**Reshma**

**Big Data Engineer**

**Phone :(469) 846-8345‬**

**Email ID : reshmatgowdact@gmail.com**

**SUMMARY:**

* Around **8+ years** of IT experience in **AWS, Cloud, Analysis, design, development, implementation, Big Data Warehouse, NOSQL, GCP, Hadoop, maintenance** and **support** with experience in developing strategic methods for deploying big data technologies to efficiently solve **Big Data** processing requirement
* Having Hands-on experience in **Talend**
* Around 5 years of experience on **BigData** using **HADOOP** framework and related technologies such as **HDFS, Teradata,** **HBASE, MapReduce, PySpark, Spark, Snowflake,** **Python,** **HBase, HIVE, PIG, FLUME, OOZIE, SQOOP, and ZOOKEEPER**
* 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
* Having excellent experience in **Data Pipeline** and **ETL Tools** such as Apache Airflow, **AWS** Data Pipeline, AWS Glue, Talend
* Expert level experience in designing, building and managing applications to process large amounts of data in a Hadoop/DevOps (GCP) ecosystem.
* Extensive experience with performance tuning applications on Hadoop/GCP and configuring Hadoop/GCP systems to maximize performance
* Experience in data analysis using **HIVE, PIG LATIN, HBASE** and custom Map Reduce programs in Java
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, **NOSQL**, 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
* Experience in writing custom UDFs in **JAVA** and **SCALA** for **HIVE** and **PIG**to extend the functionality.
* Experience with Cloudera and Horton works distributions.
* Around 3-year 2 months’ experience on **SPARK, SCALA, Python,** **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**
* Excellent experience with data processes, data quality reports
* 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**
* Extensive experience in Strategic development of a **Data Warehouse** and in Performing Data Analysis and Data Mapping from a Operational Data Store to a Enterprise Data Warehouse
* 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 AZURE Queue** 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.
* 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**
* **2**+ years hands on experience in design, develop and maintain BI Solutions using, Amazon **S3** & **Redshift**.
* Strong Experience in the Business Requirements Analysis, design, conversion of business requirements into high level functional specifications.
* Ability to interact with various groups ranging from Business Users, Business IM teams, Analysts. Had long lasting relationships with Clients in all previous engagements.

**WORK EXPERIENCE**

**Big Data Engineer  
Thermo Fisher Scientific , Carlsbad, California Nov21 to current**  
o Designed, coded and configured server side J2EE components like **AWS and Python ,SQL**  
o 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.  
o Performed necessary Transformations and Aggregation on the fly to build the common learner data model and persists the data in HDFS.  
o 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.**  
o Strong knowledge multiple cloud technologies including, **EC2, S3, RedShift, VPC, EBS, ELB, EMR, DynamoDB & Lambda**  
o Experience on Spark with Python using **Databricks .**  
o Working on Stateful Transformations in **Spark Streaming.**  
o Worked on Batch processing and Real-time data processing on Spark Streaming using Lambda architecture.  
o Good hands-on experience on Loading data onto Hive from Spark RDD’s.  
o Worked on Spark **SQL UDF’s and Hive UDF’s.**  
o Worked with Spark accumulators and broadcast variables.  
o Using decision tree as a model evaluation for both classification and regression.  
o Experienced in managing Hadoop Cluster using HORTONWORKS AMBARI  
o 5+ years hands on experience in design, develop and maintain BI Solutions using, **Amazon S3 & Redshift.**  
o Strong Experience in the Business Requirements Analysis, design, conversion of business requirements into high level functional specifications.  
o Ability to interact with various groups ranging from Business Users, Business IM teams, Analysts. Had long lasting relationships with Clients in all previous engagements.  
o Worked on Designing and Developing **ETL Workflows using Python for processing data in HDFS/HBase using Oozie.**  
o Experienced in managing and reviewing Hadoop log files.  
o Involved in loading and transforming large sets of structured, semi structured and unstructured data from relational databases into **HDFS using Sqoop imports.**  
o Working on extracting files from MySQL through Sqoop and placed in HDFS and processed.  
o Supported Map Reduce Programs those running on the cluster.  
o Configured Spark streaming to receive real time data from the **Kafka and store the stream data into AWS S3 using Python and SQL.**  
o Experience on Data Bricks, scripting language like **Python, Shell, PowerShell**  
o Worked with azure data bricks notebooks to validate the inbound/out bound from an external source like Amperity  
o Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.  
o Expertise On optimizing spark Jobs when dealing with Huge joins and data Skew.  
o Worked on Amazon Web Service(AWS) to integrate **EMR with Spark 2 and S3 storage and Snowflake**.  
o Deployed the Big Data Hadoop application using Talend on cloud AWS (Amazon Web Sevices) and also on Microsoft Azure.  
o Experienced in managing and reviewing Hadoop log files.  
o Finalize delta load for all tables and schedule jobs in Production  
o Implement the auto-loader as part of enhancements.  
o After the transformation of data is done, this transformed data is then moved to Spark cluster where the data is set to go live on to the application using **Spark streaming and kafka.**  
o Created RDD’s in Spark technology.  
o Extracting data from data warehouse on to the Spark RDD’s  
o Experience on **Spark with Python.**  
o Working on Stateful Transformations in Spark Streaming.  
  
  
**Environment**: Azure, AWS,Spark, Hive, PySpark , Spark SQL, Kafka, Horton Works, JBoss Drools, Hive, Pig, Oozie, HBase, Python, SQL, Snowflake, Big Data Warehouse, Jupiter Notebook, Visual Studio, Unix Shell Scripting.

