**Ganashree Gowrikoppal Mallegowda   
Email:** [**ganashree.gm18@gmail.com**](mailto:ganashree.gm@gmail.com)

**Contact no: 469-846-8345**

**Big Data Engineer**

**SUMMARY:**

* Having 10+ years of IT experience **in Analysis, design, development, implementation, maintenance**, **release** and **support** with experience in developing strategic methods for deploying big data technologies to efficiently solve Big Data processing requirement.
* Hands on experience on **BIG DATA** using **HADOOP** framework and related technologies such as **HDFS, GCP, HBASE, MapReduce, Spark, HBase, Databricks**, **HIVE, PIG, FLUME, OOZIE, SQOOP, and ZOOKEEPER.**
* Expert level experience in designing, building and managing applications to process large amounts of data in a Hadoop/DevOps (GCP) ecosystem.
* Experience in data analysis using **HIVE, PIG LATIN, HBASE** and custom Map Reduce programs in Java.
* Expert level experience in designing, building and managing applications to process large amounts of data in a **Hadoop/DevOps**  ecosystem
* Extensive experience with performance tuning applications on **Hadoop** and configuring Hadoop systems to maximize performance
* Design and Develop ETL Processes in AWS **Glue** to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Experience in writing custom UDFs in **JAVA** and **SCALA** for **HIVE** and **PIG**to extend the functionality.
* Experience on **Databricks**, scripting language like Scala, Shell, PowerShell
* Worked with azure **databricks** notebooks to validate the inbound/out bound from an external source like Amperity
* Experience with Cloudera and Horton works distributions.
* Around 3-year 2 months’ experience on **SPARK, SCALA, HBASE** and **KAFKA**.
* Developed analytical components using **KAFKA, SCALA, SPARK, HBASE** and **SPARK STREAM.**
* Experience in working with Flume to load the log data from multiple sources directly into **HDFS**.
* Performed Hadoop backup Strategy to take the backup of **hive, HDFS,HBase,oozie** etc.
* Experience in importing and exporting data using Sqoop from **HDFS** to Relational Database Systems (RDBMS) and from **RDBMS to HDFS**.
* Involved in creating HDINSIGHT cluster in **MICROSOFT AZURE PORTAL** also created **EVENTSHUB** and **AZURE SQL DATABASES.**
* Worked on a clustered Hadoop for Windows Azure using HDInsight and **HORTONWORKS** Data Platform for Windows.
* Built real time pipeline for streaming data using **EVENTSHUB/MICROSOFT AZUREQueue** and **SPARK STREAMING.**
* Loaded the aggregated data into HBase for reporting purpose!
* Read the data from HBase to Spark to perform Join on different tables.
* Created the HBase tables for validation, audit and offset management table.
* Created logical view instead of tables in order to enhance the performance of hive queries.
* Expertise in Extraction Transformation Loading (**ETL**) process using **Informatica** consisting of administration, data transformation, data sourcing, mapping, conversion and loading.
* Performance tuning in **Informatica** Mappings by identifying the bottlenecks and Implemented effective transformation Logic.
* Hands-on experience with **Snowflake utilities, SnowSQL, SnowPipe, Big Data model techniques using Python / Java.**
* Involved in developing **Hive DDLS** to create, alter and drop Hive tables.
* Pretty Good Knowledge on hive Optimization techniques like Vectorization and column-based optimization.
* Written oozie workflow to invoke the Jobs in predefined Interval.
* Expert in scheduling Oozie coordinator based on input data events it starts Oozie workflow when input data is available.
* On Other Hand working on POC with **Kafka** and NIFI to pull the real-time events into Hadoop Box.
* Exploring with the Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, **SPARK-SQL, DATA FRAME, PAIR RDD'S** and YARN.
* Experienced in managing Hadoop Cluster using **HORTONWORKS AMBARI**.
* Managed Linux and Windows virtual servers on **AWS** EC2 using **Puppet Server**. Configured and monitored distributed and multi-platform servers using **Puppet.**
* Experience with Infrastructure Development and Operations involving **AWS Cloud platforms**, **EC2, EBS, S3, MWAA, SNS, VPC, RDS, SES, ELB, EMR, VPC Auto scaling, Cloud Front & Formation, Elastic Cache, Cloud Watch.**
* Experience setting up and maintaining auto-scaling **AWS stacks** (using **cloud formation** and **scripting**).
* Experience in managing **DNS** using **Route 53 of AWS.**
* Excellent knowledge of **Amazon EC2, Amazon S3**, **Amazon Elastic Load Balancing** and other services of the **AWS** family.
* Automated the ingestion process through API calls to the vendor and used S3, Athena, Amazon Redshift, Airflow, and Glue
* Experience with migration to Amazon web Services **(AWS)** from Datacenter.
* Setup/Managing **CDN on Amazon Cloud** Front to improve site performance.
* Expertise on working with **MongoDB, Apache Cassandra**.
* Expertise on **Java, J2EE, Java Scripting, HTML, JSP.**
* Solid programming knowledge on **Scala, Pyspark.**
* Experience in working with **Teradata.** And making the data to be **batch processing** using **distributed computing.**
* Good working experience on Hadoop tools related to **Data warehousing** like **Hive, Pig**and also involved in extracting the data from these tools on to the cluster using Sqoop.
* Expertise in automation using **Kubernetes(EKS)** and have configured (Orchestration Tool) services in Cisco Cloud Center on Public Clouds (**AWS**, **SoftLayer**).
* Experience in orchestrating workflows using **Oozie** and **Airflow**.
* Worked with GitHub to manage source code repositories and performed branching, merging, and tagging depending on requirement.
* Expertise in creating Pods using Kubernetes and worked with Jenkins pipelines to drive all micro services builds out to the Docker registry and then deployed to **Kubernetes**.
* Having experience on Banking, Healthcare and Insurance domain application development, maintenance and testing.

