

# TIRTH SHAH

Toronto, ON • +1 (647) – 518 – 1497 • tirthsh0@gmail.com

## SKILLS

### Tools:

- Kafka, Debezium, Confluent, Airflow, Argo Workflows, Spark, AWS Pydeeque, Deltastreamer, Hudi, GCP Dataproc, Hadoop, Datadog, Jenkins, CircleCI, Helios, Docker, K8s, Helm, Ansible, GCP, AWS Redshift, SQL, JDBC, DynamoDB

### Concepts:

- ETL, Data Quality, Metadata, Neural Nets, Blockchain, Project Management, CI/CD, Agile/SDLC, UML Diagrams, OAuth, SSO, Design Patterns

## PUBLICATIONS

### (Medium)

- Getting Started with Apache Airflow
- Recognizing Presence of Private Content in Images via ML model
- Intro to Neural Nets

## EDUCATION

University of Toronto (U of T)

### Honors Bachelor of Science (June 2020)

- Computer Science, specialized in Software Engineering

## PORTFOLIO

- tirthshah.info
- github.com/tirthsh
- medium.com/@tirthsh0

## WORK EXPERIENCES (TIRTHSHAH.INFO/EXPERIENCE)

### SENIOR DATA ENGINEER

• Dec 2024 – Present

**Royal Bank of Canada (RBC)** • 88 Queens Quay W, Toronto, ON

- Designed and implemented **metadata-driven** ingestion framework using **PySpark**, **Hadoop** and **Pydantic/Yaml** models, configs, standardizing onboarding of hundreds of ESG datasets
- Built reusable **AWS Pydeeque-based Data Quality** platform and automated CDR / OSFI climate and BCAR reporting pipelines across the Climate Solutions team
- Drove engineering best practices (Github CI/CD, reusable libraries, documentation, demos), significantly improving development quality

### DATA ENGINEER

• June 2022 – Sept 2024

**theScore BET / ESPN BET** • 125 Queens Quay E, Toronto, ON

- Designed and implemented various **API** endpoints for an in-house real-time data warehouse tool, capable of ingesting batch and stream data via **Confluent** Connectors for CDC, **Kafka** topics, processed by Apache **Spark** on a GCP **Dataproc** cluster and stored as metadata on AWS **S3** which in return queryable as AWS **Redshift** External Tables to help generate and deliver daily regulatory reports to various stakeholders
- Implemented Application monitoring via **DataDog** monitors and alerting on **Slack** and **Pagerduty** of various components of the project to ensure healthy, stable production environment
- Used **Terraform** to spin up Redshift clusters, S3 buckets, AWS EC2 instances, AWS roles/users in lower environments to assist in testing

### DATA ENGINEER

• March 2020 – June 2022

**Royal Bank of Canada (RBC) Capital Markets** • 200 Bay St, Toronto, ON

- Designed, built a CI / CD framework (completely dockerized) - **automated deployment** of a **PySpark** model which generates profit and loss on trading data stored on **Hadoop** cluster, via DevOps tools – **K8s**, **Helm**, and **Ansible**

### INTERNSHIPS

- Software Developer (RBC) | Full Stack Developer (Manulife) | DevOps Engineer (TD)

## PROJECTS (TIRTHSHAH.INFO/PROJECTS)

- Simulated cryptographic attacks via **Docker** ; pushed boundaries in software vulnerability, network security & applied cryptography; attacks included Penetration testing, SQL Injection authentication, SSL stripping, ARP spoofing and packet sniffing using **Wireshark** and in-built network packages in **Python**
- Designed, developed, deployed restaurant rating app which incorporated user login system, backend database storage using **MongoDB**, interactive functionalities using **NodeJS** and Google Maps **API** with security embedded using browser cookies and sessions with Google **OAuth** using passport library; hosted app on **Heroku**