

# DELIVERING PRECISION MEDICINE WITH MACHINE LEARNING

A graphic showing the number '2X' in a large, blue, outlined font, indicating a doubling of model accuracy.

Model accuracy

## Industry

– Healthcare

## Locations

Headquarters: Chicago, Illinois, USA

“We view quality as a promise and seek to provide the best care and best outcomes. For us, the data infrastructure and tools we’re building drive towards improving and enhancing the quality of the care we provide.”

Dr. Bala Hota, Chief Analytics Officer,  
Rush University Medical Center

Rush University Medical Center, located in Chicago, Illinois, is a leading U.S. academic health system.

## Challenge

“To be data-driven is not an option anymore; it is a requirement,” said Jawad Khan, director of Knowledge Management and Data Sciences at Rush University Medical Center. “Data empowers us to understand the specific gene makeup of a given patient and what care path is best.”

However, Rush needed a modern data platform to gain deeper, and faster, insights into patient care and healthcare delivery. “We don’t believe that we can do all the projects in our pipeline without a big data platform that can support new types of data, along with the growing volume and high velocity of data we want to analyze,” said Khan.

## Solution

Rush University Medical Center is working with Cloudera and MetiStream to implement machine learning and deploy analytics-as-a-service to hospital administrators, clinicians, and researchers. With this platform, which runs on Microsoft Azure, Rush is eliminating data silos and can analyze a much wider range of data in near-real time—from genomic data to medical device data to unstructured clinical notes and text data. In fact, with Cloudera and MetiStream, Rush ingested years of clinical notes into the platform in only 36 hours.

“We started on this path for several reasons—genomics and precision medicine is a big one,” said Dr. Bala Hota, chief analytics officer at Rush University Medical Center. “Unstructured data is another. Tapping into the content of clinical notes—what doctors have written and what they mean—is also crucial. We found that adding clinical notes to models doubled the accuracy of the models.”

Added Khan, “We can write models on a larger set of data, and machine learning enables us to see exactly where the variations are and design our care paths along those variations,” said Khan. “Before we were waiting 24 to 48 hours to collect, process, and analyze data. But our platform now will enable us to stream the data in real time from any source, process it, and provide analysis on it in an almost-real-time basis.”

### Solution highlights

- Modern Data Platform: Cloudera Enterprise
- Workloads: Analytic Database, Data Science, Data Engineering, Operational Database
- Key Components: Apache Spark™, Apache Impala®, Cloudera Data Science Workbench, Cloudera Director, Kerberos
- Healthcare Analytics Platform: MetiStream Ember
- Cloud Platform: Microsoft Azure

### Applications supported

- Precision medicine

### Data sources

- Electronic medical records
- Clinical applications
- Medical devices
- Genomics

### Impact

- Delivers 200 KPIs daily to administrators for more data-driven decision making
- Doubled accuracy of machine learning models
- Ingested years of clinical notes in only 36 hours
- Will enable near-real-time analysis of data
- Positions organization to deliver precision medicine and customized treatment plans

## Implementation

As Rush implements its platform, elasticity is crucial. “We had a small budget, but we had the desire to do it,” said Dr. Hota. “Cloudera on Microsoft Azure has been a perfect solution. We can quickly spin up and down resources as our data processing needs change and we can load huge volumes of data in days that would have taken weeks on prem. Additionally, Cloudera Data Science Workbench makes the development of new models easier for our data scientists.”

Added Khan, “Cloudera Director made it easy to deploy machine learning and analytics-as-a service in the cloud with full security.”

Rush engaged MetiStream to help automate the process of extracting, processing, and applying natural language processing (NLP) to the unstructured clinical text. By combining a variety of data sets—from electronic health records, genomic, imaging, and more—they can improve care, identify errors, and operationalize analytic models and insights in days versus months. “MetiStream has been a great partner, providing frameworks and self-service tools for using big data and machine learning,” added Dr. Hota. “Their platform, Ember, will help us bring new tools to end users.”

## Results

Today, Rush’s leaders can now easily monitor more than 200 key performance indicators (KPIs) and benchmarks via executive dashboards for more data-driven decision-making.

“We view quality as a promise and seek to provide the best care and best outcomes,” said Dr. Hota. “For us, the data infrastructure and tools we’re building drive towards improving and enhancing the quality of the care we provide.”

The organization is also developing a wide range of machine learning models to improve healthcare delivery—from identifying the impact of care path deviations to improving emergency department workflows to ensuring accurate documentation in medical records for patient care, billing, and regulatory activities.

Ultimately, researchers and clinicians will be able to gain new insights that help them better customize patient treatment plans and deliver precision medicine. Rush plans to integrate these new insights into its Epic electronic medical record system to provide care management teams actionable information.

“As we build these algorithms and prediction models, we want to ensure they end up in someone’s work queue—a clinician, a registration clerk for example—so they can act on it,” said Dr. Hota.