

## Contact

Kumarch2112@gmail.com

7894054160

## Top Skills

PySpark

SQL

Hive

# Chaitan Kumar D

Senior Software Engineer for

Infosys Technologies Private Limited (C2H)

## Summary

- **Around 4 plus years** of focused experience on Software Development using Python(Pandas), MySQL, Hive, Apache spark(pyspark)
- Good in Python (Pandas), Hive, Apache Spark.
- Hands-on/technical experience with high proficiency on MySQL.
- Good exposure in Linux commands for day to day use.
- Worked on subversion like Git for version control .
- Having good skills in code review and team management .
- Good experience with building ETL pipelines.
- Excellent knowledge on writing complex SQL's.
- Work closely with different stakeholders both from business.
- Maintain traceability of requirements and be accountable for the solution through all phases of project lifecycle (design, build testing through into live implementation)
- Be responsible for all aspects of the operation including: ETL processes, managing System testing and UAT aspects.
- Participate in the resolution of Production issues collaborating with developers, functional designers, other technology and service management teams.
- Experience with or knowledge of Agile Software Development methodologies.

- Programming Language : Python
- Big Data Stack : Spark, Hive
- Database(SQL) : MySql , Oracle
- Data Analysis : Python Pandas,Numpy.
- Version Control tools : GIT
- Cloud Infrastructure : AWS Services –S3,EMR
- Scheduler : Apache AirFlow
- Other Tools &IDE : SQL Developer,Pycharm

---

## Experience

Infosys Technologies Private Limited (C2H)

Senior Software Engineer

January 2020 – Present

Bengaluru, Karnataka, India

## Education

Biju Pattnaik University of Technology, Odisha

Bachelor of Technology – B.Tech, Mechanical Engineering

Year of Passing : 2010

Jawaharlal Nehru Technological University, Kakinada

Masters of Technology – M.Tech, Thermal

Year of Passing : 2017

## **Project#1:**

**Project:** Install Base Data platform

**Role:** Pyspark Developer

**Description:** Enable an effective and productive IB selling workspace for proactive management of tech refresh and renewal sales motion. Increase executive efficiencies in Service Renewal cycles and decisions making processes through personalized and just-in-time access to customer install base data. Provide capabilities to the business from this project to support organization's goal of being able to disposition 100% of all assets set for expiration of HW/SW contract end date in FY21. Increase velocity and customer success in the Tech Refresh and Service Renewal sales motions through reduction of time to ascertain personalized install base and proper disposition.

**Source System:** Oracle, MySQL, CSV

**Tools & Technologies:** Pyspark, AWS S3, AWS EMR, GIT, Airflow

**Duration:** Jan 2020 – Till Date

### **Responsibilities:**

- Involved in complete Big data flow of the application starting from data ingestion from upstream to AWS S3, processing and analyzing the data in AWS S3.
- Involved in creating dataflow tasks to bring the data into staging layer from source systems.
- Experience with warehousing best practices and methodologies.
- Build database/ETL developer responsible for building relational and data warehousing applications, unit test, document the ETL process.
- Followed Agile & Scrum principles in developing the project.
- Used GIT as the source version control for maintaining the code repository .
- Optimising the pyspark application for better usage of resources .
- Used the parquet file format for all the data intensive computations on the large datasets for efficiency.
- Worked and learned a great deal from Amazon Web Services (AWS) Cloud services like EC2, EMR.
- Responsible for data extraction and data integration from different data sources into AWS S3 by creating ETL pipelines Using Pyspark.
- Responsible for scheduling, monitoring and executing the pyspark application through Apache Airflow.
- Involved in converting SQL queries into Spark transformations using Pyspark.
- Analyzed the SQL scripts and designed the solution to implement using PySpark .
- Environment : PySpark, Python, MYSQL, ORACLE, AWS EMR, EC2, Amazon AWS S3, Airflow, JSON, SQL

## **Project#2:**

**Project** :Contact Master (MDM)

**Role** :Python Developer

**Description** : Using power of python and mongoDB to integrate multiple source systems to single source of truth.

**Source system types:** SAP HANA, Oracle, MySQL, CSV, TXT, etc.

**Tools and Technologies used:** Python, Pandas, Numpy, MongoDB Aggregation Framework, etc.

**Duration:** Nov 2017 – Dec 2019

### **Role & Responsibilities:**

- Created an ETL pipeline using python (pandas) for mastering the data from various source systems into Single source of truth .
- Independently design, build and launch new data extraction, transformation and loading processes .
- Support existing processes running in production and implement optimized solutions with limited guidance.
- Extensively used Mongo Aggregation framework to efficiently retrieve the data .
- Automated data movements using Python scripts.
- Involved in splitting, validating and processing of files.
- Created core Python API which will be using among multiple modules.
- Uploaded statistics to MySQL for analysis and logging.
- Used GIT as the source version control for maintaining the code repository .
- Developed complex SQL queries in Oracle for testing the database functionality.
- Provided technical assistance for maintenance, integration and testing of software solutions during development and release processes.
- Created unit test/regression test framework for working/new code.
- Environment: Python(Pandas),MongoDB, GIT,Shell Scripting, PostgreSQL, MySQL, FileZilla