**WORK EXPERIENCE**

**Big data Engineer**

**TCS, Irving, TX June 2022 to till date**

**Responsibilities:**

* Supported the design, development and implementation of Big Data solutions for the Informatics Team.
* Built and automated data engineering ETL pipeline over Snowflake DB using Apache spark and integrated data from disparate sources with Python APIs, consolidated them in a data mart (star schema)
* Planned, co-ordinated analysis, design and extraction of encounter data from multiple source systems into the data warehouse relational database while ensuring data integrity.
* Orchestrated entire pipeline using apache airflow delivering daily/weekly metric email reports from the power BI server to facilitate on the go decision making for business users.
* Designed multiple Deployment strategies, Continuous Integration and Continuous Development Pipelines using jenkins to ensure zero downtime and shortened deployment cycles
* Generated the data cubes using hive, Pig, JAVA Map-Reducing on provisioning Hadoop cluster in AWS.
* Created S3 buckets(configure, policies, permissions) and used AWS S3 for storage and backup of Data in to AWS and AWS Glacier to store archive data.
* Experienced in working with spark eco system using Spark SQL and Scala queries on different formats like Text file, CSV file,transformation in GCP.
* Migrated their Big Data Platform from on-premise Hadoop to Google Cloud Platform (GCP) to one of the projects that we are working
* Wrote Spark-Streaming applications to consume the data from Kafka topics and write the processed streams to HBase
* Designed and built end-to-end Data Warehouse infrastructure from the ground up on Redshift for large scale data handling
* Open SSH tunnel to GCP-Google DataProc to access to yarn manager to monitor spark jobs
* Worked with on-premises Cloudera clusters as well as AWS cloud services to deploy big data applications.
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (**Azure Data Lake, Azure Storage, Azure SQL, Azure DW**) and processing the data in In **Azure Databricks**
* Involved in the implementation calls of each release and provided post production support activities and also tracked DEV, Testing, Pre-prod and production environments.
* Participated in weekly release meetings with Technology stakeholders to identify and mitigate potential risks associated with the releases.

