 

**Nagarjuna**

**Sr. Bigdata Engineer/Developer Email-ID :** **nagarjun.bigdata@gmail.com**

**(Big data/ Spark) Contact : (660)-851-8513**

**PROFESSIONAL SUMMARY:**

* **Having 7+ years of IT experience in Architecture, Analysis, design, development, implementation, maintenance and support with experience in developing strategic methods for deploying BIG DATA technologies to efficiently solve Big Data processing requirement.**
* Around 4 years of Experience on **BIG DATA** using **HADOOP** framework and related technologies such as **HDFS, HBASE, Map Reduce, HIVE, IMPALA, PIG, FLUME, OOZIE, SQOOP, TALEND** and **ZOOKEEPER**.
* Around 2 years of experience on apache **SPARK** and **KAFKA**.
* Experience in data analysis using **HIVE, PIG LATIN, HBASE** and custom Map Reduce programs in **JAVA**.
* Pretty good Experience with Cloudera and Horton works distributions.
* Experience in working with **FLUME, SHELL SCRIPTING** to load the log data from multiple sources directly into **HDFS**.
* Worked on data load from various sources i.e., **Oracle, MySQL, DB2, MS SQL Server, Cassandra, MongoDB, Hadoop** using **Sqoop** and **Python script**.
* Developed Spark Application by using Python (Pyspark).
* Excellent understanding /knowledge on Hadoop (Gen-1 and Gen-2) and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Resource Manager (YARN).
* I have been experienced with **SPARK SREAMING API** to ingest data into **SPARK ENGINE** from **KAFKA**.
* Developed analytical components using **SCALA, SPARK, STORM** and **SPARK STREAM**.
* Excellent understanding and knowledge of **NOSQL** database **HBASE** and **CASSANDRA**.
* Worked extensively with Dimensional **MODELING, DATA MIGRATION, DATA CLEANSING, DATA PROFILING**, and **ETL** Processes features for data warehouses.
* Implemented Hadoop based data warehouses, **INTEGRATED HADOOP** with **ENTERPRISE DATA WAREHOUSE** systems. Extensive experience in ETL Data Ingestion, In-Stream data processing, **BATCH ANALYTICS** and Data **PERSISTENCE STRATEGY**.
* Experience with creating the **TABLEAU** dashboards with relational and multi-dimensional databases including Oracle, MySQL and HIVE, gathering and manipulating data from various sources.
* Experience on monitoring, performance tuning, **SLA, SCALING** and security in Big Data systems.
* Design and document **REST/HTTP**, SOAP APIs, including JSON data formats and API versioning strategy.
* Installed and configured **JENKINS FOR AUTOMATING** Deployments and providing automation solution.
* Experience in designing both time driven and data driven automated workflows using Oozie.
* Experience working with **JDK 1.7, JAVA, J2EE, JDBC, ODBC, JSP, JAVA ECLIPSE, JAVA BEANS, EJB, SERVLETS, MS SQL SERVER**.
* Experience in J2EE technologies like **Struts, JSP/Servlets,** and **spring**.
* Good Exposure on scripting languages like **JAVASCRIPT, ANGULAR JS, JQUERY** and **XML**.
* Created a **JAVADOC TEMPLATE** for engineers to use to develop API documentation.
* Expert in **JAVA 1.8 LAMBDAS, STREAMS,** Type annotations.
* Experience in all stages of SDLC (Agile, Waterfall), writing Technical Design document, Development, Testing and Implementation of Enterprise level Data mart and Data warehouses.
* Extensive experience working IN **ORACLE, DB2, SQL SERVER** and **My SQL** database.
* Have Experience in monitoring tools like Splunk and Cloudera Manager tool.
* Ability to work in high-pressure environments delivering to and managing stakeholder expectations
* Application of structured methods to: Project Scoping and Planning, risks, issues, schedules and deliverables.
* Good Inter personnel skills and ability to work as part of a team. Exceptional ability to learn and master new technologies and to deliver outputs in short deadlines

**DOMAINS:**

* **Health Care, Banking and E – Commerce, online salvage auctioning industries**.

