DATA ENGINEER VS DATA SCIENTIST: WHAT'S THE DIFFERENCE?



The data engineer equips the business with the ability to move data from place to place, known as data pipelines. Data engineers provide data to the data science teams.

The data scientist consumes data provided by the data engineers and interprets it to say something meaningful to decision-makers in the company.

In this article, let's dive a little deeper into the roles of data engineer and data scientist.



What is a Data Engineer?

In general, the data engineers are responsible for building pipelines, architecting the back-end databases, creating queries, and more.

Responsibilities

The data engineer will often possess a degree in computer science or engineering. Their skills involve building and working with computers directly. They build databases, queries to interact with the databases, move data from one database to another, transform the data to be sent as the right type to its end point. They are the ones who build APIs.

Data engineers will use a number of computer languages to get the job done. At their level, the best language depends on the task and the equipment they are working with. Java, Scala, C++, Go, and Python may be used.

Skills

- Writing database queries
- Building database pipelines
- Building APIs
- Coding language: Java, Scala, C++, Go, Julia, Python

What is a Data Scientist?

The data scientists might have to know only a little computer coding to ingest the data from the engineer's sources, and to transform it to fit their needs.

Data scientists' skillset is founded more on having good reasoning and communication skills. Their job tends to be highly mathematical and statistical. They need to be able to:

- Create hypotheses around the data sets
- Test the hypotheses
- Put what they learn into communicable information to decision-makers

Responsibilities

Data scientists are responsible for consuming data from a source, and finding valuable information from it. Then, they are tasked with presenting the information often through a visualization.

Data scientists will have to:

- Identify relevant data sources
- Filter through structured and unstructured data
- Analyze the data for trends
- Prepare a reports and visualizations

Skills

- Strong mathematical and statistical skills

- Source, filter, clean, and verify data
- Excellent ability to reason and communicate
- Build visualizations and data dashboards

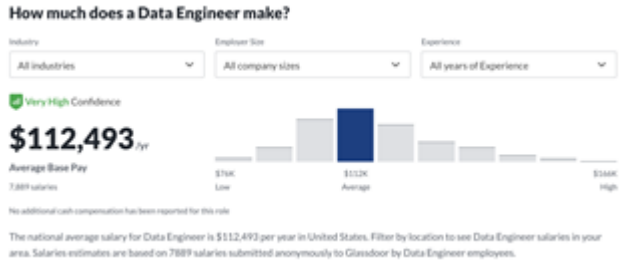
Increasingly, data scientists are adding machine learning to their skillsets, too.

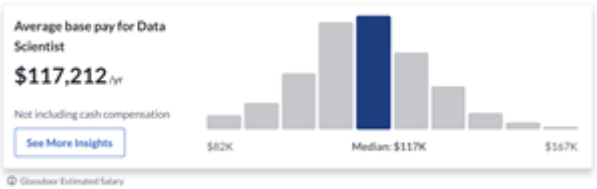
(Find out why Python is the predominant coding language for big data.)

Salary comparison

Demand for both positions is high and will be around long enough to build a career around. Their pay varies with skill-level and company. Data engineers used to make more money than data scientists, but in recent years, with the ever growing supply of people entering the field, and the growing ease to work with cloud data infrastructures, both jobs have evened out to roughly the same. Even their overall salary has dropped about $30,000 in the past 5 years.

While engineers' average salary is reported to be a little less than a data scientist's, their distribution curve is a little flatter, meaning there are more people in the position making higher salaries.





Source

Data engineers & data scientists working together

The role each plays in the company is essential. The data engineers tend to be better programmers and have a far better grasp on moving data around—after all, that's their sole job. The scientists will specialize more in data analytics and all the statistical and reporting strategies to extract meaningful information from data.

Data scientists have the more popular role because, in a way, they are the journalists of data, and create the reports for people to read. Thus, they become the face of data while the engineers are behind the scenes and make access to all the data possible for the data scientist's reports.

Data scientists' reports can also influence the data engineering team's data collection efforts. If the analysts determine a new source of information is needed, then they can tell the engineers to build a pipeline to gain access to the information, and the engineers can go develop the pipeline to give the scientists access to that information. It may be discovered that access to the information is impossible, has security issues, cost restrictions, too many inputs or unclear definitions, and they can tell the data science team that access to the information is impossible or needs to be approached

differently.

Whichever path you choose, data scientists and data engineers will be around for a long time.

Related reading

- BMC Machine Learning & Big Data Blog
- BMC Business of IT Blog
- Database Administrator (DBA) Roles & Responsibilities in The Big Data Age
- Data Science Certifications: An Introduction
- Best Books on Big Data & Data Science
- Machine Learning for Data Management