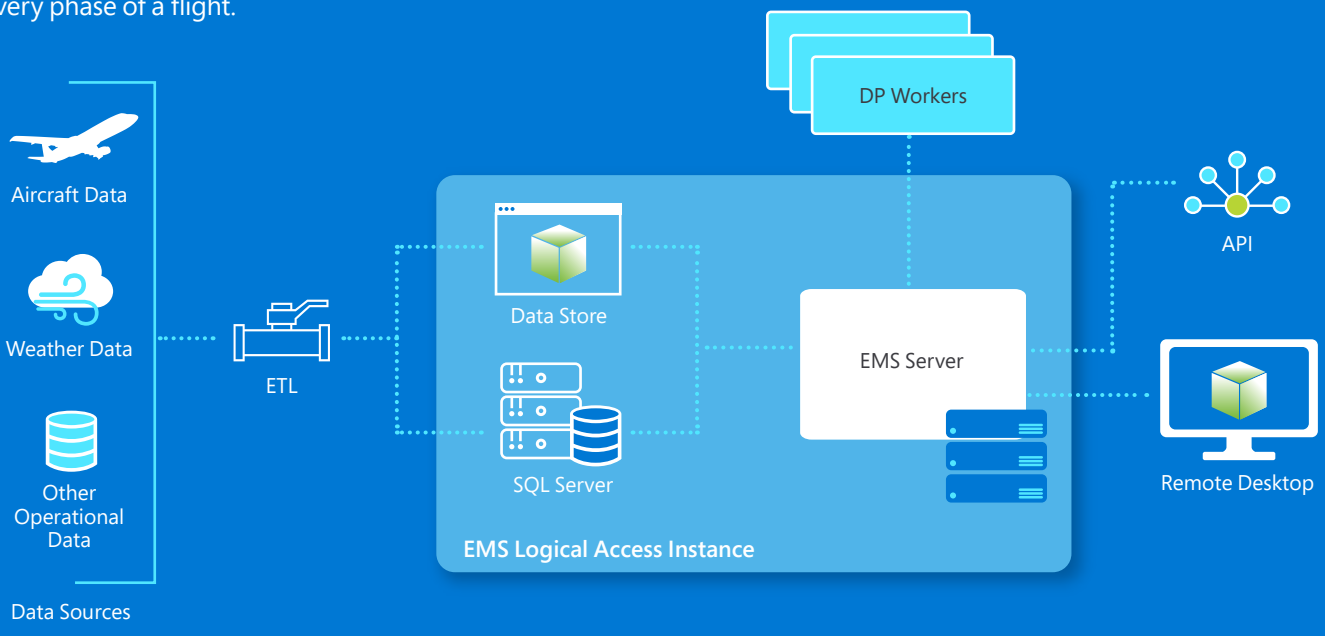


Figure 5: GE's safety analytics platform, pre-Azure Synapse

Solution review

GE's data team started by pulling the essential parts of the data ecosystem, via an API, to Azure Synapse Analytics. The data is stored in Azure Data Lake Storage Gen2, which means that the data is now available seamlessly in Azure Synapse for immediate exploration and discovery. This was a great start since it removed the previous restrictions in the platform where the data team was bound by the data scheme and data warehouse structure.

GE's engineers found that it was straightforward to ingest their EMS (flight records data) into Azure Synapse. They built a pipeline in Azure Synapse (using Azure Synapse pipelines) to load the data and transform it into a Parquet file, which was then stored in Azure Data Lake Storage Gen2. The next step was a Python script written in the integrated Jupyter notebook, running on a serverless Apache Spark pool in Azure Synapse. This notebook was used to generate condition indicators, which are floating-point aggregates where various safety and operational aspects have been measured against the time-series flight data.

Another Jupyter notebook was then used to execute a Scala script to load the resulting condition indicators into the serverless SQL pool and split them out into one fact table per condition indicator, along with a few dimension tables that describe the aircraft associated with each of the condition indicators, such as the aircraft fleet and ID.

