

# TATHAGATA MITRA

## SENIOR SOFTWARE ENGINEER 2- DATA

Data Engineer with a strong track record of designing and building scalable, high-performance data pipelines and platforms. Skilled in translating messy, complex data into clean, actionable insights. Focused on creating robust systems that are easy to maintain, monitor, and extend – with a constant eye on performance, reliability, and business impact.



✉ mitra.tathagata15@gmail.com

📱 9284264514

📍 Bengaluru, India

.linkedin.com/in/tathagata-mitra-187721186

## SKILLS



## WORK EXPERIENCE

### Senior Software Engineer 2 (SDE-3) - Data CoinDCX

11/2022 - Present

Bengaluru, Karnataka

#### Roles & Responsibilities

- Built robust ETL pipelines using **PySpark**, **SparkSQL**, and **Databricks** (on AWS S3) to ingest and process data from multiple sources, enabling clean, analysis-ready datasets for business teams.
- Designed and implemented **Data Marts** tailored to business needs, improving dashboard performance and reducing I/O overhead. Consolidated 40% of models by streamlining overlapping business logic.
- Developed an end-to-end pipeline to track **real-time holdings** across blockchains like **Ethereum**, **Bitcoin**, **Tron**, **Solana**, **Polkadot**, and more – using **REST APIs**, **RPCs**, and **Web3.py** in **Databricks**.
- Engineered a cost-saving solution (~ \$1,000/month) for automated tracking and liquidation of DeFi pool investments on Ethereum. Used **Python**, **Web3.py**, **Azure Function App**, **Databricks**, **Fireblocks SDK** and **Infura**.
- Built personalized user engagement pipelines reaching ~13M users via **MoEngage API**, delivering real-time SMS and push notifications triggered by market movements using **PySpark**.
- Built real-time market monitoring pipelines with **Spark Structured Streaming**, **Kafka**, **Delta Lake**, and **PySpark**, generating personalized insights and rule-based event triggers.
- Developed an **AI-based trading indicator** using **OpenAI**, **Python**, **Azure**, **TA-Lib** and **Streamlit**, providing market guidance using technical indicators like **MACD**, **RSI** and **Bollinger Bands**.
- Delivered a POC for an **AI chatbot** to answer user queries related to platform transactions with up to **85% accuracy**, leveraging **OpenAI APIs**, **Azure** and **Python**.
- Built a custom **archival platform** to offload transactional databases, supporting both time-based and condition-based (e.g., status) archival. Reduced DB load, lowered infra costs by **40%**, and cut maintenance downtime by **50%**.
- Mentored junior Data Engineers and led BAU projects across Finance, Operations, and Analytics, ensuring timely delivery and high code quality. Contributed to the team's growth by interviewing **50+ candidates**, actively participating in the hiring and evaluation process.

## WORK EXPERIENCE

### Senior Analyst (Data Engineering)

Exusia

04/2022 - 10/2022

Pune, Maharashtra

#### Roles & Responsibilities

- Develop pipelines using **Spark with Scala**, coded in **Databricks** to flow data from on-prem databases like Oracle and flat file sources to **Azure** cloud platform. Using the **Azure Data Factory** to orchestrate and schedule the pipelines.
- **Azure Blob Storage** containers are used as intermediate layer to perform cleansing and transformation. The **Delta Lake** in Databricks is used for better performance and storage optimization.
- Built common utilities like utility to compare two flat files and generate reports.

### Programmer Analyst (PA) - Data Engineering

Cognizant Technology Solutions

11/2020 - 04/2022

Kolkata, West Bengal

#### Roles & Responsibilities

- Implemented pipelines to ingest data from multiple sources to target using **PySpark** code written in **Databricks** notebooks.
- Read the raw data from various sources (**SQL Server**, **Teradata**, flat files with extensions like **csv**) and ingested into **Snowflake** after cleansing and transforming the data while keeping it in **Delta** tables in **S3** bucket (Inter-mediate layer). Performed the unit testing of the ingested data and validate the end-to-end flow of the process.
- Scheduling and monitoring of jobs to execute the Databricks notebooks through **Tidal Automation**. Code migration and integration using **Gitlab** and **Jenkins**.

## CERTIFICATES

Databricks Certified Data Engineer Professional (04/2025 - 04/2027)

## EDUCATION

### Master of Technology

Indian Institute of Technology, Patna

01/2023 - 01/2025

Patna, Bihar

MTech in Big Data & Blockchain

- CGPA - 8.97/10

### Bachelor of Engineering

Smt. Kashibai Navale College of Engineering, Pune (SPPU)

08/2016 - 06/2020

Pune, Maharashtra

Electronics & Telecommunication Engineering

- CGPA - 9.35/10

## ACHIEVEMENTS

### Whale of the Quarter (Performer of the Quarter) Award (01/2023 - 03/2023)

*Award presented for the outstanding performance and creating an impact as a culture champion at CoinDCX.*

### University Rank 10 (2017)

*Stood TENTH in order of merit among all the successful candidates at the First Year of Engineering (2015 Course)*

### Always Striving, Never Settling Award - Cognizant (11/2021)

*Received the award in recognition of immense hard-work and on-time delivery of required milestone.*

### Department Rank 1 (2017)

*Department of Electronics and Telecommunication, SKNCOE, Pune (First Year of Engineering)*

## LANGUAGES

### English

*Full Professional Proficiency*

### Bengali

*Native or Bilingual Proficiency*

### Hindi

*Full Professional Proficiency*

### Marathi

*Limited Working Proficiency*