

Jobe

Remote
teams

Technologies

Clients

About
company

Blog

Contact

Career

NCube > Blog > Data Engineer, Data Scientist, Data Analyst

Data Engineer vs Data Scientist vs Data Analyst: What is the Difference?




MARYNA DEMCHENKO

Published: 04 Feb 2020

DATA

Share:





For many employers data engineers, data scientists, and data analysts appear to be different names for the same role. In reality, these roles span a variety of different skill sets and responsibilities, although all of them deal with data sets and play a key role in refining data strategies.

Data engineers build, test and maintain data ecosystems. These ecosystems are essential for companies, and data scientists in particular, whose job is to analyze data in order to build prediction algorithms. As such, we can say that what data engineers do is instrumental to data scientists.

Data analysts create ad-hoc and regular reports based on past and current data in order to find answers to business questions. This role is often seen as the stomping ground for someone interested in a data-related career.

The difference between data analyst and data scientist roles is that the scope of work of data analysts is limited to numeric data, whereas data scientists work with complex data.

In this article, we have compared these three roles to provide a comprehensive answer basing on our experience and Internet resources on this topic.

What is a data engineer: profile overview

A data engineer usually has a background in one of the STEM fields and is fluent in Mathematics, Statistics, and Big Data. Some essential skills to master for this role include SQL database, ETL tools, coding, and sometimes Statistics and Maths.

[How to become a data engineer?](#) They often embark on the path of big data as traditional solution architects, working with SQL databases, web servers, SAP installations, and other systems.

What does a data engineer do, exactly?

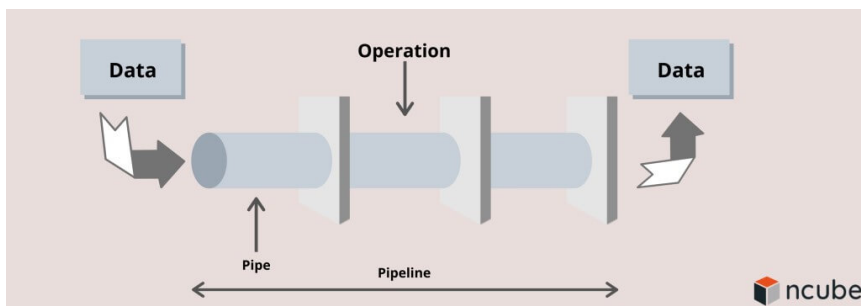
A data engineer is responsible for building, testing and maintaining the data architecture. They lay the foundation, enabling data scientists and data analysts to create new insights from data.

Furthermore, data architecture prepared by a data engineer makes the basis for further usage of data, which may include:

- Data ingestion and storage
- Algorithm creation

- Deployment of ML models and algorithms
- Data visualization

Data engineers work with raw data sets that may contain all sorts of errors: human, machine or instrument. Such data can hardly present value to data scientists. To make it usable, a data engineer needs to build reliable data pipelines, a sum of tools and processes for performing data integration. Pipelines connect data between systems and transfer data from one format into another. For this, they write customized scripts for API of external services, enrich data, implement data warehousing (or data lakes). Engineers also need to refine the pipelines continually to make sure the data is accurate and accessible. Here is what data engineering looks like, in a nutshell.



Data pipeline

A data engineer is a part of a data science team, working jointly with data analysts and data scientists. The [average data engineer salary](https://ncube.com/blog/average-data-engineer-salary) according to PayScale is 91K USD.

Data engineer skills:

Ins and outs of SQL

Data management is among the essential skills for a data engineer, and SQL is a commonly accepted standard for this activity since they work with SQL databases on a regular basis.

Database systems

Data engineers need to be fluent in SQL-based systems like MySQL, PostgreSQL Microsoft SQL Server, and Oracle Database as well as to be comfortable with NoSQL databases, including MongoDB, Cassandra, Couchbase, Oracle NoSQL Database.

