****

**Rama Teja Thummala**

**(816)-866-8301 |** [**ramtejadev@gmail.com**](mailto:ramtejadev@gmail.com) **|LinkedIn:-**[**Ramateja**](http://www.linkedin.com/in/ram-teja-781a3515a)

**PROFESSIONAL SUMMARY:**

* **Over7 years** of overall IT experience with **4 Years** of comprehensive experience as Hadoop/Spark developer using Big data technologies like Hadoop Ecosystem, Spark Ecosystem.
* In-depth understanding of **Spark** Architecture including **Spark Core, Spark SQL, Data Frames, Spark Streaming.**
* Expertise in writing **Spark RDD and Data Frames** transformations, actions, Data Frames, case classes for the required input data and performed the data transformations using Spark-Core.
* **Good knowledge** in evaluating **big data analytics libraries** and use of **Spark-SQL** for **data exploratory**.
* Experiencein writing **Hadoop Jobs**for analyzing structured and unstructured data from web logs and social networks using **HDFS**, **Hive, HBase, Pig, Spark, Kafka** and **Scala .**
* Good knowledge of Hadoop Architecture and various components such as **HDFS**, **Job Tracker, Task Tracker, Name Node, Data Node, YARN** and**MapReduce**concepts.
* Working experience on designing and implementing complete **end-to-end Hadoop** Infrastructure including **PIG, HIVE, Sqoop, Oozie, Flume**and **zookeeper**.
* Experience in using Accumulator variables, Broadcast variables and RDD caching for spark streaming.
* Experience in providing support to data analyst in running **Pig and Hive queries**.
* Experience in designing both time driven and data driven **automated workflows** using **Oozie**.
* Hands on experience in various Hadoop distributions **Cloudera** and **Hortonworks** .
* Knowledge in **shell scripts** to dump the shared data from**MySQL** servers to **HDFS**.
* Experience in performance tuning the **Hadoop cluster**by gathering and analyzing the existing infrastructure.
* Strong debugging and problem-solving skills with excellent understanding of system development methodologies, techniques and tools.
* Knowledge in **Amazon Web Services (AWS)** using **EC2** for computing **S3** as storage mechanism and Cloud services like **EBS**, **RDS** and **VPC**.
* Developed Hadoop solutions on AWS from developer to admin roles utilizing the Hortonworks hadoop stack.
* Spined an **EC2** instance of an 10 node cluster in **AWS** by my own as part of my learning for the Certification and personal Interest.
* Manaed Role Based Security and Hadoop Admin load balancing on **AWSEC2** Cluster.
* Expertise in creating Hive Internal/External tables, Hive's analytical functions, views and writing scripts in HQL.
* Experience in using Sequence files, ORC, AVRO and Parquet file formats.
* In-depth Knowledge of Data Structures, Design and Analysis of Algorithms.
* Worked in complete **Software Development Life Cycle** (analysis, design, development, testing, implementation and support) in different application domain involving different technologies varying from object-oriented technology.
* Familiar with **RDBMS** concepts and worked on **Oracle, SQLServer, DB2.**
* Ability to articulate different complex Statistical Concepts and identify **KEY INSIGHTS** from **DATA.**
* Experience in importing and exporting data from**mainframe systems** to unix with help of **NDM**.
* Hands on experience in application development using Java, RDBMS, and Linux shell scripting.
* Experience working with JAVA J2EE, JDBC, ODBC, JSP, Java Beans, Servlets.
* Exposed to build tools like **MAVEN** and bug tracking tool JIRA in the work environment.
* Have Good knowledge on the Bank Administration Institute formats (BAI Formats).
* Received several accolades from Clients.
* Exceptional ability to quickly master new concepts and capable of working in-group as well as independently with excellent communication skills.
* Facilitated knowledge transfer sessions.
* Very good at handling Onsite-Offshore model project and Strong commitment, organizational and detail oriented skills.
* Ability to work independently and as team.
* Strong domain knowledge on Banking, Insurance and Financial Domain Applications.

