USE CASES FOR APACHE AIRFLOW

By Guha Rajan

Uses cases

- ETL jobs
  - Extract, Transform and Load – Datawarehouse.

- Machine Learning pipelines.

- Delivery data to external system

- Scheduled Jobs
  - DB backups, data export etc

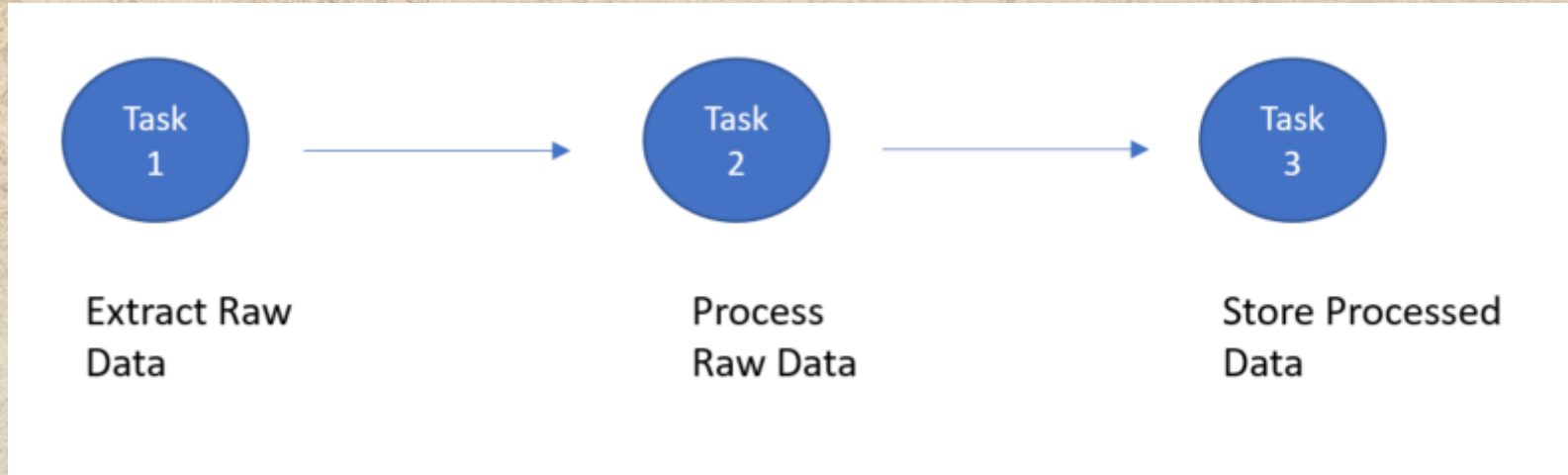- **Any automated jobs that needs to be run regularly based on dependencies, condition…..**

Why not use cron jobs?

cron is a Unix / Linux utility which schedules a command or script on your server to run automatically at a specified time and date.

Airflow is like cron, but so much more powerful:

- Dependencies between tasks
- Retries
- SLAs/error notifications
- Metrics about time to complete, number of failures etc
- Logs
- GUI / Visibility

What is task and why orchestration?

Extraction, process and transform, make it useful without manual intervention