**­RAMANA V**

469-846-8345

ramanav0929@gmail.com

**Big Data Engineer** .

Venkata Ramana Nalla has 12+ Years of extensive experience in Analysis, Architecture, Design, Development, testing of applications using Big Data and Hadoop tools, google cloud platform and AWS, SQL, No SQL databases and in business warehouse. Currently working as **Senior Data Engineer** at Verizon.

**Major Strength:**

* My core competencies include working on end-end suite of on premise and cloud Big Data technologies like **YARN, Hive, MapReduce, Spark, Kafka, Oozie, Airflow, Sqoop, NoSQL, Scala, Python, JAVA, GCP – BigQuery, Clusters, Storage in buckets and Dataflows. AWS – S3, EC2, Redshift**
* Worked as a **Developer, Data Engineer, Lead Engineer and Data Analyst**
* Extensive knowledge and expertise with data warehousing, data lakes, on premise and cloud data bases

**Summary:**

* 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, Kafka, Oozie, Sqoop, 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.
* Extensively dealt with Transaction data & Apache Kafka to fetch live stream data.
* 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.
* Good experience in writing Spark applications using Java.
* 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.
* Hands on experience in installing, configuring, and deploying Hadoop distributions in cluster environments.
* 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.
* Design and implement web applications which utilize the Flask framework
* Experience with cloud warehouse tools like Snowflake.
* Created multiple Spark Jobs using Scala for data extraction.
* Extensive experience in developing applications that perform Data Processing tasks using Oracle and MySQL database.
* 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.
* Hands on experience in installing, configuring, and deploying Hadoop distributions in cluster environments (Amazon Web Services).
* Able to work on own initiative, highly proactive, self-motivated commitment towards work and resourceful.
* Key participant in all phases of software development life cycle with Analysis, Design, Development, Integration, Implementation, Debugging, and Testing of Software Applications in client server environment, Object Oriented Experience in using various IDEs Eclipse, IntelliJ and repositories SVN and Git.
* Experience in using version control tools like Bit-Bucket, GIT etc.
* Experience in writing build scripts using MAVEN, ANT and Gradle.
* Experience in Node.js and Java script development
* Web application development experience using ReactJS, AngularJS

**SKILL SET**

|  |  |
| --- | --- |
| Big Data ecosystem Technologies: | HDFS, MapReduce, YARN, Hive, Pig, Sqoop, Flume, Oozie, Airflow, Dataflow, Zookeeper, Kafka, Apache Spark and PySpark, HBase |
| Hadoop Distribution: | Cloudera, Apache |
| Languages: | Java, Python, Scala, SQL, PL/SQL, Pig Latin, HiveQL, ANSI SQL, Node JS |
| Web Technologies: | HTML, CSS, JavaScript, XML, JSP, Restful, Bootstrap. |
| Operating Systems: | Windows, UNIX, LINUX, Mac |
| Cloud Computing: | GCP – Storage, Clusters and BigQuery. Microsoft Azure - Cloud Services (PaaS & IaaS), Azure Monitoring, Key Vault SQL Azure, AWS - S3 |
| Build Automation tools: | Jenkins |
| Version Control: | SVN, GitHub, Bit-bucket |
| Databases: | Oracle, SQL Server, MySQL, MS Access, NoSQL Database (HBase). |
| Web/Application Servers | Web logic, Web Sphere, JBoss, Tomcat, Flask |
| IDE’s & Utilities | Eclipse, PyCharm, NetBeans, IntelliJ, SQL Developer |
| Roles | Worked as lead data engineer and developer |

**Verizon**

**Irving, Texas**

**Jan 2019 till Present**

**Big Data Lead**

* Responsible for building scalable distributed data solutions using Hadoop.
* 3 years of experience as lead data engineer
* Experience in creating Kafka producer and Kafka consumer for Spark streaming which gets the data from different learning systems of the patients.
* Configured Spark streaming to receive real time data from the Kafka and store the stream data to HDFS using Scala.
* Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing.
* Worked with Spring Boot building Microservice and/or Web Apps
* Evaluated the performance of Apache Spark in analyzing genomic data.
* Performed advanced procedures like text analytics and processing using the in-memory computing capabilities of Spark.
* Implemented Spark RDD transformations to Map business analysis and apply actions on top of transformations.
* Written Storm topology to accept the events from Kafka producer and emit into Cassandra DB.
* Created POC using Spark SQL and MLlib libraries.
* Experienced in managing and reviewing Hadoop log files.
* Experience with GCP Dataproc, GCS, cloud functions and BigQuery
* Used cloud shell SDK in GCP to configure the services Dataproc, Google Cloud Storage and BigQuery
* Worked on developing ETL data flows in GCP, tested them and performed performance tuning
* Worked on pyspark extensively during extraction, transformation and loading of the data
* Experience in creating Spark infrastructure with Kubernetes and Docker
* Worked closely with EC2 infrastructure teams to troubleshoot complex issues.
* Worked with AWS cloud and created EMR clusters with spark for analyzing raw data processing and access data from S3 buckets.
* Involved in installing EMR clusters on AWS.
* Used AWS Data Pipeline to schedule an Amazon EMR cluster to clean and process web server logs stored in Amazon S3 bucket.
* Developed Python application in AWS EMR to perform two necessary steps (quality check and tokenization) before loading business data into our centralized S3 buckets.
* Refactor old Python and Flask code to ensure it follows modern principles and ways of implementation.
* Designed the NIFI/HBASE pipeline to collect the processed customer data into Hbase tables.
* Apply Transformation rules on the top of Data Frames.
* Worked with different File Formats like TEXTFILE, AVROFILE, ORC, and PARQUET for HIVE querying and processing
* Developed Hive UDFs and UDAF’s for rating aggregation.
* Developed java client API for CRUD and analytical Operations by building a restful server and exposing data from No-SQL databases like Cassandra via rest protocol.
* Created Hive tables and involved in data loading and writing Hive UDFs.
* Experience in managing and reviewing Hadoop Log files.
* Worked extensively with Sqoop to move data from DB2 and Teradata to HDFS.
* Collected the logs data from web servers and integrated in to HDFS using Kafka.
* Provided ad-hoc queries and data metrics to the Business Users using Hive, Impala.
* Worked on various performance optimizations like using distributed cache for small datasets, partition, bucketing in hive, map side joins etc.
* Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability
* Implemented Row Level Updates and Real time analytics using CQL on Cassandra Data.
* Used Cassandra (CQL) with Java API's to retrieve data from Cassandra tables.
* Worked on analyzing and examining customer behavioral data using Cassandra.
* Worked on Solr configuration and customizations based on requirements.
* Indexed documents using Apache Solr.
* Used Azure Portal, Azure PowerShell, Storage Accounts, and Azure Data Management.
* Involved in writing the scope scripts in Data Lake and HDFS to structure the Peta bytes of unstructured data stored in the Azure Data Lake.
* Deployed Resource Manager templates. Resource Manager deployment is the deployment method used by Data Factory in the Azure portal.
* Assigned the built-in contributor role on the Azure data factory resource for the user.
* Extensively use Zookeeper as job scheduler for Spark Jobs.
* Worked with BI teams in generating the reports on Tableau.
* Used JIRA for bug tracking and CVS for version control.
* Met with business/user groups to understand the requirement for new Data Lake Project.
* Worked in Agile Iterative sessions to create Hadoop Data Lake for the client.
* Defined the reference architecture for Big Data Hadoop to maintain structured and unstructured data within the enterprise.
* Lead the efforts to develop and deliver the data architecture plan and data models for the multiple data warehouses and data marts attached to the Data Lake Project.

