**Triveni Papineni**

**trivenipapineniunique@gmail.com**

**+19046892449**

**PROFESSIONAL SUMMARY**

* **8+ years** of professional experience as a **Sr. Data Engineer and Sr. Hadoop Developer in On Prem and Cloud environments.**
* Expertise in AWS EMR ETL pipeline development with **Lambdas, Step Functions, CloudWatch, Athena, Glue**.
* Strong hands on with **Apache Kafka, Spark Streaming** using **Scala** and **Python and AWS Kinesis Steaming.**
* Experience in analyzing data using **Hive SQL**, **Pig Latin** and custom **MapReduce** programs in **Java** and **Python.**
* Experience with leveraging Hadoop ecosystem components including **Pig** and **Hive** for data analysis, **Sqoop** for data migration, **Oozie** for scheduling and **HBase** as a NoSQL data store.
* Good Exposure on Apache Hadoop **Map Reduce** programming, **PIG Scripting** and Distribute Application and HDFS.
* Strong experience in migrating other databases to **Snowflake.**
* Understanding of **Snowflake cloud technology.**
* Assisted in designing & developing **data lake and ETL** using python and Hadoop ecosystem.
* Worked in **developing data lake for the GBT reporting** team.
* Professional knowledge of **AWS Redshift.**
* Experience in **Tableau** Desktop 10.x/9.x/8.x, Tableau Reader, Tableau Prep and experienced in analysis, modeling, design and development of Tableau reports and dashboards for analytics.
* Hands on experience on Tableau Desktop versions 7/8.2, Tableau Reader and Tableau Server.
* Experience in **NoSQL** database **MongoDB** and **Cassandra**.
* Designed, build and managed **ETL data pipeline**, leveraging Airflow, python, **DBT**, Stitch Data **and GCP Solutions.**
* Mentor and guide analyst on building purposeful analytics tables in **DBT** for cleaner schemas.
* Model, lify and shift custom SQL and transpose LookML into **DBT** for materializing incremental views.
* Experience in importing and exporting data using **Sqoop** from **HDFS** to Relational Database Systems and vice-versa.
* Implemented cluster services using Kubernetes and Docker to manage local deployments in Kubernetes by building a self-hosted Kubernetes cluster using **Terraform** and Ansible and deploying application containers.
* Implemented Spring boot **microservices** to process the messages into the **Kafka** cluster setup.
* Experience in Designing, Architecting and implementing scalable cloud-based web applications using **AWS** and **GCP**.
* Set up a **GCP Firewall rules** in order to allow or deny traffic to and from the VM's instances based on specified configuration and **used GCP cloud CDN** to deliver content from **GCP** cache locations drastically improving user experience and latency.
* **Developing Spark** programs using **Scala API's** to compare the performance of Spark with **Hive and SQL**.
* Implemented Spark using **Scala and Spark SQL** for faster testing and processing of data.
* Imported data from **AWS S3** and into **Spark RDD** and performed transformations and actions on RDD's.
* Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive
* Experienced in deployment of **Hadoop Cluster using Puppet tool.**
* Experience in Hadoop **Shell commands,** writing **MapReduce Programs,** verifying managing and reviewing Hadoop Log files.
* Proficient in configuring **Zookeeper, Cassandra & Flume** to the existing **Hadoop cluster.**
* In depth knowledge of **Job Tracker, Task Tracker, Name Node, Data Nodes and MapReduce concepts.**
* Experience in understanding the security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure.
* Experience in Big Data analysis using PIG and HIVE and understanding of **SQOOP and Puppet.**
* Good understanding of **HDFS Designs, Daemons, federation** and **HDFS** high availability **(HA).**
* Experience in installation, configuration, supporting and managing - Cloudera Hadoop platform along with **CDH3&4** clusters.
* Experienced in developing **MapReduce** programs using Apache Hadoop for working with Big Data.
* Experience in developing customized UDF’s in java to extend **Hive** and **Pig Latin** functionality.
* Good experience in implementing and setting up standards and processes for Hadoop based application design and implementation.
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice versa.
* Experience using middleware architecture using Sun Java technologies like **J2EE, JSP, Servlets, and application servers like Web Sphere and Web logic.**
* Familiarity working with popular frameworks likes **Struts, Hibernate, Spring MVC and AJAX.**
* Experience in Object Oriented language like **Java and Core Java.**
* Experience in creating web-based applications using **JSP and Servlets**.
* Experience in managing Hadoop clusters using Cloudera Manager Tool.
* Very good experience in complete project life cycle (design, development, testing and implementation) of Client Server and Web applications.
* Extensive experience working in Oracle, DB2, SQL Server and My SQL database.
* Hands on experience in VPN, **Putty**, **WinSCP**, VNCviewer, etc.
* Hands on experience in application development using Java, RDBMS, and Linux shell scripting.
* Ability to adapt to evolving technology, strong sense of responsibility and accomplishment.
* Developed restful Web-services using Grails framework in Python.
* Develop framework for converting existing PowerCenter mappings and to **PySpark** Jobs.
* Create **Pyspark** frame to bring data from DB2 to Amazon S3.
* Translate business requirements into maintainable software components and understand impact (Technical and Business)
* Utilized **GraphQL APIs** to connect to **AWS Aurora**
* Provide guidance to development team working on **PySpark** as ETL platform

