**Nagarjuna Putluri**

[nagarjunaputluri@gmail.com](mailto:nagarjunaputluri@gmail.com)

**(518)-595-9320**

**Relocation**: **Open**

<https://www.linkedin.com/in/nagarjuna-putluri-686899169>

**PROFESSIONAL SUMMARY:**

* **7+ years** of experience in Information Technology which includes **4 years** of experience in **Big Data technologies including Hadoop and Spark**, Around **3 years** of experience **in Java development.**
* Excellent understanding or knowledge of **Hadoop architecture** and various components such as Spark Ecosystem which includes (**Spark SQL, Spark Streaming, Spark MLib, Spark GraphX), HDFS, MapReduce, Pig, Sqoop, Kafka, Hive, Cassandra, Hbase, Oozie, Zookeeper, Flume, Impala, Hcatalog, Strom, Tez andYARN concepts like Resource Manager, Node Manager (Hadoop 2.x).**
* Designed **HIVE queries & Pig scripts** to perform data analysis, data transfer and Data distribution by implementing **partitioning, bucketing, joints.**
* Expertise in writing custom **UDFs in Pig & Hive** Core Functionality, Hands on experience dealing with **ORC, AVRO** and **Parquet** file format.
* Imported the data from relational databases into HDFS using SQOOP.
* Developed Spark jobs using Scala in test environment for faster testing and data processing and used **Spark SQL** for querying and to access **hive** tables into spark for faster processing of data. Performed map-side joins on **RDD, Spark SQL** and **Data Frames**.
* Extracted 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**.
* Hands-on experience in Amazon Web Services (AWS) Cloud services like **EC2, S3, EMR** and involved in **ETL**, Data Integration and Migration.
* Exported data to various Databases like**Teradata (Data Warehouse), SQL-Server, Cassandra using Sqoop and worked with databases like Snowflake, Teradata, Hbase, Mongo DB, Cassandra, MySQL and Oracle.**
* Experience working with **Cloudera**and **Hortonworks** distributions.
* Involved In working with **Maven** for build process.
* Extensive experience on importing and exporting data using stream processing platforms like **Flume and Kafka**
* Experience in data workflow scheduler **Zoo-Keeper** and **Oozie** to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with the control flows.
* Worked on Java Concepts like **Multithreading and Collections**.
* Worked on creating the User Interfaces using **HTML, CSS, JavaScript**.
* Used **JDBC drivers** to connect to the backend **ORACLE database**.
* Involved in Developing **Servlets** and Java Beans programming to communicate between client and server.
* Good understanding and Experience with **Agile and Waterfall** methodologies of Software Development Life Cycle (SDLC).
* Good analytical, communication, problem solving skills and adore **learning new technical,functional skills.**

**Hortonworks Certified Associate**

**Validate at:**<http://verify.skilljar.com/c/g5uc2bz23xoq>

**TECHNICAL SKILLS**:

|  |  |
| --- | --- |
| **Bigdata Technologies** | Hadoop, MapReduce, HDFS, Hive, Pig, Spark, Yarn, Zookeeper, Sqoop, Oozie, Flume, Impala, HBASE, Kafka, Storm, Amazon AWS, Cloudera and Hortonworks |
| **Build Tools** | Git, Ant, SVN, Maven |
| **Hadoop Distributions** | Cloudera, Horton works, Amazon EMR, EC2. |
| **Programming Languages** | C, C++, Core Java, shell scripting, Scala. |
| **Databases** | RDBMS, MySQL, Oracle, Microsoft SQL Server, Teradata SQL, DB2, PL/SQL, CASSANDRA, MongoDB, Snowflake, Hbase. |
| **IDE and Tools** | Eclipse, NetBeans, Tableau, Microsoft Visual Studio |
| **Operating System** | Windows, Linux/Unix. |
| **Scripting Languages** | JSP & Servlets, JavaScript, XML, HTML, Python, Shell Scripting. |
| **Application Servers** | Apache Tomcat, Web Sphere, Web logic. |
| **Methodologies** | Agile, SDLC, Waterfall. |
| **Web Services** | Restful, SOAP. |
| **ETL Tools** | Talend, Informatica. |
| **Others** | Solr, Tez, Cloud Break, Atlas, Falcon, Ambari, Ambari Views, Ranger, Knox. |

