|  |
| --- |
| **VISHAL**  **Data ENGINEER**    +1 (647)-365-2019        vishal.singh.uv.20@gmail.com          Toronto, ON |

**PROFESSIONAL SUMMARY**

* Around **5 years** of experience in Data Engineering skills and the ability to analyze **complex SQL** and **Python data sets**.
* Proficient in data visualization tools such as **Tableau** and **Power BI** to create insightful dashboards and reports.
* Experience in **data cleansing, transformation**, and **preparation** to ensure data quality and integrity.
* Skilled in using **statistical methods** and **machine learning algorithms** to analyze and interpret data.
* Ability to collaborate with cross-functional teams and communicate effectively with stakeholders.
* Experienced in **data modelling** and **warehousing** concepts using tools such as **Redshift** and **Snowflake**.
* Experienced **data engineer** with a passion for developing and maintaining complex data systems using **Hadoop**, **Spark**, and other big data technologies.
* Proven track record of designing, building, and optimizing large-scale **data pipelines** and processing frameworks to meet business needs.
* Proficient in programming languages such as **Python**, **Java**, and **SQL**, with extensive experience in **ETL** tools such as **Talend**, **Apache NiFi,** and **StreamSets**.
* Expertise in **data warehousing** and **data modelling** concepts and working with **SQL** and **NoSQL** databases such as **MySQL**, **MongoDB**, and **Cassandra**.
* Proficient in developing and maintaining **data visualization** dashboards to provide insights into data using tools like **Tableau** and **Power** BI.
* Knowledgeable in cloud-based data platforms such as **AWS**, **Azure**, and **GCP**, and experience working with related services like **S3, EC2, and EMR**.
* Skilled in implementing **data quality** checks to ensure **data accuracy** and **completeness** and detecting and resolving **data anomalies** and **inconsistencies**.
* Experience working with **machine learning models** and deploying them to production using tools like **TensorFlow**, **PyTorch**, and **MLflow**.
* Knowledge of **Agile** development methodologies, **DevOps** practices, and **CI/CD** pipelines to ensure efficient and scalable data engineering solutions.
* Experience in conducting **A/B testing** and **statistical analysis** to evaluate the effectiveness of marketing campaigns.
* Strong problem-solving skills and ability to think creatively to develop data solutions.
* Familiarity with data security and privacy measures to ensure compliance with regulations.
* Ability to work independently and manage multiple projects simultaneously.
* Experience in conducting **data quality assessments** to identify and resolve data errors.
* Skilled in using **R** and **Excel** for **predictive modelling** and **forecasting**.

**Skills**

|  |  |  |
| --- | --- | --- |
| * Python |  | * SQL |
| * NoSQL |  | * Tableau |
| * Data Modelling |  | * Data Warehousing |
| * Power BI |  | * AWS |
| * Spark |  | * Hadoop |

