**Data Engineer Job Research Report  
Zihan Ma**

**Northeastern University College of Professional Studies - ALY 6000**



**Professor: Dr. Dee Chiluiza Reyes**

**02/07/2023**

**INTRODUCTION**

The goal of this report is to give out a basic explanation of the functionality, required skills, and career prospects. I hope. It will help the applicant to gain some general ideas about Data engineering after reading this report.

**DATA ENGINEER JOB REPORT**

**Job Overview:**

What does a Data Engineer do?

According to the Coursera website, The primary job of a data engineer is to collect the data and build the storage system to store, combine, sort, and finally clean the data, making them ready for input into data analysis software or access by data analyze agents for more in-depth investigate.

Why does a company need a Data Engineer?

In *What is a data engineer and what do they do?,* the author Ben Lutkevich explains that the amount of data is very variable with the size and type of the organization. In some data-intensive industries, such as healthcare, retail, and financial services, the importance of the data will be much higher than in the other types the industry. Also, the increase in the organization's size will significantly increase the complexity of the data architecture. Then it will be necessary to hire a Data Engineer to manage these data to make them understandable, well-sorted, and stored because the raw data is completely useless.

**The Type of Data of a Data Engineer Expects to Face Every Day**

Since a Data Engineer has to be the first step of the data processing process, Data Engineer will face various raw data that an organization produces daily. A Data Engineer has to master processing many different data types and utilize multiple software at the same time, depending on the type of organization. The dremio.com listed a table that gives an example of data types produced by different industries.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Source | Applications | Data Structures | Interface | | Vendors |
| Relational databases | (operational) HR, CRM, financial planning | Tables | SQL | Oracle, Microsoft SQL Server, IBM DB2 | |
| Relational databases (analytical) | Data warehouses, data marts | Tables | SQL | Teradata, Vertica, Amazon Redshift, Sybase IQ | |
| Point C JSON databases | Web, mobile, social | JSON documents | Proprietary language | MongoDB | |
| Key-value systems | Web, mobile, social | Objects | Proprietary language | Memcached, Redis | |
| Columnar databases | IoT, machine data | Column families | Proprietary language | Apache Cassandra, Apache HBase | |
| File systems | Data storage | Files | API | Hadoop Distributed File System (HDFS) | |
| Object stores | Data storage | Objects | API | Amazon S3, Azure Blob Store | |
| Spreadsheets | Desktop data analysis | Worksheets | API | Microsoft Excel | |

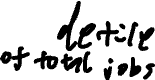
From the figure above, we can see that the data type that a Data Engineer might encounter is very. It means that it's a job that requires a lot of preparation and working experience in a similar industry.

**The number of jobs available in the US:**

By searching on multiple job-finding websites, I got the result of Data Engineer jobs available in the US listed in the table below.



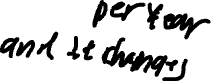
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job Types | glassdoor | linkedin | indeed | flexjobs |
| In-Person | **4734** | **163726** | **227825** | **3352** |
| Remote | **1482** | **40502** | **10891** | **2892** |
| Remote Present | **31%** | **25%** | **5%** | **86%** |



I'm more interested in remote jobs, so I additionally listed the percentage of remote positions on each website.

From the chart, we can see that although Flexjobs has a dominant percentage for remote positions, LinkedIn is the best website to apply for Data Engineer jobs for remote and in-person positions because it's a much larger base number.

**The Prediction of Future Jobs Number is w**



**Top-Five Companies for Working as a Data Engineer**

I compared the four job-finding websites above for rated of applying for a Data Engineer. The overall top five companies are:

* Accenture



* AT&T



* Wells Fargo



* Tata Consultancy Services



* Capital One



These five companies all have their advantages in some area including applier rating, wage, working quality, work load or qualifications.

**Heatmaps For Data Engineer**

**The Key Skill for Being a Data Engineer**



Include the skill listed in the second section. Data Engineer needs to master many skills because usually, Data Engineer will be the only person who deals with all the different types of data your company produces daily. Excispally in small companies –- you typically need to take the responsibility of the data analysts as well, because small companies tend to save every sent when it comes to the cost.

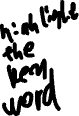


Coding: Coding will be a hard requirement for this job since you will need to process data using different languages, including but not limited to SQL, Python, Java, R, and Scala.

Relational and non-relational databases: Databases are also a hard required skill for a Data Engineer since the most common way to store an enormous amount of data is by using some kinds of databases. It will be the task you will dealing everyday as a data engerneer.



ETL (extract, transform, and load) systems: Building a database is not an ending to storing the data. You should also be familiar with how to combine/transport the data to the data warehouse using Xplenty, Alooma, Stitch, etc.



Data storage: There will be many data to store, so you will need to know how to find the most efficient and economical way to keep your company's data in a data warehouse.

Automation and scripting: You should make everything run by scripts, so you will have time to do the truly meaningful thing.



Machine learning: It is part of the job of a data analyst, but you have to have the base knowledge of machine learning, so it saves the work when data analysts need the data ready to put in the training. Or, as I mentioned before, being a data analyst and a data engineer at the same time will give you a significant advantage in finding jobs in some starting small companies.

Big data tools: Some industries are working on a much larger amount of data. As in the chart in the second section, you should know how to use Hadoop, MongoDB, and Kafka when you plan to apply for a job in these industries.



Cloud computing: As a famous future tending, more and more companies choose to use cloud services to reduce development cost. There will be many tricks can help you obtain the computing resources in a cheap way. Have a student identity is one of them, since the student usually can have cheaper or even free access to some mainstream cloud computing providers.

Data security: Data security is a very serious problem, It could cause you to lose your job or be in jail, depending on the seriousness level. It will be good to have a reliable plan to keep you away from the trouble. It will also be a good way to impress people during the conversations.

**Conclusions:**

As the future continues, people are generating more and more data in almost every industry. A professional in processing and utilizing these data will be more and more welcome in different companies. Acroading to an article on demo websites, the decision-making speed will be increased by 5 times with a data engineer, which proves the future of this Data Engineer.

**Bibliography**

*What is a data engineer?: A guide to this in-demand career*. Coursera. (2023, January 6). Retrieved February 7, 2023, from https://www.coursera.org/articles/what-does-a-data-engineer-do-and-how-do-i-become-one

Lutkevich, B., & Vaughan, J. (2021, March 16). *What is a data engineer and what do they do?* Data Management. Retrieved February 7, 2023, from https://www.techtarget.com/searchdatamanagement/definition/data-engineer

*What is Data Engineering? Responsibilities & Tools*. Dremio. (2022, January 27). Retrieved February 7, 2023, from https://www.dremio.com/resources/guides/data-engineering/

*Job search results that*. FlexJobs. (n.d.). Retrieved February 7, 2023, from https://www.flexjobs.com/search?search=Data%2BEngineer%2B&location=Remote

*Remote Data Engineer Jobs | Glassdoor*. (n.d.). Retrieved February 7, 2023, from https://www.glassdoor.com/Job/remote-data-engineer-jobs-SRCH\_KO0,20.htm

27,000+ Data Engineer jobs in United States (2,123 new). (n.d.). Retrieved February 7, 2023, from https://www.linkedin.com/jobs/search/?currentJobId=3431719056&f\_WT=2&keywords=data+engineer+

*USA jobs, employment | indeed.com*. (n.d.). Retrieved February 7, 2023, from https://www.indeed.com/q-USA-jobs.html