**PROFESSIONL EXPERIENCE:**

**Role:Hadoop/Spark Developer July 2017 to till date**

**Client:Komatsu Mining Corporation**

**Location:Milwaukee, WI**

**Project 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:**

* Ingested gigabytes of data from **S3 Buckets** into tables in **Snowflake Database**
* Experience in ingesting data into **Teradata DB** which is a relational database.
* Created Sqoop scripts to import/export data from RDBMS to S3 data store.
* The data is taken from Data lake and the raw data is in **JSON** format.
* Developed various spark applications using Scala to perform various enrichment of thesedata merged with user profile data.
* Used Storm service extensively to connect to Active MQ and KAFKA to push data to HBASE and HIVE tables.
* Setup a Multi Node Cluster Plan and Deploy a Hadoop Cluster using Hortonworks Ambari.
* Deployed multi-node development, testing and production Hadoop clusters with different Hadoop components (HIVE, PIG, SQOOP, OOZIE, FLUME, HCATALOG, ZOOKEEPER) using Hortonworks Ambari.
* Used NIFI to pull the data from different source and to push the data to HBASE and HIVE.
* Developed Applications for **Tokenization** using **Spark** with Java Framework.
* Developed **Spark-Scala Scripts** for Absolute Data Quality check.
* Involved in **data cleansing, event enrichment, data aggregation, de-normalization** and data preparation needed for machine learning and reporting.
* Used **Split Framework** which is developed using **Spark-Scala** scripts.
* Used **MPP loader** to ingest data into tables which is written in Python.
* Worked with **Parquet** format for storage which is a columnar storage.
* Design and Develop Hadoop ETL solutions to move data to the data lake using big data tools like Sqoop, Hive, Spark, HDFS, Talend etc.
* Utilized **Spark Scala API** to implement batch processing of jobs
* Trouble Shooting Spark applications for improved error tolerance.
* Fine-tuning spark applications/jobs to improve the efficiency and overall processing time for the pipelines.
* Configured and deployed hive metastore using MySQL and thrift server.
* Developed Simple to complex MapReduce Jobs using Hive and Pig.
* Involved in creating Hive tables, and loading and analyzing data using hive queries.
* Utilized **Spark in Memory capabilities**, to handle large datasets.
* Experienced in working with **EMR cluster and S3 in AWS cloud.**
* Creating tables in snowflake DB, loading and analyzing data using Spark-Scala scripts. Implemented Partitioning, Dynamic Partitions.
* Involved in Delta Migration using Sqoop Incremental Updates.
* Involved in continuous Integration of application using Jenkins.
* Git for version control and **Maven** as build tool.
* Followed **Agile** methodologies in analysis, define and document the applications, which will support functional and business requirements.

**Environment:** AWS Elastic MapReduce, Spark, Scala, Python, Hortonworks, Jenkins, Hive, Amazon S3, Sqoop, Teradata, Snowflake DB, Jupiter Notebook, Git, Maven.

**Role: Hadoop Developer Jan 2016 to July 2017**

**Client: Express Scripts**

**Location: Austin, TX**

**Project Description**:

Project objective is to assemble a Centralized Analytical information store for the whole association to have the capacity to convey all the expository necessities of all the business units. Migrated different information sources from existing Enterprise Data-warehouses to HDFS.

**Responsibilities:**

* Experience in working with migrating data from traditional **RDBMS to HDFS.**
* Ingested data into **HDFS from Teradata, MySQL using Sqoop**.
* Part developing spark application to perform ETL kind of operations on the data.
* Redesigned the existing **MapReduce jobs** to **Spark transformations** and actions by utilizing Spark RDDs, Dataframes and Spark SQL API's.
* Worked on Hortonworks 2.3distribution.
* Used Hive partitioning, Bucketing and performed various kinds of joins on Hive tables
* Created Hive external tables to perform ETL on data that is produced on daily basis
* Migrated ETL jobs to Pig scripts do Transformations, even joins and some pre-aggregations before storing the data onto HDFS.
* Validated the data being ingested into HIVE for further filtering and cleansing.
* Developed Sqoop jobs for performing incremental loads from RDBMS into HDFS and further applied Spark transformations
* Worked on loading data into hive tables from spark and used Parquet columnar format.
* Created **Oozie workflows** to automate and productionize the data pipelines
* Migrating Map Reduce code into Spark transformations using Spark and Scala.
* Collecting and aggregating large amounts of log data using **Apache Flume and Kafka** and staging data in HDFS for further analysis.
* Used Sqoop to extract and load incremental and non-incremental data from RDBMS systems into
* Hadoop.
* Worked on various enterprise data-warehouses as a part of migration project.
* Worked with **Tableau** to connect to Impala for developing interactive dashboards.
* Followed Agile Methodologies.

