

DATA DOMAINS

The data ecosystem is composed of three primary domains:

- 1. Data Engineering:** Focuses on building and maintaining data infrastructure and pipelines.
- 2. Data Science:** Involves analyzing complex data and developing advanced analytical models.
- 3. Data Analysis:** Centers on processing and interpreting data to generate actionable business insights.

Data Engineer Key Responsibilities:

- 1. Data Infrastructure Development:** Designing, building, and maintaining the architecture for large-scale data processing systems. This includes databases, large-scale processing systems, and data pipelines.
- 2. Data Integration:** Ensuring seamless data flow between various systems and platforms. They integrate data from multiple sources and ensure that it is consolidated, clean, and accessible.
- 3. ETL Processes:** Developing and managing Extract, Transform, Load (ETL) processes to collect data from various sources, transform it into a usable format, and load it into databases or data warehouses.
- 4. Performance Optimization:** Ensuring that data systems are efficient and optimized for performance, including indexing, partitioning, and tuning for speed and scalability.
- 5. Data Quality Assurance:** Implementing data validation and cleansing techniques to ensure the accuracy and reliability of data.
- 6. Collaboration:** Working closely with data scientists, analysts, and other stakeholders to understand their data needs and provide the necessary infrastructure and tools.

Data Scientist Key Responsibilities:

- 1. Data Analysis and Interpretation:** Analyzing large datasets to identify trends, patterns, and insights. This involves statistical analysis, data mining, and predictive modeling.
- 2. Machine Learning and AI:** Developing and deploying machine learning models to solve complex problems such as classification, regression, clustering, and recommendation systems.
- 3. Experimentation:** Designing and conducting experiments to test hypotheses and validate models. This can include A/B testing and other experimental frameworks.
- 4. Data Visualization:** Creating visual representations of data to communicate findings to stakeholders. This includes dashboards, reports, and other visual tools.

5. Collaboration: Working with cross-functional teams, including engineers, product managers, and business analysts, to implement data-driven solutions.

6. Research and Development: Staying up-to-date with the latest research and advancements in data science and applying them to real-world problems.

Data Analyst Key Responsibilities:

1. Data Collection and Processing: Gathering data from various sources and preparing it for analysis. This includes cleaning, transforming, and organizing data.

2. Exploratory Data Analysis (EDA): Performing initial investigations to discover patterns, spot anomalies, and test hypotheses.

3. Reporting: Creating reports and dashboards to present findings to stakeholders. This often involves using tools like Excel, Tableau, or Power BI.

4. Trend Analysis: Identifying trends and patterns in data to provide actionable insights that can drive business decisions.

5. Support Decision-Making: Providing data-driven recommendations to support strategic and operational decisions.

6. Collaboration: Working with business units to understand their data needs and provide insights that help achieve business objectives.