**Mastering Data Orchestration with Apache Airflow**

Welcome to the webinar on data orchestration with Apache Airflow! Today, we’re going to walk you through the Airflow basics (such as core components and elements of the Airflow UI), and then show you how to orchestrate your Directed Ascylic Graph.

**Overview**

Apache Airflow is an open-source platform for programmatically creating, scheduling, and monitoring processes. Data engineers can use it to create and orchestrate complicated data pipelines for large data processing. In addition, users may manage task dependencies, monitor progress, and automate error handling with Airflow. Airflow provides an easy-to-use interface for creating, visualizing, and editing workflows. It can be integrated with various big data technologies such as Apache Hadoop, Spark, and Hive. It allows users to organize and execute jobs and workflows effortlessly depending on variables like time, data availability, and external triggers.

**Introduction**

Big data orchestration using Apache Airflow is a topic that involves managing complex workflows and dependencies in big data processing pipelines. Apache Airflow is an open-source platform that allows users to define, schedule, and monitor workflows programmatically and efficiently. Workflows are defined using Python code, making it easy to modify and maintain them. Airflow also provides pre-built operators to interact with various data sources and sinks, and its web interface allows users to monitor workflow statuses and set up alerts. Airflow integrates well with popular big data technologies like Apache Hadoop, Apache Spark, and Apache Kafka, providing greater clarity and control over data pipelines.

## **What is Big Data Orchestration?**

Big Data Orchestration refers to the process of managing and coordinating various components and processes in a big data environment. It involves organizing and integrating data from multiple sources, as well as orchestrating the flow and processing of data to extract meaningful insights. This includes tasks such as data ingestion, data integration, data transformation, data storage, and data analysis. Big Data Orchestration helps enterprises ensure the smooth and efficient operation of their big data infrastructure, enabling them to extract valuable insights and make data-driven decisions. It plays a crucial role in maximizing the value and potential of big data.

## **Installing Apache Airflow**

To learn about Big data orchestration using Apache Airflow, the installation of Apache Airflow is the primary step. This involves installing required components and configuring the system. You should check system requirements, including Python 3.x and other dependencies.

Follow this [**tutorial**](https://www.notion.so/ayoadeabel/Getting-Started-with-Airflow-57b11bb93ced4deeb6702fe236ff2888?pvs=4)