

Basic ETL Pipeline

Patrick Pfenning

An Extract-Load-Transform (ETL) Pipeline is a set of processes meant to move data from an initial landing point to a production ready data system. This process pipeline works as followed:

1. Capture data from starting point. **(Extract)**
2. Massage, split, and aggregate the data to desired format. **(Transform)**
3. Export data to final location. **(Load)**

Let us now walk through a real world example using AWS. We have a customer who has csv data of sales for made in the past several hours. This data is dropped on an SFTP server which is accessible from our AWS instance. We as Data Engineers must allow our front end users to report on this data from a PostgreSQL database. ///