

GLOSSARY



Data Engineering Foundations

With Harshit Tyagi

Use the terms and definitions below to understand concepts taught in this course.

Transcript Search: note that you can search for terms spoken by the instructor during the course. To search videos, switch to the Transcript tab, then search for keywords using the [In this video](#) or [In this course](#) option.

| Term | Definition |
|------------------------------|---|
| Airflow | An Apache system for scheduling and automating processes in Python |
| Apache | A public organization committed to open-source software development and distribution |
| distributed computing | Using several hardware processors simultaneously to accomplish a computing task |
| ETL | Extract, transform, and load; A paradigm for building large data repositories from multiple sources |
| Hadoop | An Apache software library for distributed computing |
| Hive | An Apache system for large-scale SQL data warehousing |
| MapReduce | An Apache programming framework for distributed computing |
| schema | A plan for organizing data |
| Spark | An Apache system for distributed processing of large-scale data |
| SQL | Structured query language; A system for reading, processing, modifying, and writing databases |