



BREAKING DATA SILOS:

# 10 Success Stories with Starburst

---

The Modern Data Lake Analytics Platform

# Contents

<b>Optum</b> accelerates queries by 10X . . . . .	5
<b>Priceline</b> bolsters its recommendation engine and reduces data storage costs by 5X-10X . . . . .	6
<b>EMIS</b> empowers clinicians and researchers with unprecedented data analytics services . . . . .	7
<b>Genus</b> advances animal genetics and achieves 75% faster time-to-insight. . . . .	8
<b>Novant Health</b> becomes data-driven with faster performance and timely access. . . . .	9
<b>Zalando</b> Powers Data Analytics And Reduces Costs With Starburst And AWS . . . . .	10
<b>SOPHiA GENETICS</b> tightens access controls and accelerates business metrics . . . . .	11
<b>Global investment bank</b> accelerates money laundering detection with real-time analytics . . . . .	12
<b>Telecommunications giant</b> unlocks new revenue streams with a 360-degree customer view . . . . .	13
<b>Online food ordering leader</b> experiences 10X to 15X query performance improvement . . . . .	14
Add Starburst to your data arsenal . . . . .	15



Data silos continue to be a challenge, despite efforts to create a “single source of truth.”

Research from BCG found that in [more than 50% of data leaders](#), architectural complexity is a major pain point — leading companies to get lost in mountains of data, complexity, and costs.

Traditional data lakes and data warehouses attempted to address this problem by requiring all data to be landed in their proprietary systems and forcing you to access everything through there. However, because most of these systems are monolithic architectures and are built on a closed ecosystem, they are difficult to scale as your organization and your needs grow.

A modern data lake provides an innovative solution to these silos, delivering benefits like:

- Embracing open file and table formats
  - Single point of access and governance for all data in and around the lake
  - Advanced warehouse-like capabilities
  - Vendor agnostic
  - Scalable and cost-effective
- Historically, building and managing a modern data lake can be time and resource intensive.





## This is where Starburst enters the picture.

For data-driven companies, Starburst offers a full-featured data lake analytics platform, built on open source Trino. Our platform includes the capabilities needed to discover, organize, and consume data without the need for time-consuming and costly migrations.

We believe the lake should be the center of gravity for data, but it should support performant data federation to data outside the lake when needed.

With Starburst, teams can access more complete data, lower the cost of infrastructure, use tools best suited to their needs, and avoid vendor lock-in. And we have the customer stories to prove it. Keep reading to learn about 10 companies that experienced these exact results by reinforcing their analytics with Starburst.

