**Ridhwaan Rahman**

☏ 999-999-9999  Email@gmail.com

**Objective**

Highly motivated, dedicated, and analytical professional, seeking a position to effectively utilize acquired knowledge of big data engineering honed from strong industrial foundation with over three years’ experience in big data.Worked with data lakes and Big Data ecosystems such as Hadoop, Horton Works and Spark. Wide range of knowledge surrounding Hadoop and associated applications (YARN, Hive, HDFS, Kafka & Spark, etc). Experience configuring a computer cluster, to facilitate the distributed property of data stored in HDFS. Wrote SQL queries for data validation of the reports and dashboards as necessary. Utilized outstanding communication and interpersonal skills in building working relationships with colleagues and other professionals. Born and raised in Texas.

**Professional Skills**

**Engineering Fundamentals**

* Understand and perform verification evaluations and analysis of safety software requirements, design, and coding.

**Problem-Solving and Analytical Skills**

* Exemplified adeptness in formulating strategic solutions to complex situations by keenly evaluating issues and problems.

**Project Development**

* Guidance and tracking of software development through the V&V life cycle. Preparation of reports for each V&V life cycle task describing the V&V task performed and the conclusions of the task and create the overall summary report summarizing all V&V tasks completed including tasks that were performed by others.
* Conducted meetings and have experience as SCRUM master

**Group Collaboration and Teamwork**

* Collaborate with a multi-discipline team to define the requirements interfaces between hardware components, microcontrollers and software, and the operational performance requirements of the integrated system.
* Proven success in team leadership, focusing on mentoring team members and managing tasks for efficiency

**Technical Acumen**

|  |  |
| --- | --- |
| Analysis Tools | MATLAB | LabView | PSpice | Xilinx | IntelliJ | Visual Studio| PyCharm| Hortonworks |
| Lab Tools | Signal Sources | Oscilloscopes | Spectrum Analyzers |
| Operating Systems | Windows 10 Professional | MAC | Linux |
| Software Applications  Programming Languages | Microsoft Office Suite | LaTeX | PopSQL | Flume | Kafka | Github | Hadoop | Hive | Hbase  Java | Python | Verilog | SQL | PySpark | Scala | Spark | MongoDB | SparkSQL |

**Work Experience**

**CNN | Atlanta, GA**

*Data Engineer |* Oct 2019–December 2020

* Created custom producer to ingest the data into Kafka topics for consumption by custom Kafka consumers
* Created custom Spark Streaming app to method clickstream events
* Developed workflows in Oozie to modify the tasks of loading information into HDFS and pre-processing with Hive.
* Integrating streams with Spark streaming for prime speed processing.
* Developed Spark code mistreatment Scala and Spark-SQL/Streaming for quicker process of data.
* Created modules for Spark streaming in information into information Lake mistreatment Storm and Spark.
* Configured Spark Streaming to receive real time information and store the stream information to HDFS
* Configured Spark-submit command to allocate resources to all the jobs across the cluster.
* Collected log information using custom engineered input adapters and Kafka.
* Performed maintenance, monitoring, deployments, and upgrades across applications that support all spark jobs.
* Partitioning and bucketing in ranges the log file information to differentiate information on a commonplace and combination supported business needs.
* Deployed the applying jar files into AWS EC2 instances.
* Developed a task execution framework on EC2 instances mistreatment using lambda and Airflow.
* A data pipeline that connects to the Internal weather search API, using Kafka (via a python file) to produce data.
* The producer sends stores the data on the broker once it has been serialized
* The necessary data is extracted and converted into a structured DataFrame (with custom schema)
* In this format the data is stored as a table in Hive, where analysis can be performed

**Wayfair | Boston, MA**

*Big Data Engineer |* Jan 2018–Oct 2019

* Optimized nodes and performed data analysis queries on Amazon Redshift
* Extracted metadata from Hive tables with Hive QL using Impala
* Experience in Importing and exporting data into HDFS and Hive using Sqoop.
* Developed multiple ETL processes using DataFrame
* In charge of branching internal repositories for feature creation over Git
* Populating database tables via AWS Kinesis Firehose and AWS Redshift.
* Implemented AWS Lambda functions to run scripts in response to events in Amazon Dynamo DB table or S3.
* Executed Hadoop/Spark jobs on AWS EMR using data stored in S3 Buckets.
* Learned and adapt to build CI/CD pipelines using tools like Git and Jenkins
* Developed custom aggregate functions using Spark SQL and performed interactive querying.
* Developed windowing functions to track CDC
* Connected various data sources and transferred data between them using Spark and various ETL tools.
* Spined AWS EMR to process data across Hadoop clusters

**American Airline | Fort Worth, TX**

*Cloud Big Data Developer |* Jan 2017–Dec 2018

* A data pipeline that connects to the Aero-Data Rest API, using Kafka (via a python file) to produce data.
* Develop and maintain build, deployment, and continuous integration systems in cloud computing environment
* Perform analysis, implementation, and performance tuning for engineered artifacts
* Configured a flume agent for ingestion of data from source APIs and store to HDFS
* Worked with different data formats like Avro, CSV, Parquet, JSON, and sequential files
* Built Hive views on top of the source data tables
* Utilized a cluster of ten Kafka brokers to handle replication needs and allow for fault tolerance
* Used Cloudera Manager for installation and management of single-node and multi-node Hadoop cluster
* Utilized HiveQL to query the data to discover transaction trends yearly
* Transformed data from unstructured to structured data frames for data analysis
* Load and transform large sets of structured, semi-structured, and unstructured data
* Writing Hive Queries for analyzing data in HDFS using Hive
* Loaded ingested data into Hive Managed and External tables.
* Developed multiple spark jobs using SQL context in spark and connected to Hive, created a Hive table and populated
* Experience in using Kafka as a messaging system to implement real-time Streaming solutions using spark steaming.
* Used spark to load batches of Data frames
* Assist in Install and configuration of Hive, Sqoop, Flume, Oozie on the Hadoop cluster
* Exercise judgment on how to effectively communicate highly technical and complex details through the use of visualization

**iHeartRadio | San Antonio, TX**

*Hadoop Developer |* July 2015 – Dec 2016

* Installed and configured a Flume agent to ingest data from a Rest API
* Used Sqoop to migrate data from SQL to HDFS
* Involved in transforming data from legacy tables to HDFS, and HBase tables using Spark
* Developed Oozie workflows to run multiple Hive jobs
* Developed and implemented various methods to load data into HIVE tables from HDFS and Local File System.
* Migrated complex MapReduce scripts to Apache Spark code.
* Used SparkSQL module to store the data on Hive
* Designed and developed ETL workflows using Scala and Python for processing structured and unstructured data using Spark.
* Experienced in running Hadoop streaming jobs to process XML data
* Implemented security measures over a Hadoop Cluster
* Prepared scripts to automate ingestion of data in Python as needed through various sources such as API and save it to HDFS.
* Wrote Hive Queries for analyzing data in Hive warehouse using Hive Query Language.
* Installed flume in terminal and configured the source, channel and sink.

**Education**

**Master of Science in Electrical Engineering:** *The University of Texas at Dallas* May 2016

**Bachelor of Science in Electrical Engineering:** *The University of Texas at Dallas* May 2015

MAGNA CUM LAUDE | COLLEGIUM V HONORS|DEAN’S LIST HONOREE (FIVE SEMESTERS) |ACADEMIC DISTINCTION SCHOLAR