

2023

From Big Data to Bigger Insights

The Future of Cloud Scale Analytics

Creative tech for Better Change



About Devoteam

Devoteam is a leading consulting firm focused on digital strategy, tech platforms and cybersecurity. By combining creativity, tech and data insights, we empower our customers to transform their business and unlock the future.

With 25 years' experience and more than 10,000 employees across Europe, the Middle East and Africa, Devoteam promotes responsible tech for people and works to create better change.

Creative tech for Better Change



From Big Data to bigger Insights

Contents

4

What: a modern data platform

5

Why: the benefits

7

How

- Ingestion 7
- Storage 8
- Processing 9
- Analysis 10

11

Steps

- Define the business goals and use cases 11
- Identify data sources 11
- Design the data architecture 12
- Choose the right analytics tools and technologies 12
- Ingest and process the data 12
- Visualize and report on the data 13
- Monitor and optimize 13

15

Conclusion

What

In today's data-driven world, organizations generate a massive amount of data that can be used to derive valuable insights and make informed decisions. However, the ability to analyze and process this data is often limited by the infrastructure and resources available.

Cloud-scale analytics is a framework for a modern data platform that enables organizations to analyze large volumes of data in real-time, while also providing the flexibility, scalability, and security required to meet the demands of modern data analytics. It provides a comprehensive set of tools and services for data analytics, including ingestion, storage, processing, and visualization. Cloud-scale analytics is designed to scale to meet the needs of modern data analytics, allowing organizations to store and process large amounts of data quickly and efficiently.

Cloud-scale analytics is built on top of Microsoft Azure, a leading cloud platform. By leveraging the power of the cloud, the cloud-scale analytics framework provides enterprises with the ability to analyze data at scale, while also providing the flexibility and agility required to meet the needs of modern data analytics.

Why

The most logical question to ask is: why should I invest in building a data analytics cloud platform? There are plenty of reasons, so let's focus on the most common challenges enterprises face nowadays.

First, cloud-scale analytics offers **cost-effective** solutions for data processing and analysis. Traditional data processing and analysis methods can be expensive and require significant capital investment, whereas cloud-scale analytics offers a pay-as-you-go model, allowing organizations to only pay for the resources they need.

Second, it offers scalability, flexibility, and agility, which are critical for organizations that need to rapidly adapt to changing business needs. It allows them to easily **scale their data infrastructure up or down** as needed, without incurring significant costs.

Third, it provides robust **security features and governance controls**, ensuring that data is protected at all times. This is especially important for organizations that handle sensitive data, such as financial or healthcare data.

Finally, the volume of data generated by enterprises is growing at an exponential rate, and traditional data processing and analysis methods are becoming inadequate. Cloud-scale analytics enables them to process and **analyze large volumes of data** quickly and efficiently, allowing them to gain insights and make data-driven decisions in real-time.

These challenges can slow down analytics efforts, making it difficult for enterprises to gain meaningful insights and stay competitive in their respective industries. However, with the rise of cloud-scale analytics, many of these challenges can be overcome, providing organizations with the tools and capabilities they need to succeed in the world of data analytics.



How

There are several tools available for cloud-scale analytics, each with its own pros and cons. These tools provide organizations with the ability to ingest, process, store, and analyze large volumes of data, enabling them to gain insights and make data-driven decisions.

Ingestion

Ingestion is the first step in the data analytics process, and several tools are available to help organizations ingest data at scale.

[Azure Data Factory](#) is a cloud-based data integration service that creates, schedules, and manages data pipelines. This service can be used to ingest data from a variety of sources, including SaaS applications, cloud data sources, and on-premises data sources, with a variety of tools, such as Self-Hosted Integration Runtime, allowing data ingestion through a private network.

Another tool for data ingestion is [Azure Event Hub/IoT Hub](#), which provides a scalable and secure platform for ingesting and processing high-volume, real-time event data from connected devices and applications. This service can handle millions of events per second and can be used for a variety of use cases, including telemetry processing, fraud detection, and predictive maintenance.

Storage

Your current analytical data architecture and operating model can include data warehouse, data lake, and data lakehouse structures, or even models like data fabric or data mesh. Each data model has its own merits and challenges. Cloud-scale analytics helps you work from your current setup to shift your approach to data management so that it can evolve with your infrastructure. That being said, various tools can be combined under one data architecture to enable cloud-scale analytics. Let's explore the most commonly used ones.

One of the best storage options is [Azure Blob Storage](#). Azure Blob Storage is highly scalable and can handle data ingestion at a massive scale. It also provides various tiers with different performance and cost characteristics, allowing you to optimize your data ingestion for your specific needs. It is the main storage option when building a data lake, but it can also serve as a landing zone for data warehousing.

Another important one is [Azure Synapse Analytics](#). It is an analytics service that brings together big data and data warehousing. It allows you to analyze and query large amounts of structured and unstructured data using a single workspace.

A different consideration is [Azure Cosmos DB](#), which is a globally distributed, multi-model database service that is designed for massive scale and low latency. It supports multiple data models, including document, key-value, graph, and column-family data models.