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Data Technologies** | HDFS, MapReduce, Hive, Impala, Pig, Sqoop, Flume, Oozie, Zookeeper, Ambari, Hue, Spark, PySpark, Strom, Talend, Ganglia, Glue, AWS Redshift |
| **Operating System** | Windows, Linux |
| **Languages** | Java, J2EE, Scala, Python, SQL, PL/SQL, Shell Script, Iceberg, GraphQL |
| **Project Management / Tools** | MS Project, MS Office, TFS, HP Quality Center Tool |
| **Front – End** | HTML, JSTL, DHTML, JavaScript, CSS, XML, XSL, XSLT |
| **Databases** | MySQL, Oracle 11g/10g/9i, SQL Server, Teradata |
| **NoSQL Databases** | HBase, Cassandra |
| **File System** | HDFS |
| **Reporting Tools** | Jasper Reports, Tableau |
| **Cloud Technologies** | Snowflake, Snow SQL, SnowpipeAWS |
| **IDE Tools** | Eclipse, NetBeans |
| **Data ware Housing** | EDW relational/dimensional modeling, star schema/cube design, MDM hub design, ETL and BI design |
| **Data Modeling** | Relational, dimensional, Data Vault Modeling, MDM hub, ODS, Structured/Unstructured, OLTP |
| **Application Server** | Apache Tomcat, Web Logic |

**EDUCATION**

* **Bachelor of Technology** from GVR&S College of Engineering & Technology, Ap, India

**PROFESSIONAL EXPERIENCE**

**Charles Schwab, Remote June 2021 – Present**

**Sr Hadoop Developer/Data Engineer**

**Responsibilities:**

* Designing and Developing the High volume and performance Data pipelines to support the Business Intelligence and Data Science/AI Teams.
* Imported the data from different sources like AWS S3, LFS into Spark RDD.
* Developed pipelines to receive message through AWS API Gateway and write to Kinesis Stream.
* Worked and learned a great deal from AWS cloud services like EC2, S3, EBS, RDS and VPC.q
* Migration of Datawarehouse tables from On Prem to AWS and Azure clouds.
* Developed Spark scripts by using Scala & PySpark for Batch and Streaming ETL Pipelines.
* Implemented Iceberg Data Lake and Trino/Presto for querying.
* Developing Data pipelines on AWS EMR using PySpark and AWS Lambdas, Steps Functions, CloudWatch.
* Created and Loaded AWS Athena tables with Glue.
* Experience with Snowflake Multi - Cluster Warehouses.
* Received Data to Kinesis Steam in streaming mode and loaded that to S3.
* Performance tuning of complex SQL Queries, PySpark code, Streaming jobs.
* Developed ETL workflows using MS SQL Server 2016 to extract, transform, and load data from various sources into Azure SQL Data Warehouse for downstream analysis.
* Hands on experience on Tableau Desktop versions 7/8.2, Tableau Reader and Tableau Server.
* Created Oracle analytics reports with complex calculations and worked on Ad-hoc reporting.
* Developed JSON Scripts for deploying data Pipelines in Azure Data Factory (ADF) that process the data using SQL Activity
* Used Spring **Kafka API** calls to process the messages smoothly on Kafka Cluster setup.
* Utilized Hive/Hadoop to store and process large volumes of structured and unstructured data.
* Collaborated with data engineers and data scientists to design and develop big data solutions that meet business requirements and provide valuable insights.
* Designed and optimized SQL queries to extract and analyze data from large databases and wrote Python scripts to automate data processing tasks.
* Worked on end-to-end ETL process, starting from data ingestion to data transformation and finally loading the data into target systems.
* Hands on experience on Tableau Desktop versions 7/8.2, Tableau Reader and Tableau Server.

**Technology / Tools:** Spark, Spark-Streaming, Spark SQL, AWS EMR, Lambdas, Step functions, MS SQL Server, Scala, Azure, Python, JSON Script, Python Script, Tableau, Iceberg, ETL, Kafka, Hive/Hadoop, Sqoop

