

resources

Wednesday, March 1, 2023 4:21 PM

<https://madewithml.com/#mlops>

Layers of Data Infrastructure

1. Data Ingestion
2. Data Integration
3. Data Storage
4. Data Processing
5. Data Presentation

Data Ingestion	Description	Roles
Data Collection	Gathering data from various sources	Data Analyst
Data Extraction	Extracting relevant data from various sources	ETL Developer, Data Engineer
Data Transformation	Converting extracted data to desired format	ETL Developer, Data Engineer
Data Validation	Ensuring data quality and completeness	Data Analyst, Data Engineer
Data Loading	Storing data in target system	ETL Developer, Data Engineer
Data Integration	Combining data from multiple sources	Data Engineer
Data Cleaning	Cleaning and preprocessing data	Data Analyst, Data Engineer
Data Enrichment	Enhancing data with additional information	Data Analyst, Data Engineer

Data Integration	Description	Roles
Data Profiling	Analyzing the data and identifying its characteristics	Data Analyst, Data Scientist, Data Engineer
Data Mapping	Mapping the data fields from source to destination	Data Analyst, Data Engineer, ETL Developer
Data Transformation	Converting data from one format to another while maintaining its structure	ETL Developer, Data Engineer
Data Quality	Ensuring data accuracy, consistency, and completeness	Data Quality Analyst, Data Analyst, Data Engineer
Data Integration	Combining data from multiple sources to create a unified view	Data Engineer, ETL Developer, Data Analyst
Data Migration	Moving data from one system to another	Data Engineer, ETL Developer, Database Administrator
Data Synchronization	Keeping data up to date across systems	Data Engineer, ETL Developer, Data Analyst

Data Storage	Description	Roles involved
Data modeling	Designing the data schema, defining data types and relationships	Database architect, data modeler
Data indexing	Creating and managing indexes to optimize data retrieval	Database developer, database administrator
Data partitioning	Partitioning data into smaller subsets to improve query performance	Database developer, database administrator

Data backup and recovery	Ensuring the availability and recoverability of data in case of system failures or data loss	Backup and recovery specialist, database administrator
Data security	Implementing and managing security measures to protect sensitive data from unauthorized access, modification or disclosure	Security analyst, database administrator, data govern

Data Processing	Description	Roles
Data Cleaning	Identify and handle incorrect, incomplete, or irrelevant data.	Data Analyst, Data Scientist, Data Engineer
Data Transformation	Convert and standardize data from one format or structure to another.	Data Engineer, ETL Developer
Feature Engineering	Create new features or variables from existing data that can be used in predictive modeling or analysis.	Data Scientist, Machine Learning Engineer
Data Aggregation	Combine multiple data sources or individual data points into a summary or consolidated form.	Data Engineer, Data Analyst
Data Sampling	Extract a representative subset of data from a larger dataset for analysis or model training.	Data Scientist, Data Analyst
Data Visualization	Create visual representations of data to help identify patterns, trends, and insights.	Data Analyst, Data Visualization Specialist
Machine Learning	Use statistical models and algorithms to analyze and make predictions or classifications on data.	Data Scientist, Machine Learning Engineer
Natural Language Processing	Analyze and interpret human language data, such as text or speech.	Data Scientist, NLP Engineer

Data Presentation	Description	Roles
Data Visualization	Creating visual representations of data to aid in analysis	Data Analyst, Business Intelligence Analyst, Data Scientist
Dashboard Development	Designing and developing interactive dashboards	Data Analyst, Business Intelligence Analyst, Developer
Report Generation	Generating reports for stakeholders and decision makers	Data Analyst, Business Intelligence Analyst, Developer
User Interface Design	Designing user interfaces for data-driven applications	UI/UX Designer, Developer, Product Manager
Data Storytelling	Communicating insights and trends through storytelling	Data Analyst, Data Scientist, Communications Specialist
Data Presentation	Presenting data and analysis to	Data Analyst, Data Scientist,