ETL solutions

Data engineers need to have ETL tools in their toolkit to build processes to move data between systems. Examples of such technologies can be SAP Data Services, StitchData, Xplenty, Informatica, and Segment.

Data warehouse software

The ability to set up a cloud-based data warehouse and connecting data to it are essential to this role. Some of the data warehousing solutions include Amazon Redshift, Panoply, BigQuery and Snowflake.

Coding ability

Experience with Python or Scala/Java among other programming languages is valuable and in lots of cases even mandatory. Python is often used for ETL tasks. Data engineers are expected to have mastered their development skills, which is not critical for other data roles.

Big Data Tools

The most popular ones are Apache Spark, Apache Kafka, Apache Hadoop, Apache Cassandra, the first two being a common requirement. As such, it makes sense to concentrate on gaining a strong understanding of them. Knowledge of Hadoop-based technologies is a frequent requirement for this position as well.

Overview of data engineer roles and responsibilities

- Develop, construct, test and maintain architectures and processing workflows
- Build robust, efficient and reliable data pipelines
- Develop solutions for data acquisition
- Ensure architecture supports business requirements

- Develop dataset processes for data modeling, mining, and production
- Drive the collection of new data and refinement of existing data sources
- Recommend ways to improve data reliability, efficiency, and quality

What is data scientist: profile overview

A data scientist analyzes and interprets complex digital data to help business leaders make better decisions based on data.

Data scientists have profound knowledge of and expertise in math (linear algebra and multivariable calculus) which they have acquired by earning a degree in science-based disciplines.

The data scientist vs. data analyst roles have a lot in common, but the first one usually requires more advanced tech skills, such as more than one programming language, machine learning, and algorithms.

Read also: [Software Engineer Shortage in the World](#)

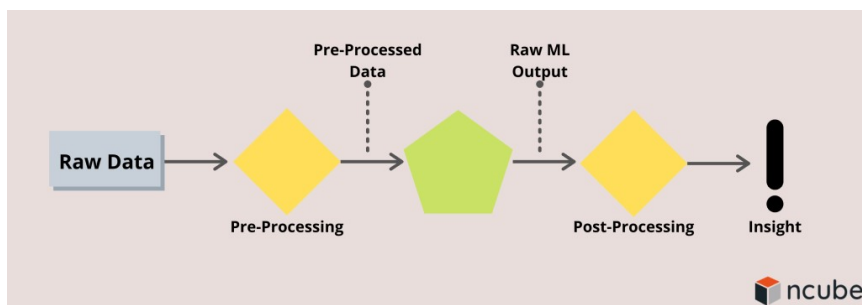
What does a data scientist do?

These professionals lean on predictive analytics, machine learning, data conditioning, mathematical modeling, and statistical analysis.

Similar to a data engineer, a data expert deals with large volumes of data by performing the following operations:

- Cleansing and collecting quality data to feed to train algorithms
- Identifying hidden patterns in data sets
- Building machine learning models
- Data visualization
- Refining business metrics by developing and testing hypothesis

The useful data is a true value for a data scientist. With this in mind, they need to explore the business domain and interact with business leaders and managers and develop general business acumen. This is done in order to formulate the questions to which the data is supposed to provide answers. However, in some companies, this element is covered by a data analyst.



Machine learning process

Despite the commonly accepted belief, building machine learning models is just one step of the process that involves a data scientist.

After post-processing model outputs, a data scientist can communicate the findings to managers, often using data visualization means. After the results have been accepted, data scientists ensure the work is automated and delivered on a regular basis.

Scientist vs. engineer: who earns more?
Comparing data scientist vs. software engineer salary: 96K USD vs. 84K USD respectively.

Skills for data scientists

R

With its unique features, this programming language is tailor-made for data science. With R, one can process any information and solve statistical problems.

Python

Python really deserves a spot in a data scientist's toolbox. Many professionals choose this language over other options such as Java, Perl or C/C++ because of its specially designed ecosystem for data science.

