Snowflake introduction

* Snowflake is a cloud based data warehouse (its mean it is a data storage tool and its cloud based tool)
* It stores structured (ex: tables, csv files) and semi structured data(ex: json files).
* Instant, unlimited scalability
* No management burden and live data sharing
* It is a multi clustered architecture
* It runs on amazon AWS, Microsoft Azure and google cloud platforms

Why snowflake

* Pay for what you use model
* No infrastructure cost
* Snowflake helps in some data transformations, create data pipes, create visual dashboards
* High scalability
* Data recovery, backup, sharing, data protection

What is airflow

Apache airflow is a open-source platform that empowers data professionals to efficiently create, schedule, and monitor tasks and workflows

Why its popular

* Pure python
* Robust integration
* Highly scalable
* Rich interface
* Open source

Where can I use airflow?

* Batch ETL pipeline
* Machine learning pipeline train/ test pipeline

What is an ETL(Extraction Transformation Loading)

* Extraction: extract the data from different data sources
* Transformation: data according to business rules. It can be cleaning, aggregations, filters
* Loading: loading the data into target system. such as data warehouse, data lake or data store