

Praveen Halapeti

+91-7829227262 | praveen.halapeti@gmail.com

[LinkedIn](#) | [GitHub](#)

PROFILE

Senior Software Engineer with 10+ years of experience in software development and data solutions, including 5+ years of experience in building large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, Azure, Databricks, and Kafka** to develop scalable multiterabyte big data solutions for Healthcare and Banking companies.

TECHNICAL SKILLS

Programming Languages: Python, SQL
Big Data Technologies: Spark, PySpark, Spark SQL, YARN, Hadoop, Hive
Streaming Technology: Kafka, Spark structured streaming
Cloud computing: Azure(ADF, Databricks, ADLS, Synapse), AWS (S3, Glue, EMR)
Data Engineering Tools: Azure DevOps, Data Modelling, ETL/ELT data Pipeline
Orchestration: Airflow, Cron
Control Systems and Documentation: Jenkins (CI/CD), GitHub, Jira, Confluence

EXPERIENCE

Wells Fargo

Senior Software Engineer

July 2020 – Present

Bengaluru, Karnataka

- Designed and implemented comprehensive data pipelines using **PySpark and Spark SQL** for batch and real-time data processing, improving data accessibility.
- Applied Spark SQL to perform complex **data transformations and aggregations**, significantly reducing data processing time for business intelligence reporting.
- Integrated and processed banking data from multiple sources including **AWS, Azure and Oracle**, handling various file formats such as JSON, CSV and Parquet.
- Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 20% decrease in processing time for handling a daily load of around **1 million records**.

Cerner Healthcare Solutions India Pvt Ltd

Associate Senior Software Engineer

June 2018 – June 2020

Bengaluru, Karnataka

- Delivered a project to migrate legacy on-premise processes to the cloud using **Big Data technologies (Spark)**, reducing processing time by 20%.
- Developed scalable data pipelines with **Python, SQL, PySpark and Azure** for healthcare analytics.
- Designed and executed advanced scheduling capabilities using **Apache Airflow** for data pipeline orchestration, reducing manual intervention time by 80% and streamlining workflow efficiency.

REConnect Energy Solutions

Software Engineer

Sept 2014 – June 2018

Bengaluru, Karnataka

- Built **ETL pipelines** to process energy consumption and production data from multiple sources.
- Extract:** Data are extracted from **APIs** (e.g. weather data, smart meter logs) or CSV files.
- Transform:** Streamlined data cleansing processes by implementing automated scripts to convert time zones and resolve missing values, resulting in a 50% reduction in manual pre-processing hours across the ETL pipeline.
- Load:** Store the cleaned data in an SQL database for querying and analytics.
- Automated the pipeline using cron jobs.

CERTIFICATIONS

Microsoft Certified: AZ-900 Fundamentals

IBM Certified: Data Engineering Professional

EDUCATION

REVA ITM

Master of Technology in Computer Networks

Bengaluru, Karnataka

July 2012 - June 2014

APPA IET

Bachelor of Engineering in Computer Science

Gulbarga, Karnataka

June 2007 - July 2011