

# Data Ingestion and Connectivity

Data ingestion and connectivity present multiple business challenges. Integrating data from multiple sources, efficiently handling large volumes of data, ensuring data quality, and maintaining data security and compliance are significant hurdles. Moreover, technical challenges arise with incompatible data formats, complex data structures, and data silos.

Tackling these obstacles requires a holistic approach that includes advanced data integration tools, scalable data storage solutions, and robust data management processes. When done correctly, businesses can see improved data accuracy, make sound decisions, and see an uptick in overall business performance.



## The Backstory

We needed to unite a global life science company's legacy customer integrations with a new target architecture. This required a deep understanding of healthcare integrations and a modular approach to seamless partner onboarding.



## The Challenge

The global life science company partners with hospitals nationwide that use EMR systems like Epic and Cerner. Our team had to create a generic data ingestion platform that connects these systems and processes, masters, and transforms data to the company's standards in hours. Another hurdle was migrating 60 petabytes of data and generating an additional 2+ petabytes of data every month.

## The Solution

To onboard more than 50 partners onto a common data platform, we implemented a transfer service mechanism using Kubernetes node pools and Google Cloud Composer orchestration. Cloud Composer's compatibility with STS enabled us to maintain managed workflows for data transfer and maintain data stores during migration. In addition, we used Google Cloud Dataproc to process, enrich, and transform incoming data.

## The Results

Our redesign of the architecture for workload handling on Google Cloud led to a 20% reduction in customer onboarding time, a doubling of workload capacity, and the ability to support up to five times the current growth by 2025. Additionally, this approach has led to a 40% reduction in engineering effort required for data pipeline maintenance, resulting in significant savings.

The logo for egen, featuring the word "egen" in a white, lowercase, sans-serif font, with a small superscript "n". The logo is positioned on an orange square background that is part of a larger geometric design consisting of overlapping light blue and dark blue shapes.