

Ablöse von Legacy Systemen aufgrund exponentiell wachsender Anforderungen

Implementierung einer Data Lake Infrastruktur in der Azure Cloud

Industry

Travel & Logistics

Services

Mainframe & Legacy, Cloud Services

Technology

Azure Cloud, DevOps

Engagement

10 months

Die Herausforderung

Der Kunde betreibt aufgrund seiner Größe und Vielschichtigkeit der angebotenen Services für interne und externe Kunden eine Vielzahl an Datenbanken und Datenverarbeitungssystemen. Aufgrund des historischen Wachstums, als auch mangelhafter Dokumentation und dem daraus resultierenden Wissensverlust sind die bestehenden Datenbank-Systeme nur mehr schwer zu warten. Diese on-premise Anwendungen bilden allerdings die Basis für die wichtigsten operativen Systeme wie Zustellung, Logistik, Adressdatenverwaltung und vielem mehr. Zudem hat in den letzten Monaten der Onlinehandel stark zugenommen und damit ist auch der logistische Aufwand sprunghaft angestiegen. Mehr als 1 Million Pakete und bis zu 20 Millionen Parcelevents pro Tag sind zu verarbeiten, Tendenz steigend. Die Datenverarbeitung erfolgt über BI-Cubes auf mehreren Ebenen verteilt. Dafür wurden über 200 stored procedures implementiert. Die bestehenden Systeme gelangen an ihre Leistungsgrenze, welche nur mit dem Zukauf von zusätzlichen physischen Servern zu beheben wäre.

Datenbank Infrastruktur Juli 2020:

6 Datenbankserver, ein always-on Microsoft SQL Server, zudem befinden sich die operativen und BI-Datenbanken auf derselben Server-Infrastruktur (Track & Trace Datenbank: 9TB, Track & Trace Replikations-Datenbank: 10 TB, BI-Datenbank: 2,5 TB)

Der Kunde wollte vor allem folgende Herausforderungen lösen:

- **Eingeschränkte Serverkapazität** – Skalierung nur durch Einkauf zusätzlicher physischer Server möglich.
- **Vorverdichtung muss die Datenmenge mit eingeschränkter Serverleistung verarbeiten** – zusätzliche Instabilität durch hohe Komplexität.
- **Starschema ist historisch gewachsen und viel zu komplex geworden.**
- **Daily und Real Time-Cube können nicht gleichzeitig laufen** -RT-Cube holt Rückstände nur schwer auf.
- **Aktuelle Reporting Services** bieten nur eingeschränkte Visualisierungsmöglichkeiten.