**CERTIFICATIONS:**

**AWS Certified Developer – Associate**

**Validation Number**:NVTXXMG1BMBQ1D5R

**Validate at**: <http://aws.amazon.com/verification>

**Hortonworks Certified Associate**

**Validate at:** <http://verify.skilljar.com/c/k6bzhq9dqcis>.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Bigdata Technologies** | Hadoop,Spark, HDFS, MapReduce, TEZ, YARN,Hive, Pig, Zookeeper, Sqoop, Oozie, IMPALA, HBASE, Kafka,Flume |
| **Hadoop Distributions** | Cloudera, Horton works. |
| **Programming Languages** | Scala, COBOL, JCL, Java,C,C++, JavaScript, XML, HTML, CSS. |
| **Databases** | MySQL, Oracle, Microsoft SQL Server, DB2, MongoDB,IMSDB. |
| **IDE and Build Tools** | Eclipse, Intellij, Sublime Text, Maven, SBT. |
| **Operating System** | Windows, Linux/Unix. |
| **Methodologies** | Agile (Scrum), Waterfall, UML, Design Patterns, SDLC. |
| **Version Control** | Git, SVN, CA Harvest. |
| **AWS services** | EC2, S3,DynamoDB, EMR, IAM,SNS, SQS,VPC,Cloud Formation, EBS, Elastic Beanstalk, Elastic Load Balancer, RDS. |
| **Others** | In sync, ISPF, DB2I, QMF, BMC Catalog Manager, Dump master, Freeze Frame, Apptune,Endevor, BMC Remedy, NDM, MQ,JIRA. |

**PROFESSIONAL EXPERIENCE:**

**Komatsu Mining Corporation,Milwaukee, WI Apr 2017 – Present**

**Hadoop/Spark Developer**

**Description:**Komatsu’s JoySmart Solutions is an IIoT-based service that helps customers optimize machine performance using machine data and analytics. The JoySmart platform ingests, stores and processes a wide variety of data collected from mining equipment operating around the globe, often at very remote locations in harsh conditions. Types of equipment monitored includes longwall mining systems, electric rope shovels, continuous miners and wheel loaders. Their data includes time-series metrics machine pressures, temperatures, currents, etc. alarm and event data, and other information from third party systems. A single machine can have hundreds to multiple thousand data metrics and generate 30,000-50,000 unique time-stamped records per minute.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop.
* Developed UDF’s using Scala scripts,which used in Data frames/SQL and RDD in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.
* Experience in designing and developing POCs in Spark using Scala to compare the performance of Spark with Map Reduce, Hive.
* Good understanding on DAG cycle for entire spark application flow.
* Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment.
* Implemented Spark Scripts using Scala, Spark SQL to access hive tables into spark for faster processing of data.
* Wrote Scala scripts to make spark streaming work with Kafka as part of Spark Kafka integration efforts.
* Used JSON and XML Serde’s for Serialization and de-Serialization to load JSON and XML data into Hive tables.
* Experience in using different HDFS file formats like Avro, RC, ORC, Parquet, Sequence files and compression formats like Codec and Snappy.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment.
* Worked in Agile Iterative sessions to create Hadoop Data Lake for the client.
* Configured Spark Streaming to get ongoing information from the Kafka and store the stream information to HDFS.
* Worked closely with the customer to understand the business requirements and implemented them.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, and Scala.
* Involved in creating Hive tables,and loading and analyzing data using hive queries.
* Written Hive queries on the analyzed data for aggregation and reporting.
* Imported and exported data from different databases into HDFS and Hive using Sqoop.
* Used HUE for Hive Query execution.
* Extensive knowledge of working on NiFi.
* Hands on design and development of an application using Hive (UDF).
* Used Sqoop for loading existing data in Oracle to HDFS.
* The collection of Necessary data to store it in one Central big data lake database This Centralized big data lake will feed into Tableau dashboard to provide clear report.

**Environment:** HDFS, Hive, Pig, Spark, Spark-Streaming, Spark,Yarn, Linux, SQOOP, Scala, Tableau, Intellij, Oracle, Git, Shell Scripting.

**Amadeus Hospitality, Portsmoth, NH Aug 2016 – Apr 2017**

**Hadoop Developer**

**Description**:  Amadeus develops open, cloud-based technology solutions for global hospitality organizations. From search and booking to check-in and departure, using latest technologies apache spark and spark streaming building the datamarts for reporting purposes.

