Praveen Kumar Chalamcharla – Technology Lead

**Contact No:** **+91 9886551987**  **Current Location:** **Bangalore**

**Email id:** **Praveen.28.1987@gmail.com** **Role Designation: Technology Lead**

**Summary**

* Aspiring **ML Engineer.**
* Pursuing Masters in **Machine learning and Data Science.**
* **Practical knowledge on Linear and Logistic Regression techniques, Logistic Regression, Decision Tree, Unsupervised Clustering**
* **Strong understanding on Inferential Statistical concepts: Discrete Probability Distributions, Continuous Probability Distributions, Central Limit Theorem, Hypothesis Testing**
* **13+ years** of experience in Big Data Ecosystem.
* **8+ years of** experience in **Cloudera Hadoop Administration**.
* **5+ years of development experience in Hadoop includes** Map Reduce using Java, Map Reduce using python as Hadoop Streaming Jobs, Hive, Avro serialization, Sqoop, Flume Interceptors, Datameer Integration Tool.
* Worked on projects in agile methodology and support model.
* Experience in Banking and Financial domain, Hi-Tech and Media-Information-Technology domains.
* Excellent understanding of Big Data architecture and various components such as HDFS, YARN, Map Reduce, Kerberos, LDAP.
* Techno-functional responsibilities include interfacing with end users, identifying functional and technical gaps, estimates, designing custom solutions, development, leading developers and producing documentation.
* Experience in working with multiple vendors and geographically distributed teams.

**Key Domain and Technical Knowledge**

* ML Libraries : scikit learn, StatsModels, MatplotLib, Seaborn, NLTK,

pandas, numpy

* Big Data technologies : Hadoop, Map Reduce, Hive, Spark, Flume, sqoop
* Programming Languages : Python, Shell, yml
* Hadoop Distributions : Cloudera, Basic understanding of Hortonworks
* Development Tools : PyCharm, Jupyter Notebooks, Visual Studio Code
* DevOps Tools : Git, Mercurial
* Domains : Payments, Banking and Financial Services, Hi-Tech,

Media Information Technology

* Schedulers : crontab, JAMS
* Security : Kerberos, LDAP, FW
* OS : RHEL, Cent OS, Windows
* RDBMS : MYSQL, MariaDB

**Academic Qualification**

* Bachelor of technology in Computer Science & Engineering with 74% in 2009 from V.R. Siddhartha Engineering College affiliated to Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.
* Pursuing Masters in Machine Learning and Data Science

**Total Work Experience:**

Company: **Infosys Limited** Period: July 2015 to present

**Projects**

**Project: AST BI**

**Duration:** December 2021 to till date

**Description: Client is financial services provider in Europe. Project is to deliver a reliable platform to BI teams who generate metrics using the financial transactional data.**

**Technology stack: HDP, JAMS, Python scripting.**

**Roles and Responsibilities:**

* + Single point of contact for Hortonworks platform.
  + JAMS Jobs monitoring
  + On demand Data migration from PR to AC environments.
  + Deployment management

**Project Title: Data Lake Platform Management**

**Duration:** FromNovember 2019 to December 2021

**Description:** An Australian telecommunications company which builds and operates telecommunications networks and markets voice, mobile, internet access, pay television and other products and services.

Project is to deliver a reliable big data analytics platform for Business.

**Technology stack: CDH 5.13, Python, Azure IaaS, Azure PaaS, Informatica BDM, IICS**

**Role and Responsibilities:**

* + Single point of contact for Cloudera platform.
  + Capacity planning
  + End user support
  + Support migration from Cloudera platform to Azure PaaS.
  + Engage Cloudera in case of cluster queries.
  + Enable HDFS linked service to help migrate HDFS data to Azure storage.
  + Engage dedicated MS account expert for any migration related queries.

**Project Title: Data Lake Platform Support**

**Duration: From** February 2018 to November 2019

**Description**: American food company formed by the merger of Kraft Foods and Heinz based in Chicago, Illinois and Pittsburgh, Pennsylvania. Client is the third-largest food and beverage company in North America and the fifth-largest in the world with $26.2 billion in annual sales as of 2017. It is responsible for over 25 brands including Kraft, Heinz, Planters, Grey Poupon, Oscar Mayer and more, of which eight have total individual sales of over $1 billion.

In 2018, Kraft Heinz launched Springboard Brands, a business focused on growing organic, natural, and "super-premium" food brands. Kraft Heinz ranked No. 114 in the 2018 Fortune 500 list of the largest United States corporations by total revenue.

**Technology stack:** Cloudera Manager, Python Scripting, Azure IaaS, Azure PaaS

**Role and Responsibilities:**

* + L2, L3 Support.
  + Actively helping user in resolving admin related issues.
  + Engaging cloudera support team in case of issues.
  + Responsible to manage data coming from different sources
  + Delivering a platform development task every month.

**Project Title: Data Platform Solution Engineering**

**Client: An Australian Insurance Client**

**Duration: From** April 2017 to January 2018

**Description**: To ensure, our partner teams adopting the data platforms have the right solutions that utilize the best practices and design templates in a consistent fashion with the following goals:

1. Be a single point of contact for platform user base
2. Evangelize and drive adoption of Data and Platform
3. Evolve Data Platform roadmap to meet our user needs
4. Enabling better solution and faster delivery of user applications.

