Narendra Reddy

**+91-9533585454 |** [**narendra.reddy009@yahoo.com**](mailto:narendra.reddy009@yahoo.com) **|** [**https://www.linkedin.com/in/narendra-reddy-752b7a60**](https://www.linkedin.com/in/narendra-reddy-752b7a60)

# ABOUT

To associate myself with an organisation, which provides me challenging tasks and opportunity to work with cutting edge technologies where I can enhance my knowledge and skills in the fields of Big Data & Data Science. Having 6+ years of work experience in Information Technology industry as a Big Data Engineer, currently associated with Lowe’s, Bengaluru as a Sr.Data Engineer.

* **Big Data with Hadoop –** Experienced in implementing big data solutions using Apache Hadoop & its components like HDFS, Hive, Sqoop, Oozie, Kafka, Apache Spark, Python, MapReduce, Apache Nifi, Cassandra, Scala, Shell Scripting, Jenkins, CICD and Weka.
* Have been implemented the Generic streaming and batch application framework which can be used by any one for streaming any type of data from Kafka topics and storing into HDFS/Hive/HBase.
* Have been implemented different parsers (Avro, Json) to parse/deserilize their respective type of data comes from kafka topics and flattening them into Human readable records (Data Frame).
* Experienced in building enterprise data lake (EDL) for historical data and performing the analytical operations to generating client requested reports for future estimations/promotions.
* Implemented the CICD enterprise workflow automation process and implemented scripts for build, deploy the data to edge nodes and HDFS using Groovy Script.
* Developed Generic scheduling framework to automating, schedule the spark, hive and bash scripts using Oozie workflows and coordinators.
* Responsible for end-to-end unit testing and document the results for each requirement story and review them with the lead/manager before moving to production.
* Translating product requirements into analytical requirements/specification, design and development required functionality.
* Experience in automating, managing and building bridge between machine learning models and Hadoop framework using python wrapper scripts.
* Strong programming, database and scripting knowledge with experience in working on SDLC, AGILE frameworks and having strong domain knowledge in Marketing and Retail industry.
* Conducted corporate trainings on Hadoop, it’s ecosystem and also actively involved in knowledge sharing sessions and trainings to help colleagues to learn the technologies.

# QUALIFICATIONS

**B. Tech, Information Technology** | **Jawaharlal Nehru Technological University**, Anantapur, India | 2009-2013

# EXPERTISE

* **Big Data:** Hadoop, HDFS, Hive, Sqoop, Oozie, Apache Nifi, Kafka and PySpark.
* **Machine Learning:** RandomForest, K-Nearest Neighbour, XG-Boost, Grid/RandomSearchCV
* **Integrations Tools:** CICD, Jenkins, git and Bitbucket
* **Python Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Sklearn and etc…
* **Programming:** Python, Spark, Scala, Groovy, Weka.
* **Functional Knowledge:** Manufacturing, Retail services, Marketing and Operations Analytics.

# EXTRA CURRICULAR ACTIVITIES

* Participated in **“River Rafting & Flying Fox”** event conducted by KPIT, 2016
* Participated in **“GO GREEN”** program organized by National Entrepreneurship Network (NEN) to encourage society to invest in green future.
* Passionate about **“Photography” –** wildlife, nature and people

# WORK EXPERIENCE

## IT Analyst – Big Data Developer (Tata Consultancy Services Pvt. Ltd, Bangalore)

**TCS** India account currently serving American multinational technology company that designs, develops and provide access to their products on subscription basis and online services as an IT Partner.

## Responsibilities

* Responsible for collecting the Business requirements, design and develop the solution.
* implemented the generic framework for streaming and batch data processing using spark which requires transition by passing parameters from HBASE.
* Have been implemented the code to deserialize the Avro data format to json and flattening the nested json record to DF using custom parser.
* Have been implemented Avro schema generation which can map the schema with record and handles the nulls and missing columns on fly.
* Developed python scripts for integrating the machine learning algorithms and Hadoop platform. which can feed the ML model using hive data and process the same.
* Automated hive logic scripts through parallel execution using oozie workflow/coordinator app, the entire workflow Automated through python and shell script/wrapper.
* Implemented data quality framework to handle data quality issues, audit framework for data integrity and primary for validating the landing/hive data.
* Worked CICD integration flow and written basic groovy script to build, test and deploy the git code edge nodes and HDFS.
* Responsible for end-to-end unit testing and document the results for each requirement story and review them with the test lead before moving to production
* Prepared High Level and Low-Level Design Documents for Streaming & Archive process by collaborating with lead.

**Technology & Tools:** PySpark, Spark Streaming, Jenkins, Groovy, HBase, Shell Scripting, Hadoop, Hive, Oozie, Tidal, Analytics, Git, CICD, Python, Scala and Agile

## Associate –Big Data Developer (Cognizant Technology Solutions, Bangalore)

**CTS** is recognized as a leading IT Services firm and India account currently works for US based product company that designs, develops and sells consumer electronics across the globe.

## Roles and Responsibilities

* Responsible to collect the client requirements for specific sources and design the solution for the same.
* Developed a shell scripts to accommodate Full loads & Incremental loads using Sqoop.
* Developed the hive scripts for creating the hive tables and performing the analytics to generate the client requested adhoc reports.
* Developed Scala code to parse Nested Json data.
* Design and Implemented the archive/purge strategy to move the older data to archive locations to increasing the query performance in partitioned tables.
* Migrated Couple of Incremental and Historical ETL workflows from HiveQL to Spark ETL for making the data available for User’s for further usage.
* Implemented data quality framework to handle data quality issues and audit framework for data integrity.
* Responsible for end-to-end unit testing and document the results for each requirement story and review them with the test lead before moving to production.
* Working on Agile methodology to meet the deadlines for full ETL cycle on client requirements.

**Technology & Tools** PySpark, Jenkins, Shell Scripting, Hadoop, Hive, Oozie, Tidal, Sqoop, Git, Python, Scala and Agile

## Senior Software Engineer – Big Data Developer (KPIT Technology Solutions, Pune)

Ahold is one of the world’s largest food retail groups a leader in supermarkets and e-commerce, and a company at the forefront of sustainable retailing.

## Roles and Responsibilities

* Involved in requirement analysis, design and development of the project specifications.
* Involved in End-to-End Ingestion Framework using Python, Sqoop, Hive, Spark and Impala for ingesting data from Oracle to Hadoop.
* Developed spark streaming application in order to connect to Kafka topics to fetch and process real-time transaction data and storing the processed data into Casandra.
* Created hive tables on Casandra to perform analytics operations and generating the reports that are requested by the client using Spark as query execution engine.
* Processed all log files which are generated from various sources to HDFS using Apache NiFi and
* Developed the Sqoop scripts in order to make the data transfer between Hive/HDFS and MySQL Source.
* Developed the Apache PIG scripts to process log files which are generating by the Mainframe server and stored processed data into the Hive tables.
* Created External/Managed Hive tables with optimization techniques like partitioning, bucketing and Serialization to improving the query execution time to fetch/retrieve the data quickly.

**Technology & Tools** Horton Works HDP.2.5., HDFS, Pig, Hive, Sqoop, Apache NiFi, Cassandra, Kafka, PySpark, Shell Scripting and MySQL.

# LANGUAGES

English: Business Fluent | Hindi: Basic | Telugu: Mother Tongue

Date:(Narendra Reddy)