

SUMMARY

AI-focused Data Engineer with 6+ years of experience building high-throughput, event-driven data platforms, real-time analytics, and automated reliability tooling across supply chain, finance, and lending domains. Blend modern data engineering with applied ML/NLP skills and cloud-native design to deliver production-grade, metric-driven solutions that unlock new automation and decision intelligence.

EDUCATION**W. P. Carey School of Business at Arizona State University**

August 2022 – May 2023

Master of Science in Business Analytics (MSBA) | GPA: 4.00 | 4.00

Tempe, AZ

Relevant Coursework: Advanced Statistics, Supply Chain Analytics, Machine Learning, Optimization Modeling, Marketing Analytics, Deep Learning, Enterprise Analytics, Natural Language Processing, Data-Driven Quality Management

Amity University Noida

July 2013 – May 2017

Bachelor of Technology (Computer Science & Engineering) | CGPA: 7.40 | 10

Uttar Pradesh, India

PROFESSIONAL EXPERIENCE**Data Engineer II | Sr. Data Analyst**

September 2023 - Present

Americold Logistics, LLC

Georgia, USA

Cloud Data Architecture & Engineering

- Architected replacement of legacy SAG system, deploying ECS Fargate cluster to ingest, map, and output WMS data to S3/XML for 3PL apps delivering real-time processing, eliminating batch delays, and scaling to support 150+ warehouses with 99.9% uptime.
- Led team of 3 engineers in Accessorials cloud migration, building end-to-end event-driven pipeline (Redshift to DynamoDB leveraging Lambda/SQS/Step Functions/Glue) scaling to 500K daily events with parallel job orchestration and real-time SNS alerting.
- Implemented AWS DMS with CDC to migrate 3PL application from on-premises Oracle to Redshift achieving zero-downtime cutover with continuous incremental sync, processing 10M+ records daily post-migration.
- Built a high-throughput AWS pipeline to capture Salesforce data via Partner Event Bus using SQS-triggered Lambda functions to efficiently ingest multi-stream revenue data processing 250K events/day into Redshift SUPER column for flexible semi-structured storage; developed dynamic PartiQL views on the table and designed a schema enabling seamless ingestion into EPM application.
- Developed Dockerized Flask API aggregating billing data from Redshift views for Finance team applications deployed on AWS ECS with Application Load Balancer, serving 10K+ daily requests at 100% SLA compliance and accelerating month-end close by 2 days.

Pipeline Automation & Data Reliability

- Engineered event-driven AWS Glue job monitoring system using Amazon DocumentDB and SNS. Triggering automatic status entries and real-time email alerts for success/failure, enabling proactive issue resolution, cutting manual monitoring by 80%.
- Created Lambda-driven Step Functions automation that automatically detects, prioritizes, and reprocesses failed billing Glue jobs from Redshift logs (last 24hrs), replacing 80+ hours of monthly manual error hunting and recovery across facilities—handling up to max-concurrency failures with smart ranking by facility/time/failed steps.
- Automated the daily SharePoint-to-S3 metrics sync using AWS Lambda integrated with Azure ADP, performing file validations and uploads without manual intervention. This replaced a 2–3 hour daily manual process, saving approximately \$41K–\$62K in labor costs. Enhanced data accuracy and reliability by triggering automated email alerts to warehouse GMs for any validation issues.
- Developed and automated an error file validation pipeline, which scans all servers, extracts error files, and compares them with the previous day's data. Integrated production log analysis and created a batch script for automated file reprocessing, packaging the solution into an executable for efficient error handling. Within 3 months, this process reduced data-related errors by ~50%.
- Created and automated an outlier report using an AWS Lambda function, integrating data from Oracle and Postgres databases. Applied z-score, IQR and other statistical measures, reducing validation time by ~1 hour daily and decreasing false positives by 20%.
- Built a comprehensive, interactive data quality tool using Streamlit, centralizing daily monitoring by integrating 5–6 key checks such as outlier detection, missing data, and error file validations. Developed dynamic dashboards enabling automated email alerts based on analysis. The tool, highly praised by management, provides real-time access to monitor data quality issues and insights.

Data Analysis, Validation & Reporting

- Built a high-performance, cross-database reconciliation framework using AWS Glue, PySpark, and Redshift Spectrum to compare Redshift and Oracle datasets via UUID-based hashing, reducing processing time from 3 hours to 3 minutes; stored results in Parquet file and created dynamic views to compare record counts and computing column-level variances with percentage differences.
- Played a pivotal role in the “Data as a Service” project with Google by designing and enhancing a robust data profiling script using Python and SQL to compare table structures, data consistency, and quality metrics between Postgres and Redshift databases; automated validations for completeness, uniqueness, and constraints proactively identifying discrepancies.
- Enhanced and automated Daily Flash report for management using AWS Lambda with advanced SQL optimizations, streamlined Data Analytics workflows, developed Python scripts for XML file comparison/modification resolving data issues, now used daily by development team to process data for 200+ warehouses.