Power BI was then used to visualize the data via interactive reports, which made it much easier for GE analysts and engineers to find abnormal events. Once an abnormal event was found, the analyst could then drill down quickly and easily into the relevant data points, which might indicate a faulty sensor or equipment that needs attention. Figure 6 shows a high-level diagram for their Azure Synapse use case.

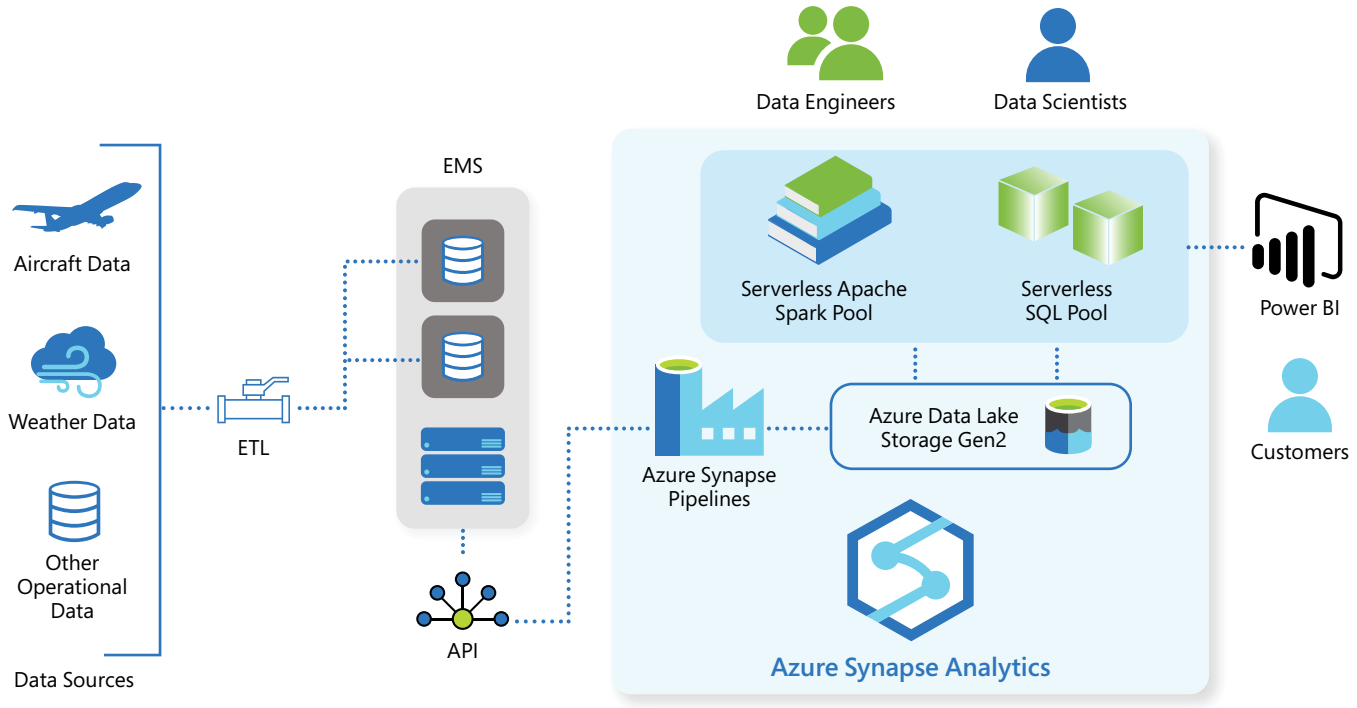


Figure 6: High-level solution design for GE predictive maintenance on Azure Synapse

Outcome

Using Azure Synapse Analytics, GE was able to scale their predictive analytics quickly and easily. Azure Synapse made it easy for them to build complex pipelines that combine many massive datasets. The outcomes for GE can be summarized as follows:

- **Using Azure Synapse Analytics made it significantly easier and quicker to build complex predictive machine learning models.** Building something similar in their previous system would have required many complex steps in many systems and across multiple environments. Azure Synapse Analytics not only created faster time-to-insight but also saved a great amount of time for the data team, which they could invest in solving real business problems rather than working around technology limitations.
- **GE found that the native integration between Power BI and Azure Synapse was extremely useful.** They can now explore data quickly and when an anomaly is found in the condition indicators report, analysts are able to do drill-down analysis of why the spikes occurred and what corrective maintenance is needed.
- **GE especially valued the limitless scale of Azure Synapse Analytics**, as they no longer need to worry about infrastructure limitations or scaling limits. This improves performance and, at the same time, can reduce costs.
- **The data team at GE was very happy with the ease of adopting Azure Synapse**, since it makes use of open source technologies and uses familiar tools and languages such as Apache Spark and SQL.

This predictive maintenance use case was based on a real-world scenario, in which GE is trialing Azure Synapse Analytics to modernize their safety analytics platform. To learn more about the customer and their use case, you can watch **this interview with GE's Senior Product Manager on Microsoft Mechanics**.

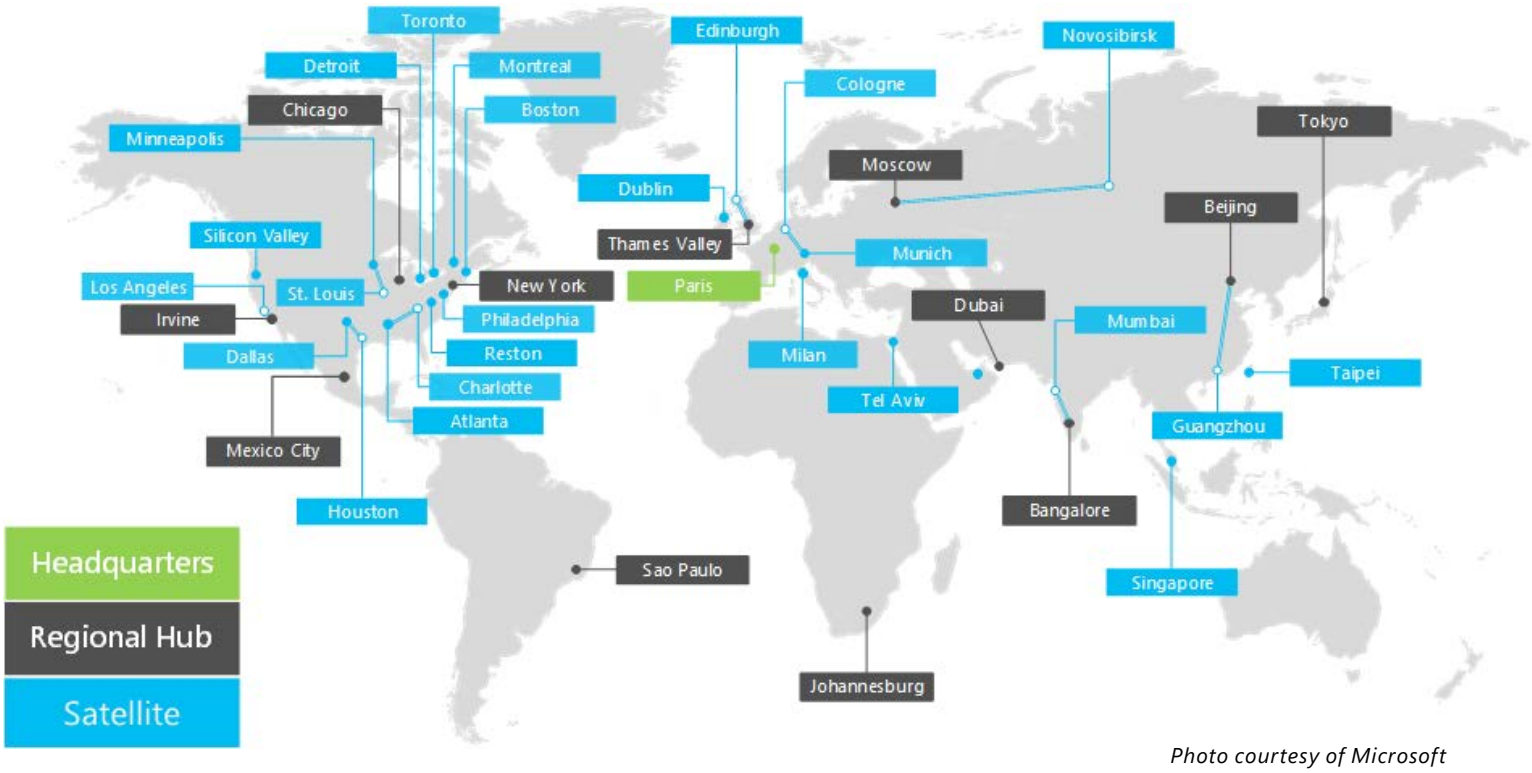


Photo courtesy of Microsoft

Use case 4: Marketing analytics (customer 360° view)

In this use case, we look at how a retail company can improve its customer experience and profitability by leveraging its data to build a 360-degree view of their customers.

Imagine a company named Contoso, a large multinational retail company that has stores across the globe. The company sells consumer goods, electronics, and personal care items through its brick and mortar stores, as well as through digital online channels (mobile and web applications). Contoso has started to use Azure Synapse Analytics to build a complete view of its customers and is aiming to use Azure Synapse as its modern data platform to enhance customer experience and increase profit.

