

# Susanth Guru

Data Engineer  
Kumbakonam, Tamilnadu, India

☎ +91 9500980912 ✉ [susanth.guru.29@gmail.com](mailto:susanth.guru.29@gmail.com) 🔗 [linkedin.com/in/susanthguru](https://www.linkedin.com/in/susanthguru) 🐙 [github.com/susanth-g](https://github.com/susanth-g)

## About

Having **6 years** of experience in documenting, designing and implementing solutions for problems of multiple usecases by analysing technologies that fit. I also have hands on experience with multiple technologies like **Hadoop, HIVE, Hbase, Spark, Cloudera, GCP and Kafka** to complete data ingestion pipe lines and provision of ingested data to the product user interface.

## Work Experience

### Paypal

Data Engineer

**Feb 2022 - Present**

Chennai, Tamil Nadu

- Designed and implemented solutions based on **Spark Streaming with Pubsub and Bigquery** for creating data ingestion pipelines.
- Developed **Bash** scripts, **SQL** scripts, **pySpark** and **Python** scripts to move or recover data from multiple sources such as log files, JSON files and HIVE parquet files.

### CanGo Networks Pvt Ltd

Lead Big Data Developer

**Nov 2019 - Feb 2022**

Chennai, Tamil Nadu

- Designed and implemented solutions based on **core JAVA** for CoreServices and SpringBoot for APIs.
- Developed spark based and **multithreaded** performant oriented applications based on **JAVA** to collect and **parse data** from various data sources such as FTP, **Command output** from SSH Session, messaging queues like Kafka and Pubsub.

### National Payments Corporation of India

Associate Enterprise Architect

**Jul 2017 - Nov 2019**

Chennai, Tamil Nadu

- Designed and implemented solutions based on **HIVE and HBASE**, **Apache NIFI** and **Apache Phoenix**.
- Developed utility jars** to help operations team for scheduled daily reports which reduced the effort and time to generate reports manually everyday.

## Technical Skills

**Languages and Scripts:** Java 8, Spring Boot, SQL, Arango QL, Shell Script, Python

**DataBase Technologies (SQL):** MySQL, Postgres, HIVE, Impala

**DataBase Technologies (NoSQL):** HBase, ArangoDB, Elastic Search

**Cloud Technologies:** GCP [ BigQuery, PubSub, CloudStorage, CloudSQL, DataProc, Kubernetes ], AWS [ S3 Bucket ]

**Big Data Technologies:** Hadoop, Apache Spark, Apache NIFI, Spark SQL, Cloudera

**Developer Tools:** VS Code, Eclipse, SublimeText 4, NP++, StarUML, draw.io

**Terminal Tools:** bash, tmux, less, vim, htop, top

**Other Technologies/Framework:** Linux, Git, SVN, JUnit, Microservices, Rserve, REST API, Kubernetes, Docker, RASA ChatBot

## Projects

**Pipe Line for BQ Data Population and Report Generation | Python, JAVA, Bigquery, Spark, Spark-UDF, PGP Enc-Dec, SFTP, Airflow**

**Role:** Data Engineer

- Developed **python** based framework **modules** as pipeline to **populate BQ tables** using queries from a file
- Developed **JAVA** based **Spark App** to Deserialize, Encrypt or Decrypt columns of data
- Developed **Spark UDFs** to Deserialize and convert the **deserialized data to columns** of a DF and then store to BQ table
- Developed **BQ Export** Script to download **large data** (~2B records - 15 cols - atleast 8char length) to local disk **as single CSV** for reports
- Developed **Email** Scripts to convert **BQ tables to multiple HTML tables** grouped on a column
- Developed scripts to **PGP encrypt or decrypt** complete **report file** before sending to destination SFTP
- Developed modules to **push and pull reports** or input files from **SFTP** directories

**Cloud Migration | Spark Streaming, Pubsub, Bigquery, Dataproc, SQL, JAVA, Python, HIVE**

**Role:** Data Engineer

- Developed a **JAVA based spark streaming** JAR with editable config file to fetch data from Pubsub then transform, validate and push into BQ tables.
- Recovered lost data in HIVE** by regenerating data from raw log files by developing a **python** script, then read and pushed the data to HIVE using **Spark** and HQL.

**NIMS - Network Inventory Management System | Core JAVA, JAVA Future, Jsch Library, ArangoDB**

**Role:** Lead and Developer

- Led a small team** of UI and API developers.
- Designed and Implemented the complete **Data collection service** using **core JAVA** to fetch and parse data from more than **24000 devices using SSH protocol** for different vendors on daily basis.

**MAXIS - OptiGo AI Product |** *JAVA Spark, Hive, Impala, MySQL, SpringBoot*

**Role:** Developer


- Implemented a **Data parser** to populate performance data from csv or xml files using **Spark SQL** into Hive database by applying pivots and transformations over dataframes as required. Optimised Spark SQL performance with persistence.
- Designed and implemented a **KQI calculator** using **spark SQL** to re-aggregate performance data based on configurations to calculate KQI and push to Hive tables in a scheduled time interval.
- Developed **CRUD API** using **SpringBoot** to configure KQI from UI.
- Worked closely with Datasience team to **invoke R code from JAVA Spring Boot with RServe**.

**Transaction Enquiry System - NPCI UPI, NACH |** *Java HTTP Servlet, Apache NIFI, Hbase, Sqoop*

**Role:** Developer

- Designed and implemented a data ingestion pipeline using **sqoop** to fetch daily data from postgres and populate into Hive. Then using **Apache NIFI** to fetch transaction data and form rowkey based on the data, then push to Hbase.
- Implemented APIs using JAVA HTTP Servlet to fetch data from Hbase using **Apache Phoenix** JDBC connection.
- Developed a scheduled **report generation JAR** to generate reports which reduced the time to generate manual reports but enabled operations to verify or analyse the data.

**Publication**

<b>Emulation of IoT Gateway</b> 	<b>Jan 2017</b>
Emulation of IoT gateway for connecting sensor nodes in heterogenous networks within lab environments.	ICCCSP – Research Gate

**Education**

<b>SSN, Anna university Affil.</b> 	<b>Jun 2015 - Jun 2017</b>
Master of Engineering (M.E.,) in Computer Science	Chennai, Tamil Nadu
<b>JJCET, Anna university Affil.</b> 	<b>Jun 2011 - May 2015</b>
Bachelor of Engineering (B.E.,) in Computer Science	Trichy, Tamil Nadu

**Languages**

Tamil - Native, English - Fluent