

# Briefing

**Airflow Hooks:** Hooks provide a reusable interface to external systems and databases. With hooks, you don't have to worry about how and where to store the connection strings and secrets in your code.

**Data Lineage:** The data lineage of a dataset describes the discrete steps involved in the creation, movement, and calculation of that dataset.

**Data Validation:** Data Validation is the process of ensuring that data is present, correct & meaningful. Ensuring the quality of your data through automated validation checks is a critical step in building data pipelines at any organisation.

**Database (Components of Airflow):**  
Saves credentials, connections, history, and configuration

**Directed Acyclic Graphs (DAGs):** DAGs are a special subset of graphs in which the edges between nodes have a specific direction, and no cycles exist. When we say "no cycles exist" what we mean is the nodes can't create a path back to themselves.

**Edges:** The dependencies or relationships other between nodes.

**End Date:** Airflow pipelines can also have end dates. You can use an `end_date` with your pipeline to let Airflow know when to stop running the pipeline. End\_dates can also be useful when you want to perform an overhaul or redesign of an existing pipeline. Update the old pipeline with an `end_date` and then have the new pipeline start on the end date of the old pipeline.

**Logical partitioning:** Conceptually related data can be partitioned into discrete segments and processed separately. This process of separating data based on its conceptual relationship is called logical partitioning. With logical partitioning, unrelated things belong in separate steps. Consider your dependencies and separate processing around those boundaries.