**Environment**: GCP, Hadoop, MapReduce, HDFS, PIG, Hive, Sqoop, Oozie, Storm, Kafka, Spark, PySpark Streaming, Scala, Cassandra, Cloudera, ZooKeeper, AWS, Solr, MySQL, Shell Scripting, Java, Tableau.

**Walmart**

**Big Data Engineer**

**Oct 2017 till Dec 2018**

* Designing, programming, testing and documentation of Cerner 3-party software applications and workflow process.
* Provide direction in architecture discussions, design sessions and code reviews to review for functional correctness, architectural maintainability and performance.
* Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing.
* Produced comprehensive architecture strategy for environment mapping in AWS that involved Active Directory, LDAP, AWS Identity and Access Management (IAM) Role for AWSAPI Gateway platform.
* Implemented Micro Services with AWS ECS Docker, code build, packaging, pipeline, provisioning, deploy, commit process and change management.
* Worked on streaming data using Kafka and Spark Streaming for Data preparation.
* Used HBase for creating Snapshot tables for updating and deleting records.
* Perform advanced procedures like text analytic and processing, using the in-memory computing of Spark using Scala.
* Implemented AWS cloud computing platform, and its many dimensions of scalability - including but not limited to: VPC (Virtual Private Cloud), EC2, load-balancing with ELB, messaging with SQS (and scalable non-AWS alternatives), auto scaling architectures, using EBS under high I/O requirements, custom monitoring metrics/analysis/alarms via Cloud Watch, the AWS API and different toolkits for instrumenting it (e.g. boto).
* Extensive experience on HANA & MSBI Stack SQL Server DBMS, **SQL Server Analysis Services (SSAS)** and **SQL Server Integration Services (SSIS)**
* Experience in Configuration Management, Cloud Infrastructure, and Automation like Amazon Web Services (AWS), Maven, Jenkins, Chef, Github, Clear Case and Tomcat.
* Develop POC on real time streaming data received by Kafka and processed the data using Spark and this data was further stored into HDFS cluster using Scala.
* Worked on cloud warehouse tools like Snowflake.
* Working on transferring data from different data sources into HDFS systems using Kafka producers, consumers and Kafka brokers.
* Written the Spouts and Bolts after collecting the real stream customer data from Kafka broker to process and store into HBase.
* Maintained and perform enhancements to the Hadoop cluster on AWS by using EC2, EMR, S3 and EBS Volume.
* Working on spinning up AWS EC2 instances, Creating IAM Users and Roles, Creating Auto Scaling groups, Load balancers and monitoring through Cloud Watch for the applications, S3 buckets, VPC etc.
* Automation Deployment of Hadoop Clusters on AWS using Ansible Playbook from end to end with all configurations.
* Auto Scaling of Infrastructure using Terraform and Chef Deployment scripts and also Continuous Deployment using Jenkins.
* Experience in chef cookbook/recipe development.
* Monitored system performance statistics like storage, CPU utilization using Zabbix.
* Using JIRA for bug tracking and Git for version control worked under Agile methodology.

**Environment**: JDK1.8, Spark, Scala, Linux, Shell Script, Terraform, CloudFormation, Cloudera, Zookeeper, Hive, HBase, Oozie, HDFS, Kafka, AWS, Chef, Ansible, Zabbix, GIT, Jenkins, Jira, Agile.

**Wipro Technologies**

**Hadoop Developer**

**Oct 2013 to Sept 2017**

* 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.

**Eastman Kodak**

**SAP BW on HANA Consultant**

**Aug 2010 to Sept 2013**

**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 CRM 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**

**Mar 2010 to Jul 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**

**Feb 2009 – Feb 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 – 2006