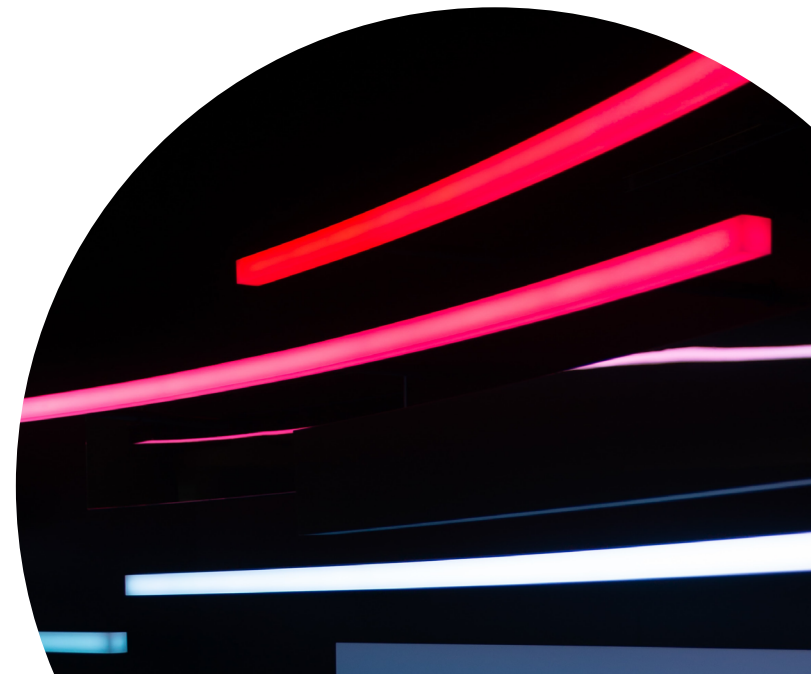
Finally, Azure offers many SQL options. One of the most popular ones is [Azure SQL Database](#), a fully managed relational database service. It provides high availability, automatic tuning, and intelligent optimization features that enable you to scale and optimize performance for your analytics workloads. It can serve as a serving layer for your analytics.

Processing

After the data is stored, it needs to be processed and transformed. Again [Azure Data Factory](#) is the usual suspect, as it can orchestrate an end-to-end ETL procedure. In the transformation phase, Azure Data Factory offers a wide range of options from built-in transformation blocks to custom-code transformations.

A tool that grows in popularity is [Azure Databricks](#), as it is a fast, easy, and collaborative Apache Spark-based analytics platform that can be used to process and analyze large volumes of data. This service provides tools to perform advanced analytics, including machine learning, streaming analytics, and graph processing.

And of course, the previously mentioned [Azure Synapse Analytics](#). Besides data ingestion, it allows the processing of large volumes of data from a variety of sources, including Azure Blob Storage and SQL Server. It also provides an integrated workspace for data preparation, data warehousing, and big data analytics.



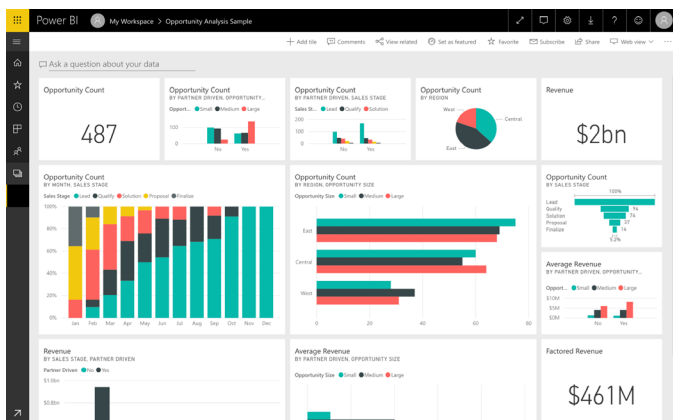
Analysis

Finally, the analysis and visualization stage. For the third time, [Azure Synapse Analytics](#) is a great option, this time for analysis, as it integrates with other Azure services such as Azure Machine Learning and Azure Data Factory, and can perform a range of data analysis tasks, including data exploration, data transformation, data modeling, and data visualization.

But what is [Azure Machine Learning](#)? It is a service that allows you to build, train, and deploy machine learning models on Azure. It provides a visual interface for designing models and supports popular programming languages like python and R.

For high-velocity, real-time data analytics Azure offers [Azure Stream Analytics](#), a service that processes streaming data from various sources like IoT devices, social media, and more. It allows you to query and analyze the data in real-time and supports integrations with various Azure services.

Finally, [Power BI](#) is a cloud-based business intelligence platform that enables data visualization and analysis. This tool allows organizations to create interactive dashboards and reports, making it easy to gain insights and make data-driven decisions.



Steps

There are several key steps involved in the process of building your data analytics platform. Each of these steps must be carefully designed and implemented to ensure that the platform is scalable, secure, and efficient.

