

CASE STUDY

Areeba Accelerates ML Model Creation with CData Virtuality

Real-time data access

Data is always available in one place, allowing Areeba to build views and aggregations for predictive analytics.

Eliminated down time

Data scientists no longer need to wait for analysts to provide the data they need.

Automated processes

Data is always up-to-date and ready to be used, thanks to scheduled automated processes.



Company

Areeba

Industry

Financial Services

Product

CData Virtuality

Country

Lebanon

Snapshot

CData Virtuality helped Areeba create a virtual data access layer to enable data architects and data scientists to connect data across multiple sources to enable a robust machine learning model.

An agile and easy-to-use data virtualization solution enabled Areeba to take its machine learning efforts to the next level.

Areeba is a leading Lebanese financial technology company that provides fast and innovative payment solutions for banks, merchants, and governments in the Middle East region. They are committed to investing in new capabilities and technologies that deliver enhanced payment experiences and more secure solutions, from biometric cards and mobile payments to SmartPOS.

Areeba is always striving for cutting-edge solutions and has started applying data science rules to transactional and operational data to improve the performance of its services. Areeba data scientists use machine learning (ML) models like time series, regression, decision tree, random forest, and clustering algorithms to learn more about customer and merchant segmentations, merchant performance, and churn. Areeba stumbled upon several challenges before finding the best data management solution to enable innovation but finally overcame these roadblocks with CData Virtuality.

The challenge: CSV chaos from scattered data

Areeba deals with huge amounts of data to apply ML models to transactional and operational data. The data comes from different sources, such as MemSQL, MariaDB, and Oracle, making it difficult to access and consolidate the information for actionable insights.

Like many data scientists, the Areeba team first tried to collect the needed data from multiple sources and in different formats, such as CSV and text files. The data was then used to build predictions and models using languages such as Python, R, or Scala. However, this process was very time-consuming, troublesome, and error-prone. They began to see that this approach would soon cause challenges for real-time analytics.

“We wished we had known CData Virtuality from the beginning when we started with our ML processes. This could have saved us a lot of time and energy. CData Virtuality is now a very essential part of our daily life as a data scientist. It helps us to capture, mix, and consume data from different sources very easily. Thereby we can save time and focus more on the end result. Exciting times ahead!”

– Khaled Eid, Data Scientist, Areeba

The solution: High-speed data access at scale

Data is the main ingredient from which ML algorithms are trained. If Areeba’s data ecosystem is not properly managed and connected, its ML strategies are rendered ineffective. Knowing that its traditional data warehousing system was inefficient and difficult to scale, Areeba began to seek out a solution that enabled data virtualization.

Although Areeba was able to gather quality data, its teams were hung up on copying and pasting the data from CSV files into a central location and formatting the data before they could start building the predictive models. That’s when their data architecture team found CData Virtuality.

Working hand in hand, the data architecture and data science teams at Areeba built a foundation and process in CData Virtuality that is efficient, scalable, and faster than ever before.

“One of the most important learnings that we got out of this journey is how important data virtualization is for the machine learning process. And this refers to all parties involved: The data architecture team as well as the data science team. CData Virtuality helped us to reduce the grunt work and eliminate idle time.”

– Bernard Bardawil, Development Lead, Areeba

The outcome: Faster, more efficient data access

Using CData Virtuality, the Areeba data architecture team built a virtual access layer on top of their existing data landscape. Using the JDBC and REST API connectors, they can now access their data in a single location without moving it all to a data warehouse. Modules can be separated by responsibilities and thereby serve different teams with different requirements – all without editing or creating new code.

Before CData Virtuality, Areeba had to code and define each individual connection to their databases, data sources, and data science/ML tools. It was a cumbersome and time-consuming task that required constant attention and maintenance. Now, they use CData Virtuality to build centralized view definitions and use APIs to connect and retrieve the data from the centralized data model. The entire process is automated and scheduled, so the data is always up-to-date and ready to be used. With this single source of truth in place, the data architecture team can manage all incoming requests in a timely manner and ensure high-quality data.

Lastly, security risks have become virtually nonexistent as the need to transfer CSV files across networks has been eliminated. The data architecture team built a Data as a Service (DaaS) concept in CData Virtuality, using data from all systems and is available in the virtual layer, which can be accessed by all services and apps.

Areeba data scientists can now build the views and aggregations needed to build predictive models in the virtual access layer, bringing in data in Python or R for ultimate performance. They have also built a connection to Tableau in CData Virtuality – without the help of a developer. Very little programming was required, except from the data architecture team. Now, the usual step of writing an SQL query on JAVA to then combine and build a JSON to finally expose it is obsolete.

In the near future, the data science team wants to expand to neural network ML and deep learning. The models that the data science team builds will be shared with the business users for feedback, which will be fed into currently existing ML models to enrich them and to learn more.

Bring your data together faster with CData

CData Virtuality offers enterprise-level performance to bring your data sources together. Gain accurate insights faster without bogging your team down with manual tasks. [Try it today.](#)



CData Software is the real-time data connectivity company. Our connectivity solutions provide easy access to live data from hundreds of on-premises and cloud applications. Consumable by any user, accessible within any application, and built for all enterprises, CData is redefining data-driven business.