Aniruddh Sharma

1857 Fairport Rd, Pickering, L1V 1T2 (CA) 416-453-4639

[asharma@pragnatyx.com](mailto:asharma@pragnatyx.com)

<www.linkedin.com/in/aniruddh-sharma-71737164>

# PROFESSIONAL SUMMARY

Experienced Solution Architect and Data & ML Engineer with a strong background in designing and implementing innovative, cost-effective, and forward-thinking technology solutions using major cloud service providers. Specializing in data engineering, machine learning and generative technologies, I transform complex data, including real-time streaming and big data, into actionable insights and create intelligent, data-driven solutions that drive business success. Proficient in developing, scaling, and monitoring AI solutions, automating ML pipelines, implementing MLOps practices, and building robust data pipelines for efficient data processing and management**,** ensuring seamless integration and operational efficiency. **Certified Data and ML Engineer in GCP and Databricks.**

# EMPLOYMENT HISTORY

## ML Engineer, Self-Employed in Pragnatyx Ltd. Toronto, ON

July. 2021 – Present

**Client:** Transportation; Ford Motor Company **Duration:** Jan 2023 - Present

Description**:** Developed an AI-driven event orchestration agent on GCP using Vertex AI and LangChain, replacing hundreds of rules within a Drools engine. This solution efficiently analyzes incoming application and system events, evaluates various data conditions and validations, and incorporates persona-based role-playing to determine the next action. Based on these insights, it intelligently invokes the appropriate microservice or validation process. As a result, development time was reduced by 50%, manual effort by 60%, and event processing efficiency improved by 35%.

Developed an enterprise RAG solution to automate data enrichment, a task previously bottlenecking hundreds of data stewards. The pipeline analyzes 100,000 existing views to retrieve relevant patterns and uses an LLM to generate correct, country-specific join conditions on the fly. This automation cut manual effort by 90% and successfully enabled a complete, zero-touch data processing workflow.

Role**:** Responsible for the architecture, design, development, and optimization of ML pipelines, ensuring robust data processing, RAG integration, and model deployment.

**Environment:** GCP Agent Development Kit (ADK), Vertex AI, Gemini, LangChain, Cloud Run

**Client:** Transportation; Ford Motor Company **Duration:** July 2021 – Jan 2023

Description**:** De-Identification (Deid) strategy is one of the main streams in Enterprise Data Lake platform delivery. Deid is a pivotal factor for democratizing data access to End Consumers and downstream applications. Developed batch and streaming pipelines for Voltage based encryption inside GCP using Dataflow; Streaming records peaked at 30K records per second and thousands of batch jobs running daily encrypting over billions of records in BQ including complex and nested columns. Developed dynamic BQ masking pipelines addressing the scale, latency, cost challenges in Voltage based pipelines. Developed a 3- tier web application for internal Users to override metadata driven generated logic for Deid entities. Implemented de-identification methods such as masking, tokenization, and generalization to protect PII while maintaining data utility.

Role**:** Responsible for architecture, design, development and optimization of de-identification pipelines.

Environment**:** GCP, GCS, BigQuery, DataFlow, Data Catalog, Dataplex, Cloud Run, Cloud Functions, Pub/Sub, Cloud SQL, Dataproc, IAM, Terraform, Java

## Sr. Solutions Architect, Amazon Web Services. Toronto, Canada

June. 2020 - June. 2021

**Client:** Financial Services; Large Canadian Bank

**Duration:** June 2020 - June 2021

**Description:** To help the customer achieve better business outcomes by building on AWS cloud platform. The goal is to deliver various technology solutions to multiple Lines of Businesses including Capital Markets, Wealth Management, Personal & Commercial Banking, Global Asset Management, Ventures and help their central IT teams including Product, Platform, Engineering, Cyber Security teams to build scalable and secure platforms. I deliveredscalable, and secure cloud architectures for Capital Markets, proactively addressing the client's anticipated surge in data processing needs within the next two years – driven by the post-COVID increase in the use of Capital Market products – by leveraging services such as EC2, EKS, and EMR

**Role:** My responsibilities included devising and executing strategies for cloud-based solutions, as well as designing and implementing highly scalable and cost-effective cloud architectures.

**Environment:** AWS, EC2, S3, EKS, EMR, Apache Spark, Glue, LakeFormation, Redshift, Dynamo DB, Lambda, SageMaker, CloudFormation, Terraform, Comprehend, Textract, API Gateway

## Solutions Architect, Deloitte. Multi City, India, Canada, US

August. 2014 - June. 2020

**Client:** Large Canadian Bank

**Duration:** Feb 2020 - June 2020 (Deloitte US)

**Project Description:** The goal of this Pilot was to test the Azure platform to deploy APIs securely for the Enterprise Data Platform. The APIs facilitated data ingestion, metadata management, and dynamic schema mapping, enabling seamless integration of diverse enterprise data sources and ensuring standardized data accessibility across the organization.