**Environment:** Cloudera Hadoop, Spark, Scala, Sqoop, Oozie, Hive,Pig, Hortonworks, Tableau, MySQL, Oracle DB, Flume.

**Role: Hadoop Developer Aug 2014 to Dec 2015**

**Client: ProSites**

**Location: Phoenix, Arizona**

**Project Description:**

Prosites is an medical and dental services where it will be generating huge amounts of data, Utilized Hadoop and tools related to it for daily analysis ofcustomer response data from different web and mobile applications collected by external systems and provide it to business users to analyze and understand consumer behavior to serve and reach customer more effectively that increased the customer effective reach rate to 80% and sales up by 15%.

**Responsibilities:**

* Experience in creating data pipeline for different events of web and mobile applications, to filter and load consumer response data in **AWS S3** bucket into Hive external tables in HDFS location.
* Involved in working with different file formats like Json, AVRO and parquet and compression techniques like snappy.
* Constructed Impala scripts for end user / analyst requirements for ad hoc analysis.
* Worked with various **Hive optimization techniques** like **partitioning, bucketing Map and join.**
* Worked with shell scripts for dynamic partitions adding to hive stage table, verifying Json schema change of source files, and verifying duplicate files in source location.
* Importing and exporting the data from HDFS to RDBMS using Sqoop and Kafka.
* Developed **UDF's** in spark to capture values of a key-value pair in encoded Json string.
* Developed spark application for filtering Json source data in AWS S3 location and store it into HDFs with partitions and used spark to extract schema of Json files.
* Used **Jenkins** for continuous integration and continuous testing.
* Used SQL for querying data from the tables which are in HDFS.
* Used **Amazon S3** buckets for data staging.
* Worked with Sqoop for ingesting data into HDFS from other databases.
* Worked with impala for massive parallel processing of queries and using HDFS as underlying storage for imapala.
* Worked with **Elastic Map Reduce** for data processing and used HDFS for data storage.
* Extensive experience working with different Hadoop distributions like **Cloudera** and Apache distributions.

**Environment:**

Hive, Spark, AWS S3, EMR, SQL, Cloudera, Jenkins, Shell scripting, Hbase, Intellij IDE, Sqoop, spark, Impala.

**Role: Java Developer Jan 2014 to Aug 2014**

**Client: Humana**

**Location: Louisville, KY**

**Project Description:**

Humana Inc., headquartered in Louisville, Ky., is a leading health care company that offers a wide range of insurance products, health and wellness services. The project is an Insurance domain application for brokers who log on to the system to view the Claims, Damages, Cancellations, etc. of their respective customers. Agents are provided with different options to search their products according to the client category, view their voluntary benefits and their Medicare plans etc.

**Responsibilities:**

* Involved in complete Software Development Life Cycle (SDLC) of the application development like Designing, Developing, Testing and implementing scalable online systems in Java, J2EE, JSP, Servlets and Oracle Database.
* Created UML Diagrams like Class Diagrams, Sequence Diagrams, Use Case Diagrams using Rational Rose.
* Implemented **MVC architectur**e using Java Spring Core.
* Implemented java J2EE technologies on the server side like Servlets, JSP and JSTL.
* Worked in Implementing Hibernate by creating hbm.xml file to configure the Hibernate to the Oracle Database.
* Involved in writing **SQL Queries**, Stored Procedures and PL/SQL for the back-end server.
* Used **HTML, JavaScript** for creating interactive User Interfaces.
* Extensively used Custom JSP tags to separate presentation from application layer.
* Developed **JSP Pages** and implemented **AJAX** in them for a responsive User Interface.
* Involved in developing presentation layer using JSP and Model layer using EJB Session Beans.
* Implemented Unit test cases by using Junit and Implemented Log4J for logging and debugging the application.
* Implemented Maven Build Scripts for building the application.
* Deployed the application in IBM Web Sphere and tested for and server related issues.
* Used Git as the repository and for Version Control. Used **Intellij as the IDE** for the development.

