

# Big Data Engineer

**Experience Should be 5+**

## **Job Responsibilities:**

Create and maintain optimal data pipeline architecture.

Assemble large, complex data sets that meet functional / non-functional business requirements.

Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, if needed re-designing infrastructure for greater scalability, etc.

Build the infrastructure required for optimal extraction, transformation, and loading of data from a wide variety of data sources using SQL and/or AWS 'big data' technologies.

Build analytics tools that utilize the data pipeline to provide actionable insights, operational efficiency and other key business performance metrics.

Work with stakeholders including the Executive, Product, Data and Design teams to assist with data-related technical issues and support their data infrastructure needs.

Work with data and analytics experts to strive for greater functionality in our data systems.

Maintain INFOSec compliance needs and adhere to them when working with Data.

## **Must Have:**

You must be proficient in **Problem Solving and Data structures** since you got to find the best way of handling huge set of data.

Advanced working SQL knowledge and experience working with relational databases, query authoring (SQL) as well as working familiarity with a variety of databases.

Experience building and optimizing 'big data' data pipelines, architectures and data sets.

Experience performing root cause analysis on internal and external data and processes to answer specific business questions and identify opportunities for improvement.

Strong analytic skills related to working with unstructured datasets.

Build processes supporting data transformation, data structures, metadata, dependency and workload management.

Working knowledge of message queuing stream processing, highly scalable 'big data' data stores.

Experience supporting and working with cross-functional teams in a dynamic environment.

Understanding of AdTech domain is a big bonus.

They should also have experience using the following software/tools:

Experience with big data tools: Hadoop, Spark, Kafka, etc.

Experience with relational SQL (MsSQL, MySQL, PostgreSQL).

Experience with timeseries data bases (KairosDB, Druid) is a plus.

Experience of NoSQL data bases (Mongo, Cassandra, Druid) is a plus.

Experience with AWS cloud services.

Experience with stream-processing systems: Storm, Spark-Streaming, etc.

Experience with object-oriented/object function scripting languages: Java , Scala, etc.