

# Soumitra Banerjee

DATA ENGINEER II

Bangalore, India

☎ (+91) 7003190619 | ✉ soumitra.banerjee@hotmail.com | 🏠 soumitrabanerjee.github.io/ | 🌐 soumitra-banerjee/

## Summary

Results-oriented Data Engineer II with 4 years at American Express, adept at constructing and deploying robust big data solutions using Databricks, Spark, Hadoop, Hive, Scala, and Java. Proven ability to strategize, execute, and deliver under tight deadlines. Passionate about innovating problem-solving approaches and rapidly acquiring new technical proficiencies.

## Experience

### American Express

Bangalore, India

DATA ENGINEER II

Sep. 2022 - Present

- **Reconstructed and debugged** PreAuth, Auth, and Settlement functionalities for Corporate Cardholders, reducing error rates by **10%** and improving transaction processing time by **30%**.
- Independently **architected** and **implemented** a comprehensive end-to-end pipeline for Accounts Payable batch processing, automating **8** manual processes and increasing processing speed by **20%**.
- Worked with vPayment team to **redesign reconciliation features** by applying new methodologies through scala-spark.
- **Spearheaded** cluster resource optimization by analyzing Spark job performance, designing a prioritized child queue system that reduced job execution time and improved resource utilization by **25%**.
- **Independently architected**, developed, tested, and integrated HashiCorp Vault, **securing 300 sensitive data points** and enhancing system security posture.

### Cognizant Technology Solutions India Pvt Ltd

Pune, India

PROGRAMMER ANALYST (DATA ENGINEER)

Sep. 2020 - Sep. 2022

- Worked on **JSON data extraction** through REST API and transformed it into tabular format using Scala, applying rigorous filter conditions.
- **Developed** a versatile Scala framework to export data from complex views into multiple file formats, enhancing project flexibility.
- **Engineered a robust job management system** in Scala, implementing job tracking, archiving, history maintenance, and status updates within the database, improving operational efficiency by **20%**.
- **Constructed** a Java application to streamline cross-environment data comparison (Dev, QA, Prod), accelerating issue resolution by more than **99%**.
- Leading the revamp of business logic for PRTS data using Scala and API Gateway, aiming to optimize user operations and **improve performance by 30%**.
- Successfully **resolved critical bugs** in Spark SQL scripts, enhancing data processing performance.
- Executed in-depth **reverse engineering** to map complex data flows across multiple systems, identifying 50 data dependencies and improving data quality.

### Cognizant Technology Solutions India Pvt Ltd

Kolkata, India

INTERN

Jan. 2020 - Apr. 2020

- Trained on Data Warehouse Under Cognizant Trainers' supervision and learnt SQL, Databases, Core-JAVA, OLAP Concept.
- Built a small OOPs based java application to store books details.

## Skills

<b>Big Data Tools</b>	Experienced: Databricks, Hadoop, Spark, Hive, Impala
<b>Programming</b>	Experienced : Java, Scala, SQL, Shell Script   Familiar : Python
<b>General</b>	Experienced : UNIX, Prompt Engineering, Data Structures & Algo, Airflow
<b>Languages</b>	Fluent : English, Hindi, Bengali

## Education

### University of Engineering and Management, Kolkata

Kolkata, India

B.TECH IN COMPUTER SCIENCE AND ENGINEERING

Jul. 2016 - Jun. 2020

- Graduated with a DGPA of 8.95

### Sri Chaitanya Techno School(CBSE)

Visakhapatnam, India

HIGHER SECONDARY (10 + 2)

2016

- Passed with overall aggregate 76.20%

### CMCL Vidya Bharati School(CBSE)

Meghalaya, India

SECONDARY

2014

- Passed with overall aggregate 85.50%