**TIAA, TX July 2019 - May 2021**

**Hadoop Developer/Data Engineer**

**Responsibilities:**

* Developed and implemented MapReduce jobs to process large data sets in Hadoop Distributed File System (HDFS)
* Designed and maintained Hive tables and partitions for efficient querying and analysis of big data
* Used Hive to do transformations, joins, filter and some pre-aggregations after storing the data to HDFS
* Implemented PySpark logic to transform and process various formats of data like XLS, XLS, JSON, and TXT. Elaborated Python Scripts to fetch/get S3 files using Boto3 module
* Utilized Sqoop for importing and exporting data between Hadoop and relational databases like MySQL and Oracle
* Configured and managed Kafka clusters to stream real-time data into Hadoop ecosystem for processing and analysis.
* Leveraged Spark, Python, and Kafka to build high-performance data processing systems that can handle large volumes of data in real-time.
* Worked with Spark to develop and deploy large-scale data processing and machine learning applications on Hadoop cluster
* Used PySpark to write data processing and transformation scripts in Python, leveraging the power of Spark's distributed computing framework
* Created custom Hadoop data processing pipelines using Pig Latin scripts for data transformation and cleansing
* Designed and implemented Hadoop security frameworks to ensure data privacy and integrity, including Kerberos and Apache Ranger
* Designed and implemented big data solutions on AWS using Spark-Streaming to process and analyze real-time data streams, and Spark SQL to query and transform large datasets stored in S3
* Worked with Hadoop ecosystem tools like Oozie and Zookeeper for scheduling and coordination of data processing workflows and distributed applications
* Worked with Various Tools such as Maven, Gradle and SBT for development Purposes
* Optimized Map Reduce Jobs to use HDFS efficiently by using various compression mechanisms

**Technology / Tools**: Hadoop, MapReduce, HDFS, Sqoop, Hive, Spark, Pyspark, Oozie, Zookeeper, MySQL, Oracle, Kafka, Python, Pig Latin, AWS, Amazon S3, Kerberos, Apache Ranger, Maven, Gradle, and SBT

**Johnson Controls, Milwaukee, WI** **Oct 2017 - June 2019**

**Hadoop Developer**

**Responsibilities:**

* Installed and configured Hadoop MapReduce HDFS; developed multiple MapReduce jobs in Java for data cleaning/preprocessing
* Created various hive external tables, staging tables and joined the tables as per the requirement; implemented static partitioning, dynamic partitioning and bucketing in hive using internal and external tables
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala
* Optimized data processing performance by tuning distributed computing frameworks, optimizing SQL queries, and profiling data processing workflows
* Installed and configured Pig as well as writing Pig Latin scripts
* Created MapReduce job using Pig Latin dialect
* Worked on Big Data infrastructure for batch processing; built scalable distributed data solutions using Hadoop
* Developed Kafka topics for loading the data from sourced HDFS system
* Build large-scale batch and real-time data processing pipe and spark streaming
* Imported data using Sqoop to load from MySQL into HDFS on a regular basis
* Crafted Hive queries for data analysis to meet business requirements
* Constructed Hive tables and worked on them using Hive QL
* Imported and exported data into HDFS and Hive using Sqoop
* Created a custom Filesystem plug-in for Hadoop so it could access files on Data Platform
* Conceived and implemented Map reduce-based large-scale parallel relation learning system
* Setup and benchmarked Hadoop/HBase clusters for internal use
* Implemented data ingestion and handling clusters in real time processing using Apache Storm and Kafka

**Technology / Tools**: MapReduce, HDFS, SQL, Hive, Spark RDDs, Python, Scala, Hadoop distribution of Horton Works, Cloudera, Pig, HBase, Kafka, Linux, XML, MYSQL, Workbench, Java, 6 Eclipse, Oracle 10g PL/SQL SQL PLUS Sub Version, Cassandra

**JDA Software, Hyderabad, India June 2015 - Sept 2017**

**Hadoop Developer**

**Responsibilities:**

* Setup and configured Hadoop MapReduce HDFS; developed multiple MapReduce jobs in Java for data cleaning and preprocessing.
* Experience installing, configuring and using Hadoop ecosystem components.
* Administration experience installing upgrading CDH3 Pig Hive HBase installation.
* Imported and exported data into HDFS and Hive using Sqoop.
* Experience creating job flows.
* Knowledge in performance troubleshooting, managing and reviewing Hadoop log files, and tuning Hadoop clusters.
* Participated in the development/implementation of Cloudera Hadoop environment.
* Loaded and transformed large sets of structured, semi structured, and unstructured data.
* Supported Map Reduce Programs that were running on the cluster.
* Involved with loading data from UNIX file system to HDFS.
* Installed and configured Hive as well as creating Hive UDFs
* Conducted the creation and loading of Hive tables with data, then writing queries which ran in map reduce mode internally.
* Setup CDH3 Hadoop cluster on CentOS platform
* Executed cluster commissioning decommissioning data node provisioning node recovery capacity planning and slots configuration.
* Constructed HBase tables to store variable data formats of Personal Identification Information (PII) data from various portfolios.
* Implemented best income logic using Pig scripts.
* Loaded and transformed large sets of structured, semi structured, and unstructured information using Zookeeper cluster coordination services.
* Exported the analyzed data into relational databases using Sqoop for visualization and reporting to the Business Intelligence team.
* Assisted with setting up QA environment and configuring scripts using Pig and Sqoop

**Technology / Tools:** Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Java, SQL, Data meter, Zookeeper, Sqoop, CentOS, SOLR, CDH3 Hadoop cluster