**Hari Krishna Chilukamari**

+1 4698468345; [saisri@mastroservices.com](mailto:saisri@mastroservices.com)

**Big Data Engineer**

**SUMMARY:**

* Around overall 6 years of IT experience and 5 years in AWS, Cloud, Analysis, GIT, design, development, Azure Data Factory, Azure Databricks, Data Lake, Kubernetes, docker, 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, Spark, Snowflake, Python, HBase, HIVE, PIG, FLUME, OOZIE, SQOOP, and ZOOKEEPER
* Hands-on experience in EDW, Data Migration, Big Data and Data Lake Testing with expertise in troubleshooting SSIS packages, Autosys and Databricks Spark Jobs
* Hands-on experience in AzurePortal, Azure Data Factory, and Big Data Cloudera /Ambari tools like HIVE, and Impala
* Experience in configuring a provider with Terraform that uses Kubernetes supported tools to build several services such as Config Map, Namespace, Volume, Auto Scaler.
* Experience in Kubernetes orchestration for Docker Containers that handle scheduling and managing and creating customized resources and upgrading components and auto-scaling containers depending on the amount of use
* 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
* Trouble shooted Azure Databricks jobs and Azure (ADF) Pipelines those deal with Spark and HDInsight (HIVE) clusters to process data into Data Lake Store
* Expert level experience in designing, building and managing applications to process large amounts of data in a Hadoop/DevOps (GCP) ecosystem
* 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.
* 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
* Build CI/CD pipeline using GitHub, Jenkins, OpenShift, Ansible and Docker
* Troubleshooted Azure Databricks jobs and Azure (ADF) Pipelines those deal with Spark and HDInsight (HIVE) clusters to process data into Data LakeStore
* Expertise in Docker including Docker Hub, Docker Engine, Docker Images, Docker Composer, Docker Swarm, and Docker Registry, and used containerization to make application platform flexible when moved to different environments.
* Involved in building Data Models and Dimensional Modeling with 3NF, Star and Snowflake schemas for OLAP and Operational data store (ODS) applications
* Experience working on Docker hub, creating Docker Images and handling multiple containers as a cluster with Docker Swarm, and Compose for various applications and several Micro services.
* Experience in using Kubernetes to provide a platform for automating deployment, scaling, and operations of application containers across clusters of hosts
* 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
* 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**

**AT&T, Dallas, TX**

**May 2023 to Current**

**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
* Managed Docker orchestration and Docker containerization using Kubernetes.
* Used Kubernetes to orchestrate the deployment, scaling and management of Docker Containers.
* 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
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, NOSQL, 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
* 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 analyse 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 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, Azure Data Factory, Databricks, Datalake, Python, Scala, Teradata, Snowflake, Big Data Warehouse, Maven, Jupiter Notebook, Visual Studio, Unix Shell Scripting.

**Big Data Developer**

**AIG, Irving, TX**

**June 2020 – May 2023**

**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-premises Hadoop to Google Cloud Platform (GCP) to one of the projects that we are working
* Evaluated and designed the software and the infrastructure for the company-wide private container cloud solution based on Kubernetes, Swarm, Docker, Python and Go.
* Managing Linux Containers using Docker and docker registry
* 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 Daturic 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.
* Analysed 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 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, Kubernetes, docker, Maven, Jupiter Notebook, Snowflake, Teradata, Visual Studio, Unix Shell Scripting.

**Software Developer**

**Tata Consultancy Services, Hyderabad, India**

**Feb 2018 – May 2020**

**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.
* 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.