Business challenges and requirements

Contoso has many different datasets, stored in disparate systems. Customer data comes from everywhere and is in everything: customer service calls, website visits, purchases in online and physical stores, and mobile app usage. These are but a small number of the channels that generate large volumes of data every minute for Contoso.

The volume of the data, as well as its varying structure and formats, made it almost impossible to process in the previous on-premises data warehouse. Besides that, the data comes from many data sources that might not have much obvious correlation with each other; how do you compare website usage data and customer purchases, or weather data and stock levels? It's a massive task not just to ingest these different datasets but to clean them and create correlations. This is extremely challenging; essentially, they are trying to create joining identifiers where there might be none.

Contoso adopted Azure Synapse to bring all this data together, to enable data discovery and data engineering in a collaborative and timely manner.

Solution review

Contoso's adoption of Azure Synapse is at the center of their data solution architecture. Figure 7 shows the high-level solution design; Contoso uses Azure Synapse pipelines to ingest data simply and quickly from different data sources to Azure Data Lake Storage Gen2. It also uses Azure Event Hubs to capture non-structured data, such as social media feeds, IoT sensors, and clickstreams. All this data is stored in Azure Data Lake Storage Gen2.

Using the Azure Synapse unified data analytics feature, the data team can now begin exploring the data straight away, without moving it or transforming it to other formats. Contoso's data engineers can now either use Azure Synapse code-free data pipelines or build custom code using Jupyter notebooks on serverless Apache Spark pools to curate, correlate, and transform data to build consumption-ready datasets. These datasets are then pushed to Power BI for interactive visualization and reporting.

