Data Engineer Job Brief

As a data engineer, you will be responsible for designing, developing, and maintaining the architecture and infrastructure needed to manage and process large and complex data sets. You will work with data scientists and business stakeholders to understand their needs and requirements and translate them into technical solutions that drive business outcomes. You will build pipelines that collect, store, process, and analyze data, ensuring its availability, accuracy, and security. You will use various tools and technologies, such as Hadoop, Spark, SQL, and NoSQL databases, to create scalable and efficient systems that support data-driven decision-making.

Data Engineer Roles and Responsibilities

This role involves the following responsibilities:

- Transforming the raw data collated into data systems.
- Testing data pipelines and observing how data is used.
- Providing carefully worked-out designs for creating big data architectures.
- Developing processes to organize, design, test, and maintain data systems.
- Exploring ways to acquire data and staying up-to-date on industry trends.
- Interpreting patterns and trends in data.
- Collaborating closely with teams to help solve any queries related to ongoing projects.
- Ensuring that the systems designed align with the objectives of the business.
- Dealing with incomplete or messy data and deciding how data should be managed, accessed, and stored.

Data Engineer Requirements and Skills

An ideal candidate for the role of Data Engineer should have the following skills:

- Bachelor's or Master's degree in Computer Science, Data Science, or a related field is preferred.
- Strong knowledge of programming languages and other technical skills.
- Experience in Big Data Stacks operations.
- Knowledge of SAP systems (ECC).
- Knowledge of how applications are used to manage both structured and unstructured data.
- Understanding of DevOps and automation scripting for repetitive tasks.
- Extensive hands-on experience in Unix/Linux.
- Expertise in big data technologies, such as Hadoop, Spark, and MapReduce.
- Experience with cloud-based data infrastructures like AWS (Amazon Web Services), Azure, or Google Cloud.
- Ability to integrate multiple data sources into one system.
- Understanding of various code versioning tools like Git, SVN, etc.
- Knowledge of Airflow, Flink and Hive, Presto, EC2, EMR, and ETL.
- Experience in PL/SQL and API Framework, Java Spring Boot, and Web Development.
- Knowledge of any RDBMS like PostgreSQL or Oracle, SQL Server, etc.

- Development, planning, and maintenance of data architecture.
- Great problem-solving and multitasking skills.