![data virtuality logo]

An agile and easy-to-use data integration solution enables Areeba to take machine learning to the next level

![areeba logo]

Areeba is a leading Lebanese financial technology company that provides smarter, faster, and innovative payment solutions for banks, merchants, and governments in the Middle East region. Areeba is committed to investing in new capabilities and technologies that deliver enhanced payment experience and more secure solutions from biometric card and mobile payment to smartPOS.

As an innovative company, Areeba is always striving for cutting-edge solutions and has started to apply data science rules on transactional and operational data in order to improve the performance of their services. Areeba stumbled upon several challenges before finding the dream solution, but finally overcame these with the Data Virtuality Platform.

Challenges

Large Amounts of Data in Various Data Sources

In order to apply machine learning models on the transactional and operational data, Areeba has to deal with huge amounts of data. Furthermore, all the data lies in different sources, such as MemSQL, MariaDB, and Oracle. This makes the challenge even more complex.

Manual Export of Data Resulting in CSV Chaos

Like many data scientists, the data science team at Areeba first tried to collect the data that is needed from multiple sources and in different formats, such as csv and text files. The data was then used to build the 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 kind of approach would cause challenges for real-time analytics in the near future.

Action

Looking for a Scalable Solution

Areeba knew that traditional data warehousing system were inefficient and difficult to scale. Thus, they began to seek out a solution which enabled data virtualization. At the end of the day, data is the main ingredient from which machine learning algorithms are trained. Although Areeba was able to gather data with good quality, they were hung up on copying and pasting the data from csv files into a central place and formatting the data even before they could start building the predictive models. That's when their data architecture team found Data Virtuality Platform.

Result

Working hand in hand, the data architecture/analyst team and data science team at Areeba built a foundation and process in Data Virtuality which is efficient, scalable, and faster than ever before.

● The Data Architecture Team

Using the Data Virtuality virtual layer, the data architecture team began to integrate the data from the different data sources into a central place, using the JDBC and REST API connectors. In this virtual layer, the data architects could separate the modules by responsibilities and thereby serve different teams with different requirements without editing or creating new code. Furthermore, the whole 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, not only can the data architecture team manage all incoming requests in a timely manner, but they can also ensure a high quality of the 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 Data Virtuality, using data from all systems and is available in the Data Virtuality layer, which can be accessed by all services and apps. Without Data Virtuality, Areeba would need to code the connection to the databases as well as the query definitions for every single connection between data sources and the data science/machine learning tool. But with Data Virtuality in place, Areeba now builds centralized view definitions in the Data Virtuality Platform and uses APIs to connect and retrieve the data from the centralized data model.

Instead of coding the connections to databases and query definitions separately from each datasource to different data science/machine learning tools, Areeba builds centalized view definitions in Data Virtuality and uses APIs to connect and retrieve the data from the centralized data model.

From the front-end perspective, Areeba connects from R to the Data Virtuality engine via a JDBC connector. They stream data of 40 to 50 million transactions this way.

Next to Data Virtuality and R, Areeba is using Spark. They configured calls in Spark so that data is retrieved in partitions and not in just one query. Handling large amounts of data from different data sources isn't a challenge anymore.

