VAIBHAV SEHGAL

+1(438) 928-1730 | Canada | vaibhavsehgaljobs@gmail.com | [LinkedIn](#) | [Github](#) | [My Portfolio](#)

EDUCATION

Master of Applied Computer Science, Concordia University, 3.7 GPA   Sept 2022 - June 2024
Relevant Coursework: Big Data Analytics, Algorithm Design Techniques, Artificial Intelligence, Data Structures

Bachelor of Technology, Computer Science & Engineering,
Guru Gobind Singh Indraprastha University, 3.9 GPA   Aug 2017 - July 2021

SKILLS

Languages	Python, Scala, Java, SQL, HTML, CSS, Javascript
Libraries	Pandas, Numpy, Matplotlib, Scikit-Learn, NLP, Tensorflow, Keras, Pytorch
Big Data	Apache Spark, Hadoop, Airflow, Hive, Kafka, Snowflake, Databricks, Delta Lake, Talend, Informatica
Cloud	AWS(S3, Lambda, Glue, Redshift), Azure, Jenkins, PowerBI, Docker, Confluence, Data Modelling

EXPERIENCE

Data Engineer Intern - AI   Sept 2023 - May 2024
Intact Financial Corporation   *Montreal, Quebec*

- Developed multiple data pipelines for ingesting Personal Lines, Commercial Lines Pega data and SAS tables as part of the Cloud data enablement project.
- Developed an automated Snowflake-based ETL pipeline for Oracle to EDP migration, utilizing the SAS-Snowflake connector to manage large-scale migrations with advanced transformation logic, boosting efficiency.
- Created stored procedures and dynamic views in Snowflake for advanced querying and user-specific data insights.
- Crafted tag-based dynamic masking policies to provide controlled access to the users as part of the Data Governance initiative to implement role based controls.
- Streamlined CI/CD by creating a unified Jenkinsfile for multiple environments and configurations, improving efficiency and accelerating deployment with new pipeline stages and deployable objects.
- Contributed to PL project by implementing advanced logic for auditing and rollback capabilities for AWS S3 and audit tables through Delta Live Tables, ensuring precise and dynamic audit management.

Data Engineer   Oct 2021 - Aug 2022
Evalueserve   *Gurugram, Haryana*

- Engineered an automated batch ETL pipeline for transforming and aggregating data to a central hub(AWS S3), utilizing time-based policies for on-demand target-level processing.
- Developed a versatile PySpark script complemented by a nested configuration file, enabling dynamic handling of diverse data sources, formats, and time-window aggregations.
- Leveraged AWS Glue for script scheduling on the client's infrastructure, achieving a 30% reduction in manual processes for the project.

Data Analyst Intern   Aug 2021 - Oct 2021
SpreadIt   *New Delhi, India*

- Optimized data consistency and integrity with advanced data cleansing techniques.
- Analyzed data and delivered actionable insights via Power BI dashboards and reports, driving growth.
- Identified and resolved root causes using SQL and Python, resulting in significant cost savings.

PROJECTS

Movie Recommendation System ↗ - Python, Pandas, Scikit Learn, Matplotlib, EDA, Surprise, Plotly
Created a precise movie recommendation system using Numpy, Pandas, Scikit-learn's Tfidf Vectorizer, and collaborative algorithms like SVD/ALS for enhanced suggestion accuracy.

Flight Tracker ↗ - Angular, Javascript, HTML, CSS, MongoDB
Designed a real-time flight tracking web application using MVC architecture, integrating Table Data Gateway and Data Mapper patterns for optimized data retrieval and manipulation.