Hadoop

Although the knowledge of this tool is rather nice-to-have than mandatory, Hadoop increases the value of a data scientist, especially if they have experience with Hive or Pig. Cloud tools such as Amazon S3 may also come in handy.

SQL

Speaking one language with databases is essential for data scientists. As such, they must be proficient in SQL to be able to get information from databases using query instructions without having to write custom code.

Algebra, Statistics, and ML

Data scientists do have versatile skill sets. They excel at linear algebra and calculus and have sufficient coding skills. Of course, there are superstars that excel at both, but most data scientists gravitate towards mathematics.

Data visualization tools

The amount of data in the corporate world is huge. They require conversion to easier-to-understand formats. As a rule, people better perceive data in the form of graphs and charts.

Business acumen

Understanding the domain and the business tasks that the company faces seems to be a starting point for the success of one in this role.

Communication skills

Companies that are looking for a strong data scientist need a person who can clearly and freely convey technical results to non-techies, such as marketers or sales specialists.

Overview of data scientists' responsibilities

- Apply quantitative techniques from fields such as statistics, econometrics, optimization, and machine / deep learning toward the solution of important business problems from many areas of the automotive and mobility industry
- Utilize statistical approaches to build predictive models
- Enable evidence-based decision making by extracting insights from structured and unstructured data sets
- Identify new and novel data sources and explore their potential use in developing actionable business insights
- Explore new technologies and analytic solutions for use in quantitative model

development

- Design and develop customized interactive reports and dashboards
- Help maintain and improve existing models

What is a data analyst: profile overview

What is data analyst, exactly? According to Technopedia's data analyst definition, it's one who deciphers numbers and translates them into words to explain what data tells.

Landing a data analyst job doesn't require a strong math background. However, they can't fare well in this role without comprehension in statistics, data pre-processing, data visualization and EDA analysis, and of course, proficiency in Excel.

The most valued skills for data analysts are a deep understanding of the business area and presentation skills. Tech skills like programming language SQL, R, Python and machine learning are desirable but not a must.

What does a data analyst do?

Guided by business questions, data analysts (sometimes called big data analysts) explore

CATEGORIES

Expertise	▼
Industry	▼
Technology	▼

TABLE OF CONTENTS

We compared data engineer vs data scientist vs data analyst
Data engineer profile overview
Data engineers' scope of work
Skills for data engineer
Overview of data

data to glean information for questions posed by businesses.

Data analysts are engaged in retrieving relevant data from various sources and preparing it for further analysis. Basing on the analysis, a data analyst needs to make conclusions, complete reports and supports them with visuals. Along with reports, they need to explain what differences in numbers mean when looked at from month to month or across various audiences.

Thus, we can see that the scope of work of data analysts is aimed at analyzing and describing the past or previous strategies based on past or current data, while data scientists focus on creating forecasts to create the future strategies.

The scope of work for a data analyst:

- Collecting data basing on a specific request from leaders
- Familiarizing with the parameters of the data set (types of data, how it can be sorted)
- Pre-processing: making sure data is free of errors
- Interpreting data and analyzing ways it solves the business problem
- Drawing conclusions from the analysis

engineers
responsibilities

Data scientist profile
overview

Data scientists' scope of
work

Skills for data scientist

Overview of data
scientists'
responsibilities

Data analyst profile
overview

Data analysts' scope of
work

Skills for data analyst

Overview of data
analysts' responsibilities

How are these roles
connected?

Ask our specialis

- Visualizing and presenting the findings to the managers

Core skills for data analyst

Statistics

Having a background in different areas of statistics is absolutely necessary for a data analyst. The knowledge of stats makes exploring data easier and helps in avoiding logical errors. Additionally, data analysts can't do without tools of statistical analysis like SPSS, SAS, Matlab.

SQL

Similar to their counterparts, data analytics use databases to extract data for analysis from the data warehouse. This makes SQL a frequently used tool in the toolbox of these professionals.

Microsoft Excel

A deep understanding of Excel and its advanced features is vital for this role. Needless to say that it's more than just a spreadsheet. Its methods are go-to for quick analytics and working with light databases. However, learning R or Python is essential when working with big data sets.