Define the business goals and use cases

Clearly define your business goals and the specific use cases you want to address with your analytics solution. Identify the key business problems you want to solve, and the data required to solve them.

Identify data sources

Identify the data sources that you will use to gather data for analysis. These may include data from various sources such as on-premises databases, cloud-based storage, IoT devices, and third-party APIs.

“68% of organizations could demonstrate quantitatively that data-driven decision making has a positive impact on business.”

IDC '22

Design the data architecture

Design the data architecture for your analytics solution, which includes deciding on the storage technologies that will be used. Enterprises can have different needs, based on their size or applications. For example, for a small organization with limited resources, a centralized operations model mixed with some business subject matter experts might fit its situation. While for larger organizations with an autonomous business unit, a distributed operating model such as data mesh or data fabric might better address its needs. Microsoft sets the pillars for building a **well-architected platform**, taking into account factors such as reliability, security, cost optimization, operational excellence, and performance efficiency, to make the correct decisions.

Choose the right analytics tools and technologies

Choose the right analytics tools and technologies that will help you process and analyze the data. Cloud-based analytics tools such as Azure Data Factory, Azure Stream Analytics, Azure Synapse Analytics, and Azure Machine Learning can be used to ingest, transform, and analyze the data.

Ingest and process the data

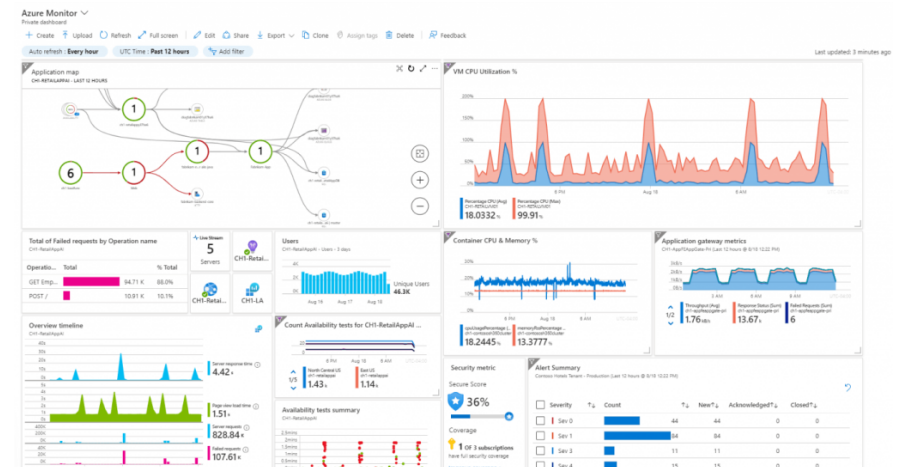
Once your data sources and analytics tools are in place, the data ingestion and processing can begin. This involves setting up pipelines and workflows to move data from the source to the destination and applying transformations and aggregations as necessary, with tools such as Azure Data Factory, Azure Databricks, or Azure Synapse Analytics.

Visualize and report on the data

Visualize the data and create reports that provide insights and recommendations to your stakeholders. Use tools such as Power BI to create dashboards and reports that are easy to understand and communicate.

Monitor and optimize

Continuously monitor and optimize your analytics solution to ensure that it is performing as expected. Azure offers built-in tools for monitoring, such as Azure Monitor.





Conclusion

Analytics has become a critical component of many modern businesses. As data continues to grow in volume, velocity, and variety, organizations need to be able to collect, store, process, and analyze it in a scalable and cost-effective manner. Cloud-scale analytics solutions enable them to unlock insights from their data and make informed decisions based on that data.

While the implementation can be complex, the benefits of such an environment are numerous. Enterprises can gain insights from their data that were previously impossible, make data-driven decisions, and remain competitive in today's data-driven economy.

In conclusion, cloud-scale analytics is a critical tool for modern businesses, and organizations that can effectively implement and utilize it can reap significant benefits. By carefully planning and executing a cloud-scale analytics strategy, they can unlock the power of their data and make informed decisions that drive business success.

Trusted Microsoft partner in EMEA



Sized for agility and scale

Our 1,200 Microsoft experts in the EMEA region are ready to transform your data into impact.

Build your cloud data foundations.

Unlock the power of your data.

Unleash your exclusive insights anywhere, anytime.

Our areas of expertise:

- Data strategy
- Data architecture
- Cloud native data platform
- Data migrations
- Data integration
- Data security and compliance

Our areas of expertise:

- Data governance
- Data Catalogue & Lineage
- Data quality
- Data democratization
- DataOps
- Master Data Management
- Data products

Our areas of expertise:

- Modern Visualization
- Self-Service BI
- Streaming analytics
- Artificial Intelligence

Georgios Gkenios

Senior Data Engineer

georgios.gkenios@devoteam.com



Stefan van Eijden

Business Lead Microsoft

stefan.van.eijden@devoteam.com



Ties van de Voort

Business Lead Microsoft

ties.van.de.voort@devoteam.com



Creative tech for Better Change