**Environment**:  Hadoop , Oozie , GCP, Zookeeper , Cassandra , Sqoop , ETL, Azure, Databricks, Hive , HBase , HDFS , Flume , Tableau, GIT, Kafka ,Snowflake

**BigData Developer**

**AT&T, Dallas, TX Aug 2020 to May 2022**

**Responsibilities:**

* Developed Fully automated Configuration driven data pipeline using **Spark**, **Hive**, **HDFS**, **SQL**Server, oozie and azure File storage to load the client data into Mirror databases.
* Performed necessary Transformations and Aggregation on the fly to build the common learner data model and persists the data in **HDFS**.
* Explored the usage of Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, **Spark SQL** and **Spark Yarn**.
* Involved in converting **Hive/SQL** queries into Spark transformations using Spark **RDDs**
* and Scala
* Designed and implemented data pipelines consisting of launching several Spark clusters equipped with **Glue** that read the datasets from various data sources and perform transformations, analytics and finally store results to application
* Experienced in working with spark eco system using Spark SQLand Scala queries on different formats like Text file, CSV file,transformation in **GCP**
* Expertise On optimizing spark Jobs when dealing with Huge joins and data Skew.
* Written oozie workflow to invoke the Jobs in predefined Intervals.
* Expert in scheduling Oozie coordinator based on input data events it starts **Oozie** workflow when input data is available.
* ETL pipelines in and out of data warehouse using combination of **Python** and Snowflakes SnowSQL Writing SQL queries against **Snowflake**.
* Developed Robust application with the help of **Drools**-**engine** In-order to execute complex rules on top of customer data.
* Pretty Good knowledge On the Horton works administration and security things such as Apache Ranger, **Knox** Gateway, High Availability.
* Performed Hadoop backup Strategy to take the backup of hive, **HDFS**, **HBase**, oozie etc.
* Used Hive to analyze the Partitioned and Bucketed data and compute various metrics for reporting.
* Performed hive performance tuning aspects like **Map** join, cost-based optimization and column level statistics.
* Created logical view instead of tables in order to enhance the performance of hive queries
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Databricks 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.
* Responsible for estimating the cluster size, monitoring and troubleshooting of the **Spark databricks** cluster.
* Involved in developing Hive **DDLS** to create, alter and drop Hive tables.
* Experience in creating **UDF's**, **UDAF's** for Hive and Pig.
* Extensively used Pig for data cleansing and **HIVE** queries for the analysts.
* Performed Benchmark between Hive and **SparkSQL**.
* Responsible for Designing and configuring Network Subnets, Route Tables, Association of Network **ACLs** to Subnets and Open **VPN**.
* Responsible for Account management, **IAM** Management and Cost management.
* Designed AWS Cloud Formation templates to create **VPC**, subnets, **NAT** to ensure successful deployment of Web applications and database templates.
* Creating **S3** buckets also managing policies for **S3** buckets and Utilized S3 bucket and Glacier for storage and backup on **AWS**.
* Experience to manage **IAM** users by creating new users, giving them a limited access as per needs, assign roles and policies to specific user.
* Act as technical liaison between customer and team on all **AWS** technical aspects.
* Working with two different datasets one using **HiveQL** and other using **Pig Latin.**
* Experience on **moving the raw data** between different systems usin**g Apache Nifi.**
* Implemented **ELK** (Elastic Search, Log stash, Kibana) stack to collect and analyze the logs produced by the spark cluster.
* Good working experience on Hadoop tools related to **Data warehousing** like **Hive, Pig**and also involved in extracting the data from these tools on to the cluster using Sqoop.
* Working experience with cloud infrastructure of AWS **(Amazon Web Services)** and computing AMI virtual machines on **Elastic Compute Cloud (EC2).**
* Worked on **AWS opswork, AWS cloudformation** and **cloud foundry**.
* Worked on **AWS**
* **Code deploy, AWS lambda , AWS RDS**

**Environment**: Azure, Spark, Hive, Spark SQL, GCP, Kafka, Snowflake ,Elastic Search,chefmi Horton Works, JBoss Drools, Hive, Pig, Oozie,GCOP, HBase, Pyspark, Scala, Maven, Jupyter Notebook, Databricks.