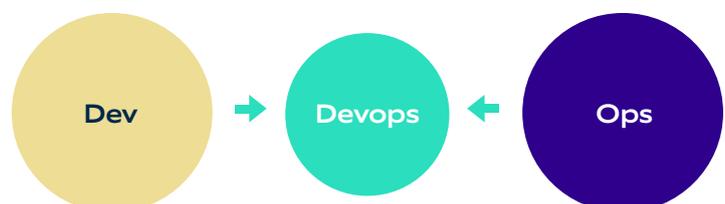


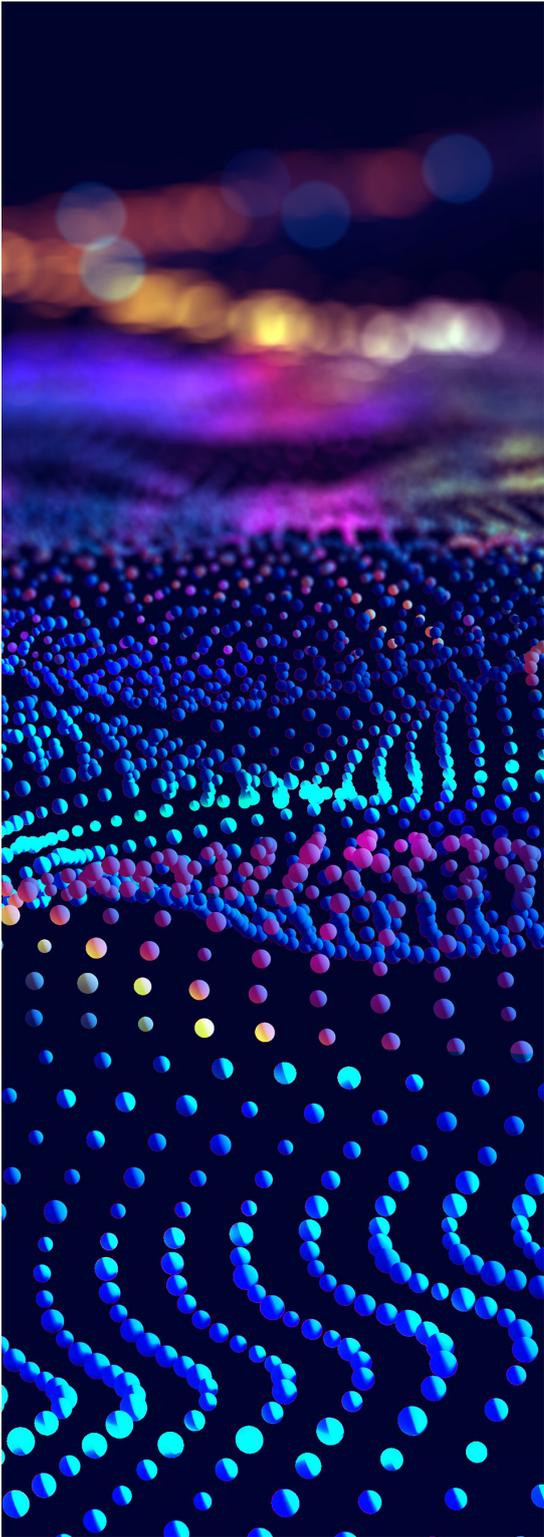
Die Lösung

Nagarro empfahl den Aufbau einer Infrastruktur für einen Big Data Lake in Azure durch ein dezidiertes DevOps Team in Zusammenarbeit mit dem Projekt-Team.

Das Setup für die Zusammenarbeit mit dem Projekt Team entstand aus der im Unternehmen bestehenden Team-Topologie. Ein zentrales DevOps Team war unter anderem für den Support und das Enabling der Projekt-Teams bei der Implementierung von CI/CD Pipelines und Cloud Infrastruktur zuständig. Sie kümmerten sich um die Standardisierung der CI/CD Pipelines, ein cloud-basierendes Ressourcen Deployment mittels Infrastructure as Code (IaC) und die Standardisierung einer entsprechenden Cloud Architektur. Aufgrund eines Azure Cloud-first Ansatzes des Unternehmens war Azure Technologie die erste Wahl. Nagarro nutzte die Komponenten Synapse Workspace, um die Data Warehouse Infrastruktur aufzubauen. Eine Azure Data Factory Instanz wurde für die Aufbereitung der Daten eingesetzt.

Skalierbarkeit und Sicherheit waren ein wesentlicher Faktor bei der Entwicklung der Infrastruktur. Daher wurde von Anfang an darauf geachtet, dass die Infrastruktur auf „Knopfdruck“ auch in weiteren Azure Regionen vollautomatisiert ausgerollt werden konnte. Ein virtuelles Netzwerk Konzept mit entsprechenden Subnetzen stellt sicher, dass der Zugriff auf die Systeme nur von berechtigten Ressourcen oder Gruppen möglich ist. Die Zugriffsrechte werden pro Stage verschärft, sodass eine Vollautomatisierung auf Infrastruktur- und Applikationsebene „erzwingen“ wird. Diagnostics Data wird für die Azure Data Factory und dem Synapse Workspace in einem Log Analytics Workspace gesammelt und mittels Dashboards aufbereitet und zur Verfügung gestellt.





Business Impact

Folgende maßgebliche Verbesserungen wurden erzielt:

- Modernisiertes Konzept, um Daten zu speichern, durch die Nutzung von Azure und dem Cloud Potenzial.
- Effizientere Datenverarbeitung: Datenpipelines in der Azure Data Factory verarbeiten neue Daten nahezu in Echtzeit.
- Interne Systeme bieten einen standardisierten Zugriff mittels API's.
- Reports können aufgrund der zentralisierten Daten attraktiver aufbereitet werden. Aufgrund der Data Lake Struktur ist es möglich eine Vielzahl an Reports zu generieren und weitere Reportingtools zu integrieren.

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