Arnob Kumar Dey

[Senior Data Engineer]

https://dataspec.dev

arnobkrdey@gmail.com

(+91) 9748895343

https://linkedin.com/in/arnobde/

Summary

- Around 10+ years of combined experience in cloud computing and data engineering.
- M-TECH in Data Science and Engineering from BITS Pilani.

- ***Technical skills:***

AWS	● ● ● ● ○
SQL	● ● ● ● ○
PySpark	● ● ● ● ○
Python	● ● ● ● ○
ETL	● ● ● ● ○
Airflow	● ● ● ○ ○
Terraform	● ● ● ○ ○
Databricks	● ● ● ○ ○

- ***Functional skills:***
  Successfully lead teams distributed across multiple locations and met project goals within the given timeline.

Certifications

- AWS Certified Developer Associate.
- Databricks Certified Data Engineer Associate.
- Oracle Certified Associate in Java Programming.
- Cisco Certified Network Associate (Routing & Switching).

Experience

1. Senior Data Engineer at **Intuitive Technology Partners (India) Pvt. Ltd** [https://www.intuitive.ai/]
   March'2025 till January'2026

   Client: S&P Global

   Project: Recall Modernization and Data Mart Development

   Key Contributions:

   - Led the end-to-end modernization of a legacy system (*Recall*) by migrating it from VMware to a fully managed AWS environment, improving scalability, fault tolerance, and operational efficiency.
   - Architected and developed a centralized data warehouse, *Recall Data Mart*, enabling standardized, efficient data delivery to multiple downstream applications and analytics platforms.
   - Implemented scalable ETL pipelines using AWS Glue with support for modern table formats like Apache Iceberg, improving data query performance and schema evolution capabilities.
   - Designed cost-effective, serverless data processing workflows using EMR Serverless, Lambda, SNS, and Apache Airflow (MWAA) for orchestration, reducing maintenance overhead.
   - Developed automated data ingestion, validation, and transformation logic using Athena, S3, and Oracle DB, ensuring robust and accurate data availability.
   - Leveraged Terraform for infrastructure-as-code to provision and manage AWS resources securely and reproducibly.

   ***Tech Stack*:** AWS Glue (Iceberg), EMR Serverless, Apache Airflow (MWAA), Athena, Lambda, S3, Oracle DB, SNS, Terraform, Python, PySpark

   Internal Projects_____

   Data Validation Framework with Great Expectations

   - Designed and implemented a custom data validation framework using the *Great Expectations* Python library to enforce data quality standards across multiple pipelines handling FHIR (Fast Healthcare Interoperability Resources) data.
   - Deployed the framework in both Databricks and AWS, ensuring seamless integration with existing ETL workflows.
   - Automated AWS deployment using Terraform scripts, enabling reproducible infrastructure setup and compliance with organizational standards.
   - Delivered a scalable solution that improved data reliability, schema consistency, and governance across internal projects.
   -

   Tech Stack: Great Expectations, Python, Databricks, AWS (S3, Glue, Lambda), Terraform

2. Senior Data Engineer at **Kruti Solutions** [*https://www.krutisolutions.com*]
   Oct'2023 till Feb'2025

   Key Contributions:

   - Designed and maintained robust, production-grade ETL pipelines from AWS to Snowflake, enabling real-time analytics and reporting.
   - Developed and optimized Databricks notebooks and jobs to support data transformation and enrichment for analytics teams.
   - Collaborated closely with data analysts and business stakeholders to understand reporting requirements and ensure data accuracy and completeness.

   Tech Stack: AWS, Snowflake, Python, SQL, PySpark, Databricks

3. Senior Data Engineer at **EXL Service (Inductis (India) Pvt Ltd)** [*https://www.exlservice.com*]
   Internal Designation: Sr. Manager
   May'2023 till September'2023

   Client: Prudent

   Key Contributions:

   - Enhanced the capabilities of the Enterprise Data Platform by introducing new data pipeline features and improving operational performance.
   - Developed serverless and event-driven ETL pipelines using AWS Glue, Lambda, and Step Functions integrated with AWS RDS, S3, DynamoDB, and Athena.
   - Created automated monitoring and alerting systems using CloudWatch to ensure data reliability and minimize downtime.

   Tech Stack: AWS (Glue, Lambda, S3, Athena, DynamoDB, RDS), Python, SQL, Airflow, Terraform

4. Lead Developer at **Delta Airlines Global Technology Hub LLP** [*https://www.delta.com*]
   October'2021 till May'2023

   Project Name: Operational Experience (OE)

   Key Contributions:

   - Led the design and implementation of a cloud-native scheduling optimizer for airline pilots, automating compliance with FAA regulations.
   - Built a high-performance infrastructure using AWS CDK for provisioning, Lambda for orchestration, and SageMaker for model deployment.
   - Integrated various AWS services including Glue, Step Functions, Athena, RDS, SQS, SES, and CloudFront to support scalable and secure application workflows.