**CERTIFICATIONS:**

* **Cloudera Certified Hadoop and Spark Developer(CCA-175**)
* **Bachelor’s in ECE from Jntu-Hyderabad, 2010.**

**TECHNICAL SKILLS:**

**Big Data Technologies :** Hadoop2.x & 1.x, Hive0.14.0, Pig0.14.0,

Oozie4.1.0, Zookeeper3.4.6, Impala2.1.0,

Sqoop1.4.6, MapReduce2.x, Tez0.6.0,

Spark1.4.0, Flume1.5.2, HBase0.98.0,

Solr4.0.0, Kafka0.8.0, YARN.

**Software &Tools :** Eclipse, Putty, Cygwin, Hue**,** JIRA, IntelliJ IDEA, Net Beans, Jenkins, Aginity.

**Distributions :** Cloudera, Horton works

**Monitoring Tools :** Cloudera Manager, Ambari

**Cloud Technologies :** Aws, Azure.

**Java Technologies :** Core JAVA, JSP, Servlets, spring, Hibernate, Ant, Maven

**Programming Languages :** JAVA, SQL, Pig Latin, HiveQL, Shell Scripting, Python, Scala

**Databases :** NoSQL (HBase), Oracle 12c/11g, MySQL, MS SQL Server

**Testing Methodologies :** JUnit, MRUnit

**Operating Systems :** Windows, Linux (RHEL, CentOS, Ubuntu)

**ETL Tools :** Tableau, Pentaho, Talend

**PROFESSIONAL SUMMARY:**

**Client: PG&E, SFO-CA, May 2017 – Till date**

**Role: Sr BIG DATA Engineer/ Hadoop Developer**

**Project: Data Lake Creation.**

**Responsibilities:**

* Being Part of Data Service team, Actively involved in moving the data from relational databases(Teradata) to the Hadoop.
* Performed data quality checks on data as per the business requirement.
* Performed data validation on target table in compared to the source table.
* Achieved high throughput and low latency for ingestion jobs leveraging the **Sqoop**
* Transformed the raw data from traditional data warehouse and loaded into stage and target tables.
* Fine tuned the Hive Queries for huge tables to realize low latency inserts.
* Developed the Pig 0.15.0 UDF's to pre-process the data for analysis and Migrated ETL operations into Hadoop system using Python Scripts3.5.1.
* Installed/Configured/Maintained Hortonworks Hadoop clusters for application development.
* Optimally stored the data in hadoop using file formats like avro and parquet.
* Automated the ingestion and transformed components by creating **oozie** workflows.
* Closely worked with Data Modelling team to realize business requirements.
* Performed the incremental imports successfully and made the table in hive consistent.
* Performed Hive partioning and bucketing to reduce the disk I/O.
* Performed windowing and analytical functions in hive to optimize the transportation logistics.
* Consumed the flat files from **Apache Kafka** vendors and imposed hive schema on top of it to correlate

with the tables in data lake.

**Environment -**

**Kafka 1.10.0, MariaDB10.1.21, Data lake, Teradata, Data warehouse, Python, HDFS2.6.0, Hortonworks, Hadoop2.6.0, Spark1.4.0, Zookeeper3.4.9 , Hive 0.14.0, ava8.**

