Definitions

- **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 cant create a path back to themselves.
- **Nodes:** A step in the data pipeline process.
- **Edges:** The dependencies or relationships other between nodes.

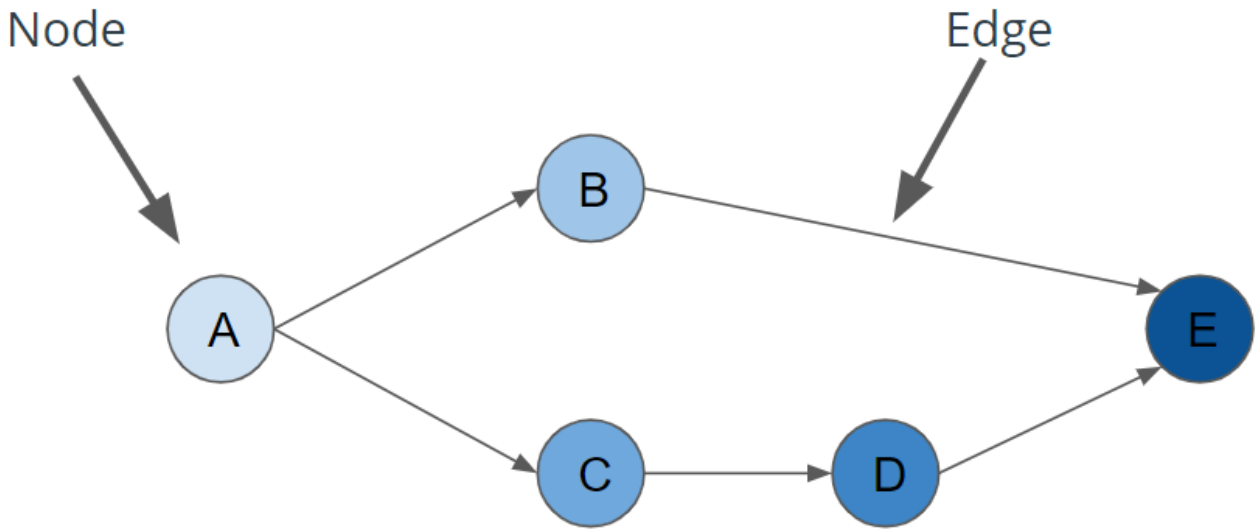

Diagram of a Directed Acyclic Graph

Common Questions

Are there real world cases where a data pipeline is not DAG?

It is possible to model a data pipeline that is not a DAG, meaning that it contains a cycle within the process. However, the vast majority of use cases for data pipelines can be described as a directed acyclic graph (DAG). This makes the code more understandable and maintainable.

Can we have two different pipelines for the same data and can we merge them back together?

Yes. It's not uncommon for a data pipeline to take the same dataset, perform two different processes to analyze the it, then merge the results of those two processes back together.