

PRATEEK DUBEY

58 BAYSHORE ROAD, BAYSHORE PARK, SINGAPORE - 469981 | +65-93804965 | prateekdubey12@gmail.com

[LinkedIn](#) | [GitHub](#) | [Website](#) | [Medium](#)

PROFILE

Data & AI Engineering Leader with **13 years of experience** architecting enterprise-scale Data & AI solutions. Proven track record leading high-performing teams and driving digital transformation initiatives resulting in significant cost savings and efficiency improvements. Expert in Data, AI & ML Engineering, Data Management & Governance, Cloud-Native Architectures, & DevOps.

SKILL & COMPETENCIES

- ♦ **Big Data processing frameworks:** Hadoop, Spark, Flink, Kafka
- ♦ **Cloud Data Platforms:** Databricks, Snowflake, Azure Synapse, Redshift
- ♦ **Data formats:** Parquet, Delta Lake, Iceberg, Avro, JSON, YAML
- ♦ **Data Management & Governance:** Apache Ranger, Apache Atlas, Amundsen, Datahub, Open Metadata, Great Expectations
- ♦ **Cloud:** Amazon Web Services, Azure, Google Cloud Platform
- ♦ **On-Premise:** Cloudera CDH, Ceph, Bind9, Metal LB, Rancher Kubernetes (RKE and K3s)
- ♦ **ML-Ops:** Kubeflow, KServe, MLflow, Seldon, Feast
- ♦ **Programming & IDE:** Python, Pandas, Polars, Unix, Shell scripting, SAS, DBT, SQL Mesh, Jupyter, VSCode, Pycharm
- ♦ **Observability:** Grafana, Prometheus, Loki, Thanos, Promtail, Elasticsearch, Kibana, Logstash, SigNoz, Open telemetry
- ♦ **Orchestration:** Control-M, Apache Oozie, Apache Airflow, Dagster
- ♦ **Relational, NoSQL, Graph Databases:** DB2, Informix, MySQL, Postgres, Teradata, HBase, Trino, Druid, DuckDB, Presto
- ♦ **AI tools:** Ollama, Claude Code, Cursor, Vector DBs, RAG, Agentic AI, Langfuse, Langchain, LangGraph, etc
- ♦ **CI/CD and DevOps:** Git, Ansible, Gitlab CI, GitHub Actions, Spinnaker, Argo CD, Tekton
- ♦ **Infrastructure as a Code and Containers:** Terraform, CloudFormation, Azure ARM, Docker, Kubernetes, Helm

PROFESSIONAL EXPERIENCE

EY (AI & DATA) <https://ey.com/>

DIRECTOR OF AI & DATA ENGINEERING | SINGAPORE | SEPTEMBER 2025 - PRESENT

- ♦ Principal AI & Data Engineer in Technology Consulting division. Leading a team of 10+ AI, Data & Software Engineers
- ♦ Working with clients in ASEAN across Gen AI, Agentic AI and Real-time data solutions and architecture
- ♦ Building our Agentic AI Accelerator Platform - Decision OS to bring Agents to our enterprise customers with end-to-end Agent Observability, Governance, Guardrails
- ♦ Delivering GenAI projects for FairPrice Group (POS AI Knowledge Assistant) and Temasek (AI Fluency Assessment Platform)

CHUBB (CHUBB LIFE) <https://chubb.com/>

HEAD OF DATA ENGINEERING & ARCHITECTURE | SINGAPORE | JUNE 2025 – SEPTEMBER 2025

- ♦ Partnered with global senior leadership to define and execute Chubb Life's enterprise-wide Data & AI strategy, shaping technology roadmaps to drive measurable business impact and foster a data-driven culture.
- ♦ Built, coached, and scaled a high-performing regional team of over 50+ Engineers establishing a COE for DE and applied AI.
- ♦ Directed the implementation of an advanced, cloud-native data platform, creating scalable frameworks for ingestion, transformation, and governance that accelerated the delivery of enterprise-scale data solutions.
- ♦ Championed innovation by leading the development of applied ML solutions, including RAG based AI assistants that empowered business stakeholders with self-service analytics. Currently pioneering the use of AI Agents to automate and scale DE and ML.

