

Data Engineer VS Data Scientist

[Christopher Bolard](#)

[Want to read more Data Science and Engineering article? Follow Digital Source on LinkedIn!](#)

Big Data has changed the face of the world!

With 2,3 trillion gigabytes of data created each day, companies have access to a broad range of information on their users, market and much more.

This Data allows them to constantly improve their product/service.

Companies have understood the opportunity that Big Data represents. The soar of Data Engineer and Data Scientist jobs show it to us.

In 2011, *Harvard Business Review* has elected **Data Scientist the sexiest job of the 21st century** to underline the success of the profession!

However, this job field being not being fully mature yet, Data jobs are still subject to misunderstandings. It appears for many as a blurry technical ‘thing’ which could potentially implement their product or service.

This misunderstanding could result in failure in a good use of resources. Let’s get back onto the fundamentals of these professions and decrypt the value of each.

I — What are the differences between a Data Engineer and a Data Scientist?

1- Understand the hierarchy of the Data Process.

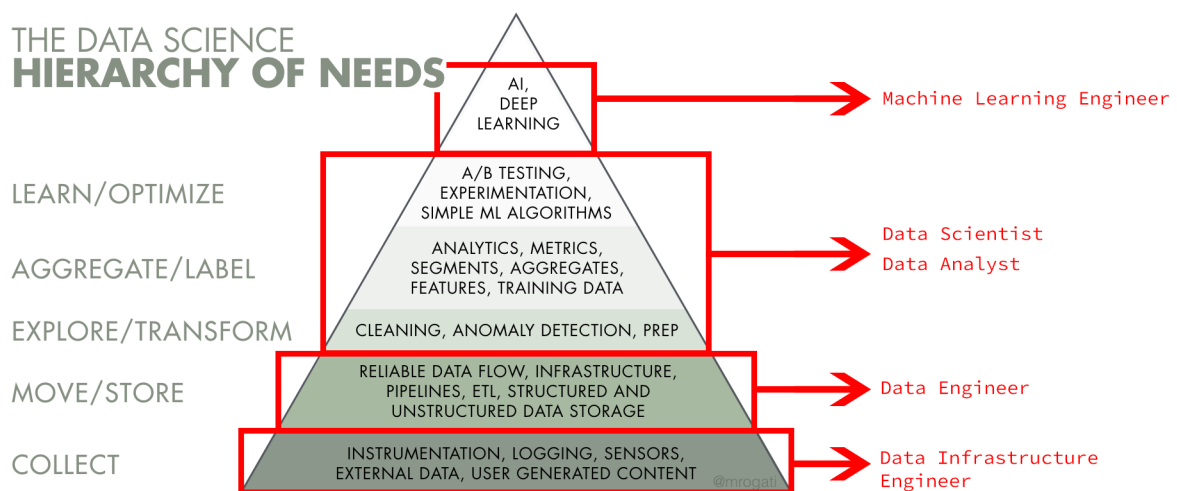


Fig.1 — THE DATA SCIENCE HIERARCHY OF NEEDS — Created by [Monica Rogati](#)

When a company makes a product/service, they need valuable information to understand their market, get information about their competitors, create a product, understand customers, and much more.

Before the world starts to collect data to feed what we call Big Data, businesses have no choice to:

- Do their own analyses
- Use their intuition for taking decisions

To resume, you needed experience (or luck) to meet market expectations.

Modern technologies are the 4th revolution. **Big Data offers a tremendous amount of new possibilities to create better businesses, organizations, charities, and so on...**

But taking advantage of Big Data requires processes, organization, tools and most importantly: people able to manage it.

And depending on what you want to do, Data Engineers and Data Scientists will be crucial for certain tasks of the process.

The “Data Science Hierarchy of Needs” pyramid is a great reflection of a complex data process.

Data Engineers collect relevant Data. They move and transform this Data into “pipelines” for the Data Science team. They could use programming languages such as **Java, Scala, C++ or Python** depending on their task.

(read the article about [Programming languages in Data Engineering](#))

Data Scientists analyze, test, aggregate, optimize the data and present it for the company.

(read the article about [Programming languages in Data Science](#))

Some companies with advanced processes complete their teams with **AI Engineers, Machine Learning Engineers or Deep Learning Engineers**.

See “*the Data Science hierarchy of needs*” **fig.1**

It becomes quite understandable that all these tasks have to be divided and given to specific Data professionals.

2 — Data Engineer — The technical part of data — Design — Build — Arrange.

Data Engineers are specialized in 3 main data actions: to design, build and arrange Data “pipelines”.

They are sort of Data Architects.

