**VENKATA RAMANA N**

[saisri@mastroservices.com](mailto:saisri@mastroservices.com)

469 846 8345

**Big Data Engineer** .

14+ Years of extensive experience in Analysis, Architecture, Design, Development, testing of applications using Big Data, Hadoop Tools, Machine Learning, SAP, SQL, No SQL databases and in business warehouse.

* Very good experience in developing and deploying enterprise-based applications using major components in Hadoop ecosystem like Hadoop 1.X and 2.X, YARN, Hive, MapReduce, Spark, Oozie, Scala, Python.
* In-depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, MapReduce programming paradigm.
* Worked with major distributions like Cloudera (CDH 3, 4&5) and Hortonworks Distributions.
* Experience in handling large datasets using Partitions, spark in-memory capabilities, Broadcasts in Spark with Scala.
* Experience in developing data pipeline using Sqoop to extract the data from weblogs and store in HDFS and accomplished developing custom MapReduce programs and using HiveQL for data analytics.
* Hands on experience in various Bigdata application phases like data ingestion, data analytics and data visualization.
* Experience in converting Hive/SQL queries into Spark transformations using Scala and experience in ETL development using Spark SQL, RDD.
* Implemented pre-defined operators in spark such as map, flat Map, filter, reduceByKey, groupByKey, aggregateByKey and combineByKey etc.
* Experience in Oozie and Airflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows.
* Good experience in SQL and PL/SQL Programming languages to writing complex queries, functions, stored procedures, and triggers.
* Experienced with different scripting language like Python and Unix/Linux Shell Scripts.
* Experienced in data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka.
* Optimization of Hive queries using best practices and right parameters and using technologies like Hadoop, YARN, Python, PySpark.
* Develop ML-pipeline using Machine Learning APIs algorithms and regression techniques to develop predictive models as part of supervised and unsupervised machine learning to train the data and store the models to provide the personalized score
* Good understanding of automatic machine learning (AutoML) tools, Google Cloud AutoML and H2O.ai
* Experience with cloud warehouse tools like Snowflake.
* Created multiple Spark Jobs using Scala for data extraction.
* Experience working on NoSQL databases including HBase and MongoDB.
* Strong knowledge on performance tuning Hive queries and troubleshooting various issues related to Joins, memory exceptions in Hive.
* Very good understanding of Partitions, bucketing concepts in Hive and designed both internal and external tables in Hive to optimize performance.
* Experience in using different columnar file formats like Avro, RCFile, ORC and Parquet formats.
* Trained and certified on SAP Application Associate - Reporting / Modeling and Data Acquisition with SAP BW/4HANA
* Worked extensively on SAP BW and SAP BW on HANA

**SKILL SET**

* HDFS
* MapReduce and YARN
* Apache Spark
* SQL
* MongoDB
* GCP
* AutoML
* Hive
* Oozie
* Scala
* Python
* Tableau
* SAP BW/4HANA
* SAP BW 7.4/7.5 on HANA
* SAP Analytics Cloud
* Service-Now

**Big Data Engineer**

**Verizon**

**May 2022 till Present**

* Design and develop custom, scalable, reusable and resilient applications to integrate various components, increase consistency, automate tasks, alerts, assist in monitoring/diagnosing and processing of data in the Data Lake using Hadoop components in Hortonworks Hadoop cluster using Spark and Scala.
* Build data pipelines to collect data points from various internal and third-party data sources and cleanse, transform and process the data and store the results in near real-time to perform exploratory analysis and discover data acquisition opportunities
* Build ETL jobs for address standardization of business and third party raw data using Verizon Locus API
* Perform initial triage, identify immediate and permanent fix based on the root cause analysis for all the production defects and provide hot fixes
* Perform data analysis to make data fit into the business model. By building a data ecosystem that serves the organization’s needs using Hive, Jupyter, Python, Scala and Spark
* Design Oozie workflows and automate the data pipelines using Oozie Coordinator and screwdriver
* Migrate data from Network Data Lake to Verizon Corporate Grid cluster using Apache Nifi, Spark, Scala, Hive and Hadoop Distributed File System
* Used Google Cloud Platform (GCP) to store the data in buckets and analyze the performance of the spark submit jobs and queries through yarn, Jupyter and Big Query.
* Used machine learning libraries like fuzzywuzzy for adhoc matching and pattern matching to identify closest location for a given address
* Also used JaccardNetric, JaroMetric and JaroWinklerMetric for string similarity
* Conduct exploratory data analysis, feature engineering, and model selection.
* Implement and optimize machine learning algorithms using Python, TensorFlow, and Scikit-Learn.
* Develop and deploy scalable models on cloud platforms (e.g., AWS, GCP).
* Monitor and evaluate model performance, making adjustments as needed.
* Develop test cases using Scala Test library to perform unit testing and capture the result sets against the jobs developed in lower environments by creating the mockup data
* Adapt SAFE agile environment and maintain the user stories and tasks in the rally tool
* Interact with business users and business analysts in the daily SCRUM and weekly audit meetings andf analyzing BRD documents and incorporating CART, SOX, CCPA and GDPR security rules in the application

**Environment**: GCP, Big Query, Yarm Resource Manager. Hadoop, MapReduce, HDFS, Hive, Oozie, Spark, Scala, , HortonWorks, ,MySQL, Shell Scripting.

