

Sudhesh Prabhu

Sudheshat94@gmail.com | Bengaluru, IN | [linkedin.com/in/sudhesh-prabhu](https://www.linkedin.com/in/sudhesh-prabhu) | [sudheshprabhu.com](https://www.sudheshprabhu.com)

DATA PROFESSIONAL

PROFESSIONAL SUMMARY

Data professional with **10 years of experience** in banking and finance across data ownership, data science and data engineering. Currently **Data Owner for Digital Communications at JPMorgan Chase**, a product that sends more than **5 billion alerts and notifications** to customers across the bank, partnering with product, technology and data technology teams to publish data from system-of-record application databases to the **AWS S3 data lake** and **Snowflake**. Background in building end-to-end pipelines with **Apache Spark (PySpark)**, **Hive, Airflow and Snowflake** on AWS, and in customer segmentation using **K-means clustering and supervised learning**.

KEY SKILLS

Data Ownership • Data Pipeline Design & Optimization • ETL Development • Data Lake & Data Warehouse Architecture • Data Migration • Customer Segmentation • Machine Learning • Regulatory & Compliance Reporting • Agile Methodologies • Cross-team Collaboration • Mentoring • Problem-solving

TECHNICAL SKILLS

- **Big Data:** Apache Spark (PySpark), Hadoop, Hive, Airflow
- **Data Science:** K-means clustering, supervised learning, Python
- **Languages:** Python, SQL, UNIX shell
- **Cloud Computing:** AWS (S3, EMR, Glue, Athena, Redshift)
- **Databases:** Snowflake, CockroachDB, Teradata, Redshift
- **Tools:** GitHub, Jenkins, CI/CD, VS Code, Jupyter Notebook, Ab Initio, Tableau

PROFESSIONAL EXPERIENCE

Data Owner Jul '26 - Present
JPMorgan Chase Bengaluru, IN

- Data owner for **Digital Communications**, which sends more than **5 billion alerts and notifications** to customers across the bank
- Work closely with **Product Tech and application teams** to publish data from **system-of-record (SOR) application databases** to the **AWS S3 data lake**, and help develop the pipelines that do it
- Partner with the **Data Technology team** to publish data from the data lake to **Snowflake**
- Key contributor to migrating the alert preferences application from a legacy system to **Alerts 2.0**, including the move to **CockroachDB**, a new pipeline to the data lake and Snowflake, and every stage of testing

Data Scientist Jan '25 - Jun '26
JPMorgan Chase Bengaluru, IN

- Built **customer sub-segments** using **K-means clustering** followed by **supervised learning** techniques, used primarily for **marketing and customer targeting**
- Built the customer datasets used as input for clustering, using **PySpark on AWS EMR** with data from **Snowflake**, enriched with bureau and credit history via **Amazon Redshift**
- Worked on every step of the segmentation process, from data preparation and clustering through to delivering the final segments

Senior Data Engineer Dec '22 - Dec '24
JPMorgan Chase Bengaluru, IN

- Engineered end-to-end **PySpark pipelines on AWS EMR**, integrating data from 10+ sources into clean, unified datasets for **customer clustering and segmentation**
- Delivered production datasets powering **real-time Tableau dashboards** on **digital customer behavior**, improving stakeholder decision-making speed by **30%**
- Led the migration of customer data from **on-premise Hadoop (Hive) and Teradata** to **Snowflake**, improving the scalability and agility of analytics workloads
- Worked across **PySpark, Hive, SQL, Snowflake, Teradata** and **Airflow**, with **AWS S3, Athena, EMR, Redshift** and **Glue**

Data Engineer Jul '22 - Nov '22
Tiger Analytics Bengaluru, IN

- Created and maintained scalable data pipelines for the **Dynamic Promo Generation** application at **Lands' End**, enabling **automated, personalized marketing campaigns**

- Worked with the client's Product Manager to refine business requirements and implement pipeline enhancements
- Built pipelines with **Spark (Python)** on **AWS** using **S3, EMR, Athena and Glue**

Data Engineer

Jul '16 - Jul '22

Wells Fargo

Bengaluru, IN

- Progressed from **Apps Programmer/Analyst** to **Data Engineer**, contributing to the **Capital Mortgage Operational Data Store (CMODS)**
- Worked in an Agile team with business stakeholders to **gather requirements** and **design, build and enhance data pipelines**
- Developed ETL with **Ab Initio** and provided **Level 3 production support**, including on-call ownership of critical batch operations
- Built pipelines with the **Hadoop ecosystem (HDFS, Hive, Sqoop)**, **Spark (Python)**, **SQL**, **Teradata** and **Unix**, scheduled with **Autosys** and visualized in **Power BI/Tableau**
- Mentored junior developers on technical implementation and domain knowledge, accelerating onboarding and delivery

EDUCATION

Post Graduate Certification in Data Science and Engineering

Jan '24 - Feb '25

International Institute of Information Technology (IIIT) Bangalore

Bengaluru, IN

- Secured **CGPA 3.7/4**

Bachelor of Engineering in Computer Science and Engineering

Aug '12 - Jul '16

R.V. College of Engineering, Bangalore

Bengaluru, IN

- Secured **CGPA 8.8/10**

KEY PROJECTS

- **Alerts 2.0 Migration | Tech Stack: CockroachDB, AWS S3, Snowflake, Python | 2026**
 - Migrated the alert preferences application from a legacy system to **Alerts 2.0**, collaborating closely with **product, technology and data technology** teams for a smooth transition
 - Moved the application from its legacy database to **CockroachDB** and built a new pipeline to the **AWS S3 data lake** and on to **Snowflake**
 - Involved in **every stage of testing** across the application, data lake and Snowflake
- **Customer Sub-Segmentation | Tech Stack: Spark, AWS, Snowflake | Jan '25**
 - Built **PySpark pipelines on AWS EMR** to prepare customer data from **Snowflake** as input for clustering, encrypting sensitive attributes and transferring data securely to **AWS S3**
 - Enriched datasets with **bureau and credit history** via **Amazon Redshift**, improving feature richness
 - Created sub-segments using **K-means clustering** followed by **supervised learning**, delivering segmentation-ready data back to Snowflake for **personalized marketing**
- **Digital Channel Submission Rate | Tech Stack: Spark, AWS, Hadoop | Jan '24**
 - **Migrated legacy Hive workflows** to an **automated PySpark pipeline on Hadoop**, cutting processing time from **12 hours to under 2**
 - Implemented **Apache Airflow** for scheduling and orchestration, and integrated optimized datasets into **Snowflake**
 - Delivered curated data to **Tableau dashboards** for real-time visibility into digital metrics like payments and account openings
- **Dynamic Promo Application | Tech Stack: Spark, AWS | Jul '22**
 - Developed **PySpark pipelines on AWS EMR** to calculate sell-through scores from inventory, demand and sales data in **S3**
 - Applied business rules to set SKU-level promotional discounts, automated with **AWS Glue triggers**, stored in **Amazon RDS** and cached in **Redis**
- **Capital Mortgage Operational Data Store | Tech Stack: Ab Initio, Teradata, Hive | Jan '21**
 - Integrated mortgage data from Parquet files, mainframe systems and Teradata into a consolidated **Teradata data mart**
 - Reported **Qualified Financial Loans to the FDIC** and optimized batch workflows to meet the **4:45 AM EST processing deadline**
 - Built **Fair Value accounting pipelines** in **Ab Initio**, contributing to a rules-based engine capturing **P&L components** of non-agency RMBS loan sales

CERTIFICATIONS

AWS Certified Data Analytics – Specialty • AWS Certified Developer – Associate • AWS Certified Cloud Practitioner