Data Engineers often have a computer engineering or science background and system creation skills.

“Data pipelines are sequences of processing and analysis steps applied to data for a specific purpose. They’re useful in production projects, and they can also be useful if one expects to

encounter the same type of business question in the future, so as to save on design time and coding. For instance, one could remove outliers, apply dimensionality reduction techniques, and then run the result through a random forest classifier to provide automatic classification on a particular dataset that is pulled every week.”

Colleen Farrelly, Data Scientist/Poet/Social Scientist/Topologist (2009-present)

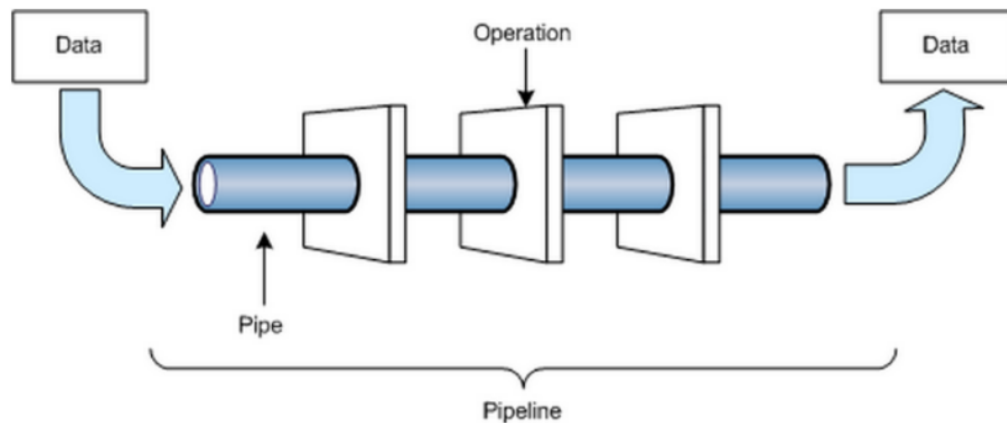


Fig.2 — Pipeline created from raw data to end results data.

What tasks have a Data Engineer in a company?

- Design the big data infrastructure and prepare it to be analyzed.
- Build complex queries to create “pipelines”.
- Arrange any problems in the programmed system.

What competencies wait from a Data Engineer?

- Logical mind
- Knowing what data to extract
- Management and organizational skills
- Working with cross-functional teams

3 — Data Scientist — Analyse — test — create — present

Data Scientists have normally 4 main tasks in a company. He analyses, tests, creates and presents them to the team.

Data Scientists have a math and statistical background. They are also comfortable with creating machine learning and artificial intelligence models.

What tasks have a Data Scientist in a company?

- Work on clean data
- Find solutions with the data available
- Communicate analyzes with the team
- Work onto solution problem and get some

What competencies wait from a Data Engineer?

- Good communication skills.
- Good analysis.
- Good hypothesis.
- Broad knowledge in different techniques in machine learning, data mining, statistics, and big data infrastructures.
- Be a problem solver.

As you can see, Engineers and Scientists require a different skillset and their profile are different.

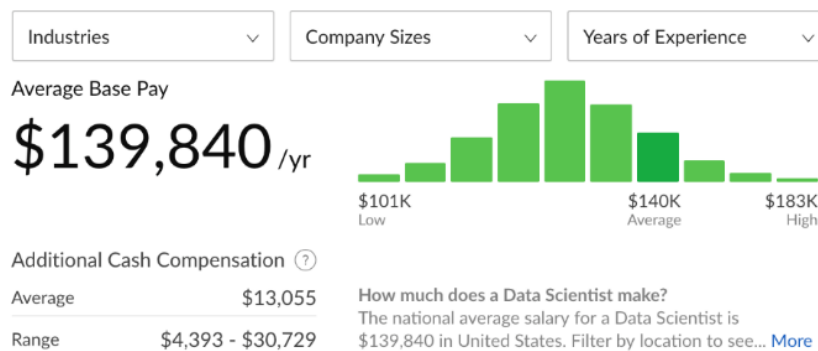
Also, Data Scientists must have very good communication skills to present Data and propose decisions based on their work.

4 — What about job openings and salaries in all that?