**Bigdata Developer**

**Alcons Lab, Fort Worth, TX Sept 2019 to July 2020**

**Responsibilities:**

* Used Bash Shell Scripting, Sqoop, AVRO, Hive, Pig, Java, Map/Reduce daily to develop ETL, batch processing, and data storage functionality.
* Used Pig to do data transformations, event join sand some pre-aggregations before storing the data on the HDFS.
* Developed an enterprise application, Spring/Hibernate ORM framework for backend, Spring MVC for middle layerImporting and exporting data into HDFS and Hive using Sqoop.
* Worked on Amazon Web Service (AWS) to integrate EMR with Spark 2 and S3 storage and Snowflake.
* Configured Spark streaming to receive real time data from the Kafka and store the stream data into **AWSS3** using Scala.
* Extensively used various Spring modules – **AOP** (logging/messaging), **DI**, Auto wiring, Inheritance, JDBC templates, used DAO pattern for fetching data from databases through Hibernate and carry through business logic.
* Exploited **Hadoop MySQL**-Connector to store Map Reduce results in **RDBMS**.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Worked on loading all tables from the reference source database schema through Sqoop.
* Worked on designed, coded and configured server side **J2EE** components like **JSP**, **AWS** and JAVA.
* Collected data from different databases (i.e. **Oracle**, **MySQL**) to **Hadoop**
* Used Oozie and Zookeeper for workflow scheduling and monitoring.
* Worked on Designing and Developing **ETL** Workflows using Java for processing data in HDFS/HBase using Oozie.
* Experienced in managing and reviewing Hadoop log files.
* Involved in loading and transforming large sets of structured, semi structured and unstructured data from relational databases into **HDFS** using **Sqoop** imports.
* Working on extracting files from **MySQL** through Sqoop and placed in HDFS and processed.
* Supported Map Reduce Programs those running on the cluster.
* Cluster coordination services through **Zookeeper**.
* Involved in loading data from **UNIX** file system to **HDFS**.
* Created several Hive tables, loaded with data and wrote Hive Queries in order to run internally in **MapReduce**.
* Developed Simple to complex MapReduce Jobs using **Hive** and **Pig**.
* Works closely with project manager to develop work plan for **Data** Warehouse projects and keep the manager aware of any issues
* Supporting the development group by providing API's and performing middle-tier application development
* Providing analytical network support to improve quality and standard work results
* Providing accurate estimates for project development and implementation. Work with management to meet those estimates.
* Developed Kibana Dashboards based on the Log stash data and Integrated different source and target systems into Elastic search for near real time log analysis of monitoring End to End transactions.
* Develop framework, metrics and reporting to ensure progress can be measured, evaluated and continually improved
* Perform in-depth analysis of research information for the purpose of identifying opportunities, developing proposals and recommendations for use by management
* Develop, refine and scale data management and analytics procedures, systems, workflows, and best practices.

**Environment**: AWS, Spark, Java, Hive, Stream-sets Spark SQL, Kafka, EMR, Snowflake, Elastic Search, Nebula, Hive, Pyspark, Scala, Maven, Jupiter Notebook, Visual Studio, Unix Shell Scripting.