TEMUS (AI & DATA) <https://temus.com/>

DATA ENGINEERING MANAGER/ STAFF DATA ENGINEER | SINGAPORE | APRIL 2024 – JUNE 2025

- ♦ Managed, mentored and coached a team of 20+ Data Engineers covering Data Engineering, MLOps, AIOps, Data & AI Architecture
- ♦ Data architecture lead guiding both Public and Private sector clients in implementing standardized architectural frameworks and methodologies. Ensuring consistency in design approaches across projects to facilitate the creation of scalable, reusable solutions.
- ♦ Delivered Industry Partnerships Platform for Ngee Ann Polytechnic using LLM's to solve conversational AI on Azure GCC
- ♦ Worked with FWD Insurance as a Lead Engineer & Architect to rollout LLM powered voice to voice conversational AI avatars for training sales agents. Scaled LLM application on EKS to serve 1000s of users across South East Asia

- ♦ Worked with Temasek Trust & Laconic on Lakehouse design on Azure & AWS using Synapse, EKS, ADF, Spark, DBT, OM
- ♦ Led pre-sales, BD, proposals, tech consulting for public and private sector clients. Successfully closed deals worth 500k SGD.
- ♦ Developed internal solutions to automate ETL development using LLMs and data governance using GX & Open Metadata

GLOBAL FASHION GROUP (GSF DATA) <https://global-fashion-group.com/>

DATA ENGINEERING MANAGER/ STAFF DATA ENGINEER | SINGAPORE | NOVEMBER 2021 – MARCH 2024

- ♦ Worked with Dafiti (LATAM) on data strategy, restructuring, prioritisation, roadmap planning, resolving customer pain points, and adopting data mesh architecture by treating data as a product for business deliverables
- ♦ Managed a team of eight (8) Data & MLOps Engineers based out of Vietnam (FTE) and India (Contractual) covering Data Engineering, DevOps, MLOps, Solution Architecture, Data Governance, Observability and Cloud for GSF Data
- ♦ Staff Engineer/ Architect for Global Data Platform using OSS CNCF, Linux & AI, and ASF tech stack on Kubernetes (AWS EKS). GDP is a Central Data Platform for GFG, Zalora, The Iconic and Dafiti to run their workloads including Model Serving.
- ♦ Technical Leader for GDP (Infrastructure as a Service), DaaS (Data as a Service) projects. Worked closely with regional teams from Zalora (SEA), and The Iconic (ANZ) to deliver data and ML solutions for Pricing Engine, Recommendation, Personalisation.
- ♦ Responsible for setting quarterly OKR's, weekly 1:1s, technical delivery of projects, setting up roadmap for Data Platform, etc.
- ♦ Helped Zalora, Iconic, Dafiti migrate their Seller Dashboards to GDP to build a subscription based data as a product.
- ♦ Led the efforts to shutdown Snowflake and Redshift Cloud DWHs to cut cost and migrate to Trino backed Lakehouse.
- ♦ Ingested GA3 (The Iconic, Dafiti) and Segments (Zalora) events into GDP for behavioural data recommendation.
- ♦ Improved GDP observability and availability using Loki, Thanos, Grafana, and Prometheus with Slack and Email alerts.
- ♦ Incorporated Git-Ops methodology, setup CI/CD processes using GitHub Actions, Argo CD, Ansible and Terraform.
- ♦ Setup Data Governance & Quality for our Iceberg Lakehouse backed by Trino using Apache Ranger & Datahub and GX.
- ♦ Setup ML-Ops solution using Mlflow (Experimentation), Feast (Feature Store), KServe (Model Serving), Kong (API Gateway).

