**Vinay**

**Sr. Bigdata AWS Cloud Engineer**

**Email:** [gvinayk1826@gmail.com](mailto:gvinayk1826@gmail.com)

**Phone: +1 (304) 808-1826**

**SUMMARY:**

* 9+ years of experience in IT which includes Analysis, Design, Data Mapping, Datastage, Development of Big Data using Hadoop, AWS, Python, data Lake, Scala, GCP, Machine Learning and development of web applications using data base and data warehousing development using My SQL, Oracle.
* Experience in creating applications using Spark with Python.
* Experienced in Apache Spark and developing data processing and analysis algorithms using Python.
* Strong working experience on Big Data Analytics with hands on experience in installing, configuring, and using ecosystem components like Hadoop Map reduce, HDFS, HBase, Zookeeper, Hive, Sqoop, Pig, Flume, Cassandra, Kafka and Spark.
* Building data processing triggers for Amazon S3 using AWS Lambda functions with Python.
* Worked on the Spark SQL and Spark Streaming modules of Spark and used Scala and Python to write code for all Spark use cases.
* Strong knowledge of Data Extraction, Data Mapping, Data Conversion and Data Loading process using UNIX Shell scripting, SQL, PL/SQL, and SQL Loader.
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data lake store using Azure Data factory.
* Good Understanding of Hadoop architecture and Hands on experience with Hadoop components such as Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce concepts and HDFS Framework.
* Expertise in preparing Interactive Data Visualization's using Tableausoftware from different sources.
* Strong experience in analyzing large amounts of data sets writing PySpark scripts and Hive queries.
* Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
* Extensively worked on AWS services like EC2, S3, EMR, FSx, Lambda, Cloud watch, RDS, Auto scaling, Cloud Formation, SQS, ECS, EFS, DynamoDB, Route53, Glue etc.
* Experience in Data load management, importing & exporting data using SQOOP&FLUME.
* Experience in dealing with log files to extract data and to copy into HDFS using flume.
* Experience in integrating Hive and Hbase for effective operations.
* Experience in Impala, Solr, MongoDB, HBase and Spark, Kubernetes.
* Expertise in Waterfall and Agile - SCRUM methodologies.
* Experienced with code versioning and dependency management systems such as Git, SVT, and Maven.
* Experience in Testing and documenting software for client applications.
* Good experience in using Data Modelling techniques to find the results based on SQL and PL/SQL queries.
* Experience working with different databases, such as Oracle, SQL Server, MySQL and writing stored procedures, functions, joins, and triggers for different Data Models.
* Experienced in handling different file formats like Text file, Avro data files, Sequence files, Xml and Json files.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Cloud Technologies | Azure, AWS |
| Azure Skills | Azure Data Lake, Data Factory, Databricks, Azure SQL Databases, Azure SQL Datawarehouse |
| AWS Skills | S3, EC2, EMR, Glue, Athena, Aurora, SNS, Crawler |
| Big data Skills | HDFS, Hive, MapReduce, Pig, Sqoop, Flume, Oozie, Hadoop distribution, and Hbase, Spark, Spark Streaming, Yarn, Zookeeper, Kafka, ETL |
| Programming Languages | Python, Spark, R, Scala, Terraform, Angular |
| Databases | Aurora DB, Oracle, MySQL |
| Job Scheduler Tools | Airflow |
| CICD | Bit Bucket, Jenkins |
| IDE | PyCharm, Visual Studio |