Research Assistant | Supply Chain Department

July 2023 – September 2023

Arizona State University

Arizona, USA

- Spearheaded a project focused on enhancing process efficiency within the supply chain department leveraging analytics, automation techniques, Lean Six Sigma (LSS), and Robotic Process Automation (RPA) to optimize operations.
- Leveraged Analytics principles to design and implement automated data pipelines for report maintenance, and classroom assignment updates. Orchestrated workflows with Apache Airflow to optimize process efficiency and minimize manual effort.

Data Scientist Intern | Capstone Project

January 2023 – May 2023

Axon Enterprise

Arizona, USA

- Built an NLP-based stacked binary classification model to classify customer reviews, reducing manual effort by 3 business days per month. This resulted in an 11% boost in the Net Promoter Score, reflecting improved customer satisfaction across all services.
- Delivered an interactive and user-friendly Tableau Dashboard and Streamlit web application, assisting stakeholders effortlessly analyse the consumer behaviour across all categories, highlighting Axon's strengths and pinpointing areas for improvement.

Business Analyst

April 2020 – March 2022

Better.com

Gurgaon, India

- Streamlined and optimized data pipelines for evaluating income, assets, credits, and other KPIs, enhancing the accuracy of mortgage option identification. This reduced mortgage processing time by 20% and significantly improved customer satisfaction.
- Developed machine learning models to accurately predict loan eligibility of approximately \$2.5 million per week in purchase loans by collaborating with underwriting team, boosting approval rate by 15% and driving higher loan origination volumes.
- Prepared and validated monthly dashboards using SQL, applying Star schema principles for optimized data modeling. Providing actionable insights on market trends and revealing a 37% increase in refinance loans triggered by the COVID-19 pandemic.
- Executed risk analysis on 1,500 loans per week using predictive analytics and risk rating models. Validated data integrity through automated pipelines, evaluating key metrics like loan-to-value and debt-to-income ratios to support informed decision-making.
- Automated workload allocation involving cross-functional collaboration with data engineers and product management team for integration of a Python script with an in-house tool for process improvement, increasing team efficiency and productivity by 20%.

Project Engineer

September 2017 – December 2019

Credit Suisse | Wipro Technologies

Pune, India

- Used ETL tools like Informatica for data extraction and transformation, incorporating APIs to integrate and retrieve data from various sources. Performed necessary data augmentations based on critical financial indicators, resulting in standardized operations.
- Ensured data quality through rigorous validation and verification post-ingestion, leveraging Snowflake schema principles for structured data integration. Conducted market analysis across EMEA and APAC regions, providing insights for strategic decisions.
- Designed automated Tableau dashboards, saving ~20 hours per month in reports generation, providing recommendations on areas for improvement, and evaluating data impact issues, resulting in a 15% increase in efficacy in decision-making processes.
- Improved operational efficiency of workforce by optimizing various tasks, including developing a generic file comparison script with Python, resulting in saving 2 business days of work in manual data comparison.

PROJECT EXPERIENCE

Hotel Reservation Prediction

November 2022 – December 2022

- Using statistical modeling techniques like logistic regression, built a model to predict hotel booking cancellations with a 94% recall and 72% accuracy. Implemented a user-friendly web application using Streamlit for seamless access to the model's predictions.

Project Mentor and Reviewer | Udacity.com

August 2020 – February 2022

- Mentored ~85+ students and reviewed ~1000+ projects worldwide under analytics and product management domains. Provided critical insights and constructive feedback, helping students graduate from respective certification programs.

Assessing the Impact of New Landing Page on User Conversions: A/B Testing

January 2020 - February 2020

- Conducted A/B testing to evaluate the impact of a new landing page on user conversions. Statistical and Quantitative analysis (Z-test, 95% confidence level) revealed no significant difference in conversion rates. Hence, failing to reject the null hypothesis.

SKILLS & ACTIVITIES

- **Languages:** Python, PySpark, SQL, R
- **Tools:** Tableau, Looker, Microsoft Excel, IBM SPSS, Google Analytics, Minitab, Power BI, Snowflake, Apache Airflow, Git, Informatica, Bash
- **Technical Skills:** Data Analytics, Data Visualization, Business Intelligence, Machine Learning, Natural Language Processing, Data Engineering, Data Architecture, AWS, Redshift, Data Warehousing
- **Certifications:** Data Analytics, and Marketing Analytics Nanodegree (Udacity.com), NLP with Python bootcamp (Udemy.com), Tableau Fundamentals (Datacamp.com), Data Analyst in PowerBI Track (Datacamp.com), Profession & Associate Data Engineer Track (Datacamp.com), AWS Cloud Practitioner, AWS Data Engineer Associate
- Member of **Beta Gamma Sigma**, The International Business Honor Society
- Director of Professional Services

Rotaract Club of Delhi South | August 2020 – July 2022