**Bigdata Developer/Java Developer**

**XACTWARE SOLUTIONS / VERISK ANALYTICS**

**LEHI, UT Jan 2018 to Aug 2019**

**Responsibilities:**

* Designed the Endpoints (Controllers), Business Layer, DAO Layer using Hibernate/JDBC template using Spring IOC (Dependency Injection).
* Part of the full-cycle development of an Enterprise J2EE application, testing Servlets and JSPs in the typical MVC2 Architecture, involved in various phases of Software Development Life cycle (SDLC) and followed Agile-Scrum Methodology.
* 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 java for hive and pig and worked on reading multiple data formats on HDFS using Scala
* Used Scala collection framework to store and process the complex consumer information.
* Used Scala functional programming concepts to develop business logic.
* Designed and implemented Apache Spark Application (Cloudera)
* Importing and exporting data into HDFS Sqoop and Flume and Kafka.
* Troubleshoot and debug Hadoop ecosystem run-time issues.
* Used Oozie Scheduler systems to automate the pipeline workflow and orchestrate the map reduce jobs that extract
* Worked with Hue GUI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Worked with BI team in the area of Big Data Hadoop cluster implementation and data integration in developing large-scale system software.
* Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Flume, Oozie Zookeeper and Sqoop.
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshoot managing and reviewing data backups and Hadoop log files.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Extensively involved in Installation and configuration of Cloudera distribution Hadoop, NameNode, Secondary NameNode, JobTracker, TaskTrackers and DataNodes
* Worked with Spring MVC transaction management, Spring/Hibernate Batch transactions and Hibernate cache concepts and used Spring MVC and different View Controllers to implement business logic, Validation Framework.
* Developed REST web services clients to consume web services from third-party applications and injected into Presentation tier.
* Developed the business layer of the application through EJBs, web services SOAP, WSDL and web services clients
* Expertise in Oracle DB activities and good knowledge on backup variants of schemas and architecture.
* Worked on data analysis, SQL programming and query design/optimization, metrics development, and web-based reporting development.
* Experienced in developing applications using Java, JSP, Servlets, JavaBeans, JSTL, JSP Custom Tag Libraries, JDBC, SQL, AJAX, JavaScript, XML
* Designed a website based on Vehicle Rentals (Three-layered architecture/MVC and DAO patterns), wrote business logic in Java EE (which includes Servlets, JSP, JDBC concepts) and used Oracle as Database.
* Used Servlet API for HTTP request/response and SQL scripts to make the CRUD operations on Oracle database and Java Script for making validations and was part of end-to-end web development, GitHub for source code control.
* Installed, configured and management of several application and database servers in multiple Linux environments (RedHat, CentOS, Ubuntu) and migrated few Pega applications to higher versions.

**Environment:** JAVA/J2EE, JSP, CSS, JavaScript, Python, Junit, AJAX, Servlets, Hibernate, Spring, SASS, XML, EJB, React JS, Restful Web Services, Jenkin, Eclipse, IntelliJ, Bootstrap, WebSphere, Oracle, HTML, JMS, CSS, XSLT, DOM, and JDBC.

**CIT Health Care – ValueTrack**

**Module Lead - .net and SQl Developer**

**MindTree Ltd - India Aug 2010 to May 2015**

**Responsibilities:**

* Involved in requirement analysis, design, coding and implementation.
* Designing prototypes and development
* Migration of VC++ modules to C# and Enhancements.
* Understanding of SQL and Relational Database queries in SQL.
* Advanced SQL and SQL programming Automated the administrative tasks Using SQL Server Agent, event notifications, and scheduled maintenance plans according to Microsoft best practices.
* Used Crystal Reports to provide a print preview of Financial Reports and Monthly Statements.
* Carried out integration testing & acceptance testing
* Proficient in SQL Programming and vast experience in creating complex stored procedures, triggers, views, and user defined functions on SQL servers to meet the requirement of the applications.
* Developed use case diagrams, class diagrams, database tables, and mapping between relational database tables.
* Involved in specification analysis and identifying the requirements.
* Participated in design discussions for the methodology of requirement implementation.
* Involved in preparation of the Code Review Document & Technical Design Document
* Good Experience on SDLC (Software Development Life cycle).

**EDUCATION:**Bachelor of Engineering in Instrumentation Technology from Malnad College of Engineering, Hassan, Visveswaraiah Technological University in the year 2006.