**GDAI**

GDAI (Generic Data AI) provides a framework for managing data ingestion, transformation, and deployment in a standardized and scalable manner. This seminar will cover the core components of GDAI development, including ingestion frameworks, data transformation, and orchestration.

**End-to-End Data Flow in GDAI**

The data processing in GDAI involves the following key stages:

1. Data Ingestion via MDMF
2. Data Transformation using ePrep (PySpark-based)
3. Deployment using Databricks Asset Bundles
4. Orchestration via Apache Airflow
5. Repository Management at the Product Level

**MDMF - Ingestion Framework**

**Purpose:**

The MDMF (Metadata-Driven Management Framework) standardizes data ingestion from storage.

**Key Features:**

* Configurable ingestion pipelines
* Metadata-driven processing using a configuration file

**Data Transformation with ePrep**

**Overview:**

ePrep is a transformation framework built using PySpark to ensure scalability and efficiency in data processing.

**Key Features:**

* Configuration-driven transformation
* Modular and reusable transformation logic
* Supports data quality checks and testing

**Orchestration with Apache Airflow**

Apache Airflow is used to orchestrate the entire data pipeline, ensuring smooth execution of ingestion, transformation, and deployment steps.

**Capabilities:**

* Triggers and manages ingestion, transformation, and deployment
* Monitors pipeline execution with logging and alerting
* Uses **MDMF – AzureDataFactoryRunPipelineOperator** for ingestion
* Uses **EPREP - EprepDatabricksRunNowOperator** for transformation

**Conclusion**

This seminar provides an in-depth understanding of the GDAI development framework, covering data ingestion, transformation, and orchestration. By leveraging tools like MDMF, ePrep, and Airflow, organizations can build scalable, efficient, and standardized data pipelines.

Hackerrank questions:

1. A screenshot of a computer program

AI-generated content may be incorrect.

2. A computer screen shot of a program

AI-generated content may be incorrect.