

Data Process

Master Data Management: Collaboration between business and IT to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of shared master data assets

Data Governance: System for defining who within an organization has authority and control over data assets and how they will be used. Encompasses the people, processes, and technologies required to manage and protect data assets

Data Standards: Documented agreements on representation, format, definition, structuring, tagging, transmission, manipulation, use, and management of data. Make it easier to create, share, and integrate data by making sure there is clear understanding of how data are represented and received.

Data Quality: The planning, implementation, and control of activities, that apply quality management techniques to data. Ensures data accuracy, completeness, consistency, integrity, timeliness, uniqueness, validity, and accessibility.

Data Access: On-demand authorized ability to retrieve, modify, copy, or move data from systems. Can perform these functions in any location and with data in motion or rest.

Data Pipeline: Set of data processing elements connected to a series, where the output of one element is the input of the next one.

Data Driven Decisions

Business Intelligence

Descriptive Analysis

- used to describe the basic features of the data in a study
- used to present quantitative descriptions in a manageable form

Common Questions

- What is my time to hire last year?
- What is the time to cert last year?
- What is my workforce breakout?
- What are my attrition rates by quarter?

Tools Used To Accomplish Low/No Code

SAP/HANA Service Now Tableau PowerBi

Data Needs

Can be used on small as well as large datasets. Can be done adHoc. Can be connected to data as needed. Can only be generalized to the population set.

Products

Interactive and Static Reports and Visualization, Dashboards, Generalized analysis

AIML

Predictive Analysis

- trying to reach conclusions that extend beyond the immediate data alone
- make inferences from our data to more general conditions

Common Questions

- How successful will our new hires be?
- Can you predict when cert backlogs will occur?
- What will our workforce look like in 3 years?
- How many employees will attrit in the next 18 months? Why?

Tools Used To Accomplish High Code

BERT TensorFlow Sagemaker Keras R/Python

Data Needs

Requires large datasets to be effective. Needs constant access to data in order to maintain predictive performance. Can be inferred against new data.

Products

Decision Trees, Continuous NLP Models, Predictive and Explicit Analysis, Simulations, What-if models

Data Roles

Data Consumer: Individuals using the data. Important the data is good enough for them to do their job. Responsible for defining what makes the data good enough to use.

Data Analyst: Examine large data sets to identify trends, develop charts, and create visual presentations to help make strategic decisions. Utilize well defined data to draw meaningful insights and solve problems.

Data Scientist: Design and construct new processes for data modeling and production using prototypes, algorithms, predictive models, and custom analysis. Estimate the unknown by asking questions, writing algorithms, and building statistical models.

Data Steward: Responsible for the quality of a defined dataset on a day to day basis. Knows the data and business needs and rules that govern it.

Data Engineer: Helps make mechanisms and interfaces for the smooth flow and access to reliable data. Build and maintain the frameworks.

Data Architect: Help define the data principles and standards by translating the business requirements into technical requirements. Visualize and conceptualize data frameworks

Data Owner: A Senior individual that has authority to make decisions about business term definitions, data quality, accessibility, and retention requirements as they tie to the business needs.