According to Glassdoor, the average base salaries in US (updated Sep 26, 2018) are :

Data Scientist Salaries

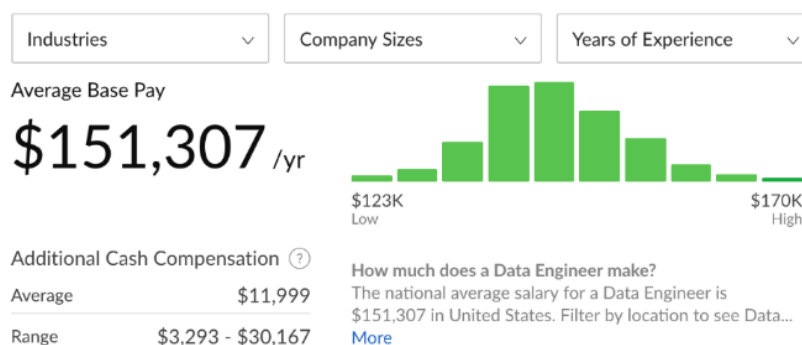
3,703 Salaries Updated Sep 26, 2018



- **For a Data Scientist: \$139K / year on average**

Data Engineer Salaries

1,356 Salaries Updated Sep 26, 2018



- **For a Data Engineer: \$151 / year on average**

“The number of job openings for data engineers is almost five times higher than the number of job openings for data scientists. This makes sense as most organizations need more data engineers than data scientists on their team” according to Glassdoor.

II- Data Engineer vs Data Scientist: what is the state of the Data job market?

1 — Data scientists: A growing sector

Data Scientist is a **dream work** on the paper.

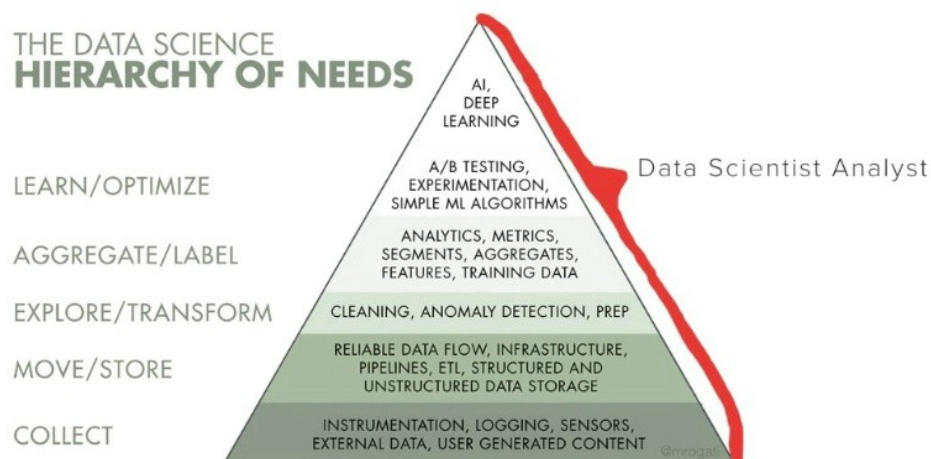
- A good salary
- A challenging job where you have to solve complex problems

However, when they work in little structures, data scientists could be transformed as multitask employee.

When Data Scientists have to deal with all the Data Hierarchy, it can become difficult to do the work as they are not Data Engineer or Software Engineer.

It can result whether in a devaluation of the profession or a waste of resources for the company.

Sometimes, being a Data Scientist in a company could look like that:



As a result, studies show that in 2017, **24.0% of Data Scientists have changed job.**

For sure, the Data Science job market is a flourishing environment which permits to change for the project employee like the most.

However, it also shows that a large amount of Data Scientists try to find a better place on the market.

2 — The Data Engineers penury

Data Engineers have become a rare commodity.

Glassdoor makes a census of more than 107K Data Engineers job opening. This has been so demanded that everybody is now touched by this penury:

“Even the hottest Silicon Valley companies are unable to achieve a one-to-two ratio. [...] You don’t have enough engineering talent out there. It’s very expensive.” says Tomer Shiran, the CEO and co-founder of Gremio, a developer of big data middleware.

Why recruiters have difficulties to find data engineers today?

For example, in the Netherlands, recruiters are searching for native speaking Engineers with skills on very specific programming tools.

However, most of them face long-term waitings before they find their talent.

So how to find a data engineer?

You should know first that it will require a good process.

Having a good candidate list of Data Engineers is essential to select the one that you need.

However, the process of finding Data Engineers need time and energy that companies doesn’t necessarily have.

One goal of recruitment agencies is to fulfill this gap between demand and offer.

By searching everyday best recruits on a specific field, there are able to answer to this major issue.

[Are you a company looking for a Data Engineer? Digital Source is a recruitment agency specialized in Big Data. Contact us by clicking here and let’s discuss your recruitment need.](#)

[Looking for a new opportunity in Big Data? Click here and look at new job opportunities.](#)