In-Memory database is used to materialize some data from Data Lake and from Master Data Management, which is needed for high speed access. Other data which do not need high speed access resides in the Data Lake and MDM. In general, we can talk about a virtualized data lake. Different areas in this virtual data lake use different storage and processing technologies, so that different types of data access requirements can be served in the best possible way.
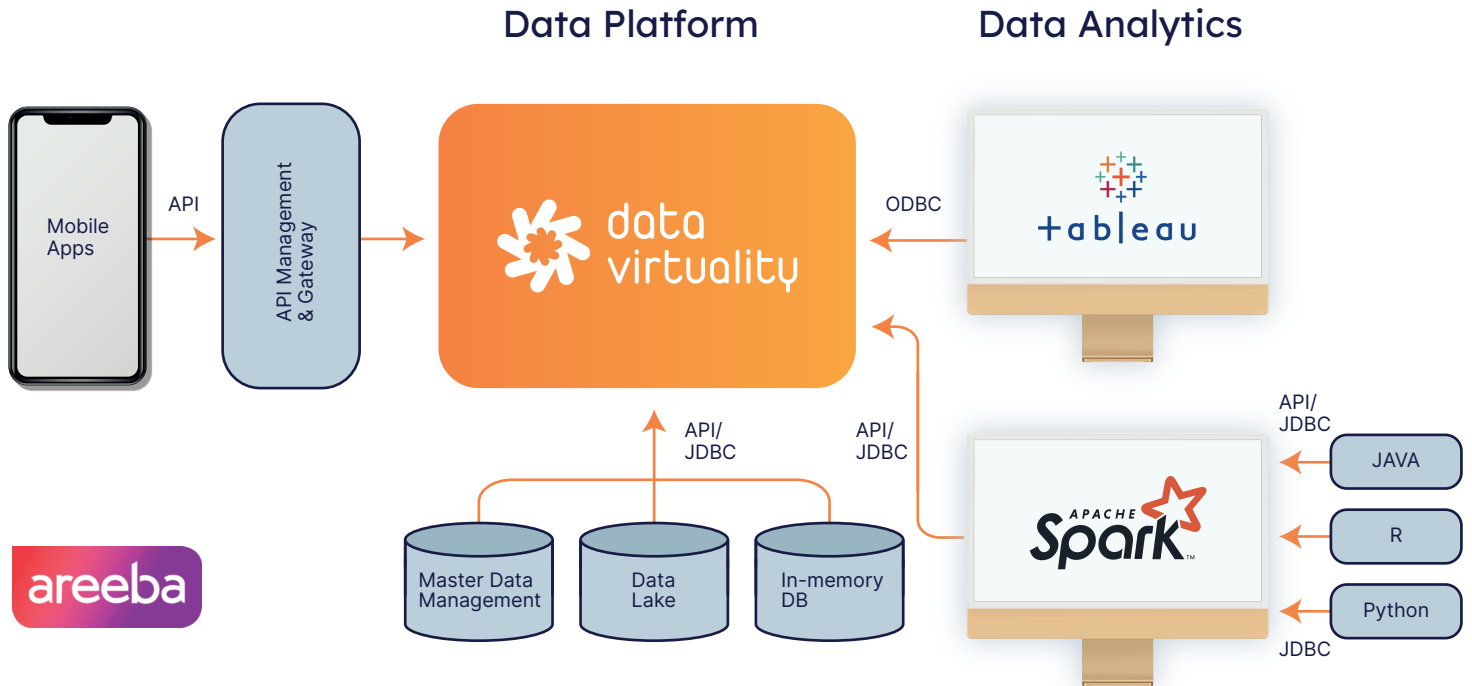
● The Data Architecture with Data Virtuality

Data Platform   Data Analytics

Mobile Apps —API→ API Management & Gateway → data virtuality ←ODBC— tableau

Master Data Management | Data Lake | In-memory DB —API/JDBC→

API/JDBC → Apache Spark

JAVA —API/JDBC→
R →
Python —JDBC→

The architecture with Data Virtuality ensures a single point of truth to manage all incoming requests in a timely manner and ensure high data quality.

Data Virtuality provides data integration solutions that help companies to easily connect and manage their data from multiple data sources such as APIs, databases and flat files. The revolutionary single source of data truth platform combines data virtualization and automated ETL. In this way not only is data management simplified but data integration efforts are significantly reduced - by up to 80%.

● The Data Science Team

The data science team no longer has to copy and paste their data from different csv files. Today, they go directly to the virtual layer to get any data that they need. This has provided two big benefits:

1. Elimination of Idle Time: The data scientist does not have to wait for the data analyst team to provide the data they requested.

2. Real-time Data from Multiple Sources with the right format: Data is always available in one single place.

The data scientists can now build the views and aggregations that are needed to build the predictive models in the virtual layer. They can then bring the data in Python or R for ultimate performance.

So, what kind of models does the data science team build and for what do they use them?

The data science team at Areeba applies machine learning to learn more about different segmentations such as customer and merchant, merchant performance, and churn. Models that are used are time series, regression, decision tree, random forest and different clustering algorithms.

As the information won from of these learnings are also important for the business users, the data science team built a connection to Tableau in Data Virtuality. This could be done without the involvement of any developer.

Very little programming was required, except from the data science team. The usual step of writing a SQL query on JAVA to then combine and build a JSON to finally expose as JSON is obsolete now.

● Benefits for:

The Data Architecture Team

• Serve different teams with different requirements without coding
• DaaS concept in Data Virtuality
• Centralized view definitions and API to communicate with data science/machine learning tools
• Easily handle large amount of data from different data sources

The Data Science Team

• One central place to
    - access the data
    - build views and aggregations for predictive models
• Very little programming required

● The Future

Now that Areeba has a scalable and reliable solution in place, they have even bigger projects in mind. In the near future, the data science team wants to expand to neural network machine learning and deep learning. The models that the data science team builds are shared with the business users for feedback. This feedback will be fed into currently existing machine learning models to enrich them and to learn more.

data virtuality