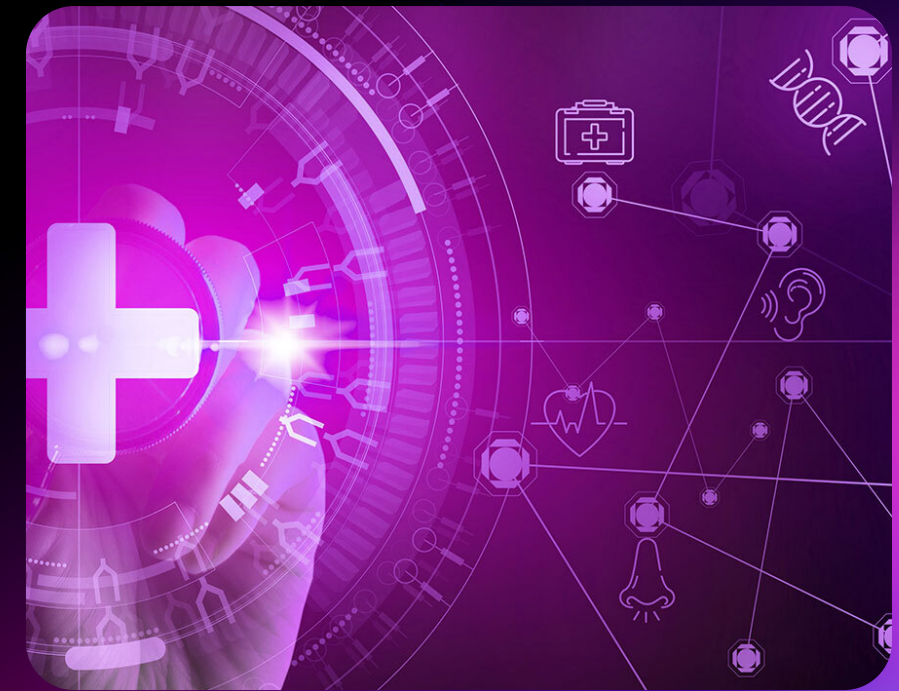


# Optum

## Optum accelerates queries by 10X

### Overview

Information technology service provider Optum is dedicated to shaping a healthcare system that gives patients a complete view of their health, providing them with personalized insights that lead to improved outcomes.



“Our data lake backbone was on a traditional Hadoop infrastructure. While that approach had its day, it’s not flexible. We needed to scale out and separate our compute from our storage without moving the data.”



Mike Prior, Principal IO Engineer

### Challenge & Solution

Optum’s data warehouse solution, Hive, could not support the growing demand for Optum’s analytics systems. This led to a poor end-user experience and an inability to bring on new workloads. To solve this challenge, Optum replaced Hive with Starburst deployed on Hadoop. “Providing users with one endpoint is so much easier. They can use the same familiar tools, but everything is happening faster,” shares Prior.

### What Starburst Delivered

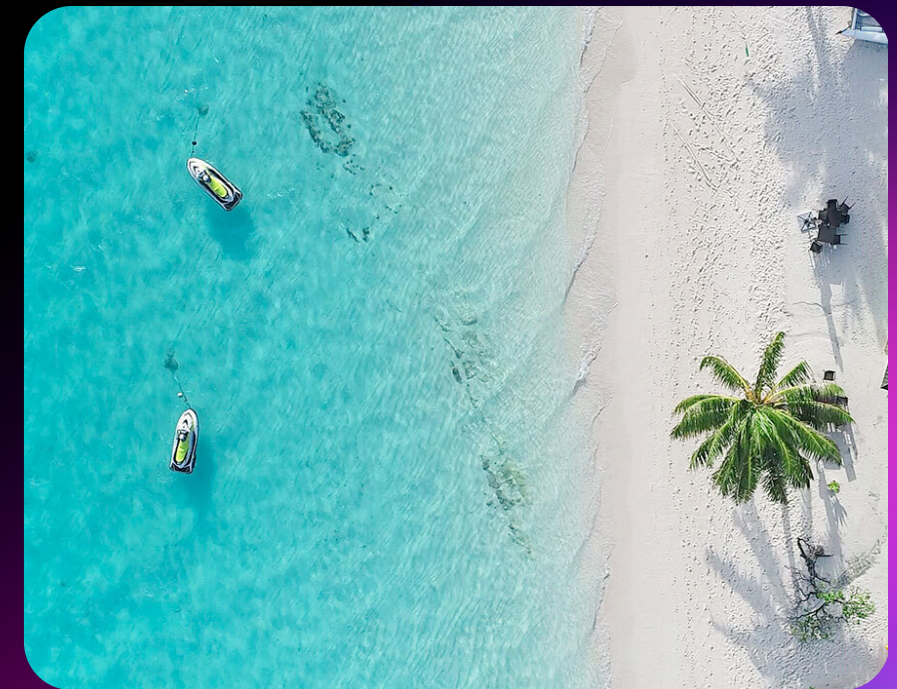
- 10X faster queries
- 30% reduction in infrastructure costs
- \$8 million in projected savings



# Priceline bolsters its recommendation engine and reduces data storage costs by 5X-10X

## Overview

Priceline is part of the world's largest travel company, Booking.com. Though still a small company with under 1000 employees, Priceline makes a big impact, helping consumers save over \$1B a year on travel by letting customers name their own price for empty hotel rooms and airplane seats.



“One of the key reasons we are on this journey is to democratize data. By using Starburst, we can make the data available for all users and, then, make more informed decisions with data sets that we never explored in the past. We can see the hidden data and leverage it to provide an even better customer experience.”