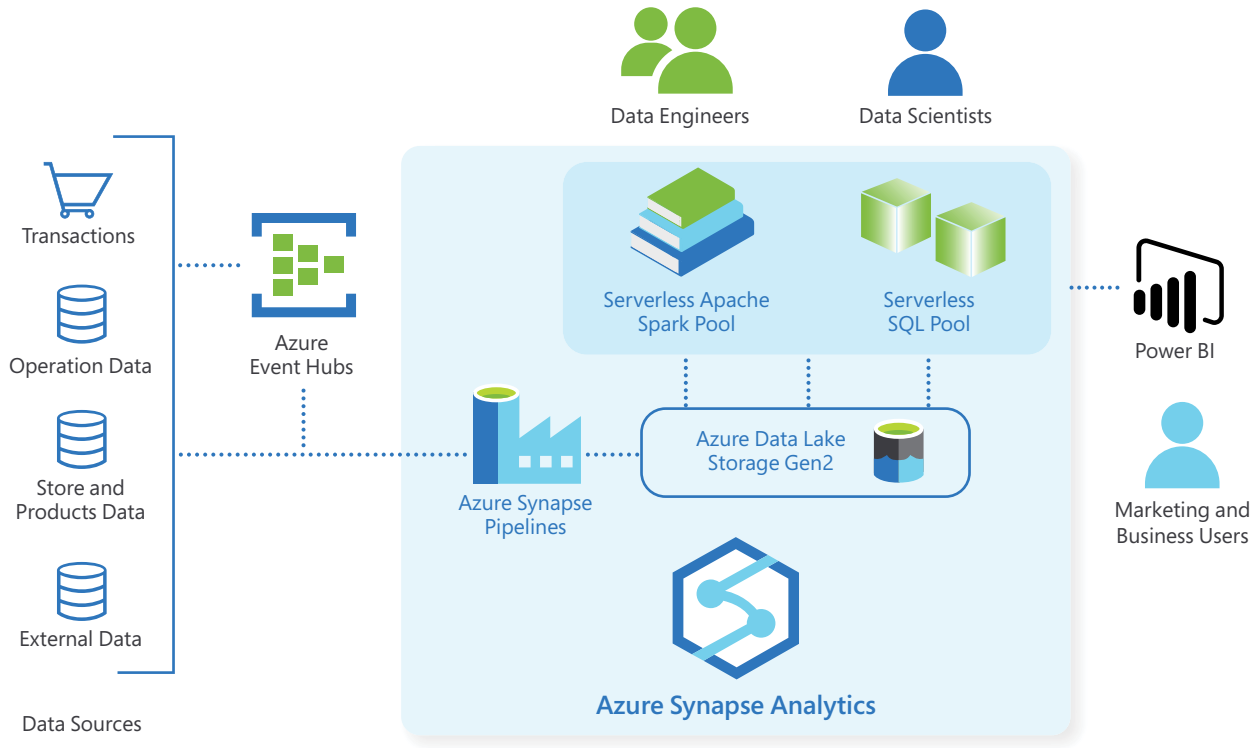
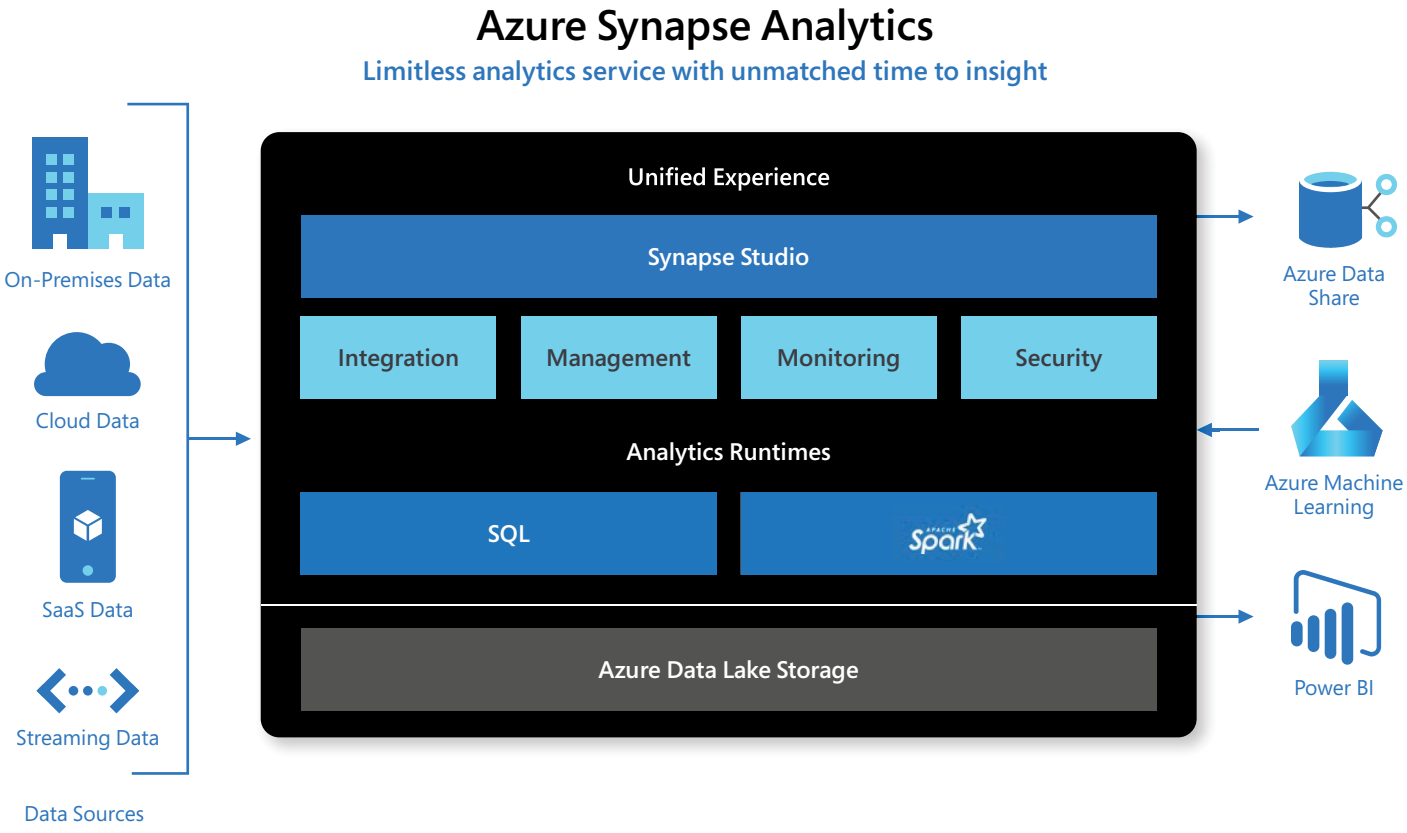


