**SANTOSH M NAGARAJA**

**Junior Big Data Developer**

[**Santosh.nagaraja@wavicledata.com**](mailto:Santosh.nagaraja@wavicledata.com)

**SUMMARY**

* Experience working on Talend to perform Big data integration.
* Experience in developing multiple Talend jobs to automate tasks such as loading data from cloud servers such as S3 buckets into Distributed file system HDFS.
* Migrated jobs from legacy ETL tools to Talend for performance improvement.
* Experience in creating Talend jobs that are reusable which ensures faster execution and low memory usage.
* Experience in scripting HIVE queries to perform data partition and bucketing to help reduce the time invoking the MapReduce programs.
* Experience writing Sqoop commands to automate tasks to import data from multiple sources at timely order as required by the client.
* Experience deploying spark jobs both on-premise and Cloud
* Experience converting batch Hive/ Impala queries to Pyspark to run faster and efficiently on the cluster
* Experience Optimizing Spark jobs to run efficiently
* Experience working on Big data technologies for data analytics using tools such as Sqoop, Hive, Pig, Spark, and NoSQL databases.
* Experience working in Cloud Migration Projects
* Experience coordinating and handling offshore Teams
* Experience developing and deploying automation tasks using python, bash scripts and crontab.
* Experience in developing MapReduce programs using Python to perform simple ETL operation on data stored on HIVE tables.
* Experience working with Jenkins and Groovy to deploy pipelines
* Experience in working with Python, SFTP and FTP to automate data load and data ingestion
* Databricks Certified Apache Spark 2.4 Developer
* Cloudera Certified CCA175 Hadoop and Spark Developer.

**EDUCATION**

University of Bridgeport, Bridgeport, CT **Aug 2018 - May 2018**

Master’s in computer science (Big data and data Analytics)

GPA – 3.48/4.0

**SKILLS**

**Programming Languages:** Python3, C, C++, Java, SQL, HTML, and JavaScript

**Big data tools:** MapReduce, Hadoop 2.5x, Sqoop, Hive, Spark, and Kafka

**Software:** Visual studio, PyCharm, VMware, Atom, Eclipse, RStudio, and Talend.

**Scheduling Tools**: Automic, UC4, Cron-tab

**Programming Languages:** Python, Pyspark, Hive, Impala, Bash, Shell, Scala SQL.

**EXPERIENCE**

**Wavicle Data Solutions, Chicago, IL May 2018 - Present**

**Junior Big Data Developer:**

* Extracted transactional form S3 and processed the data for daily, weekly, monthly and yearly transactions using Talend with Spark and finally loaded the processed data into Redshift.
* Building ETL pipelines using Talend open studio, AWS Redshift, S3, HDFS (AWS EMR) based on the requirements specified by Clients.
* Developed multiple Spark and MapReduce jobs in Talend to perform ETL or ELT based on business logics.
* Creating, designing and deploying ETL jobs and pipelines using Talend Bigdata edition, according to the requirements specified by client.
* Working on scripting extensive Hive queries to perform transformations, partition and bucketing on huge data sets depending on the required fields to be used by downstream systems and reduce the execution time for invoking MapReduce.
* Developed multiple Talend jobs to automate tasks such as loading data from cloud servers such as S3 buckets into Distributed file system HDFS.
* Worked on a pilot project to develop a Model to detect objects in a video stream in real-time and send real-time alerts using open source Machine learning model "YOLOv2", Label Box, AWS SES, AWS SNS, OpenCV 3.4 and Python 3.6 to perform video analytics.
* Worked on creating Chat Bot for a help desk to help users in FAQ's and resetting account credentials, and managing requests using Amazon services like Lex, ASK, Connect, DynamoDB all communications with each other using Lambda functions coded in Python 3.6. and integrated the bot in various platforms (Skype, Amazon connect and Slack) to interact with the users.

**Client: Cars.com**

* Working alongside the Data Science team, by pre-processing and enriching the dataset for training and testing ML models.
* • Working on creating data pipelines using Spark, Hive, and Jenkins.
* • Converting Batch Hive/Impala queries to PySpark and optimizing them to run faster and efficiently on the cluster.
* • Converting on-prem jobs to cloud jobs.
* • Testing ML models using ML dashboards for their performance and efficiency.

**Cognizant Technology Solutions, Chennai, India Sep 2015 – Mar 2016**

**PROGRAM ANALYST**

* Professionally trained in SQL, Excel, and Python used to perform data analytics.
* Worked extensively on Excel using advanced functions such as Macros, Pivot Tables and VLOOKUP's on sales data to perform data analytics and visualize the results using visualization tools.
* Retrieved data from multiple databases through SQL queries and prepared the data to analyze trends and patterns in sales using data visualization tools.
* Presented detailed reports about the meaning of gathered data to members of management and help them by identifying scenarios utilizing modification in the data.

**Seize Code Technologies, Bangalore, India Jan 2014 – Sep 2015**

**Junior Data Analyst**

* Worked on writing PySpark scripts and Hive scripts to perform ETL jobs based on business logics.
* Worked on automating Sqoop using Sqoop scripts and Crontab to handle bulk importing of relational data from multiple sources into Hive or HDFS.
* Written extensive Hive queries to do transformations, partition and bucketing on the data depending on the business logic to be used by the downstream system and reduce the execution time for invoking MapReduce programs.
* Analyzed and identified huge data sets for patterns and implemented predictive analysis trends and improved e-commerce electronic sales by 15% compared to the previous quarter.
* Developed classification and regression predictive models to identify various patterns in the sales data to help improve future sales and target a set of customers.