**Hadoop Developer**

**Apple Inc**

**Mar 2018 till Apr 2021**

* Strong understanding of Hadoop eco system such as HDFS, MapReduce, HBase, Zookeeper, Pig, Hadoop streaming, Sqoop, Oozie and Hive.
* Installed and configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster.
* Collected and aggregated large amounts of web log data from different sources such as webservers, mobile and network devices using Apache Flume and stored the data into HDFS for analysis.
* Installed and configured Hadoop MapReduce, HDFS, developed multiple Map Reduce jobs in java for data cleaning and processing.
* Involved in making Hive tables, stacking information, composing hive inquiries, producing segments and basins for enhancement.
* Involved in migrating tables from RDBMS into Hive tables using SQOOP and later generate particular visualizations using Tableau.
* Analysed substantial data sets by running Hive queries and Pig scripts.
* Created Partitions, Buckets based on State to further process using Bucket based Hive joins.
* Involved in transforming data from Mainframe tables to HDFS, and HBase tables using Sqoop.
* Defined the Accumulo tables and loaded data into tables for near real-time data reports.
* Created the Hive external tables using Accumulo connector.
* Written Hive UDFs to sort Structure fields and return complex data type.
* Used distinctive data formats (Text format and ORC format) while stacking the data into HDFS.
* Involved in creating Shell scripts to simplify the execution of all other scripts (Pig, Hive, Sqoop, Impala and MapReduce) and move the data inside and outside of HDFS.
* Creating files and tuned the SQL queries in Hive utilizing HUE.
* Experience working with Apache SOLR for indexing and querying.
* Created custom SOLR Query segments to optimize ideal search matching.
* Worked with NoSQL databases like HBase in making HBase tables to load expansive arrangements of semi structured data.
* Acted for bringing in data under HBase using HBase shell also HBase client API.
* Designed the ETL process and created the high-level design document including the logical data flows, source data extraction process, the database staging, job scheduling and Error Handling
* Created ETL Mapping with Talend Integration Suite to pull data from Source, apply transformations, and load data into target database.
* Developed Flume Agents for loading and filtering the streaming data into HDFS.
* Experienced in handling data from different data sets, join them and pre-process using Pig join operations.
* Developed Map-Reduce programs to clean and aggregate the data
* Developed HBase data model on top of HDFS data to perform real time analytics using Java API.
* Implement counters on HBase data to count total records on different tables.
* Experienced in handling Avro data files by passing schema into HDFS using Avro tools and Map Reduce.
* Experienced in handling different types of joins in Hive like Map joins, bucker map joins, sorted bucket map joins.
* Experienced import/export data into HDFS/Hive from relational data base and Tera data using Sqoop.
* Handling continuous streaming data comes from different sources using flume and set destination as HDFS.
* Integrated spring schedulers with Oozie client as beans to handle cron jobs.
* Experience with CDH distribution and Cloudera Manager to manage and monitor Hadoop clusters
* Actively participated in software development lifecycle (scope, design, implement, deploy, test), including design and code reviews.
* Involved in story-driven agile development methodology and actively participated in daily scrum meetings.

**Environment**: Hadoop YARN, Spark-Core, Spark Streaming, Spark SQL, Scala, Kafka, Hive, Sqoop, Amazon AWS, HBase, Teradata, Power Center, Tableau, Oozie, Oracle, Linux.

**Big Data Lead**

**General Electric**

**Oct 2015 till April 2018**

**Responsibilities:**

* Worked on SAP BW 7.5 on HANA and HANA Studio
* Developed end-end solutions with the best approaches to fit the requirements of the users
* Created solutions with complete data flows, infoobjects, HANA views, data sources, Advanced DSO, Open ODS views, Composite Providers and developed reports on them
* Created information views - attribute and calculation views and generated reports on them
* Developed a solution for Demand Planning to extract the data from APO systems, used an advanced DSO and Composite Provider  and developed reports using BWMT and created a process chain
* Developed a solution with HANA view as a data source and created BW objects on them
* Developed Bex Reports using BWMT with Formulas, Selections, Structures, Calculated Key Figure, Restricted Key Figure and also Used Variables with User Entry/Default Value and Customer Exit
* Designed and developed HANA based models to replace existing BW objects and saved 50 GB space
* Written SQL scripts to create tables, compare the data between tables and replaced old and huge one
* Worked on upgrading the data providers after upgrading the system to BW 7.5 on HANA
* Modelled mixed scenarios involving BW objects and HANA views
* Developed reports on the views and data providers and created CKFs, RKFs, formulas, variables and written user exits to populate the data
* Worked on ABAP debugging to find issues while testing solutions and written ABAP code to fix them
* Attended BW/4HANA trainings from SAP
* Created Customer Exits to populate the values on selection screen, used SAP function modules

