

# Jaygovind Sahu

Senior Data Engineer

mr.jaygovind.sahu@gmail.com +1-609-789-7641 LinkedIn

Senior Data Engineer with 16 years of experience building data platforms that deliver measurable business outcomes across financial services, retail, and streaming media. At Amazon, led development of a marketing analytics platform that reduced campaign costs by 30% in its first year and an ML-driven pipeline projected to cut costs by a further 25%. At Jornaya, re-architected a high-scale pipeline serving 50 vendors — cutting AWS spend by 55% while processing a billion records in under 30 minutes. Consistently bridges data infrastructure and business strategy, enabling leadership teams to make faster, better-informed decisions.

## EXPERIENCE

### Data Engineer

Jul 2025 – Present

Netflix

- Building something awesome.

### Data Engineer

Mar 2022 – Jul 2025

Amazon

- Developed a robust data platform to measure and enhance the efficiency of marketing campaigns across multiple channels, including offline (billboards, newspapers, gas station TVs), social media (Facebook, Instagram), and search engines (Google, Bing). The pipeline ingested and analyzed data from various vendors to generate key metrics such as cost per click, cost per application, impressions, and GRP — leading to a 30% reduction in marketing costs in one year after launch.
- Built a data pipeline using Airflow to standardize addresses for millions of records via an address standardization API, benefiting multiple pipelines across domains. Automated process replaced a manual one, reducing standardization time by ~90% and eliminating manual errors.
- Partnered with a third-party data provider to integrate hundreds of millions of sensitive records into the data lake. Designed and implemented a pipeline to transform and prepare the dataset for machine learning, and developed a reusable framework for running regression and classification models — expected to reduce marketing costs by an additional 25%.

### Data Engineer

Apr 2021 – Mar 2022

Vanguard

- Implemented and maintained data pipelines using AWS EMR, Glue, and other AWS products for big data analytics, enabling successful functioning of ~15 Tableau dashboards used by financial and business analysts for strategic decision-making by leadership teams.

### Data Engineer

Dec 2019 – Apr 2021

Jornaya (Verisk Marketing Solutions)

- Led development and enhancement of a core product requiring processing of billions of records from clients and building reports for users — using ~70 EMR clusters, ~10 Glue jobs, and AWS Lambda functions for serverless architecture.
- Optimized Apache Spark and AWS EMR configurations for a data pipeline serving ~50 vendors, saving more than 55% in AWS costs in one year while lowering the failure rate by ~90%, resulting in faster and more resilient delivery of results.
- Designed and implemented a data pipeline to asynchronously call an external API endpoint, transform the response, and write datasets to Amazon S3 — using AWS SNS (queuing), Lambda (API calls with multi-processing and transformation), and Step Functions (orchestration). The pipeline processed 1 billion records in ~25 minutes, enabling timely delivery of critical client reports.

### Data Developer

Feb 2010 – Dec 2019

Tata Consultancy Services

-

Built data solutions for 2 major clients in the banking and financial domain, contributing to more than 15 successful projects and numerous ad-hoc analyses for business users. Developed data pipelines across technologies ranging from legacy IBM Mainframes (COBOL, JCL, DB2) to modern stacks using Python, Apache Spark, and AWS products.

## PROJECTS

---

### The Data Domain Blog

A personal blog that curates news and articles in Data Engineering, Data Analytics, and AI/ML from various sources and RSS feeds. The backend uses Airflow to automate content discovery, retrieval, and summarization using LLMs from OpenAI and Anthropic. WordPress APIs are used to auto-publish articles with concise summaries and links to originals. Apache Airflow · Python · OpenAI API · Anthropic API · WordPress APIs · AWS

[datadomain.blog](https://datadomain.blog)

### Data Engineer's Guide

Dataguide.dev is an open-source knowledge base providing high-level technical resources for the data engineering community. It features curated guides on system design, data modeling, and interview preparation. Built to eliminate paywalls, the platform offers deep dives into modern architectures like Lakehouse and CDC, supporting engineers in career advancement React

[dataguide.dev](https://dataguide.dev)

## EDUCATION

---

### Veer Surendra Sai University of Technology

Bachelor of Technology, Electrical Engineering

## SKILLS

---

**Languages:** Python, SQL, COBOL, JCL

**Data & Pipelines:** Apache Airflow, Apache Spark, DBT, AWS Glue, AWS EMR, AWS Lambda, AWS Step Functions, AWS SNS, Amazon S3

**Cloud & Infrastructure:** Amazon Web Services (AWS), Terraform

**Analytics & Modeling:** Data Engineering, Data Modeling, Data Warehousing, Data Analysis, Machine Learning (Regression, Classification)

**Tools & Platforms:** Tableau, WordPress APIs, OpenAI API, Anthropic API, DB2

## CERTIFICATIONS

---

**AWS Certified Data Analytics – Specialty** Jul 2023