   Tech Stack: AWS CDK, Step Functions, SageMaker, Lambda, Glue, Athena, SQS, SES, RDS, DynamoDB, CloudWatch, Route53, Python, FastAPI, PySpark

5. Cloud Deployment Engineer at **Adobe Systems India Pvt Ltd** [*https://www.adobe.com*]
   Since October'2017 till October'2021

   Key Contributions:

   - Provisioned end-to-end cloud environments in AWS and Azure for Adobe Experience Manager (AEM) and Adobe Connect clients.
   - Automated cloud infrastructure setups implemented custom features, and integrated ML solutions to enhance product capabilities.
   - Delivered ML solutions for insurance damage prediction and customer behaviour analytics using Python and Jupyter.

   Tech Stack: AWS (EC2, S3, ALB, Route53, IAM), Azure, Python, SQL, AEM, Jupyter, ETL, Nagios, New Relic.

   ___Projects_____

   Project Background:
   AXA customers/onsite survey managers upload images of vehicles involved in accidents to from their site hosted through AEM which then stores the images in AWS S3. AXA would then manually import these images to their on-premises storage for manual inspection by analysts and later determine final insurance settlements with regards to claims from their customers.

   Project Requirements:
   Develop and train a machine learning model that would access images directly from AWS S3 and predict extend (in percentage) of damage to the vehicles with best possible accuracy.

   Tech stack: Python, IDE: Jupyter Notebook

   Project Background:
   Adobe AEM enterprise customers are entitled several additional services which enables customers to get the best returns on their investments. Examples – Cloud Manager, Site Analytics & Targeting etc. Inhouse development of these additional services requires significant time, effort and investment. Hence, it's necessary to understand current adaption rate for these services by the customers and predict expected future adaption rate.
   Project Requirements:
   Customer records are stored ServiceNow. Collect and build training dataset.
   Identify relevant columns. Cleanup dataset using data preprocessing technique.
   Build a suitable model to meet the expectations.

   Tech stack: Python, IDE: Jupyter Notebook, SQL

6. Cloud Administrator at **Aptean India Pvt Ltd** [*https://www.aptean.com*]
   Since September'2015 till September'2017

   Key Contributions:

   - Set up scalable and secure AWS Cloud environments for small to medium-sized applications and services.
   - Designed and implemented load balancing using Elastic Load Balancers (ELBs) to ensure high availability and fault tolerance.
   - Managed end-to-end SSL certificate lifecycle, including CSR generation and integration with ELBs and EC2 instances.
   - Configured and maintained DNS routing and domain management using Amazon Route 53.
   - Handled identity and access management using AWS IAM, and led VPC provisioning and maintenance for network isolation and security.
   - Monitored cloud infrastructure health and alerts using Nagios, PagerDuty, and OpsGenie, improving incident response times.
   - Wrote and optimized SQL queries for inventory database insights and reporting.
   - Administered Windows Active Directory including user, group, and OU management to support enterprise access policies.
   - Conducted regular Qualys vulnerability scans to support PCI compliance and reduce security risks.

_____Projects_____

Project Background:
DHL uses Aptean's SCM (Supply Chain Management) product Tradebeam. Countries around world keeps updating their custom laws concerned with items (categorized using HS code) being shipped to or from their respective countries. For some countries these updates are consolidated in either .csv format or pdf format which is then parsed by Tradebeam.
Before Tradebeam can parse these documents, it needs to be validated by individual country managers who after validation would send emails with the list of sheets that they have validated. This process is not efficient and time consuming.

Project Requirements:
Eliminate the need for email communication and manual production deployment by automating the process flow. Available list of sheets to be validated are collected by EMF (Event Management Framework) which will produce an XML file which will be read using java and populate the list in UI using JSP. The country managers after validation would approve the sheets on the UI. The approved and validated sheets will be processed and deployed in production automatically.

Implementation Language: Java
Tools: EMF

Academic Qualification

1. M-Tech in Data Science and Engineering from Birla Institute of Technology in the year 2021 [Regular offline – Work Integrated Learning Program - WILP].
2. B-Tech in Electronics & Communication Engineering (ECE) from Academy of Technology under West Bengal University of Technology in the year 2013 [Fulltime].

Personal Details

Languages Known : English, Hindi and Bengali (Mother Tongue)
Relocation Preference : Bengaluru, Gurgaon, Mumbai, Hyderabad, Pune, Chennai, Kolkata

Nationality : Indian
Date of Birth : 02-May-1987

US Visa : Available (B1/B2, Multiple Entry, valid till April, 2028)

Permanent Address : B-13/64, Kalyani, Dist.: Nadia, West Bengal, Pin-741235

Date: Jan 29th, 2026

Arnob Kumar Dey