**PROFESSIONAL EXPERIENCE:**

***Role: SrBigdata AWS Cloud Engineer February 2022 –Present***

***Client: CapitalOne, Chicago, IL***

**Responsibilities:**

* Created various data pipelines usingSpark, Scala and SparkSQL for faster processing of data
* Created data mappings based on the business requirements and translated them to technical solutions
* Designed number of partitions and replication factor for Kafka topics based on business requirements and worked on migrating MapReduce programs into Spark transformations using Spark and Scala, initially done using python (PySpark)
* Good understanding of Spark Architecture with Databricks, Structured Streaming. Setting Up AWS and Microsoft Azure with Databricks, Databricks Workspace for Business Analytics, Manage Clusters In Databricks, Managing the Machine Learning Lifecycle
* Performed Analysis on Data mapping between different data models
* Utilized machine learning algorithms such as linear regression, multivariate regression, Naive Bayes, Random Forests, K-means, & KNN for data analysis
* Created Data Quality Scripts using SQL and Hive to validate successful das ta load and quality of the data. Created various types of data visualizations using Python and Tableau.
* Extensively worked with Avro and Parquet, XML, JSON files and converted the data from either format Parsed Semi Structured JSON data and converted to Parquet using Data Frames in PySpark.
* Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis
* Worked with a team to migrate from Legacy/On prem environment into AWS.
* Stored and retrieved data from data-warehouses using Amazon Redshift.
* Experienced in analyzing and Optimizing RDD's by controlling partitions for the given data
* Experienced in writing live Real-time Processing using Spark Streaming with Kafka
* Used HiveQL to analyze the partitioned and bucketed data and compute various metrics for reporting
* Worked with querying data using SparkSQL on top of Spark engine.
* Involved in managing and monitoring Hadoop cluster using Cloudera Manager.
* Used Python and Shell scripting to build pipelines.
* Developed data pipeline using SQOOP, HQL, Spark and Kafka to ingest Enterprise message delivery data into HDFS.
* Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive.
* Developed and executed a migration strategy to move Data Warehouse from an Oracle platform to AWS Redshift.
* Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Assisted in creating and maintaining technical documentation to launching HADOOP Clusters and even for executing Hive queries and Pig Scripts.
* Assisted in Cluster maintenance, cluster monitoring, adding and removing cluster nodes and installed and configured Hadoop, Map Reduce, HDFS, developed multiple Map Reduce jobs in SQL for data cleaning and pre-processing.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS.
* Automated and monitored complete AWS infrastructure with terraform.
* Created data partitions on large data sets in S3 and DDL on partitioned data.
* Converted all Hadoop jobs to run in EMR by configuring the cluster according to the data size.

***Role: Bigdata Engineer February 2019 – December 2021***

***Employer: OathMap LLP, Hyderabad, India***

***Client: XILINX (Contractor)***

**Responsibilities:**

* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala
* Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Hive.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Involved in Migrating the platform from Cloudera to EMR platform
* Extensively used open source tools - R Studio(R) and Spyder (Python) for statistical analysis and building the machine learning
* Extensively involved in developing Restful API using JSON library of Play framework.
* Developed Storm topology to ingest data from various source into Hadoop Data Lake.
* Developed web application using HBase and Hive API to compare schema between HBase and Hive tables.
* Connected to AWS s3 using SSH and ran spark-submitjobs
* Developed Python Script to import data SQL Server into HDFS & created Hive views on data in HDFS using Spark.
* Expert in Troubleshooting MapReduce Jobs.
* Created scripts to append data from temporary HBase table to target HBase table in Spark.
* Developed complex and Multi-step data pipeline using Spark.
* Worked on Big Data Integration and Analytics based on Hadoop, SOLR, Spark, Kafka, Storm and web Methods technologies.
* Populated HDFS and Cassandra with huge amounts of data using Apache Kafka.
* Involved in creating ETL flow using Pig, loading with data and writing Pig Latin queries which will run internally in Map Reduce way.
* Involved in writing Unix/Linux Shell Scripting for scheduling jobs and for writing pig scripts and hive QL.
* Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run MapReduce jobs in the backend.
* Assisted in exporting data into Cassandra and writing column families to provide fast listing outputs.
* Used Zookeeper for providing coordinating services to the cluster.
* Worked with Hue UI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Developed and designed system to collect data from multiple portal using kafka and then process it using spark.

