LTIMindtree

Grunt work out. $1.5 Mn savings in, with Scintilla

SAS to PySpark Migration at work
for global healthcare player

About the client

Our client, one of largest integrated healthcare (payer and provider) organisation , is a leading provider of healthcare technology services, specializing in digital health solutions that enhance patient engagement and optimize healthcare operations.

They leverage advanced technologies to offer innovative, data-driven services, assisting healthcare organizations in improving outcomes and operational efficiency.

Challenge

The client faced several challenges with their Healthcare Effectiveness and Data Information Set (HEDIS) business application that was based on a legacy statistical analysis system (SAS). These challenges fuelled their need for an enterprise-wide cloud transformation journey and migration from SAS to PySpark, a more modern, cos-effective and scalable data processing system. One of the critical pieces of this modernization program is migration of ETL logic from SAS ETL scripts to PySpark Their challenges included:

- **Need for performance optimization** as their legacy SAS system was inefficient.
- **Needed a more cost-effective system** with open standards and easier application hosting.
- **Need for system scalability** to handle increasing volumes of data and user traffic.
- **Burden of high maintenance costs of SAS scripts** that were difficult to update.
- **Lack of collaboration** capabilities in their development environment.
- **Need to convert over 10,000 legacy SAS scripts** and more than 500K lines of code, into production-ready PySpark Notebooks in Databricks.
- **Approximately 55% of the scripts were highly complex,** adding another layer of difficulty.
- **Orchestrating the volume of jobs** based on interdependencies and managing versioning and continuous integration for ongoing changes.
- The entire healthcare industry faced an unprecedented cyber security event that disrupted the entire care delivery system and took the client's New York HEDIS system out of action putting their Star ratings at risk.
- **Performance dip** handling high volume data fetch and transformations.
- **Lack of version control** mechanism.
- **Lack of governance** and observability.
- **High license tool cost.**
- **Multiple tools** and tech in platform for ETL purpose.

Why are enterprises migrating from SAS to PySpark?

There is a shift underway across industries with many enterprises migrating from SAS to PySpark because of PySpark's advantages including:

- Cost-effectiveness
- Open standards and formats
- Simplified application hosting
- Rapid implementation and time-to-value
- Scalability
- Better storage
- Disaster recovery
- Collaboration
- 99% uptime
- Agility
- Ease of integration
- Security
- Less maintenance
- Task automation
- Flexibility
- Mobility

LTIMindtree's solution

We utilized LTIMindtree's Scintilla automation product along with several unique and innovative approaches for a swift and successful cloud migration of the client's HEDIS business application from SAS to PySpark, including:

Automated code conversion using Scintilla: This automation converted the majority of the existing SAS code to PySpark. It saved time, manual effort and reduced the risk of errors. Code analysis using Scintilla identified areas of complexity, inefficiency, and potential bottlenecks.

Swift migration: Migrated 75% of the existing SAS scripts to PySpark within 6 months using Scintilla with some manual effort for the rest of the scripts, which was 30% faster than expected.

Containerization: Ensured that the new PySpark code was portable, scalable, and easy to deploy.

Cloud-based development tools: Enabled collaboration, version control, and continuous integration.

Data-driven decision-making: This approach ensured that the new PySpark code met the required standards.

PySpark training: Ensured the team had a solid understanding of the new technology and its capabilities.

Code refactoring: Refactored the existing SAS code to make it more modular, efficient, and scalable.

Unit testing: Ensured that the new PySpark code was functioning correctly and met requisite standards.

Integration testing: Ensured that the new PySpark code worked seamlessly with the existing system.

Performance optimization: Optimized the new PySpark code's performance to ensure that it met the required performance standards.

Collaboration and communication: The team worked closely together to ensure that everyone was aligned and aware of the progress, challenges, and changes.

Tech stack

During the cloud migration project for the SAS to PySpark migration, we utilized PySpark, Databricks, Python, Jupyter Notebooks, Git, Docker and LTIMindtree's Scintalla—an SAS code conversion to PySpark including the lineage for all SAS scripts.

Business benefits

Our cloud migration and modernization using Scintilla led to several benefits for the client:

- **$1.5 million** saved annually by reduction in TCO.

- Reduced the HEDIS application's total execution cycle from **60 hours to ~15 hours per month,** giving the client time to validate data and proceed with next steps in a timely manner.

- Retained nearly **$800 million** in Star's revenue across all their geographical locations (PA, WV, DE, and Western NY).

- **Helped the client's NY Health Maintenance Organization (HMO) achieve a rating of 4.5 Stars,** an accomplishment achieved only once in the past 7 years.

- **60%** increase in scalability allowing the business to process larger datasets and make more informed decisions.

- **25%** reduction in maintenance costs due to the improved efficiency and scalability of the new code.

- **Enhanced data quality** due to implementation of data validation and data cleansing techniques, resulting in more accurate and reliable data.

- **Increased flexibility and collaboration** due to a cloud-based development environment, allowing team members to work from anywhere.

- **Improved performance and efficiency** due to the optimized PySpark code.

- **Better decision-making** by using data-driven insights, resulting in improved business outcomes and increased competitiveness.

- **Increased customer satisfaction** by delivering high-quality data and insights, resulting in improved customer loyalty and retention.

Conclusion

In the healthcare sector, there is shift towards utilizing the cloud for data-driven decision-making underway. This SAS to PySpark migration provides a wide range of benefits for healthcare enterprises and the customers they serve. Leveraging Databricks to usher in a new era of cloud-based healthcare will be a key differentiator for modern healthcare enterprises.

Testimonial

" As the healthcare industry continues to evolve, we're seeing a significant shift towards data-driven decision making. With the increasing complexity of patient data, regulatory requirements, and the need for personalized care, healthcare organizations are looking for ways to modernize their analytics infrastructure. Our experience with migrating from SAS to Databricks on cloud has been transformative. By leveraging the scalability and flexibility of the cloud, we've been able to unlock new insights from our data, improve patient outcomes, and reduce costs. "

-Head of enterprise data analytics

- -

" The Databricks platform has enabled us to streamline our analytics workflows, integrate disparate data sources, and deploy machine learning models at scale. The results have been remarkable–we've seen significant improvements in patient engagement, reduced readmissions, and enhanced operational efficiency.

As the healthcare industry continues to navigate the challenges of data growth, regulatory compliance, and innovation, we believe that cloud-based analytics platforms like Databricks will play a critical role in shaping the future of healthcare. Our migration from SAS to Databricks has been a key enabler of our digital transformation, and we're excited to see the impact it will have on our patients and our business. "

-Director of data engineering

Please email us at **Databricks.SBU@ltimindtree.com** to get in touch with us.

LTIMindtree