**AISHWARYA S. WAGH**

+1 519-781-1684 |[aishwaryaswagh@gmail.com](mailto:aishwaryaswagh@gmail.com)|[LinkedIn profile](https://www.linkedin.com/in/wagh-aishwarya/) |Toronto, Ontario, Canada

**SUMMARY**

* Self-motivated and result-driven engineering professional with over 5 years of rich experience in software design and development, data engineering.
* Demonstrated strong problem-solving abilities and effective communication through extensive work on customer-facing applications and diverse projects.

**PROFESSIONAL EXPERIENCE**

**Royal Bank of Canada, Toronto, ON Jun 2022 - Present**Senior Data Engineer

* Orchestrated and implemented **ETL / ELT** pipelines in **Python** using **Airflow** and **RunDeck** to ingest, clean and transform large volumes of real time and historical data from various sources.
* Fulfilled data migration requirements using **Python, SQL, DBT** ensuring timely delivery of accurate results. Proactively provided validation scripts to ensure data integrity.
* Designed Self-Serve tool using **Spark** and **Scala** for entity resolution and data enrichment that **reduced communication overhead** and saved **8 hours per week** for each request.

**Tata Consultancy Services, Mumbai, India Jun 2017- Mar 2021**  
Software Engineer Data platform

* Designed and maintained **Java** web applications for multi commodity exchange and trading platform at **National commodities and Derivatives Exchange e-Marketing Limited**. Contributed to requirement gathering, solution design, development, and enhancements in the application while working in an **Agile** Environment.
* Developed scalable distributed applications using Big Data Technologies such as **Apache Spark, Hive** to replace existing ETL processes for electricity trading sector clients which improved the processing speed by ~42%.
* Translate business propositions into quantitative queries and collect, clean the necessary data.

**TECHNICAL SKILLS**

**Programming languages:** Scala, Java, Python, R

**Big data technologies:** Hadoop, HDFS, Apache Spark, MapReduce, Hive, Kafka, PySpark, Snowflake, Data Build Tool (DBT)

**Databases:** Microsoft SQL Server, MySQL, PostgreSQL, Cassandra

**Data orchestration and workflow management:** Airflow, Kubernetes, Docker, RunDeck, Dagster, OpenShift Container Platform OCP 4, Zookeeper.

**Others:** Microsoft Excel, GitHub, GitHub Actions, Linux, Data modelling, ETL, Parquet, Avro, YAML, SageMaker, Helios (CI/CD)

**PROJECTS**

**Enterprise Client - Merchant Resolution**

* Developed and implemented **ETL** pipelines in **Scala** for the ingestion of daily transaction data, optimizing the resolution and enrichment of merchant and client information. Leveraged **Spark** for improving scalability and performance optimization ensuring efficient data processing.
* Utilized **Apache Airflow** for orchestrating and managing the ETL workflows, ensuring reliability and scheduling flexibility.
* This significantly impacted **decision-making** processes by providing accurate & comprehensive insights into merchant identities, brands, and locations, thus enabling informed **strategic partnerships** and targeted promotional **campaigns**.

**Data Migration**

* Migrated **Teradata** based data warehousing infrastructure to **Snowflake** **cloud** data platform.
* Oversaw the end-to-end migration process, including schema conversion, data extraction, transformation, and loading, leveraging Helios for **CI/CD** automation.
* Leveraged expertise in **DBT**, **SQL** and cloud data technologies to ensure seamless transition, optimizing performance, scalability, and cost efficiency in the new Snowflake environment.

**Fraud Guardian**

* Incoming transactions are compared with historical trends for the relevant client and others showing similar patterns.
* Utilized **AWS S3** to store historical transaction data securely. Implemented a near real-time fraud detection system employing **lambda** **architecture**, leveraging **Kafka** streaming for incoming transactions. Leveraged **AWS SageMaker** for advanced machine learning models to enhance fraud detection accuracy.
* Utilized **SQL**-based rules to identify potentially fraudulent activities.
* Incorporated a dynamic rule modification feature for on-the-fly adjustments.

**Automated Twitter response system**

* Application replies to tweets using natural language processing within 5 seconds. It captures live tweets, funnels them into a **Kafka** topic, undergoes processing via a **Scala ETL pipeline**, and interfaces with an **NLP** model to craft responses.

**EDUCATION**

**Master of engineering (ECE) | University of waterloo| waterloo Jan 2021 - May 2022**

**Relevant coursework:** Distributed and network centric computing, Data & knowledge modelling and Analysis, Tools of Intelligent Systems Design, Statistics for Data Analytics, Methods and tools for software engineering, Foundations of software engineering.

**Bachelor of engineering (ECE) |Government College of engineering |India Aug 2014 - Jun 2017**

**Relevant Coursework:** Discrete mathematics, Object oriented programming, Operating systems, Mathematical Statistics.