**Client: ASURION, Nashville-TN April 2016 – April 2017**

**Role: Hadoop developer/Big Data Engineer.**

**,**

**Responsibilities:**

* Developed data pipeline using **FLUME, SQOOP, HIVE AND JAVA MAPREDUCE** to ingest customer behavioral data and financial histories into HDFS for analysis.
* Used **HIVE** to do transformations, event joins, filter boot traffic and **SOME PRE-AGGREGATIONS** before storing the data onto HDFS.
* Extensive experience in ETL (Talend) Data Ingestion, In-Stream data processing, **BATCH ANALYTICS** and **Data PERSISTENCE STRATEGY**.
* Worked on Designing and Developing ETL (Talend) Workflows using Java for processing data in HDFS/Cassendra using Oozie.
* Developed multiple Spark jobs in PySpark for data cleaning and preprocessing.
* Expertise with the tools in Hadoop Ecosystem including **PIG, HIVE, HDFS, MAP REDUCE, SQOOP, KAFKA, YARN, OOZIE, AND ZOOKEEPER**. Hadoop architecture and its components.
* Involved in integration of Hadoop cluster with spark engine to perform **BATCH** and **Streaming** operations.
* Explored with the **SPARK,** improving the performance and optimization of the existing algorithms in Hadoop using **SPARK CONTEXT, SPARK-SQL, DATA FRAME, PAIR RDD'S, SPARK YARN**.
* Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive.
* Experienced in running Hadoop streaming jobs to process terabytes of formatted data using Python scripts
* Import the data from different sources like **HDFS/Hbas**e into **SPARK RDD**.
* Developed **SPARK CODE** using **SCALA** and Spark-SQL/Streaming for faster testing and processing of data.
* Developed **KAFKA PRODUCER** and consumers, HBase clients, SPARK and Hadoop Map Reduce jobs along with components on HDFS, Hive.
* Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Involved in developing **HIVE DDLS** to create, alter and drop Hive tables and storm.
* Create scalable and high-performance web services for data tracking.
* Involved in loading data from **UNIX file** system to HDFS. Installed and configured Hive and also written Hive UDFs and Cluster coordination services through Zoo Keeper.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
* Experienced in managing Bigbucket for java and python code.
* Experienced in managing Hadoop Cluster using **CLOUDERA MANAGER TOOL**.
* Involved in using **HCATALOG** to access Hive table metadata from Map Reduce .

**Environment**: MAP REDUCE, YARN, HIVE, PIG, CASSENDRA, OOZIE, Talend, SQOOP, SPLUNK, KAFKA, ORACLE 11G, CORE JAVA, CLOUDERA, AKKA, ECLIPSE, PYTHON, SCALA, SPARK, PYSPARK, SQL, PYTHON, HORTONWORKS. TABLEAU, BIG BUCKET, UNIX SHELL SCRIPTING.

**Client: Trans Union, Bloomington, IL July 2015 – March 2016**

**Role: Big data Engineer/Hadoop Developer**

**Responsibilities:**

* Designed a data warehouse using Hive.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Using **Hive, Map-reduce**, and loaded data into HDFS.
* Responsible for building Hadoop clusters with hortonworks Distribution and integrate with Pentaho Data Integration (PDI) server.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.
* Monitored workload, job performance and capacity planning using Cloudera Manager.
* Worked extensively with **SQOOP** for importing metadata from Oracle.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Extensively used Pig for data cleaning.
* Created partitioned tables in Hive.
* Worked with business teams and created Hive queries for ad hoc access.
* Evaluated usage of **Oozie** for Workflow Orchestration.
* Mentored analyst and test team for writing Hive Queries.
* Experience in writing Map Reduce programs with Java API to cleanse Structured and unstructured data.
* Experience in **RDBMS** such as **Oracle, Teradata**
* Worked on loading the data from **MySQL & Teradata** to **HBase** where necessary using Sqoop.
* Consumed the flat files from **Apache Kafka** vendors and imposed hive schema on top of it to correlate

with the tables in data lake.

* Launching Amazon EC2 Cloud Instances using Amazon Images (Linux/ Ubuntu) and Configuring launched instances with respect to specific applications.
* Gained very good business knowledge on claim processing, fraud suspect identification, appeals process etc.

**Environment:** Hadoop, Map Reduce, HDFS, Hive, Teradata, Pig, Sqoop, AWS, Kafka, Java, python, Hortonworks, Oozie, MySql.

