

Abhinav Ahuja

• Hyderabad, India • sourcewrite@gmail.com • +918699033319

Profile

Data Engineer with **8+ years at Amazon**, specializing in architecting and delivering large-scale data and analytics solutions. Proven expertise in building and optimizing ETL pipelines with PySpark, Redshift, and Athena, and in automating platform operations using AWS microservices and infrastructure as code. **5+ years mentoring mid-level engineers** in Amazon Last Mile Operations—focusing on design, reliability, and operational excellence; supporting a 5,000+ person org across Data & Applied Scientists, BI Engineers/Analysts, and Program & Operations Managers. **Advanced through four roles at Amazon while delivering 5 VP- and Director-level goals.**

Technical Skills

AWS, Data Engineering, System Design, Statistics, Redshift, Pandas, Pyspark, Glue, Sagemaker, Streamlit, IAM, CDK, Athena, S3, Lambda, ECS, Docker, CloudFormation, EMR, MageAI, DevOps, SQS, SNS, Quicksight. Python, SQL, JavaScript, TypeScript.

Experience

Amazon
Data Engineer II

Hyderabad
April 2024 - Present

- Engineered DETSO—a comprehensive data engineering platform leveraging Redshift, MageAI, and AWS CDK—to automate cluster monitoring, governance, and orchestrate 100 Big Data pipelines. Delivered near real-time dashboards, a 120-minute Query Killer, automated scaling, and advanced alerts that enhanced cost transparency and boosted cluster efficiency, significantly reducing monthly infrastructure costs against a \$3M annual Redshift budget.
- Developed a GenAI-powered SQL Converter using Streamlit, AWS Bedrock, Claude 3.5 Sonnet, and LangChain to automate complex SQL conversions between Redshift, Spark SQL, and PySpark—containerized with Docker and deployed via AWS ECS/CDK—cutting analyst effort by ~2 weeks per query set, accelerating ETLM migration to Spark and freeing critical Redshift resources.
- Deployed a 4-node RA3.4xlarge Redshift cluster using CloudFormation to automate setup and configure complex IAM roles for seamless integration with Amazon Internal Orchestrator—onboarded 200 datasets and optimized workload performance with WLM tuning and ETLM Unload to S3, boosting high-performance analytics for the Sciences team.
- Engineered a SageMaker-integrated EMR solution by developing CloudFormation templates and a one-click SageMaker Service Catalog that empowers BI teams and developers to launch Spark clusters directly from Jupyter Notebooks—simplifying cluster deployment for efficient querying of large datasets and streamlining the developer experience.
- Engineered Athena Over Hubble (AOH) to enable serverless SQL querying of 1,000+ Andes3 datasets by creating dedicated Athena Workgroups and automating cost tracking with AWS Glue and Lambda—streamlining access management and data sharing across BI, Sciences, Program, and Tech teams while enhancing cost transparency organization-wide.
- Delivered VP-level Andes3 Migration three months ahead of schedule by automating 12K+ ETLM profile

transitions (via Datanet APIs and Python, saving 2K+ manual hours) and migrating 1K+ tables from Andes2 to Andes3 using Subscription API and Andes SDK—eliminating redundant load jobs and boosting Redshift cluster performance for real-time data sharing.

Amazon Hyderabad
Business Intelligence Engineer II July 2022 - March 2024 (1 year 9 months)

- Engineered scalable data pipelines and advanced algorithms for Global Address Management (GAM) to launch in Turkey, Luxembourg, and Japan—enhancing a Bi-LSTM parser by correcting 9,000 token errors to boost F1 score from 52% to 69% and deploying a Spark-based MinHash LSH pipeline that raised matcher recall from 35% to 54% and de-duplication from 3% to 49%, reducing processing time from weeks to 6 hours and significantly improving address intelligence and operational efficiency.
- Achieved full compliance and operational excellence by resolving 700+ vulnerabilities (including critical Sev-2 Log4js escalations) and deprecating outdated AWS resources—saving \$30K/month—while streamlining AWS account management and driving comprehensive L&D initiatives (training sessions, video tutorials, and a centralized repository) to enhance team proficiency and operational efficiency.

Amazon Hyderabad
Business Intelligence Engineer I June 2020 - June 2022 (2 years)

- Built a fully automated Redshift–S3–Cradle–Spark–QuickSight pipeline to process 1B+ shipment records over one year—classifying into 25+ root-cause categories with 4-hour ETL updates and alerts, cutting EU misplanned rates from 1.51% to 0.40% and uncovering delivery-station improvements to boost routing efficiency.
- Engineered a large-scale analysis pipeline using SQL, Python, and statistical modeling to process 1B stops over six months—isolating 420K ExCo (Exclusive Containerization) vs. Non-ExCo stops across 8K Unit Places and uncovering weighted average savings of 93 seconds per stop and 7 seconds per package—driving strategic routing updates that enhanced driver efficiency and network performance.
- Engineered an end-to-end AWS pipeline (Andes → Cradle → S3 → Lambda → Glue Crawler → Glue Spark → Quicksight) that processes 7B+ daily signals across 30 types for 50K+ delivery vehicles—delivering near real-time, self-service dashboards to flag firmware anomalies and optimize telemetry hardware performance.

Amazon Hyderabad
Business Analyst Support Jan 2018 - May 2020 (2 years 5 months)

- Optimized Fleet Edge deployment by engineering a serverless AWS-based pipeline—leveraging Spark RDD for 7B+ data points and Rtree indexing for geospatial computations—to process 700M+ traversals against 200M map segments, reducing computation time from weeks to hours and evaluating 66M+ DSP allocation combinations, which drove \$80.4M in savings and improved on-road efficiency for 5,000 delivery vehicles within a \$33M operational cost framework.
- Engineered a scalable AWS-based data architecture and custom image labeling tool using SageMaker and Python, enabling manual auditors to tag multiple objects within 2 days and automating batch processing for new

uploads—accelerating POD (Photo on Delivery) model training and reducing time-to-market while ensuring strict compliance with AWS internal microservices for Restricted Data usage.

Amazon
Data Support Engineer

Hyderabad
Mar 2017 - Dec 2017 (10 months)

- Built a scalable Spark workflow to analyze WAPE errors at trip and segment levels, improving segment-level WAPE by 400 bps and revealing that trips >500m accounted for 50% of errors.
- Converted R-based route optimization to a Python solution with multiprocessing, reducing per-station run time from 6 hours to 6 minutes and automating ETL to deliver multi-region results in under 4 hours.

Early Career

May 2014 - Sep 2016

Hindu Council UK (London, Remote) • Electric City (Chandigarh) • I-next Solutions (Chandigarh)

Program Manager • Sales Manager • Web Developer

Oversaw web hosting and CMS site builds, managed outsourced/freelance developers, and led sales team segmentation and brand-engagement initiatives.

Honors & Awards

Just Do It Award (Q1-Q2 2024), Innovation Maven (Q3-Q4 2023), Service Champion (Q3-Q4 2017) – Amazon

Education

Punjab Technical University
Bachelors in Electronics & Communications Engineering

Skills & Interests

Language: English, Hindi

Interests: Spending time with my dog and cats, playing games like Dota, Satisfactory and Hearts of Iron IV.