**Technology stack:** Shell scripting, Cloudera Manager, House Keeping scripts developed in python and shell scripting

**Role and Responsibilities:**

* + Responsible for performance tuning, backup and recovery with respect to Hadoop systems
  + Involving in transferring data between RDBMS and HDFS using Sqoop
  + Responsible for troubleshooting issues in the execution of Map Reduce jobs by inspecting and reviewing log files
  + Creating internal and external Hive tables and defined static and dynamic partitions as per requirement for optimized performance
  + Installing, configuring, expanding and maintained Hadoop cluster.
  + Sizing hardware requirements for large & small clusters including storage capacity, memory & processing power along with OS level tweaks

**Project Title: Platform As A Solution**

**Client: One of the leading providers of Credit Cards in USA**

**Duration: From** November 2015 to April 2017

**Description**: Client is an American multinational financial services corporation headquartered in Foster City, California, United States. It facilitates electronic funds transfers throughout the world, most commonly through Visa-branded credit cards and debit cards.

Built and owning four highly robust Cloudera clusters. Participated in multiple maintenance activities including CDH upgrade, CDSW upgrade, providing support during OS upgrade, and remediating security vulnerabilities. Implemented HA on all the clusters. Participated in multiple vendor calls on upcoming releases and planning the Cluster maintenances. Suggestions to Cloudera on the product and engaging Cloudera Management for getting up to date plans and upcoming releases of the product.

**Technology stack:** Shell scripting, Cloudera Manager, House Keeping scripts developed in python and shell scripting

**Role and Responsibilities:**

* + Actively helping user in resolving admin related issues.
  + Engaging cloudera support team in case of issues.
  + Managing cluster during qualys patching activity.
  + Gathering requirements with business users, analysis and design.
  + Responsible to manage data coming from different sources
  + Involved in creating Hive tables, loading with data and writing hive queries which will run internally mapreduce jobs
  + Automated all the jobs, for pulling data from SFTP server to load data into Hive tables, using CRON jobs.
  + Loading data into the test HDFS, analytical test outputs, performing the analysis on the output records.

Company: **Capgemini** Period: February 2015 to July 2015

**Projects**

**Project Title: Radian6**

**Client: One of the leading Information technology Company in USA**

**Duration: From** February 2015 to July 2015

**Description**: Radain6 survey data is coming for different media provider like Facebook, twitter, linkdin in form of xml files. These unstructured data are processed using Map reduce and stored in hdfs. Once the data is available in csv format in hdfs then it is loaded in Vertica database based on the data model.

Business team then generates report and analyze the sentiment on these data as per the requirement by using statistical analysis tool Tableau.

**Technology stack:** Sqoop, Hadoop Streaming, Hive, Avro

**Role and Responsibilities:**

* + As an Individual contributor, I am involved in analysis of raw and sentiment data and then create mapping document.
  + Write Map reduce program to process these data in Hadoop cluster.
  + Implemented framework to automate the end to end process

Company: **Tata Consultancy Services** Period: September 2013 to February 2015

**Projects**

**Project Title: Feed Onboarding**

**Client: One of the leading Investment Banking in Europe**

**Duration: From** September 2013 to February 2015

**Description**: This project is mainly for the re-platforming of the current existing system on various databases, servers and data warehousing systems. Focus on a new solution technology called Big Data Hadoop which can able to process large date sets (i.e. Terabytes and Petabytes of data) in order to meet the client requirements with the increasing competition from various banks.

**Technology stack:** Datameer Integration Tool, Pentaho, Hadoop Streaming using Python, Flume

**Role and Responsibilities:**

* + Gathering requirements with Business users, Analysis and design.
  + Create Import, Workbook and export jobs to load data from various source systems to Hadoop platform using Datameer tool.
  + Transform the structured data as per requirement and load to HIVE tables.
  + Importing and exporting data into HDFS and Hive using Sqoop.
  + Responsible to manage data coming from different sources
  + Working extensively on writing program to convert it into the structured format.
  + Load and transform large sets of structured, semi-structured data.
  + Load data from relational database to Hadoop platform using Datameer.
  + Involved in loading data from UNIX file system to HDFS using Datameer.
  + Involved in creating Hive tables, loading with data and writing hive queries which will run internally mapreduce jobs
  + Automated all the jobs, for pulling data from SFTP server to load data into Hive tables, using Datameer Cron jobs.

Company: **Virtusa India Pvt Ltd** Period: March 2010 to August 2013

**Projects**

**Project Title: Integrated metric System**

**Client: A Multinational mass media and Information Firm**

**Duration: From** March 2010 to August 2013

**Description**: Project was to develop a metric system capable of analyzing various datasets. Most of them did not conform to any specific schema. This project aims to move all data from individual servers to HDFS as the main storage and management system and then perform analysis on these HDFS datasets. Used scheduled jobs to move the data periodically into HDFS. Once the dataset is inside HDFS, Map Reduce, Pig and Hive used to perform various analysis.

**Technology stack:** Java Map Reduce, Big Data Hadoop Ecosystem, Hive, Mysql

**Role and Responsibilities:**

* + Everyday interaction with customers and onsite counterpart on the progress of deliverables, targets to attain before next meeting. I also involve in Coding, development environment testing, preparing the test results and test plan document, Test results are verified with the customer and getting approvals for the same, co-ordination with change management team and implementation process closure.
  + Monitoring the production daily jobs, providing the immediate solution to production errors and amends, making on-time coding changes and making emergency implementation with quick turnaround time. Incident process closure.
  + Adding values to the customer business by identifying repeated similar incidents and providing best solutions.
  + Loading data into the test HDFS region, coding, testing, preparing analytical test outputs, performing the analysis on the output records.
  + Change Request preparation, preparing unit Test cases, perform unit testing, unit Test output results preparation, BVT(Business validation Template), Impact analysis

**Personal Information**

* Gender : Male
* Marital Status : Married