SINGTEL (EMERGING TECH) <https://www.singtel.com/>

DATA ENGINEERING MANAGER/ LEAD DATA ENGINEER | SINGAPORE | AUGUST 2020 – NOVEMBER 2021

- ♦ Built Telco Cyber Data Lake & Platform using Hybrid Cloud for Cyber Analytics and Threat Intelligence for Singtel Cybersecurity.
- ♦ Lead Architect and Engineer for Telco CDL Kubernetes Platform running On-Premise using Rancher and on AWS using EKS.
- ♦ Processed TBs of data using Spark on K8s stored in Kafka, Ceph and S3 using our Data and ML Platform on AWS and On-Premise.
- ♦ Built Streaming & Batch data pipelines using Apache Kafka and Spark Streaming. Orchestrated data pipelines using Airflow.
- ♦ Setup requisite logging and monitoring for our system and services using Elasticsearch, FluentD, Grafana, and Prometheus.
- ♦ Setup up DevOps, CI/CD processes using Gitlab CI, Argo CD, Ansible and Terraform.
- ♦ Setup Data Governance for Kafka, Ceph, Hive, & S3 using Apache Atlas, Apache Ranger and Amundsen.
- ♦ Setup ML-Ops solution for Telco CDL by leveraging Mlflow to track Experiments, log metrics & parameters.

GRAB (ML & AI INFRASTRUCTURE) <https://www.grab.com/>

DATA SCIENCE LEAD | SINGAPORE | SEPTEMBER 2018 – JULY 2020

- ♦ Technical lead for ML Infrastructure team. Responsible for looking after Cloud Architectural design solutions over AWS and Azure for Data Science. Involved in ML-Ops, DevOps, Cloud, Data Engineering, and Data Science.
- ♦ Led a team of ML Engineers to build, manage and support ML Platforms on AWS and Azure along with supporting teams in their ML projects by building jobs related to Data Engineering, Feature Engineering, Model Training, Model Deployment etc.
- ♦ Data Science PIC for vendor engagement with AWS, Microsoft and Databricks and also owner for all our DS Cloud Accounts.
- ♦ Implemented Data Governance and Infrastructure Governance on our ML Platforms on AWS and Azure.
- ♦ Worked with GrabFood and Azure CAT team to develop GrabFood restaurant behavioural recommendation system on Azure Cloud Platform thus increasing Click through Rate by 170% compared to existing recommendation systems at Grab.
- ♦ Developed a Multi-User Jupyterhub and Kubeflow Data Science ML Platform on AWS EKS for ML and Analytics.
- ♦ Automated Cloud Infrastructure using Terraform and Azure ARM. Used Helm as a K8s package manager on EKS and AKS.
- ♦ Led the team to Architect and develop multiple solutions for our ML Platform like Spark UI Proxy, Jupyterhub resource tracking, AWS monitoring, Datadog monitoring for EKS, Agartha – Docker build Automation using Git Pull Requests.

ROYAL BANK OF SCOTLAND (ENTERPRISE SOLUTIONS) <https://www.rbs.com/>

SENIOR DATA ENGINEER | GURUGRAM, INDIA | JULY 2017 – AUGUST 2018

- ♦ Worked as part of Cloud Engineering and was responsible for migrating Athena SAS DWH onto AWS for Risk & Finance Solutions.
- ♦ Designed Data Lake on AWS which was utilized by Data Engineering & Analytics for ETL, Analytics and Machine Learning.
- ♦ Automated AWS Cloud Infrastructure using Terraform for Risk and Finance Solutions.
- ♦ Designed and developed a Python Parser to auto-convert HiveQL codes into equivalent Pyspark (Spark SQL) jobs to leverage the Spark capabilities on AWS EMR, thus reducing conversion time by over 90%.
- ♦ Reduced job workflow creation time by 80% through an automated Oozie workflow creation framework.

- ♦ Designed services for seamless monitoring like monitor Active EMR Clusters and Sagemaker instances running across all regions. Used Boto3 library and deployed solution on Lambda. Business notifications configured via SES and scheduled via CloudWatch.
- ♦ Enabled Amazon Sagemaker for Risk Modelling (Machine Learning) to leverage Spark on Sagemaker both locally and through remote EMR Cluster. Persisted Sagemaker Jupyter Notebooks in S3 instead of local EBS volume.
- ♦ Installed and setup Apache Airflow to schedule our Core Spark ETL Workflows.
- ♦ Worked on a framework for orchestration & monitoring of our Core EMR Cluster using Lambda, CloudWatch, & SNS.

