Rahul Pitta

LinkedIn: **https://www.linkedin.com/in/rahulpitta/** GitHub: **https://github.com/rahulbittu**

1100 W Corral Ave, Apt 66, Kingsville, TX-78363 | 469-601-5140 | rahulpitta4@gmail.com

---

Professional Summary

A versatile programmer with strong programming skills offers a solid understanding/use of algorithms and analytics as well as proven technical leadership and project management skills. Leverages comprehensive technical expertise for software engineering, data management, and developing innovative business solutions for corporate customers. Currently preparing for Cloudera's **"CCA Spark and Hadoop Developer Exam (CCA175)"**

CORE COMPETENCIES

Big Data processing	Hadoop Eco System	Data cleaning
ETL Pipelines	Streaming Data	Apache Spark
Redshift, MySQL, MongoDB	Extract, Transform, Load (ETL)	Handling Different formats of Data
Data Warehousing	Amazon Web Services (AWS)	Data Mining
Software Development	Agile Methodologies	Data Analytics
Sqoop, Flume	Python, Scala, Java	Hive, HBase, Kafka, SparkSQL

Work Experience

GGK Technologies   Hyderabad, India

Big Data Engineer   June 2015 – August 2017

- Involved in creating data-lake by extracting customer's data from various data sources to HDFS which include data from Excel, databases, and log data from servers.
- Developed SQOOP jobs to import data in Avro file format from RDBMS to HDFS.
- Performed inter/intra-cluster copying using distributed copy (DistCp).
- Implemented SparkSQL to access Hive tables into Spark for faster processing of data.
- Worked with various HDFS file formats like Avro, Sequence File, Parquet and various compression formats like Snappy, bzip2.
- Involved in performance tuning of Spark Jobs using Cache and Persist.

Delkin Labs IT   Hyderabad, India

Hadoop Developer   October 2013 – June 2015

- Designed and Developed data migration from legacy systems to Hadoop environment.
- Imported the data from RDBMS to HDFS & Hive and performed incremental imports using Sqoop Job in various formats such as Avro, Text and Parquet formats.
- Involved in processing xml & JSON file formats, created partitioning of the data and implemented bucketing in Hive for performance optimization, loaded data with complex data types such as Maps, Arrays and Structs into Hive tables.
- Experience in building Pig Latin scripts to extract, transform and load data onto HDFS.

Skills Set

Programming Languages: C, C++, C#, Java, Python, Scala
Frameworks: Hadoop Eco System (HDFS, Sqoop, Pig, Hive, HBase, Strom, Spark, Kafka).
Database Skills: Oracle, SQL, MySQL, NoSQL
IDE/Platforms: Eclipse, Net beans, IntelliJ IDEA, Visual Studio, Windows, Linux, Unix, MAC-OS

Education

Master of Science, Computer Science: Texas A&M University – Kingsville, TX, United States   Expected Graduation: May 2019
Achievements: Instate scholarship for 2017-2018 by Javelina Scholarship   **CGPA: 4.0**