**Role:** Responsible for architecture, design, development and optimization of metadata-driven automated data pipelines.

**Environment:** Azure, Azure API Management, Azure Functions, Azure App Service, Java

**Client:** Financial Services; Large Canadian Bank & Google

**Duration:** Aug 2018 – Jan 2020 (Google Cardinal)

**Project Description:** Client selected Google to suggest the best practices for architecture, design, and development of Data Lake. The goal is to develop de-identification of sensitive data within structured datasets before persisting it to a technically standardized zone.

**Role:** Collaborated with Google Customer Engineer and Ski. Designed and executed PoC's to test Dataflow scalability for encrypted files. Suggested alternative approaches to build the data pipeline and contributed to its overall design. Led a multi-vendor development team to deliver event-based orchestration, audit framework, multiple dashboards, HA/DR, multitenant solution. I was responsible for architecture, design, development, and post production support.

**Environment:** GCP, Dataflow, Stackdriver, IAM, Deployment Manager, BigQuery, Cloud Functions, PubSub, Google Cloud Storage, Data Studio, Compute VM, Load Balancer, Java, NodeJS, Python

**Client:** Financial Services; Large Canadian Bank

**Duration:** Oct 2016 – May 2018 (Deloitte Canada)

**Description:** The goal was to design and develop a metadata drive ingestion framework for Data Lake by reducing the overall cost of ingesting data while implementing leading data management practices for metadata management, data lineage, data quality etc.

**Role:** Responsible for architecture and low-level design. Led multi-vendor development team for implementation of complex files, relational databases, and mainframes file ingestion, including data verification and integrity checks across all ingestion patterns. Optimized Spark and Hive jobs on the multi-tenant cluster.

**Environment:** CDH, Spark, Hive, Sqoop, Cloudera Navigator, Mainframe, Oracle, Oozie, Scala, Python, Java, MapForce, SyncSort

**Client:** Financial Services; Large Canadian Bank

**Duration:** Sep 2015 – Oct 2016 (Deloitte Canada)

**Project Description:** The goal was to enhance analytics capabilities to address key business questions. The scope of assignment included delivering customer journey analytics, first time home buyers well as innovative applications of graph analysis on Hadoop in customer segmentation and fraud detection. This involved utilizing Spark for scalable machine learning models, including linear regression and clustering techniques, to process and analyze large datasets effectively.

**Role:** Responsible for collaborating with Lead Data Scientist to understand requirements of creations of new features and models. Responsible for feature creations using Hive. Optimized Hive queries using best practices, increasing container sizes dynamically for large queries. Created linear regression models using Spark ML.

**Environment:** CDH, Hive, Spark, R, RevR, Scala, Spark ML

## Team Lead, IBM. Multi City, India, Spain, US

August. 2004 - August. 2014

**Client:** Telecommunication; Large European Company

**Duration:** Feb 2008 – Aug 2014

**Project Description:** The project involved a COTS product Clarify widely used in Telecom Industry for CRM. It mainly caters to the Provisioning and Customer Support area in accordance with eTOM framework.

**Role:** Team Lead

**Environment:** Clarify 10.1, C, Tuxedo, Tibco, Sybase

**Client:** Energy & Utilities; Large US Company

**Duration:** Sep 2004 – Feb 2008

**Project Description:** CIS was a mission critical, highly integrated application. It supported Usage calculation, Billing, Cash processing, Financial transactions, Order entry, scheduling and completion, Collections, and Assistance programs in a client-server environment.

**Role:** Team Lead/Developer

**Environment:** Shell Scripts, C, MFC, Win32 API, DB2

# EDUCATION

## National Institute of Technology, Kurukshetra

B.Tech, Electronics & Communication, June. 2004

# CERTIFICATIONS

* GCP Professional ML Engineer
* GCP Professional Data Engineer
* Databricks Data Engineer Associate
* Databricks Data Engineer Professional
* Databricks Generative AI Associate
* Databricks Machine Learning Associate

# SKILLS

* Programming Language:
  + Java, Scala, Python, Shell Scripting
* Cloud Platforms:
  + Databricks, Google Cloud Platform, AWS Cloud, Azure
* ML Stack:
  + MLflow, Vertex AI, LangChain, SparkML, Tensorflow
* Hadoop Distributions:
  + GCP Dataproc, AWS EMR, Cloudera, Hortonworks
* Distributed Processing Framework:
  + Apache Beam, GCP Dataflow, Apache Spark, Spark Streaming, Pubsub, Mapreduce
* Data Warehouse
  + BigQuery, Redshift
* Infrastructure as Code:
  + Terraform
* NoSql:
  + HBase, BigTable, Dynamo DB
* Schedulers:
  + Airflow, Autosys, Oozie
* Serverless Functions:
  + Cloud Run, Cloud Functions, Lambda