

Legacy modernization for exponential growth

A data lake infrastructure implementation on Azure Cloud

Industry

Travel & Logistics

Services

Mainframe & Legacy, Cloud Services

Technology

Azure Cloud, DevOps

Engagement

10 months

The challenge and the initial situation

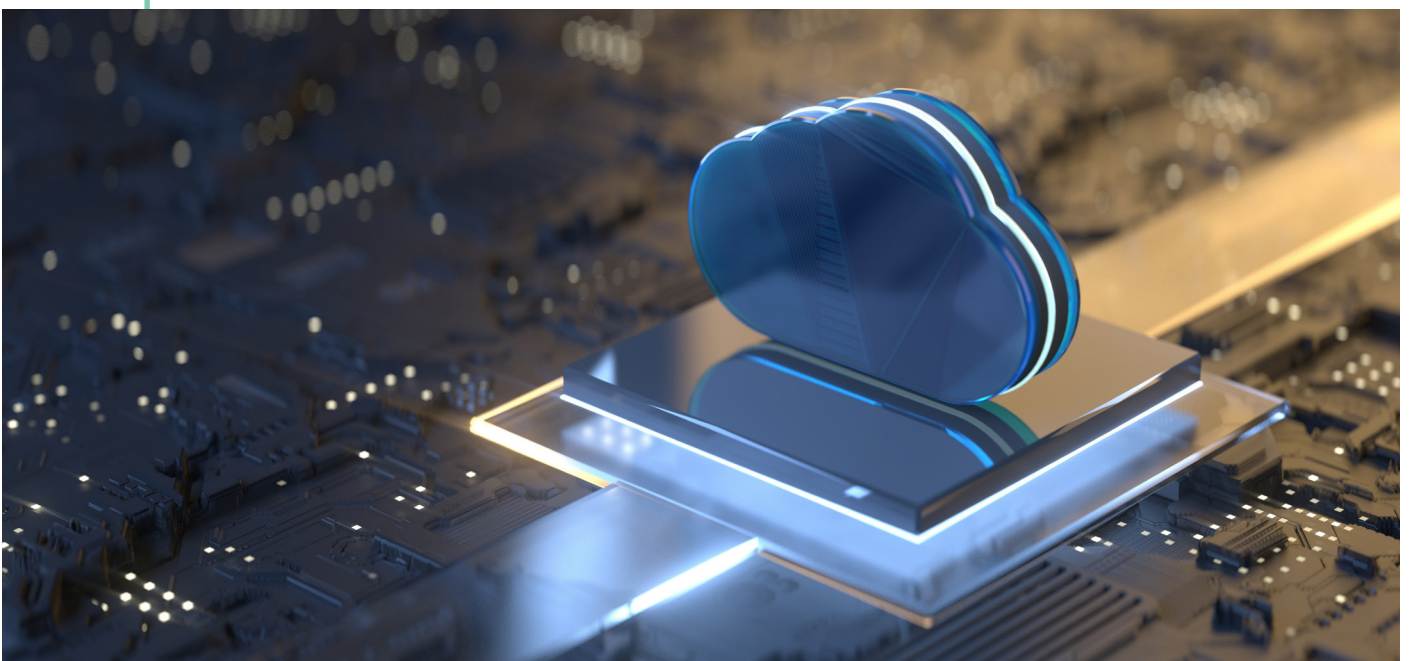
A client in the logistics domain faced challenges due to its size and the complexity of the services it offered customers, and the use of multiple databases and data processing systems. Due to historical growth, insufficient documentation, the database systems were difficult to maintain. These on-premise databases and tools are the foundation for the most critical operational systems such as delivery, logistics, address data management, etc. Additionally, with the rise in the popularity of online shopping, the logistical effort has also increased by leaps and bounds. More than 1 million packages and up to 20 million parcels are processed per day, and the number is rising. The data processing is distributed over several levels via BI (Business Intelligence) cubes. More than 200 stored procedures have been implemented for this purpose. The existing systems had reached its performance limits that could only be remedied with additional physical servers.

Database infrastructure July 2020:

6 database servers, an always-on Microsoft SQL Server, in parallel the operational and BI databases were located on the same server infrastructure (Track & Trace database: 9TB, Track & Trace replication database: 10 TB, BI database: 2.5 TB)

The client wanted to solve the following challenges in particular:

- **Limited server capacity:** Scaling-up was only possible by purchasing additional physical servers.
- **Instability due to high complexity:** Precompression processed the data amount with limited server capacity
- **Complexity of historical data:** Star schema has grown historically and become far too complex.
- **Lack of bandwidth:** Daily and Real Time Cube cannot run at the same time -RT Cube catches up with backlogs only with difficulty.
- **Limited monitoring:** The reporting services offered only limited visualization option





The Solution

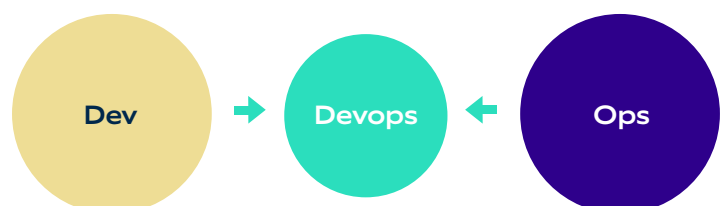
Nagarro recommended building an infrastructure for a Big Data Lake in Azure by a dedicated DevOps team in collaboration with the project team.