**Eastman Kodak**

**Manufacturing**

**SAP BW on HANA Lead**

**August 2010 to Sept 2015**

**Responsibilities:**

* BW 7.4 on HANA upgrade checks, cleaned up space, retired 3.x objects like infospokes, infopackage groups
* Tested the Infoproviders, Rules, Data Sources, Bex reports during and after the upgrade BW 7.0 to BW 7.4 on HANA
* Converted the DSOs and Infocubes to HANA optimized objects using the transaction RSMIGRHANADB
* Implemented end-end SAP BW solutions, extracting data from SD, CRM and FI modules
* Designed and created Attribute, Analytical and Calculation views over HANA optimized objects.
* Designed and developed mixed views with HANA tables and BW on HANA objects
* Designed and created Advanced DSOs for persistency in SD and FICO applications
* Designed and developed composite providers over cubes and DSOs in SEM and SA applications
* Designed and developed Open ODS views for mixed modelling on HANA and BW data.
* Recreated and moved the Bex reports on to Analysis for office
* Designed and developed data models from ECC into BW and SAP HANA.
* Designed and Created reports Query Designer, Web application Designer (WAD), Business Intelligence 4.0 Web Intelligence (WEBI) and Integration of BO 4.0 WEBI with SAP BEx Query using IDT
* Designed and created crystal reports based out of BW reports
* Designed and developed reports with BOAO (excel)
* SAP Lumira - Data discoveries, visualization and analysis and created story boards and shares result with business and senior management
* Weekend Migrations – Weekly queueing up of BW transports and migrating them to production during Saturday maintenance windows
* Year End Activities – Freeze master data loads from ECC and flat files during the year end and reload the master data and transaction data after completion of FMH updates on ECC
* Cut over activities – Deleted and reloaded the data across applications SD, SA and SEM after the migrations.
* Performed pre-post cutover activities during Service Pack upgrade, Stack upgrade and Security upgrade
* Created SAP OSS notes and coordinated with SAP for complex issues during the upgrades and patch implementations
* Created User Exit variables to be used in queries and APDs
* Converted Infopackage Groups to Process Chains
* Automation through process chain and Tivoli Workload Scheduler
* Fixing batch failures
* Worked on different kind of User tickets, few of them were,
* Bex and BOAO Report issues
* Data reconciliation issues
* Delta invalidations and debugged routines
* Team meetings with Kodak managers, Wipro managers, onsite team and offshore team

**Sealed Air**

**Manufacturing**

**SAP BW Consultant**

**March 2010 to July 2010**

**Responsibilities:**

* Performed analysis, design, development and implementation of customized Infocubes and DSOs
* Created info objects, data sources, DSOs, transformations, transfer rules and Infocubes
* Used function module based extractor to extract the data from R/3.
* Created multiprovider for better performance
* Created a process chain for automating the whole process
* Developed reports on Multiprovider, created variables, cell formatting, formulas, calculated key figures and restricted key figures.

**Kone Corporation**

**Engineering and Services**

**SAP BW Consultant**

**June 2009 – February 2010**

**Responsibilities:**

* Installation of Business content data sources and activation of Standard BW objects.
* Design, develop and implement SAP BI 7.0 objects including Extraction, Transformation and Load processes,
* Data store objects, Cubes and Reports.
* Data Extraction from R/3, Creating Info packages, Scheduling and Monitor settings.
* Worked on Logistics Cockpit extractions for uploading the data from R/3 to BW and initialized delta and uploaded transaction data from R/3 systems by V3 updates.
* Generic data sources based on views.
* Development of objects as per report's requirements.
* Transporting objects from development to Quality and to Production by raising SD Ticket.
* Involved in creation of design document for info Objects, info sources, info providers, and Queries.
* Involved in System Integration Testing.
* Flat file extraction
* Involved in creating process chain to automate the flat file extraction using ABAP code.

**Education**

* Masters in IT from Jawaharlal Nehru Technological University from 2006 - 2008
* Bachelors of Technology in Computer Science from Jawaharlal Nehru Technological University from 2002 – 2006sss