**Experience**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Shape  Description automatically generated | **senior Data Engineer**, CIBC Bank   |  |  |  | | --- | --- | --- | | Toronto, ON |  | *January 2022 - Present* |      |  |  |  | | --- | --- | --- | | * Designed and implemented data ingestion pipelines using **Hadoop** and **Spark** to extract, transform, and load data from various sources, including structured, semi-structured, and unstructured data. |  | * Developed and maintained data models and architectures using **SQL** and **NoSQL** databases such as **MySQL**, **MongoDB**, and **Cassandra**, and optimized query performance using indexing and partitioning techniques. | | * Implemented **data processing** and **streaming** **pipelines** using **Kafka** to enable real-time data processing and analytics and designed and implemented custom Kafka connectors to integrate with other systems. |  | * Developed and maintained custom **Python** libraries and packages for **data processing** and **machine learning**, and deployed machine learning models to production using **Python** and **Spark**. | | * Designed and developed custom **Power BI** and **Tableau** dashboards to provide real-time insights and visualizations and integrated them with the data processing pipelines using **REST APIs** and webhooks. |  | * Designed and implemented **data warehousing** systems using **Snowflake, Redshift,** and **BigQuery**, and optimized query performance using techniques such as star schema design and query optimization. | | * Deployed data processing and analytics systems on **AWS** and **Azure** using services such as EMR, HDInsight, and Data Factory, and implemented security measures such as VPCs, NSGs, and RBAC. |  | * Conducted **data modelling** and **schema design** using **ER diagrams** and **UML** and implemented schema evolution and data partitioning to enable scalability and flexibility. | | * Conducted data profiling and analysis using tools such as **Pandas**, **Dask**, and **PySpark**, and developed custom **Python scripts** and **SQL** queries to automate data analysis and reporting. |  | * Developed and maintained **disaster recovery** and **business continuity plans** for data processing and analytics systems and conducted regular disaster recovery tests and simulations to ensure system availability and readiness. | | * Implemented change data capture (CDC) mechanisms to capture and process real-time data changes, improving data accuracy and timeliness. |  | * Developed and maintained a **data lake architecture** to store and process large volumes of data, enabling efficient data storage and processing. | | * Implemented data partitioning techniques to improve data processing efficiency and reduce query response times. |  | * Developed and maintained **data pipelines** to integrate third-party data sources, improving data completeness and providing additional insights for business stakeholders. | | * Developed and maintained a **data visualization** platform to provide real-time insights into key performance metrics, improving decision-making processes for business stakeholders. |  | * Conducted **data profiling** and **analysis** to identify and prevent fraudulent activities, resulting in a reduction in fraudulent transactions. | | * Developed and maintained a **disaster recovery plan** to ensure data availability in case of system failure, ensuring business continuity. |  | * Conducted **performance tuning** and **optimization** of data pipelines, resulting in a reduction in processing times and improved overall system performance. | | * Provided technical guidance and mentorship to junior data engineers, ensuring best practices and high-quality code standards are maintained. |  | * Worked with business stakeholders to identify and prioritize data needs, resulting in the development of data-driven solutions to address key business challenges. . | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Data Engineer,**  RBC Bank   |  |  |  | | --- | --- | --- | | Toronto, ON |  | *May 2019 – December 2021* |      |  |  |  | | --- | --- | --- | | * Conducted data analysis using **SQL** and **Excel** to support business operations and developed and maintained dashboards and reports to provide insights to stakeholders. |  | * Conducted **data quality assessments** to identify and resolve data quality issues, resulting in a reduction in data errors. | | * Developed and maintained **predictive models** using **R** and **Excel** to forecast trends and identify patterns. |  | * Conducted **A/B testing** and statistical analysis to evaluate the effectiveness of marketing campaigns, resulting in an increase in conversion rates. | | * Developed and maintained data visualization dashboards using **Tableau** and **Excel** to provide real-time insights into key performance metrics, improving decision-making processes for business stakeholders. |  | * Developed and maintained **data pipelines** using **SQL** and **Excel** to extract, transform, and load data from various sources. Conducted ad-hoc data analysis and reporting to support business operations and identify areas for improvement. | | * Worked with business stakeholders to identify key metrics and develop custom reports and dashboards to track business performance. |  | * Conducted **data quality assessments** and implement **data cleansing** and **transformation** techniques to improve data accuracy and completeness. | | * Developed **predictive models** using **R** and **Excel** to forecast trends and identify patterns and present findings to business stakeholders. |  | * Developed and maintain **data visualization dashboards** using **Tableau** and **Excel** to provide real-time insights into key performance metrics, enabling stakeholders to make data-driven decisions. | | * Worked with cross-functional teams to identify business requirements and develop data solutions to address those needs. |  | * Developed and maintained **data dictionarie**s and data documentation to ensure consistency and accuracy across different data sources. | | * Stayed up to date with the latest data analysis techniques and tools, and continually seek to improve data analysis and reporting processes. |  | * Developed and maintained automated reporting processes using **VBA** and other tools to streamline reporting workflows. | | * Conducted root cause analysis to identify the underlying causes of business problems and develop recommendations to address those issues. |  | * Provided regular updates and presentations on data insights and trends to business stakeholders and communicate data findings to non-technical audiences. | | * Conducted **data segmentation and clustering** to identify customer segments and develop targeted marketing campaigns. |  | * Analyzed and interpreted data to identify opportunities for process improvement, cost reduction, and revenue growth. | |

**Education**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | **Bachelor of Engineering in Computer Science** |  |  |  |  |  |  | | --- | --- | --- | | Uttar Pradesh Technical University, Uttar Pradesh, India |  |  | |
|  |  |
|  |  |

**Academic projectS**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  |  |  |  |  | | --- | --- | --- | | Predictive analytics for customer churn: |  |  | |
|  |

* Analyzed customer data for a telecom company to identify customers who are likely to churn.
* Used techniques such as **logistic regression, decision trees, or random forests** to build a predictive model.
* Evaluated the model's performance using metrics such as **accuracy**, **precision**, **recall**, and **F1 score**.
* The project required **data** **collection**, **cleaning**, **preprocessing**, and **analysis** skills.
* It also required knowledge of **machine** **learning** and **statistical** **techniques**.
* Programming skills in languages such as **Python**, **R, or Java** are needed, as well as proficiency in **data analysis** and **visualization tools.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  |  |  |  |  | | --- | --- | --- | | Sentiment Analysis of Social Media Data Using NLP and Machine Learning: |  |  | |
|  |

* Collected data from social media platforms such as Twitter and analyzed it to determine the sentiment of users towards a particular topic.
* Used techniques such as **natural language processing (NLP), machine learning**, and **text analytics** to classify tweets or posts as positive, negative, or neutral.
* Visualized the sentiment distribution using tools such as **Matplotlib** or **Tableau**.
* Theprojectrequired **data collection, cleaning, preprocessing, and analysis skills.**
* It also required knowledge of **NLP, machine learning, and statistical techniques**.
* Programming skills in languages such as **Python, R, or Java** are needed, as well as proficiency in **data analysis** and **visualization tools.**