**Environment:**

java, J2EE, EJB, Servlet, JSP, JSTL, Spring Core, Spring MVC, Hibernate, HTML, CSS, JavaScript, AJAX, Oracle, Stored Procedures, PL/SQL, Junit, Log4J, Maven, WebSphere, Git, Intellij

**Role: Java Developer Apr 2012 to Nov 2013**

**Client: Value Momentum**

**Location: Financial District, Hyderabad**

**Project Description:**

Value Momentum provides Software and Services to Insurance, Healthcare & Financial Services firms. Worked as a Junior Java developer for LIC life insurance client to develop multiple applications to handle their customers registration, insurance policies and renewals. These applications are for the end client and for the client customers.

**Responsibilities:**

* Actively participated in Designing and defining phases of the Application development.
* Followed Agile methodologies in analysis, define and document the applications, which will support functional and business requirements.
* Develop Use Case Diagrams, Object Diagrams and Class Diagrams in **UML** using Rational Rose.
* Participated in gathering Requirement analysis, Design, Coding, Implementation and Maintenance of this application follow the complete SDLC life cycle along with the team.
* Worked with **JDK 1.3** and worked with core java concepts like Multithreading, Collections, Generics and Serialization.
* Designed and developed frontend using **Servlet, JSP, HTML, CSS and JavaScript**.
* Created tile definition, Structs-Config files and validation files for the application using **STRUTS framework.**
* Implemented Action Classes and Action Forms using Struts.
* Used JDBC drivers to connect to the backend **ORACLE database**.
* Involved in implementing Unit Test scripts using Junit.
* Used ANT as Build tool and deployed the application using **ANT in Apache Tomcat**.
* Used **IBM ClearCase** for version control and workspace management

**Environment:** Agile, JDK 1.3, Struts, Oracle DB, UML, Junit, ANT, IBM ClearCase, Servlet, JSP, HTML, CSS and JavaScript

**Role: Java Developer May 2011 to Apr 2012**

**Client: Eicher Pinnacle**

**Location: Nacharam, Hyderabad**

**Project Description**:

Worked with Eicher Pinnacle, which is a pioneer in manufacturing stand-by gensets, offering a wide range of specifications available in market. Pinnacle is specialized in the field of manufacturing, installation, commissioning, preventive maintenance, servicing of Diesel Generating Sets. Worked as a Junior Java Developer to develop the web portal for the Pinnacle.

**Responsibilities:**

* Understanding User requirements and participating in design discussions, implementation feasibility analysis and documenting requirements.
* Using Rational Rose, Developed Use Case, Class, Activity and Sequence UML Diagrams.
* Worked on Java Concepts like Multithreading and Collections.
* Developed JSP’s and servlets and have used internal tools like content Management to organize JSP.
* Worked on creating the User Interfaces using **HTML, CSS, JavaScript.**
* Implemented Ajax in the User Interface for more responsive **front-end GUI.**
* Involved in Developing **Servlets and Java Beans** programming to communicate between client and server.
* Participated in designing the architecture of the schemas in **MySQL.**
* Written and implemented SQL queries in the application like views and triggers.
* Integrated **Log4J** into the application for Debugging and logging purposes.
* Performed Unit Testing with Junit, Integration Testing and System Testing.
* Deployed and tested the application Using **Apache Tomcat**.

**Environment:**

Java, J2EE, EJB, Servlet, JSP, JSTL, HTML, CSS, JavaScript, AJAX, Oracle, Stored Procedures, PL/SQL, Junit, Log4J, MySQL, Git, Intellij, Apache Tomcat.

**EDUCATION:**

Bachelor of Technology in Electronics and Communication Engineering