Virtusa builds a Managed Data Lake with Next-gen Analytics on AWS for a leading investment bank in the US

Challenges

- Our client was facing many challenges with its existing legacy data architectures and infrastructure as it looked to build a real-time view into operations and customer behavior.
- Can't move data across silos
- Can't deal with dynamic data and real-time processing
- Can't deal with format diversity and change rate
- Complex ETL processes
- Difficult to find the people with adequate skills to configure and manage these systems
- Can't integrate with the explosion of available social and behavior tracking data

Solution

- Virtusa built a Data Lake to store and analyze massive volumes and heterogeneous types of data. Features include:
- Decoupled storage and computed by making S3 object-based storage; not a fixed tool to cluster the data lake
- Leveraged a variety of AWS tools for data ingestion depending on data structure
- Used a centralized data catalog
- Implemented the right cloud security controls to protect data on AWS

Value delivered

- Our client is now able to store and analyze all its data - structured and unstructured - in one centralized location at low cost and can now:
- Quickly ingest data without needing to force it into a predefined schema, enabling ad-hoc analysis by applying schemas on read, not write.
- Use any tool in the ecosystem thereby increasing the flexibility of the right tool for the job
- Future-proof its architecture. As new use cases and new tools emerge, our client now can plug and play the current best of breed.

vírtusa Engineering First