# Data Engineering Overview

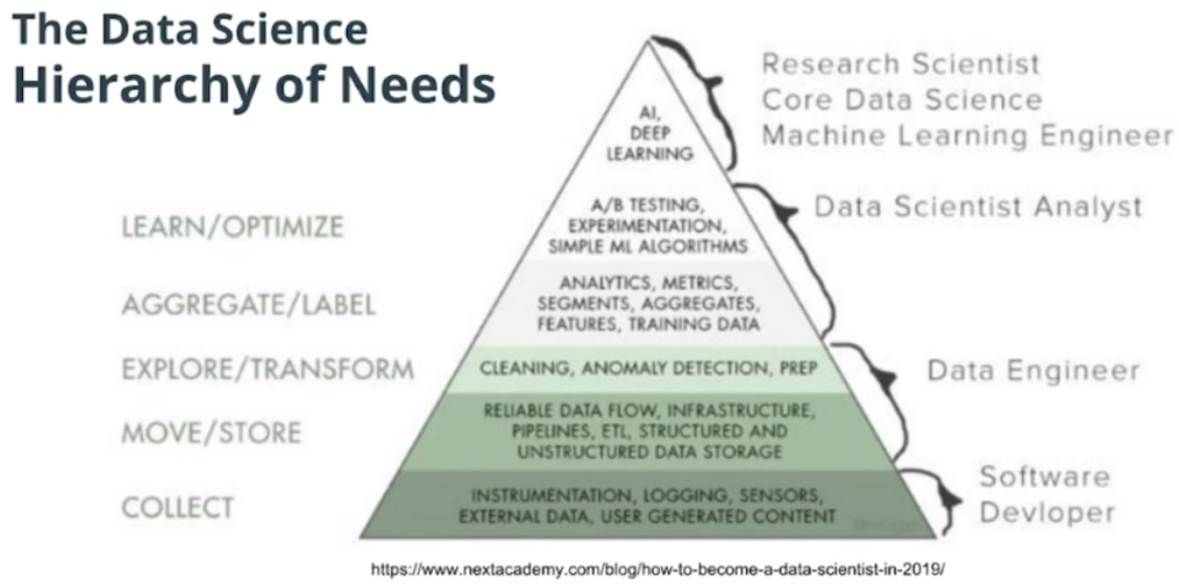
Saturday, April 6, 2019 3:01 AM

## What is Data Engineering?

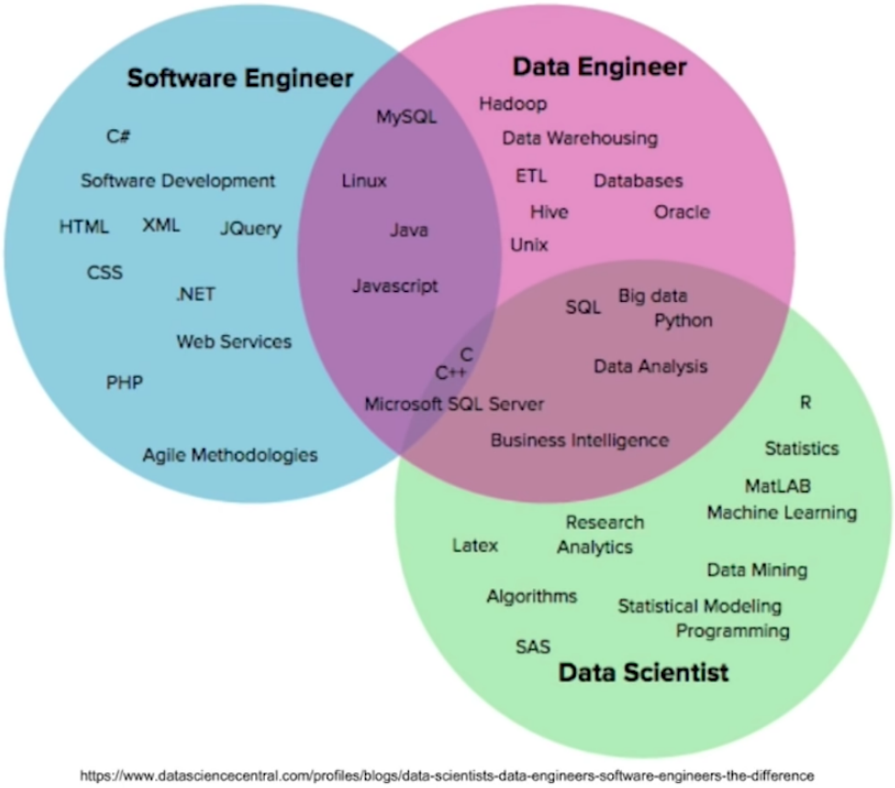
All the processing needed to make data available to users.

## Where does Data Engineering Fit?

Basically between Data Science and Software Development




https://www.nextacademy.com/blog/how-to-become-a-data-scientist-in-2019/




https://www.datasciencecentral.com/profiles/blogs/data-scientists-data-engineers-software-engineers-the-difference

## What do Data Engineers Do?

- Operational Monitoring – Make sure nothing has failed unexpectedly
- Write a lot of code (ETL, Spark, Airflow etc) - There are a lot of tools that fall into some broad categories
  - ETL – Database (relational or not)
  - Spark – Distributed Data Processing

    - Airflow – DAG Processing (pipelines)
- Tools aren't standardized so there is a lot of variance
- Soft skills in working with non-technical people are very important so the right questions are asked up front.

Evolution of Data Engineering - https://medium.com/analytics-and-data/on-the-evolution-of-data-engineering-c5e56d273e37
- Pivoting towards more Software Engineering
    - "The role of the data engineer is no longer to just provide support for analytics purpose but to be the owner of data-flows and to be able to serve data both to production and for analytics purpose."

Data Evolution Epochs - https://learn.panoply.io/hubfs/Data%20Engineering%20-%20Introduction%20and%20Epochs.pdf
- There are some long roots, as people have been wanting to report on data pretty much forever.

Tool Landscape - https://datafloq.com/big-data-open-source-tools/os-home/
- It's very crowded at present, and multiple major vendors in most spaces (as well as open source)