



Amar Prakash Pandey

Senior Data Engineer

Pune, India | 8765598189 | amar.om1994@gmail.com | <https://amarpandey.me>

Summary

Experienced Data Engineer with **7+ years** of expertise in **designing and optimising scalable data pipelines, data warehousing solutions, and batch processing workflows**. Proficient in handling **petabyte-scale datasets** using modern technologies like **Apache Spark, Airflow, and dbt**. Skilled in building **cloud-native solutions** on **AWS and GCP**, with a strong focus on **performance optimisation, cost efficiency, and data governance** across various industries.

Actively contributed to projects like **Apache Airflow, CRI-O, and DuckDuckGo**, while also maintaining personal open-source projects.

Experience

Sahaj Software
Senior Data Engineer

March 2021 - June 2025
Pune, India

Developed and maintained end-to-end batch pipelines to process audience and frame-level data for out-of-home (OOH) advertising, enabling accurate and timely delivery of insights to campaign optimization and analytics teams. Ensured data quality and performance for effective ad targeting and impact measurement.

- Designed and optimized ETL pipelines** with **Apache Spark and PySpark**, processing terabytes of data and executing large-scale joins to enrich advertisement analytics.
- Enhanced Spark job performance by 50%** through advanced shuffle partition tuning and resource optimization, significantly reducing processing time.
- Implemented and managed** workflow orchestration with **Apache Airflow**, ensuring efficient sequencing, automated retries, and SLA compliance for batch processing.
- Cut S3 storage costs by 60%** by implementing efficient data lifecycle policies and optimizing file formats for cost-effective cloud storage.
- Automated deployment of **Airflow DAGs, Apache Spark/PySpark jobs, and microservices** using **GitLab CI/CD**, eliminating manual errors and ensuring seamless promotion across dev and production environments.

Tata Consultany Services
System Engineer

December 2017 - Feburary 2021
Chennai, India

Developed a **pricing platform** that streamlined and automated pricing updates for thousands of components. My contributions focused on **optimizing workflows, improving system reliability, and enhancing deployment efficiency**.

- Developed APIs** for Cummins' pricing platform using **Java, Spring Boot, and MySQL**, enabling real-time pricing updates for thousands of components.
- Automated pricing workflows**, reducing update time from **one day to near real-time**.
- Migrated Oracle trigger and function logic** to a service layer for **better logging and maintainability**.
- Automated deployment using CI/CD pipelines with Git, **Jenkins, and GitLab**, improving efficiency and reliability.

Skills

Programming Languages <div><div></div><div></div><div></div><div></div><div></div></div> Java, Python, Scala, SQL	Big Data Technologies <div><div></div><div></div><div></div><div></div><div></div></div> Apache Spark, Pyspark, Apache Hadoop	ETL Tools & Orchestration <div><div></div><div></div><div></div><div></div><div></div></div> Apache Airflow, dbt (Data Build Tool)
Cloud Technologies <div><div></div><div></div><div></div><div></div><div></div></div> AWS S3 EMR, ECS, ECR, Lambda, RDS, Google Cloud Platform BigQuery	Databases <div><div></div><div></div><div></div><div></div><div></div></div> Oracle, PostgreSQL	DevOps <div><div></div><div></div><div></div><div></div><div></div></div> Docker, Terraform, Jenkins, Gitlab CI/CD

Projects

spot-optimizer

January 2025

<https://github.com/amarlearning/spot-optimizer>

Developed an open-source tool to **optimize AWS Spot Instance usage, reducing cloud costs** through **automated instance management** and dynamic scaling.

geohashviz.com

July 2024

<https://geohashviz.com/>

GeohashViz is a web-based visualization tool that helps users explore **geohashes** interactively. It allows users to encode locations into geohashes, visualize different precision levels, and understand spatial hierarchies. Built using **React** and **Leaflet.js**, the tool is designed for geospatial analysis and learning.

Finger Detection and Tracking

April 2017

<https://github.com/amarlearning/Finger-Detection-and-Tracking>

A computer vision project that detects and tracks fingers in real time using **OpenCV**. It enables gesture-based interactions by identifying finger movements from a live camera feed.

Awards

Senior Data Engineer

March 2024

Recognized for exceptional contributions at **Sahaj.ai**, receiving multiple commendations for assisting projects, building useful tools, and consistently going above and beyond.

System Engineer
Cummins

June 2019

Awarded Star Performer of the Year at **TCS** for outstanding performance on the **Cummins** project.

Software Developer
Google Summer of Code

August 2017

Completed Google Summer of Code (2017), contributing to the **Teammates** project at the **National University of Singapore**.