Data visualization tools

Data analysts need to be able to create visual representations of complex data sets to make them easy for others to understand. To that end, they gain comprehension of available visualization tools such as Tableau, Infogram, QuickSight, Power BI and more.

Typical data analyst responsibilities

- Provide source-to-target mappings for data sets
- Perform testing and validation of data sets
- Collaborate with leaders and managers to determine and address data needs for various company projects
- Determine the meaning of data and explain how various teams and leaders can leverage it to improve and streamline their processes
- Write and apply data quality rules
- Create data quality dashboards and KPI reports about data
- Document structures and types of business data

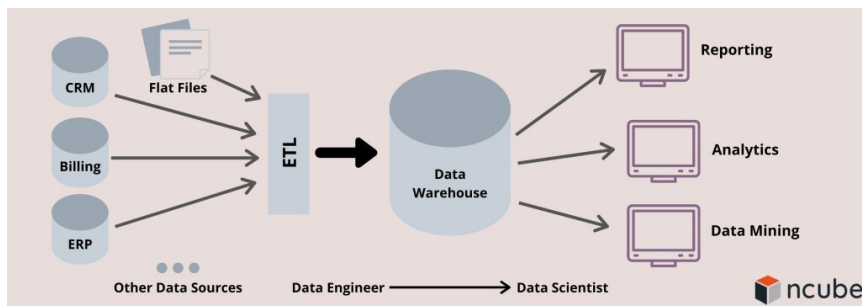
Read also: [What is the Difference Between Developer and Architect](https://ncube.com/blog/data-engineer-data-scientist-data-analyst-what-is-the-difference)

Data Engineer vs Data Scientist vs Data Analyst: How they all fit together?

Comparing the roles of data analyst vs data scientist, we can see that the first are focused on building reports and interpreting numeric data so that managers and business leaders can understand and use it. Data scientists deal with complex data from various sources to build prediction algorithms, while data engineers prepare the ecosystem so these specialists can work with relevant data.

Data engineer	Data scientist	Data analyst
Developing and maintaining database architecture that would align with business goals	Collecting and cleansing data used to train algorithms	Data pre-processing, collection and documentation
Building pipelines for communication between systems	Sifting through data to identify hidden patterns	Reporting based on previous or current data

Deployment of machine learning algorithms and models	Building predictive and prospective ML models	Statistical data analysis and interpretation
Data warehousing solutions	Refining business metrics by developing and testing hypothesis	Identifying data trends or patterns over certain periods of time



How data science engineer vs. data scientist vs. data analyst roles are connected

If we take a look at the difference between data engineers and data scientists in terms of skills, the first gravitate towards software development, DevOps and maths. Data scientists are usually strong mathematicians with a programming background and a good deal of business acumen. Data analysts are valued for statistics proficiency and also business acumen.

From our experience, we can say that at different companies these roles may incline towards a different set of skills. For example, a data scientist can use maths for 75%, machine

learning for 20% and deal with business needs 5% of the time. Of course, there are superstars that have a profound knowledge of all three fields but they are rare.

DATA

DEVOPS

Share:

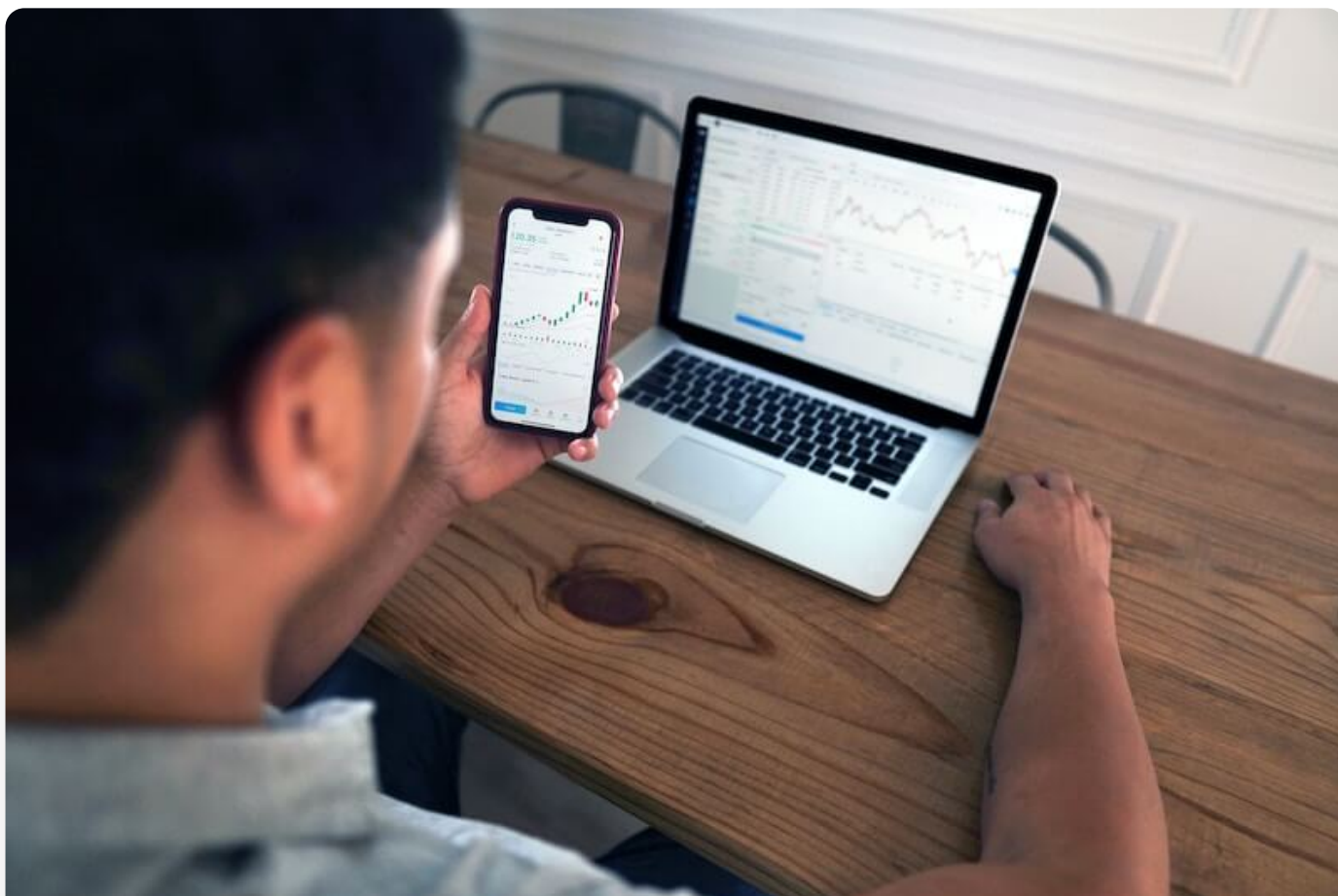
Recommended articles



Building Virtual Teams

Offshore Development Center vs. In-House Team: Which one is Best for You?

