

Defining New Terms

The video above includes references to a few terms that you may not be familiar with. Below are some definitions that you might find useful.

Extract Transform Load (ETL) and Extract Load Transform (ELT):

"ETL is normally a continuous, ongoing process with a well-defined workflow. ETL first extracts data from homogeneous or heterogeneous data sources. Then, data is cleansed, enriched, transformed, and stored either back in the lake or in a data warehouse.

"ELT (Extract, Load, Transform) is a variant of ETL wherein the extracted data is first loaded into the target system. Transformations are performed after the data is loaded into the data warehouse. ELT typically works well when the target system is powerful enough to handle transformations. Analytical databases like Amazon Redshift and Google BigQ."

Source: [Xplenty.com](#)

This [Quora post](#) is also helpful if you'd like to read more.

What is S3?

"Amazon S3 has a simple web services interface that you can use to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites."

Source: [Amazon Web Services Documentation](#).

If you want to learn more, start [here](#).

What is Kafka?

"Apache Kafka is an **open-source stream-processing software platform** developed by LinkedIn and donated to the Apache Software Foundation, written in Scala and Java. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds. Its storage layer is essentially a massively scalable pub/sub message queue designed as a distributed transaction log, making it highly valuable for enterprise infrastructures to process streaming data."

Source: [Wikipedia](#).

If you want to learn more, start [here](#).