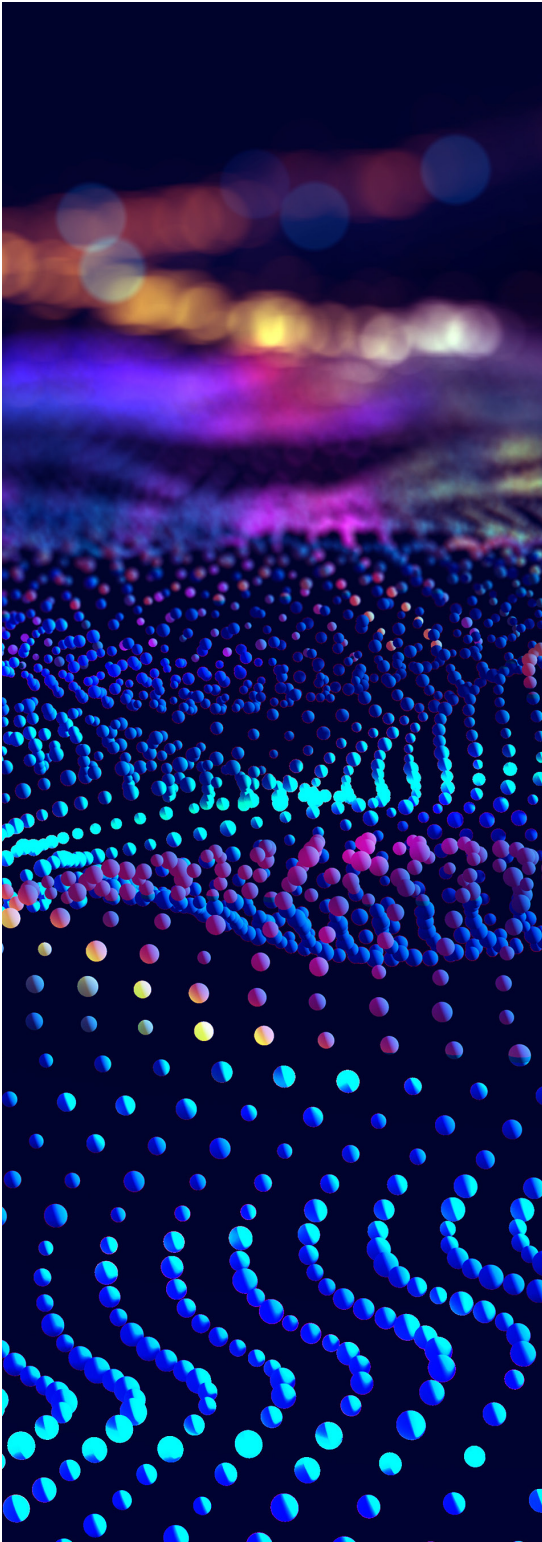
The setup for the collaboration with the project team arose from the existing team topology in the company.

A central DevOps team was responsible for supporting and enabling the project teams in the implementation of CI/CD pipelines and cloud infrastructure. They undertook the responsibility to standardize the CI/CD pipelines, a cloud-based resource deployment using Infrastructure as Code (IaC) and the standardization of a corresponding cloud architecture.

Azure technology was the first choice due to the company's Azure cloud-first approach. Nagarro used the Synapse Workspace components to build the data warehouse infrastructure. An Azure Data Factory instance was used to prepare the data. Scalability and security were key factors in the development of the infrastructure. Therefore, it was ensured from the very beginning that the infrastructure could be rolled out fully automatically in other Azure regions at the 'push of a button.'

A virtual network concept with corresponding subnets ensures that access to the systems is only possible for authorized resources or groups. The access rights have been tightened per level to ensure complete automation is 'enforced' at the infrastructure and application level. The diagnostics data is collected for the Azure Data Factory and the Synapse Workspace in a Log Analytics Workspace and processed and made available via dashboards.





Impact to Business

An overall success, the modernization led to significant improvements including:

- **Modernized approach:** The data is stored by leveraging Azure and the cloud potential.
- **Efficient data processing:** Data pipelines in the Azure Data Factory process new data in near real-time.
- **Standardization:** Internal systems provide standardized access via APIs.
- **Continuous Monitoring:** Reports can be prepared more attractively due to the centralized data. Due to the data lake structure, it is possible to generate many reports and integrate further reporting tools

About Nagarro

In a changing and evolving world, challenges are ever more unique and complex. Nagarro helps to transform, adapt, and build new ways into the future through a forward thinking, agile and CARING mindset. We excel at digital product engineering and deliver on our promise of thinking breakthroughs. Today, we are 12000 experts across 27 countries, forming a Nation of Nagarrians, ready to help our customers succeed. www.nagarro.com