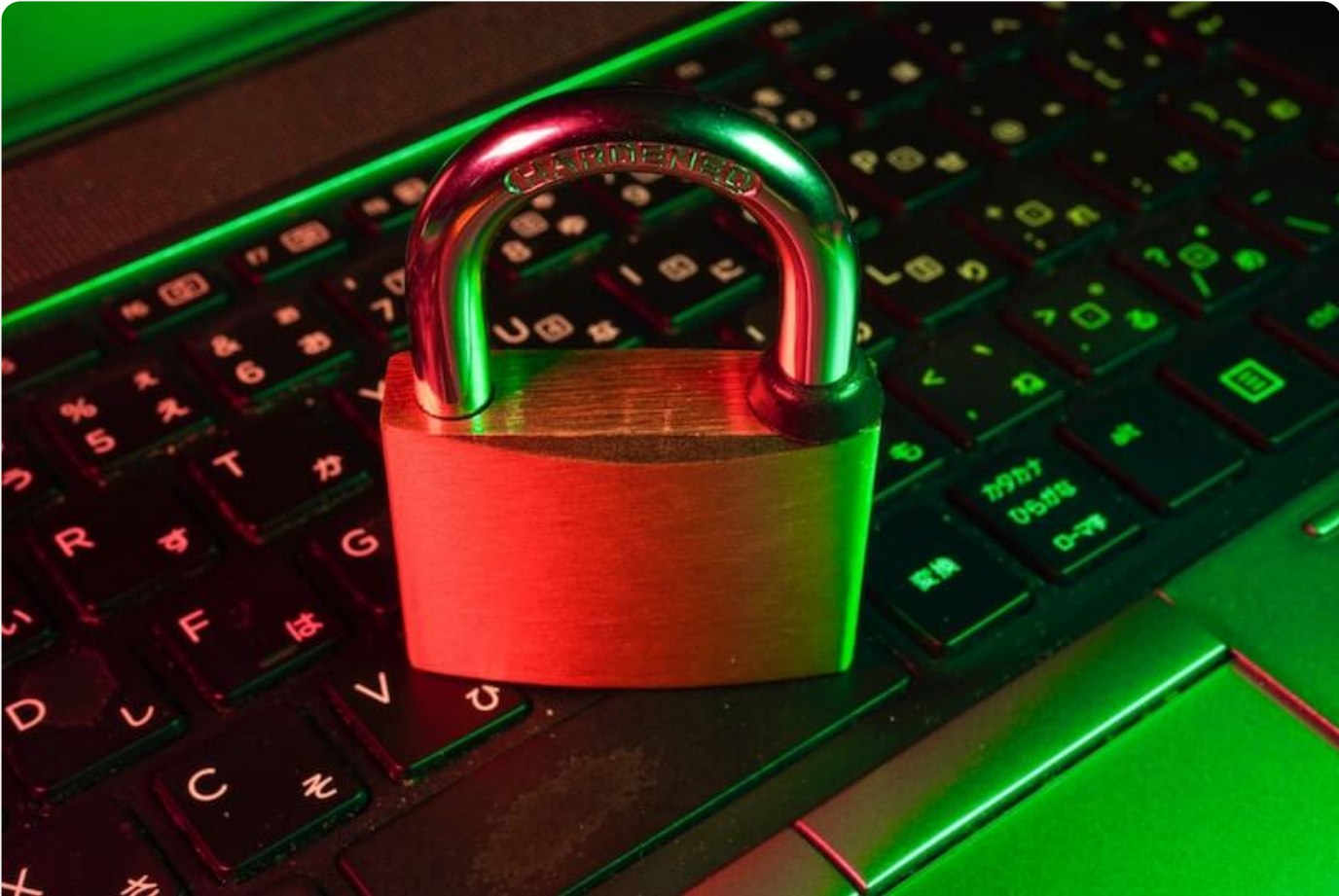
MARYNA DEMCHENKO | 18 JUN 2023



Investment applications

How to Create an Investment App

YULIA BORUSIUK | 14 JUN 2023



Fintech security

Security: Best Practices for Fintech Application Development

YULIA BORUSIUK | 12 JUN 2023

All articles

WHAT WE DO

Remote teams
Custom software
development

COMPANY

Europe

London, ECIV, UK

About us

HIRE DEVELOPERS

Java Android
Kotlin .Net
Python JavaScript
PHP iOS
Ruby Scala
Go

SUCCESS STORIES

Flightright
Encore Capital
FRST
Data-analytics
CrossEngage

BLOG

EVERBETICE

+380 97 982 08 07

Career

EXPERTISE

Ukrainian developers

Privacy Policy

Blockchain

Software development

Terms of use

IoT

Outsource

AI

Management

Building virtual teams