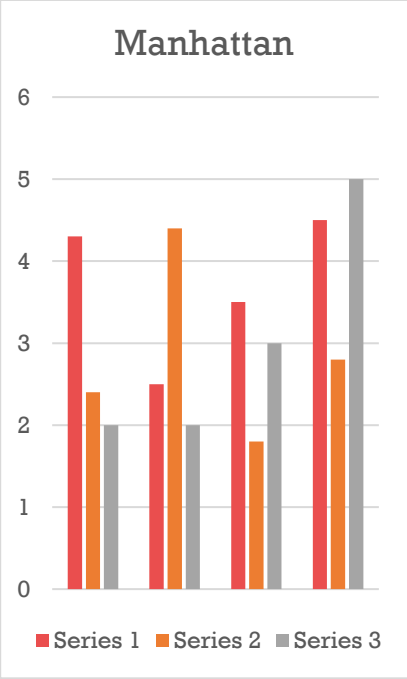
DB RESEARCH

WWW.DBRESEARCHINC.CO

TAG LINE

Manhattan: an actuarial data science platform



STORYBOARD

ABOUT MANHATTAN

Manhattan is an actuarial data science platform which allows you to setup secure data lakes easily.

By using data catalogs, a simple and cost-effective way to categorize your data, Manhattan automatically identifies the data formats and schema and builds a metadata repository by eliminating the need to manually define and maintain schema, so it's much easier to discover and understand your data sets.

This gives the most relevant data for analysis and maximizes the value of your data throughout the organization.

Within a few clicks you will be able to add your external datasets into the data lake. Once your job is ready you can categorize into data catalogs.

With robust search capabilities, which can search from a wide range of data sources, its capabilities include search by tags, description, keywords, and business terms. It quickly identifies related databases, tables, columns and Hadoop directories & files.

Set up access control for different users, and enforce encryption without affecting data quality so that users realize the full potential of their enterprise data assets.

PROBLEMS TO SOLVE

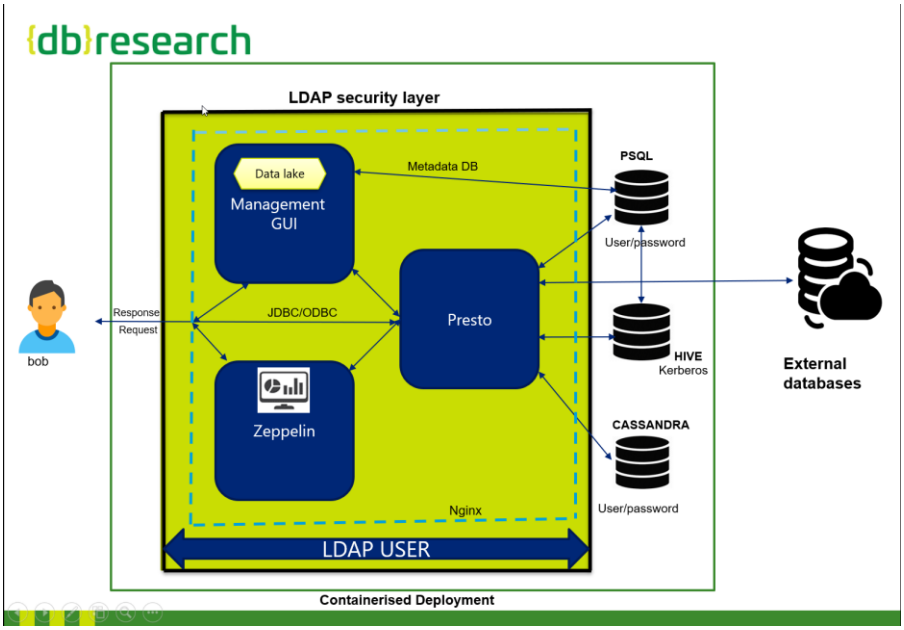
- Lack of reliable data creates difficulties for actuaries to determine accurate metrics and actionable insights in group health insurance risk pools.

- It keeps track of data that they have acquired for use with appropriate metadata.
- Run analysis on this data to create new insights managed as derived data with its own metadata.

WHY MANHATTAN?

- With this platform, the actuaries will be able to find the data they need easily without spending hours searching and prepping it for analysis.
- Act on the insights they've derived based on models that are defensible with academic rigor.
- Reduce the time to value ratio from their actuarial data science investments.

DATA FLOW



Consider a user Bob wants to fire a query into Manhattan

1) Manhattan will first check the authorization permission for the user BOB in LDAP
2) After successful authentication user will able to fire the query into Zeppelin and can then analyze their data sets

IF BOB IS A DATA CONSUMER

1) He can also manage the data catalogs on Management GUI
2) Find out what data they have and perform a robust search on it.

IF BOB IS A DATA ENGINEER

1) He is able to ingest data into the data lake.
2) Perform robust searching

IF BOB IS AN ADMIN

1) He can add or remove users from LDAP
2) Manage docker containers, start or stop it
3) See the activity log
4) Mange datasets
5) Discover Hadoop conf files
6) Add external databases into the Presto catalogs

IF BOB IS A DATA OWNER

1) He is able to do cataloging and see which data is sensitive or not.
2) Add tags and descriptions on the data catalogs
3) Perform robust searching
4) Encrypt Parquet data fields