Figure 7: Azure Synapse for Marketing Analytics

Outcome

Azure Synapse Analytics has enabled Contoso to unite its data, developers, and business users in ways that were not possible before. Azure Synapse has simplified Contoso's data ingestion and data processing and made it easy for Contoso to have a central data store that holds all operational and historical data and that can be refreshed in near real time.

Azure Synapse also simplified Contoso's data exploration and discovery, removing the need to transform data from one format to another or move data to other systems. This enables the Contoso data team to experiment, by mapping and correlating the different datasets to produce curated datasets that are ready for consumption.



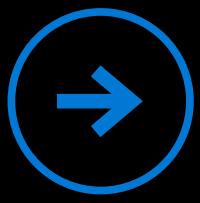
Azure Synapse is built on a modern architecture that is very performant (up to 14 times faster than comparable services), costs up to 94% less than other services, and offers more features that are critical for data analytics projects such as Integration, Management, Monitoring, and Security.

Next steps

Today's significant economic and health challenges can be turned into growth opportunities when faced with great leadership, a clear vision, and intelligent data analytics. Organizations need, in these times, to lean on data and science, as well as the will and knowledge of their teams. Empowering teams with the right tools is critical to ensure that they can collaborate, discover, and innovate using data analytics.

As we have discussed, Azure Synapse Analytics offers a limitless, unified, and powerful analytics service that enables businesses to start small and scale as needed. You only pay for what you use, and only when you need it.

Azure Synapse can unite your data, your developers, and your business users, with the only cloud analytics service that provides everything you need in a single solution. Azure Synapse puts the power of data analytics in your team's hands.



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.