***Role:Bigdata Engineer December 2017 – February 2019***

***Employer: OathMap LLP, Hyderabad, India***

***Client:AT&T (Contractor)***

**Responsibilities:**

* Configured real-time streaming pipeline from DB2 to HDFS using Apache Kafka.
* Followed Agile Scrum methodology that included iterative application development, weekly Sprints and stand up meetings
* Acquire data from Transactional source systems to Redshift data warehouse using spark and AWS EMR
* Creating External and Managed Hive tables and working on them using HiveQL.
* Analysis &documenting of new/Existing system.
* Validated the Map reduce, Pig, Hive Scripts by pulling the data from the Hadoop and validating it with the data in the files and reports.
* Provided a new Web Service and Client using Spring-WS to get the alternate contractor details.
* Used JIRA for the issue tracking and bug reporting.
* Expertized in implementing Spark using Scala and Spark SQL for faster testing and processing of data responsible to manage data from different sources.
* Implemented Spark using Scala and Spark SQL for faster testing and processing of data
* Collaborated with the team using GIT, GitHub, SourceTree version control platform.
* Implemented AWS solutions using EC2, S3, RDS, EBS, Elastic Load Balancer, Auto-scaling groups
* As part of another project I made use of Amazon Web Services (AWS).
* Bottle micro-framework implemented with REST API and MongoDB (NoSQL) as back end database.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Developed Spark code to using Scala and Spark -SQL for faster processing and testing.

***Role: Data Engineer January 2016 – November 2017***

***Employer: OathMap LLP, Hyderabad, India***

***Client: Headstrong (Contractor)***

**Responsibilities:**

* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala
* Developed Real time data processing applications by using Scala and Python and implemented Apache Spark Streaming from various streaming sources like Kafka, Flume and JMS
* Developed multiple POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Analysed the SQL scripts and designed the solution to implement using Scala.
* Developed analytical component using Scala, Spark and Spark Stream.
* Developing UDFs in Python for hive and pig and worked on reading multiple data formats on HDFS using Scala.
* Troubleshoot and debug Hadoop ecosystem run-time issues.
* Created POC to store Server Log data in MongoDB to identify System Alert Metrics.
* Monitored Hadoop cluster job performance, performed capacity planning and managed nodes on Hadoop cluster.
* Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop.
* Performed analysis on the unused user navigation data by loading into HDFS and writing MapReduce jobs. The analysis provided inputs to the new APM front end developers and lucent team.
* Wrote MapReduce jobs using Python.
* Wrote Pig scripts to run ETL jobs on the data in HDFS and further do testing.
* Used Hive to do analysis on the data and identify different correlations.
* Involved in HDFS maintenance and administering it through Hadoop API.
* Design and implement MapReduce jobs to support distributed processing using python, Hive and Apache Pig.
* Create Hive external tables on the MapReduce output before partitioning, bucketing is applied on it.
* Loaded the load ready files from mainframes to Hadoop and files were converted to ASCII format.
* Configured Hive Server (HS2) to enable analytical tools like Tableau, Qlikview and SAS to interact with Hive tables**.**

***Role: Java Developer September 2013- December 2015***

***Client: OathMap LLP, Hyderabad, India***

**Responsibilities:**

* Responsible for analysis of business user needs, documentation of requirements and translation into system requirement specifications
* Responsible for system design, architecture, and implementation of Blockchain-based inter-carrier dispute resolution and settlement software solutions.
* Followed SDLC-Agile/Scrum methodology in the application development.
* Design and Develop Java & J2EE based core components using Core Spring Framework and other sub-projects like Spring Security, Spring data, MVC, Spring boot (Micro services), Spring Cloud, Spring Batch
* Developed RESTful microservices using Java Spring, AWS S3 and PostgreSQL to extract, transform and load data to a NodeJS RESTful API.
* Developed views controller and model components implementing JavascriptMVC.
* Maintained and extended a NodeJS RESTful API used for secure communication with the Hyperledger Fabric Blockchain network using Express and the Hyperledger Fabric SDK.
* Used Angular 8 to implement Single Page Application (SPA) architecture and client side MVC.
* Used CSV Reader to convert the csv files into JSON format which required to create as Task in SQL Database.
* Created Angular 8 Directives, Components, Pipes, and Injectables across the whole system, such as to build the custom filter.
* Worked with Angular 2 and typescript as part of migration from Angular and vanilla javascript to Angular 2 and React.
* Developed Single Page applications using AngularJS using components, directives using TypeScript.
* Web Interface is designed using J2EE, XML/SOAP, WSDL, Web Services, JDBC and EJB.
* Integrated Angular Client applications with newly created and modified Java and NodeJS RESTful API endpoints.