EXL (BARCLAYS DECISION ANALYTICS) <https://www.exlservice.com/>

SENIOR DATA ENGINEER | GURUGRAM, INDIA | DECEMBER 2016 – JULY 2017

- ♦ Worked on migrating SAS Datawarehouse onto Big Data platform - Cloudera for Risk Solutions vertical of Barclays.
- ♦ Designed an On-Premise data lake using Cloudera CDH, which was utilized by Data Engineering & Analytics for ETL workloads.
- ♦ Developed Real-Time, and batch data pipelines using Spark streaming, Pyspark, HBase.

AON (BUSINESS INTELLIGENCE) <https://www.aon.com/>

SENIOR DATA ENGINEER | GURUGRAM, INDIA | NOVEMBER 2013 – DECEMBER 2016

- ♦ Worked on multiple End to End Big Data project deliverables involving data ingestion, ETL, designing a data lake, data migration, creating unit test plans, proof of concepts etc.
- ♦ Designed and developed HRBPO Attendance Tracking System using Hive and Hadoop used for tracking logging hours of CSR's.
- ♦ Carried out POC's on varied Hadoop technologies like HDFS Snapshots, High Availability of HDFS/Hive/YARN/HUE, NoSQL database - HBase, Active Directory implementation on Hadoop Servers, Hive on Spark, Scala on Spark etc.
- ♦ Responsible for production ETL workload job monitoring using Control-M, ETL fixes and ETL enhancements raised by business stakeholders and users. Designed and developed ETL codes in PIG Latin, HiveQL, Unix and SQL.

PROJECTS

DATA, ML & AI ENGINEERING

- ♦ **Chubb, Singapore:** Led Data Engineering efforts to support business critical dashboards like Daily Sales, Agency Metrics, Customer 360, etc. Designed and architecture digital consumption of Customer 360 data via Mobile, Web, CRM, etc.
- ♦ **Temus, Singapore:** Led various Data and AI Engineering efforts on AWS and Azure for Public and Private sector clients to support business use cases around optimisation, digitalisation, migration, etc.
- ♦ **Global Fashion Group, Singapore:** Led the Data Engineering track to support ML use-cases around recommendation, personalisation, etc. Supported Zalora's pricing engine, data as a service, Datajet (ML as a service) products and solutions.

DATA & ML PLATFORMS

- ♦ **Global Data Platform, GFG, Singapore:** Led the development of GDP on AWS EKS establishing an end-to-end Data & ML infrastructure to support diverse workloads across GFG, Zalora, and Iconic. Support data ingestion from varied data sources for behavioural data recommendations, search, personalisation, inventory replenishment, etc.
- ♦ **Telco Cyber Data Lake Platform, Trustwave, Singapore:** Led the development of Telco CDL and Platform using hybrid Kubernetes infrastructure with Rancher for on-premise and AWS EKS for cloud deployments. Processed TB's of network logs from Singtel for Threat Intelligence and Analytics.
- ♦ **AWS and Azure ML Platforms, Grab, Singapore:** Led the development of ML Platforms, leveraging AWS EKS and Azure Databricks to support production ML workloads for Grab Transport and Grab Food. Use cases include recommender systems, surge pricing, food order batch queuing, personalisation, etc.

DATA & ETL MIGRATIONS

- ♦ **Athena Re-platforming, RBS, India (SAS DWH Migration to AWS):** Worked as a Spark Developer to successfully migrate 2000+ ETL scripts from SAS to Pyspark on AWS EMR, Glue, & Athena.
- ♦ **Barclays Bureau, EXL, India (SAS DWH Migration to Cloudera):** Worked as a Spark Developer to successfully migrate 1000+ ETL scripts from SAS to Pyspark on Cloudera's Hadoop Platform.
- ♦ **Informix Shutdown, AON, India (Legacy DWH Migration to Cloudera):** Worked as a Hadoop Developer to successfully migrate 2000+ ETL scripts from Unix/ SQL to Pig Latin on Cloudera's Hadoop Platform.

EDUCATION

UNIVERSITY OF PUNE, INDIA (2009 – 2013)

BACHELOR OF ENGINEERING, COMPUTER ENGINEERING, FIRST CLASS WITH DISTINCTION

Place: Singapore

Date: 1st May, 2026