Sachin Gopalakrishna Menon,  
Senior Director of Data

## Challenge & Solution

Priceline launched a major data transformation initiative to leverage vast amounts of streaming and historical datasets stored across different cloud and on-prem systems. Most of Priceline's data processing was driven by a traditional ETL approach, which caused a 1-2 day delay in getting answers to their questions.

Priceline selected Starburst Enterprise to democratize data access and give its users a high-performance, cost-effective way to transform all its raw data (directly from its data lake) into novel insights.

## What Starburst Delivered

- 5X to 10X lower storage costs
- Near real-time access to data rather than 1-2 days
- Improved customer satisfaction with a more personalized booking experience and more accurate recommendations



# Healthcare technology leader, EMIS, empowers clinicians and researchers with unprecedented data analytics services

## Overview

EMIS is one of the largest suppliers of electronic health records software in the UK. The company was founded more than 30 years ago, but its mission has remained the same — to develop the best possible technology to help healthcare professionals and ultimately, to improve patient health.



“To get this kind of information in the past would have involved several weeks of copying and analyzing data. With EMIS-X Analytics, powered by Starburst Enterprise, researchers get the information in almost real-time.”



Richard Jarvis,  
Chief Technology Officer

## Challenge & Solution

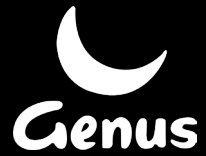
EMIS needed to provide faster access to patient health data for analysis in a secure and compliant way.

Clinical data is continuously streamed into EMIS’ cloud data lake, ensuring that it includes the very latest information, including patient medications, allergies, symptoms and more. To underpin its transformative analytics suite, EMIS-X Analytics, EMIS chose Starburst as its data lake analytics platform.

## What Starburst Delivered

- Join data simultaneously for EMIS-X Analytics to perform at scale
- Data discovery reduced from 48 hours to less than four
- Pharmaceutical companies able to fill clinical trials faster





# Genus advances animal genetics and achieves 75% faster time-to-insight

## Overview

Genus researches and develops innovative technologies that support a more sustainable food system while helping farmers with quality genetics to help them produce animal protein and meet the growing global demand for food.



“With Starburst, we have accelerated data discovery, simplified data pipelines, and have a unified query layer across all data sources. These three points are critical to what we do.”



Patrice Linel, Senior Manager of Data Science & Data Engineering

## Challenge & Solution

With specialized databases for genetic information, a hybrid and multi-cloud data platform, and a spectrum of legacy databases for the various functions of the business, Genus' data engineers were burdened with managing complex ETL pipelines that took weeks to run.

Genus deployed Starburst Enterprise to establish inter-connectivity of datasets and speed up data discovery by querying directly from its data lakes (Amazon S3 and Azure Data Lake Storage).

## What Starburst Delivered

- Accelerated animal genetic improvement through data discovery
- 75% faster time-to-insight
- Up to 150X faster analytical queries, data product creation, and data product validation



# Healthcare leader Novant Health becomes data-driven with faster performance and timely access

## Overview

Novant Health is a network of physician practices, hospitals, outpatient, diagnostic imaging centers, and other medical facilities. In addition to employing world-class physicians and caregivers, the organization relies on advanced technology solutions to optimize patient care and the overall patient experience.



“We needed to point our customers directly to data where it lives... and give them the capability to query that data and ultimately deliver insights for the organization. Starburst hits the sweet spot for us because it delivers exactly what we need, and it’s built on proven technology that has been out there in the market and tested at large companies.”



Justin Byrd, VP of Data Platform & Integration

## Challenge & Solution

Novant Health ingests and stores a variety of data relating to patient health, as well as clinical and financial information and more, in multiple systems, such as Azure Data Lake Storage. The organization needed to be able to extract insights from its distributed data to enhance the patient experience and drive better business outcomes.

Novant Health deployed Starburst Enterprise to function as an abstraction layer between its varied data sources and the analysts, data scientists, and other end users.

