

Cloud Based Data Warehouse (HIPAA)

Designing a data warehousing solution that provides a single source of truth for all data



AWS, Docker,
Java, Spring,
Python



\$30K



1 month

Challenge

Our client has identified a need for a solution that would act as a Data Warehouse (DW) and single source of truth for the observational data gathered in the different queries performed by another tool they were using but were not pleased with.

Until now they gathered information from three different sources while having historical data as a set of data files. The process was done manually by copying data to a local machine and running data manipulation scripts. This was error prone and time consuming. This created a risk to data integrity, a possible compliance gap, divergent data sets (no single source of truth), and a data silo inaccessible to all teams.

Solution

We designed a solution that would provide a single source of truth for all data, with the system that logs all changes to the data and track code changes. The solution was designed to run on the cloud without a need for on premise infrastructure. The top-level scope of the new solution was to:

- Design the DW for the RA USA registry so that adding other registries in the future is supported,
- Include historical RA data migration into the warehouse, and
- Include logic so that the switch-over from current data sources could be performed.

Benefits

- The automated process would reduce risk of human errors and latency of updates
 - The solution was designed to easily keep track of the regulatory and make changes towards reaching the compliancy
 - The client was able to process all historical data and leverage them in further research
 - We implement a concept of continuous delivery in the design of the solution so that long-term the client would have the setup to reduce the cost, time and the risk of delivering changes to the solution's environment with a possibility to easily add more environments in the future.
-