**Client: NCR Corporation, Atlanta, GA Sep 2014 – June 2015**

**Role: Big data/Hadoop Developer**

**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.
* Used Pig to do data transformations, event joins and some pre-aggregations before storing the data on the HDFS.
* Exploited Hadoop MySQL-Connector to store Map Reduce results in **RDBMS**.
* Used Ambari for Managing Hortonworks Distribution of Hadoop, especially for fault-tolerant workflows & error handling etc.
* 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 EMR and JAVA**.
* Collected data from different databases (i.e. Oracle, MySQL) to Hadoop.
* Working on **Azure** environment like deploying.
* 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 Zoo Keeper.
* 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 Map Reduce.
* Developed Simple to complex Map Reduce Jobs using Hive and Pig.

**Environment:** Apache Hadoop, AWS EMR, Map Reduce, HDFS, Hive, Java, python, SQL, PIG, Zookeeper, Java (jdk1.6), Flat files, Oracle 11g/10g, MySQL, Windows NT, UNIX, Sqoop, Hive, Oozie, HBase.

**Client: Woolworths Retail, Hyderabad-India. November 2013 – July 2014**

**Role: JAVA/J2EE Developer**

**Responsibilities:**

* Designed Use Case and Sequence Diagrams according to UML standard using Rational Rose.
* Implemented Model View Controller (MVC-2) architecture and developed Form classes, Action Classes for the entire application using Struts Framework.
* Performed client-side validations using JavaScript and server-side validations using in built Struts Validation Framework.
* Implemented the data persistence functionality of the application by using Hibernate to persist java objects to the relational database.
* Used Hibernate Annotations to reduce time at the configuration level and accessed Annotated bean from Hibernate DAO layer.
* Worked on various SOAP and RESTful web services used in various internal applications.
* Used SOAP UI tool for testing the RESTful web services.
* Used HQL statements and procedures to fetch the data from the database.
* Transformed, Navigated and Formatted XML documents using XSL, XSLT.
* Used **LAMBDA EXPRESSION OF JAVA 1.8** features extensively to remove the boiler plate code and to extend the functionality.
* Used a LAMBDA EXPRESSION to improve Sack Employees further and avoid the need for a separate class.
* Used JMS for asynchronous exchange of message by applications on different platforms.
* Developed the view components using **JSP, HTML**, Struts Logic tags and Struts tag libraries.
* Involved in designing and implementation of Session Facade, Business Delegate, Service Locator patterns to delegate request to appropriate resources.
* Used JUnit Testing Framework for performing Unit testing.

**Environment:** Struts 2.0 Hibernate 3.0, JSP, JDK 1.7, RAD, JMS, CVS, JavaScript, XSL, XSLT, lambda expression, Servlets 2.5, Web Sphere Application Server, Oracle 10g.

**Client: Syntel, Hyderabad - IN March 2010 – October 2013**

**Role: Java Developer**

**Responsibilities:**

* Involved in the analysis, design, and development and testing phases of Software Development Life Cycle (SDLC)
* Designed and developed framework components, involved in designing MVC pattern using Struts and spring framework.
* Responsible for developing Use case, Class diagrams and Sequence diagrams for the modules using UML and Rational Rose.
* Developed the Action Classes, Action Form Classes, created JSPs using Struts tag libraries and configured in Struts-config.xml, Web.xml files.
* Involved in Deploying and Configuring applications in Web Logic Server.
* Used SOAP for exchanging XML based messages.
* Used Microsoft **VISIO** for developing Use Case Diagrams, Sequence Diagrams and Class Diagrams in the design phase.
* Developed Custom Tags to simplify the JSP code. Designed UI screens using JSP and HTML.
* Actively involved in designing and implementing Factory method, Singleton, MVC and Data Access Object design patterns.
* Web services used for sending and getting data from different applications using SOAP messages. Then used DOM XML parser for data retrieval.
* Wrote **JUNIT** test cases for Controller, Service and DAO layer using **MOCKITO, DBUNIT**.
* Developed unit test cases using proprietary framework which is similar to JUNIT.
* Used JUnit framework for unit testing of application and ANT to build and deploy the application on Web Logic Server.

**Environment**: Java, J2EE, JDK1.7, JSP, Oracle, VSAM, Eclipse, HTML, Junit, MVC, ANT, Web Logic.