## What Starburst Delivered

- 60% faster access to data
- 7 distributed datasets accessible through one tool
- Curated data sets are available to end users in minutes instead of months

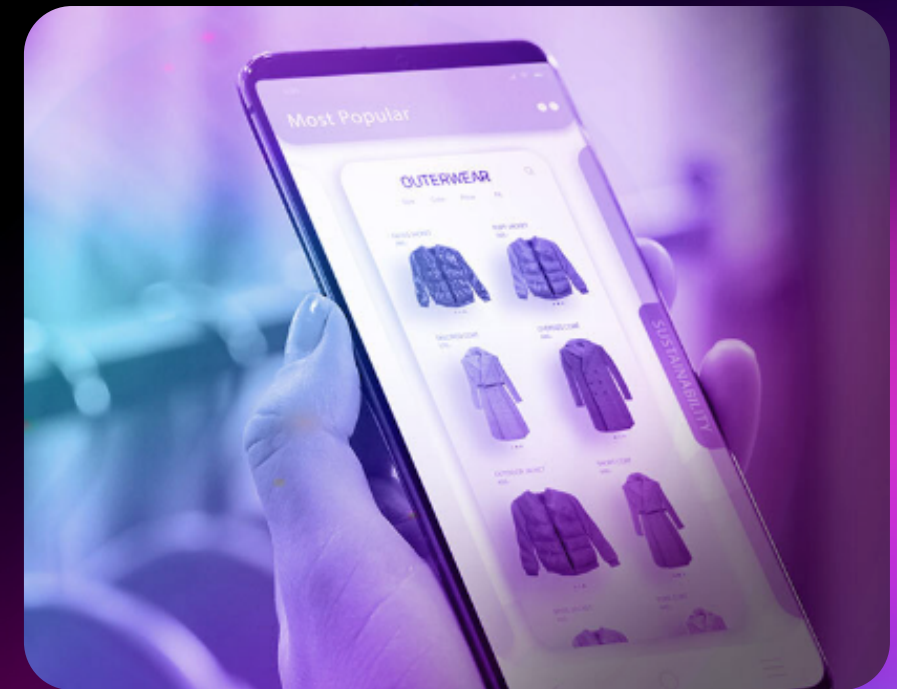




# Zalando powers data analytics and reduces costs with Starburst and AWS

## Overview

Zalando SE, Europe's leading online platform for fashion and lifestyle, boasts \$5B in annual revenue, largely from selling brand-name clothing and footwear online.



“The decision to deploy Starburst Enterprise was made simpler because it has proven to be a reliable, fast, and stable query engine for S3 data lakes.”



Alberto Miorin,  
Engineering Lead

## Challenge & Solution

When transitioning from a legacy data warehouse to an AWS cloud data lake, Zalando needed to give its buyers and business analysts a more efficient way to extract value from this distributed data to launch its Customer 360 program.

The company's initial choice was Trino, but the open source deployment was missing key enterprise features, including security and support. Ultimately, Zalando chose to launch a Customer 360 program with Starburst powering its analytics engine on Amazon S3.

## What Starburst Delivered

- 50% reduction in infrastructure costs
- Improved buyer recommendations
- Compliance with GDPR requirements
- A complete Customer 360 program that improved internal success metrics like increased % of wallet



# SOPHiA GENETICS tightens access controls and accelerates business metrics

## Overview

Working with more than 780 hospitals and research institutions in over 70 countries, SOPHiA GENETICS enables its customers to outsource their bioinformatics operations by providing them with both a cloud-based, Software-as-a-Service analytics platform and unprecedented insights from the global network. This way, SOPHiA GENETICS' customers can focus on what they do best — advancing research, treatment decisions, and drug development efforts.



“One of the core missions of my team is to make the data mesh happen while still maintaining everything that we need to maintain in terms of policies and data privacy constraints. Starburst is making my life a lot easier by creating the first mesh platform for business metrics, that we can start operating within.”



Alexander Seeholzer,  
Director, Data Services

## Challenge & Solution

SOPHiA GENETICS advances data-driven medicine through a pioneering global network of healthcare institutions. Accessing highly sensitive customer data cross-regionally, while maintaining HIPAA and GDPR compliance, had become increasingly difficult with the ever-changing regulatory environment.

Starburst replaced the incumbent solution as the analytics platform for SOPHiA's decentralized data mesh architecture. The Starburst Stargate feature provides global data access while maintaining security and compliance of PII data. Seamless access data between its cloud data lake and cloud data warehouse has increased data sharing exponentially. This further enables the discovery and evaluation of new models of healthcare analytics at SOPHiA GENETICS.