**Responsibilities:**

* Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs.
* Devised procedures that solve complex business problems with due considerations for hardware/software capacity and limitations, operating times and desired results.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Provided quick response to ad hoc internal and external client requests for data and experienced in creating ad hoc reports.
* Migrated an existing on-premises application to AWS and Implemented 10 node cluster in AWS with help of the EMR in the test environment to implement in the cloud.
* Knowledge on Amazon EC2 Spot integration & and Amazon S3 integration.
* Managed Amazon redshift clusters such as launching the cluster by specifying the nodes and performing the data analysis queries.
* Optimizing the EMRFS for Hadoop to directly read and write in parallel to AWS S3 performantly
* Experience inAWS services like EC2 and S3 for small data sets.
* Used Cloud watch logs to move application logs to S3 and create alarms based on a few exceptions raised by applications.
* Extract real time feed using Kafka and Spark Streaming and convert it to RDD and process data in the form of Data Frame and save the data as Parquet format in HDFS.
* Responsible for building scalable distributed data solutions using Hadoop.
* Handled Imported of data from various data sources, performed transformations loaded data into HDFS.
* Installed Oozie workflow engine to run multiple spark jobs.
* Exploring with Spark, improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, and Pair RDD's.
* Imported and exported data from different databases into HDFS and Hive using Sqoop.
* Experience migrating MapReduce programs into Spark transformations using Spark and Scala.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data.
* Involved in converting Hive/SQL queries into Spark transformations using Spark SQL,Python and Scala.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Extensively worked on the core and Spark SQL modules of Spark.
* Used Spark API over Hadoop YARN to perform analytics on data in Hive and Spark SQL.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Create views over HBase table and used SQL queries to retrieve alerts and meta data.
* Worked with HBASE NOSQL database.
* Helped and directed testing team to get up to speed on Hadoop Data testing.
* Worked on loading and transforming large sets of structured, semi structured and unstructured data.
* Used Oozie Workflow engine to run multiple Hive and Pig jobs.
* Created stored procedures, triggers and functions to operate on report data in MySQL.

**Environment:** Cloudera, Spark, Spark-Streaming, Spark, Yarn, SQOOP, Scala, Oozie, Tableau, Intellij, Oracle, Git, Shell Scripting.

**BNY Mellon,**[**New York City, NY**](https://www.google.com/search?rlz=1C1CHBF_enUS777US777&q=New+York+City&stick=H4sIAAAAAAAAAOPgE-LSz9U3MCorKDatUOIEs-ONLMy0tLKTrfTzi9IT8zKrEksy8_NQOFYZqYkphaWJRSWpRcUAKrUwnEUAAAA&sa=X&ved=0ahUKEwi4o4qymMzZAhVRhOAKHesFBesQmxMIiwIoATAb) **Nov 2014 - Jul 2016**

**Hadoop Developer**

**Description:**BNY Mellon acquired lot of business over period and hence there were duplication of processes. There are 17 different file transfer servers to send reports, feeds from BNY Mellon to users and external systems. This project was aimed at building a single solution that reports and monitors the 17 systems.

**Responsibilities:**

* Involved in review of functional and non-functional requirements.
* Installed and configured Hadoop Mapreduce, HDFS, Developed multiple MapReduce jobs in java for data cleaning and pre-processing.
* Develop MapReduce jobs for the users. Maintain, update and schedule the periodic jobs which range from updates on periodic MapReduce jobs to creating ad-hoc jobs for the business users.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Experienced in defining job flows.
* Programming spark streaming on the real time data and processed the data by applying the business logic with accurate results.
* Experienced in managing and reviewing Hadoop log files.
* Extracted files from SQLDB through Sqoop and placed in HDFS and processed.
* Experienced in running Hadoop streaming jobs to process terabytes of xml format data.
* Load and transform large sets of structured, semi structured and unstructured data.
* Responsible to manage data coming from various sources.
* Supported Map Reduce Programs those are running on the cluster.
* Created NDM Jobs in Mainframe in order to copy the daily SOR files from Mainframe to Edgenode (Unix
* Involved in loading data from UNIX file system to HDFS.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally inmap reduce way.
* Developed a custom File System plug in for Hadoop so it can access files on Data Platform.
* Designed and implemented Mapreduce-based large-scale parallel relation-learning system
* Setup and benchmarked Hadoop/HBase clusters for internal use.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Provided quick response to ad hoc internal and external client requests for data and experienced in creating ad hoc reports.
* Facilitated knowledge transfer sessions.

