

Building a simple data pipeline on AWS

Mark Durbin

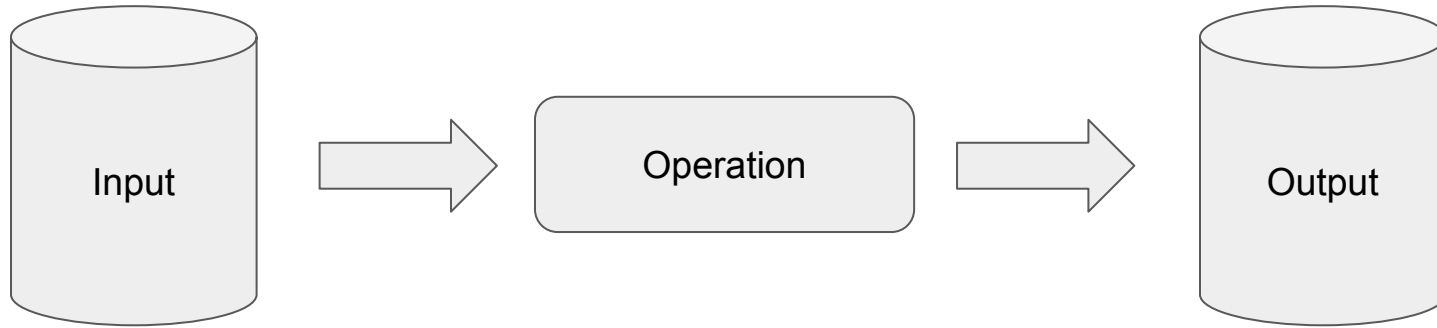
May 3, 2023

Overview

After today's demo, you should be able to follow along with the steps available in the GitHub repository (<https://github.com/smarkdurbin/aws-simple-data-pipeline>) to build your own simple data pipeline on Amazon Web Services (AWS).

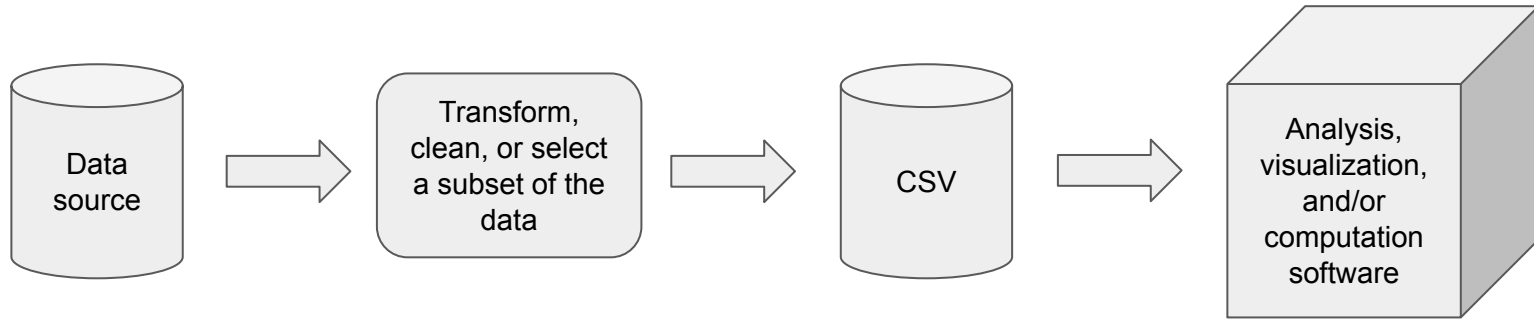
What is a data pipeline?

A data pipeline ingests data from a source, manipulates the source data, and stores it elsewhere.



How are data pipelines used?

Data pipelines are used to prepare data for analysis, presentation, and/or computation.



How can you build a data pipeline?

You can build a data pipeline on just about any computer that you can write and execute your own software on.

If you have a computer with Python installed, you can create a Python script to read from a data file like JSON, manipulate the data, and write the data to an output format like CSV. You can run your data pipeline on a schedule using Scheduled Tasks in Windows, or Cron in most Linux distributions.

Demo and Questions

Thank you!