

# PARTHASARATHI SWAIN

## SENIOR DATA ENGINEER

Data Engineer with 9.5 years of experience in Software development and Hadoop Ecosystem, Building Data Pipelines, Use of Data warehousing tools on cloud in BFSI domain.

✉ partha.hadoop.pyspark@gmail.com

📍 Bangalore, India

☎ +91 - 7795909536

🌐 [linkedin.com/in/parthass](https://www.linkedin.com/in/parthass)

## INDUSTRY EXPERIENCE SUMMARY

### Sr. Data Engineer

#### Genpact

07/2021 - Present

The company wanted to improve the insurance management system

##### Responsibilities :

- Design and building Data warehouse on **Snowflake** in **AWS** cloud.
- Build pipelines using **Streamsets** and processed **Big Data** (20 TB daily) related to Financial Health, Credit/Debit Transactions, customer data.
- Create **SnowPipe** with Auto-Ingest Enabled & S3 event notification to automate Snowpipe.
- Applying business logic using **Pyspark** & schedule the jobs using **Control-M**.

### Data Engineer

#### Dell

05/2019 - 05/2021

The company wanted to automate the inventory management system

##### Responsibilities :

- Build scalable ETL pipelines using **Streamsets** & schedule jobs by using **Control-M**.
- Translate business proposal into quantitative queries & collect/clean necessary data by handling **DBNULL**, **datatype conversion**, **column level mapping**.
- Automate **Snowpipe** data loads using **AWS SQS** notifications for an **S3** bucket.

### Data Engineer

#### Intuit

11/2018 - 05/2019

##### Responsibilities :

- Design, develop & support data processing pipelines using **Streamsets** & **Pyspark** to process the customer eligibility data per market.
- **Webscrapping** using **Beautifulsoup**.
- Evaluate the workflow & increase the efficiency of data pipelines that process over 10TB of data daily.

### Data Engineer

#### ITC Infotech

04/2017 - 07/2018

##### Responsibilities :

- Process data by using **SparkSQL** and stored it into **HDFS**.
- Support all round the clock for 10+ applications.
- Create and manage workflows using **control-M** to schedule & monitor customer targeting jobs built on **PySpark** and **Hive**.

### Senior Software Engineer

#### HCL Technologies

09/2013 - 12/2016

##### Responsibilities :

- Expertise in **production support** processes such as outage/incident.
- Write code in **HTML** for displaying message in front page during any maintenance period or any release period using LogViewer tool.
- Web development using **Java(1.6)**, create & expose web API using **Flask**.

## TECHNICAL SUMMARY

Python

Java

Pyspark

Spark

Hadoop

HDFS

Hive

Streamsets

Snowflake

Control-M

SQL

Numpy

Pandas

Data Wrangling

GIT

AWS

Beautiful Soup

Flask

Web Scraping

Heroku

MySQL

Linux Terminal

ETL Pipelines

Docker

Postman

Machine Learning

## EDUCATION SUMMARY

B.Tech from Biju Pattnaik University of Technology, Odisha, 2011 | CGPA: 8.01

## CERTIFICATES

Oracle Certified Java(1.6) programmer(OCJP) with 88%  
*Full Professional Proficiency*

Completed SQL Community Class program from iNeuron  
*Full Professional Proficiency*