**Environment:**Eclipse, Oracle 10g, Hadoop, Hive, HBase, Linux, MapReduce, Java (JDK 1.6), Cloudera, MapReduce, DataStax, IBM DataStage 8.1, Oracle 11g / 10g, PL/SQL, SQLPLUS, UNIX Shell Scripting.

**INautix Technologies,Chennai, India Jan 2012- Oct 2014**

**Java Developer**

**Description:**BNY Mellon (NYSE symbol: BK) is a global leader in providing a comprehensive array of services that enable institutions and individuals to manage and service their financial assets in more than 100 markets worldwide. We strive to be the global provider of choice for investment management and investment services and be recognized for our broad and deep capabilities, superior client service and consistent outperformance versus peers. Our global client base consists of financial institutions, corporations, government agencies, endowments and foundations and high-net-worth individuals.

**Responsibilities:**

* Individually worked on all the stages of a Software Development Life Cycle (SDLC).
* Extensively used core Java,Servlets,jsp and XML.
* Extensively applied various design patterns such as MVC, Front Controller, Factory, Singleton, DAO etc. throughout the application for a clear and manageable distribution of roles.
* Used JavaScript code, HTML and CSS style declarations to enrich websites.
* Implemented the application using Spring MVC Framework which is based on MVC design pattern.
* Writing stored procedures to process the data and store it in an organized manner.
* Wrote SQL queries, stored procedures, and triggers to perform back-end database operations.
* Developed SQL stored procedures and prepared statements for updating and accessing data from database.
* Fine tuning SQL queries for maximum efficiency to improve the performance.
* Used MQ Series for integrating with other legacy systems.
* Developed application service components and configured beans using (applicationContext.xml) Spring IOC
* Designed User Interface and the business logic for customer registration and maintenance.
* Integrating Web services and working with data in different servers.
* Involved in designing and Development of SOA services using Web Services.
* Understanding the requirements from business users and end users.
* Working with XML/XSLT files.
* Experience creating UML class and sequence diagram.
* Experience in Creating Tables, Views, Triggers, Indexes, Constraints and functions in SQLServer2005.
* Worked in content management for versioning and notifications.
* Involved in configuring and deploying the application with Tomcat.
* Involved in Unit testing, Integration testing and User Acceptance testing.
* Utilizes Java and SQL day to day to debug and fix issues with client processes.

**Environment:**Java,python, J2EE, HTML, CSS,Java Script, Struts, Hibernate, Eclipse IDE.

**INautix Technologies,Chennai, India Sep 2010- Dec 2011**

**Mainframe Developer**

**Description:**Modify Treasury Edge Event Manager Process to interface with GSF to obtain information previously provided by AIS for the AIS Cash Position Even. The current process that generates the AIS Cash Position Event (scheduled event) must interface with GSF in exactly, the same way Treasury Edge will interface with GSF when an Investment Cash Position Report is manually requested. The request and GSF’s response should be identical between the two processes, so that GSF would not be required to know and differentiate between the two requests. The report requested manually via Treasury Edge should be generated identically via Event Manager.

**Responsibilities:**

* Involves in designing and creation of new tables for the development
* Involved in conversion of databace for IMS DB to DB2.
* Worked on Backend implementation using MVC framework.
* Auto creation for jcl modules from the COBOL program
* Creation of code blocks for the application
* Developed programs for accessing the database and creating XML format to send the request for wed application.
* Written SQL code blocks using cursors for shifting records from various tables based on checks
* Written procedures and triggers for validating the consistency of metadata.
* Knowing the existing logic for end to end for the migration and analysis of impact on existing modules.
* Involved in requirement analysis, coding, testing and release of all the elements related to the project.
* Preparation of Detailed Design Document.
* Preparation of elements migration requests and checklist.
* Followed coding and documentation standards

**Environment:** COBOL, DB2, DB2 Utilities, JCL, Sort, MQ.

**EDUCATION:-**

**Bachelor of Technology in Electronics and Communications Engineering**

S.A.S.T.R.A University – Tamil Nadu, India