**Data Engineer (SE) – Big Data technologies**

**Mandatory Skills**

* Having over all 5+ years of IT development experience in Database, Java or any programming languages.
* Minimum 3+ years of Developing and testing experience in the Apache Hadoop framework, HDFS, Map Reduce, Hive, Flume, Sqoop, Oozie, Spark, Spark Streaming and Scala.
* Experience in Big Data and batch/real-time ingestion and analytical solutions leveraging transformational technologies.
* Experience in Java, spring, Elasticsearch and Cassandra.
* Understand the source systems data, gather & able to transfer the data into HDFS as RAW and Decomposed layer (including writing Oozie workflow, coordinators, Sqoop and Flume).
* Ability to identify Unit testing.
* Experience and knowledge of a wide variety of testing methods and tools covering functional, performance, regression, system integration, smoke testing of Business Intelligence, ETL, BigData ( Ingestion, Modeling, Serving layer) and Business Analytical landscape.

**Additional Skills/Qualification**

* Having exposure on the Cloudera Manager, Cloudera Navigator, Hue, Shell Scripting, Apache Pig and Python
* Good understanding in XSD, XML and JSON schemas
* Integrating Hadoop with Oracle and NoSQL databases
* Experience in DevOps tools ( CI/CD)
* Use technical tools or native programming languages, to find, identify and diagnose solution defects based on data models or data extracts. Write advanced SQL queries to verify the integrity of databases including data transformations

**Competencies / Agile Skills**

* Transparency and Collaboration
* Self-driven and Self-learning
* Problem solving and decision making
* End-End solution
* Appetite to develop in multiple platforms
* Excellent attention to detail

**Data Engineer (SSE) – Big Data technologies**

**Mandatory Skills**

* Having over all 7+ years of IT development experience in Database, Java or any programming languages.
* Minimum 5+ years of working experience in the Apache Hadoop framework, HDFS, Map Reduce, Hive, Flume, Sqoop, Oozie, Spark, Spark Streaming and Scala.
* Experience in Big Data and batch/real-time ingestion and analytical solutions leveraging transformational technologies.
* Experience in Java, spring, Elasticsearch and Cassandra.
* Experience in data mining, data modelling, data processing, machine learning, natural language processing or information retrieval.
* Understand the source systems data, gather & able to transfer the data into HDFS as RAW and Decomposed layer (including writing Oozie workflow, coordinators, Sqoop and Flume)
* Capability to design and develop independently.
* Experience and knowledge of a wide variety of testing methods and tools covering functional, performance, regression, system integration, smoke testing of Business Intelligence, ETL, BigData ( Ingestion, Modeling, Serving layer) and Business Analytical landscape.

**Additional Skills**

* Good understanding about Lambda architecture and Big Data analytics framework.
* Having exposure on the Cloudera Manager, Cloudera Navigator, Hue, Shell Scripting, Apache Pig and Python
* Good understanding in XSD, XML and JSON schemas
* Integrating Hadoop with Oracle and NoSQL databases
* Use technical tools or native programming languages, to find, identify and diagnose solution defects based on data models or data extracts. Write advanced SQL queries to verify the integrity of databases including data transformations
* Degree in Software Engineering, Computer Information Systems, Commerce, Business Administration, Operations Research or a related field

**Competencies / Agile Skills**

* Transparency and Collaboration
* Self-driven and Self-learning
* Problem solving and decision making
* End-End solution
* Appetite to develop in multiple platforms

**Data Engineer (STE/PSE) – Big data technologies**

Big Data engineers to evaluate, design, develop, test and maintain big data solutions within organisations on Enterprise Data Hub (Helix) platform. A Big Data engineer builds large-scale data processing systems, is an expert in data modelling & processing solutions and should be able to work with the latest Big Data and NoSQL database technologies.

Because we work on the cutting edge of a lot of technologies, we need someone who is a creative problem solver, resourceful in getting things done, and productive working independently or collaboratively. This person would also take on the following responsibilities:

* Understand the source systems data, gather & able to transfer the data into HDFS as RAW and Decomposed layer (including writing Oozie workflow, coordinators, Sqoop and Flume).
* Exposure in handling to massive quantities of data by taking advantage of both batch and speed methods.
* Strong knowledge of and experience with statistics; potentially other advanced math as well and good working knowledge of SQL ability.
* Deep knowledge in data mining, data modelling, data processing, machine learning, natural language processing or information retrieval.
* Experience in processing large amounts of structured / unstructured data.
* Capability to design and document solutions independently, do Code reviews, Ensure standard Quality Assurance activities are completed on time.

**Mandatory Skills**

* Having over all 10+ years of IT development experience in Data, Database, JAVA or any programming languages.
* Minimum 5+ years of working experience in the Apache Hadoop framework, HDFS, Map Reduce, Hive, Flume, Sqoop, Oozie, Spark, Spark Streaming and Scala.
* Minimum of 5+ years of experience in Big Data and batch/real-time ingestion and analytical solutions leveraging transformational technologies.
* Minimum of 5+ years of experience in JAVA, spring, Elasticsearch and Cassandra.
* Good understanding about Lambda architecture
* Architects Big Data analytics framework.
* Translates complex functional and technical requirements into detailed architecture, design, and high performing software
* Codes, tests, and documents new or modified data systems to create robust and scalable applications for data analytics.
* Implements security and recovery tools and techniques as required.
* Ensures all automated processes preserve data by managing the alignment of data availability and integration processes.

**Additional Skills**

* Having exposure on the Cloudera Manager, Cloudera Navigator, Hue, Shell Scripting, Apache Pig and Python
* Good understanding in XSD, XML and JSON schemas
* Exposure on Graph Databases - Neo4j
* Technical Leading Skill and exposure as Enterprise Architect will be value added
* Worked on IT security and standardization for Hadoop related technologies
* Integrating Hadoop with Oracle and NoSQL databases
* Exposure to MSTR, BI.BO, Tabulo and other data visualizing tools

**Competencies / Agile Skills**

* Technology Doers / Full stack developers
* Drive Efficiency and leaders in cutting edge technologies
* Technical Capability development
* Transparency and Collaboration
* Self-driven and Self-learning
* Problem solving and decision making
* End-End solution
* Appetite to develop in multiple platforms

**Data Engineer – Technology stack**

* Data Ingestion (Batch/HDFS)
  + HDFS, Sqoop, Oozie, Flume, Oracle, Scripts, FTP and Hive
* Data Modelling and Processing (Batch/HDFS)
  + HDFS, Ooze, Hive, Pig, Python, Spark, Scala, Exadata and Data modelling
* Data Serving and (Speed Ingestion, Modelling & Processing)
  + HDFS, Kafka, Flume, Impala, Spark, Scala, Java, Spring, Spring boot, MSTR, Elasticsearch and Cassandra