**Big Data Engineer**

**AT&T, Dallas, TX**

**Nov 2020 to Oct 2021**

**Responsibilities:**

* Developed Fully automated Configuration driven data pipeline usingSpark, 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.
* Worked with data pipelines using Big Data stacks like Apache Spark with Python and AWS cloud.
* worked on various data mart projects, migration projects (Hive to Spark with Scala or PySpark), productionizing ML , **CI/CD pipelines,** etc. Tech stacks wise, I was working on Apache Spark with Python and Scala, AWS cloud - EMR, Lambda, Step Functions, CloudWatch, etc
* Explored the usage of PySpark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark SQL and Spark Yarn
* Experienced in working with spark eco system using Spark SQLand Scala queries on different formats like Text file, CSV file,transformation in GCP
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* 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.
* 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 Hortonworks 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.
* 2+ years hands on experience in design, develop and maintain BI Solutions using, Amazon S3 & Redshift.
* Strong Experience in the Business Requirements Analysis, design, conversion of business requirements into high level functional specifications.
* Strong knowledge multiple cloud technologies including, EC2, S3, RedShift, VPC, EBS, ELB, EMR, DynamoDB, Lambda, Route 53.
* Ability to interact with various groups ranging from Business Users, Business IM teams, Analysts. Had long lasting relationships with Clients in all previous engagements.
* 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.
* 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.
* 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.

**Environment**: Azure, AWS,Spark, Hive, PySpark, GCP, Spark SQL, Kafka, Horton Works, JBoss Drools, Hive, Pig, Oozie, HBase, Python, Scala, Teradata, Snowflake, Big Data Warehouse,Maven, Jupiter Notebook, Visual Studio, Unix Shell Scripting.

**Big Data Developer**

**AIG, Irving, TX**

**June 2017 – Sept 2020**

**Responsibilities:**

* Importing and exporting data into HDFS and Hive using Sqoop.
* Used Bash Shell Scripting, Sqoop, AVRO, Hive, Pig, Java, Map/Reduce daily to develop ETL, batch processing, and data storage functionality
* Migrated their Big Data Platform from on-premise Hadoop to Google Cloud Platform (GCP) to one of the projects that we are working
* 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 layer.
* Worked on Amazon Web Service(AWS) to integrate EMR with Spark 2 and S3 storage and Snowflake
* Open SSH tunnel to GCP-Google DataProc to access to yarn manager to monitor spark jobs.
* Configured Spark streaming to receive real time data from the Kafka and store the stream data into AWS S3 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.

**Environment**: AWS, Big Data Warehouse,Spark, Java, Hive, Stream-sets Spark SQL, Kafka, EMR, Snowflake, Nebula, Hive, Python, GCP, Scala, Maven, Jupiter Notebook, Snowflake, Teradata, Visual Studio, Unix Shell Scripting.

**Big Data Developer**

**State Auto Insurance, Columbus, Ohio**

**Sept 2016 to May 2017**

**Responsibilities:**

* Invoking an external UDF/UDAF/UDTF python script from Hive using Hadoop Streaming approach which is supported by Ganglia.
* Setting up the work schedule using oozie.
* Identifying the errors in the logs and rescheduling/resuming the job.
* Able to handle whole data using HWI (Hive Web Interface) using Cloudera Hadoop distribution UI.
* Deployed the Big Data Hadoop application using Talend on cloud AWS (Amazon Web Sevices) and also on Microsoft Azure.
* Involved in Designing and Developing Enhancements product features.
* Involved in Designing and Developing Enhancements of CSG using AWS APIS.
* Enhance the existing product with newly features like User roles (Lead, Admin, Developer), ELB, Auto scaling, S3, Cloud Watch, Cloud Trail and RDS-Scheduling.
* Created monitors, alarms and notifications for EC2 hosts using Cloud Watch, Cloud trail and SNS.
* Involved in Designing the SRS with Activity Flow Diagrams using UML.
* Employed Agile methodology for project management, including: tracking project milestones; gathering project requirements and technical closures; planning and estimation of project effort; creating important project related design documents and identifying technology related risks and issues.
* As per as business requirements we use Talend to integrate the data on cloud and make it accessible to the offshore medical team.
* Working with Informatica 9.5.1 and Informatica 9.6.1 Big Data edition. Scheduling the jobs
* After the transformation of data is done, this transformed data is then moved to Spark cluster where the data is set to go live on to the application using Spark streaming and kafka.
* Created RDD’s in Spark technology.
* Extracting data from data warehouse(TeraData)on to the Spark RDD’s
* Experience on Spark withScala/Python.
* Working on Stateful Transformations in Spark Streaming.
* Worked on Batch processing and Real-time data processing on Spark Streaming using Lambda architecture.
* Good hands-on experience on Loading data onto Hive from Spark RDD’s.
* Worked on Spark SQL UDF’s and Hive UDF’s.
* Worked with Spark accumulators and broadcast variables.
* Using decision tree as a model evaluation for both classification and regression.
* Collaborated with the infrastructure, network, database, application and BI teams to ensure data quality and availability.

**Environment:**HDFS cluster, Hive, UNIX, Teradata, Snowflake, Apache Nifi, Pig, Sqoop, Oozie, MapReduce, Talend, Big Data Warehouse,Python.

**Software Developer**

**Squires Solutions, Hyderabad, India**

**Feb 2013 – April 2016**

* 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.
* 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, SQL, XML, Linux (Redhat, CentOS), Oracle DB, JDBC, SQL, Snowflake, AJAX, JavaScript, XML.