## What Starburst Delivered

- 900% increase in users accessing production data
- 10-15% greater data availability
- A single view into globally-dispersed data sets

# Global investment bank accelerates money laundering detection with real-time analytics

## Overview

As one of the largest banking and financial services corporations in the world, this company's mission is to make its client's lives better with modern financial solutions to contribute to a growing economy. The multinational investment bank serves millions of people and institutions in over 100 countries.



"I've talked to a lot of vendors over the last 18 years, and nothing came close to what Starburst can do. This is something we had been looking for some time and I'm very happy that we found Starburst."



Anonymous,  
Head of AML Technology

## Challenge & Solution

This global financial institution had limited data access due to global data sovereignty requirements. This led to delays in cleansing and transferring large data sets, incomplete or sparse anti-money laundering (AML) analytics, and increased operations costs due to data duplication.

The bank needed to modernize its data lake, based on Hadoop, to improve the performance and speed of analytics. By replacing Hive with Starburst, the bank can now access data from anywhere without needing to duplicate or move it. This provides them with timely access to data, even from countries with strict data sovereignty rules.

## What Starburst Delivered

- Millions of dollars saved in averted fines
- Instant access to analytics across 100+ countries
- Petabytes of data queried where it resides



# Telecommunications giant unlocks new revenue with a 360-degree customer view

## Overview

This telecommunications company is the leading pay-TV and cable TV company in the United States. With more than 15 million subscribers, this customer retains and ingests tremendous volumes of data across various platforms.



“When end users are going into on-prem or cloud environments, they will be presented with all the data sets they have access to, irrespective of where the data is located. This offers a huge value to our end users.”



Anonymous, Director of Software and Engineering

## Challenge & Solution

The CMO wanted to run campaigns to upsell the existing subscriber base. With the billing data in Teradata and user data in a large Hadoop Cluster, being able to query the data to run this sales campaign would be impossible.

This telecommunications giant selected Starburst to migrate its data off Teradata to Amazon S3 and establish universal data access. With Starburst, end users are able to easily run queries against different data sources.

## What Starburst Delivered

- >\$200 million in new subscription revenue
- >250 TB of data ingested daily
- Reduced time to insight from 18 months to 5 weeks

# Online food ordering leader experiences 10X to 15X query performance improvement

## Overview

This online food ordering and delivery company is one of the largest in the U.S. with over 20 million customers. Its platform connects customers with hundreds of thousands of merchants to place online orders, have their purchases delivered directly to their door, and track orders through a web application.



“With Starburst, the future of data analytics looks bright. We hope to achieve our vision of a unified query engine and a secure single point of access to all of our data with Starburst Enterprise.”



Anonymous, Data Platform  
Engineering Manager

## Challenge & Solution

With customer order data residing in both Databricks and Snowflake, the analytics team lacked the ability to easily access both data sets simultaneously. Workarounds led to higher TCO and slower time to insight.

The company deployed Starburst as the single point of access to process over 10,000 queries and 250TB of distributed data daily. Of the company’s analytics workloads, 95% are served through Starburst.

## What Starburst Delivered

- 10-15x query runtime improvement in Amazon S3
- Significantly lower TCO
- Faster time-to-insight with real-time reporting and visualizations





## Add Starburst to your data arsenal

The challenges posed by data silos and architectural complexity continue to plague organizations in their quest for a reliable data lake analytics platform. While traditional data lakes and data warehouses offer promising solutions, the difficulties associated with being locked into a proprietary data ecosystem limits customers from using the architecture that meets their specific business needs.

Starburst bridges the gap between data lakes and advanced analytics by enabling companies across all industries to activate the data within and around their data lake. Starburst offers the easiest way to build and manage your modern data lake with the following capabilities:

- Deploy anywhere
- Ecosystem of connectors
- Scalability
- Performance and flexibility
- Resource elasticity
- Security and compliance
- Domain expertise

The modern data lake is here.  
Visit [starburst.io](https://starburst.io) to get started.

