Homa Zadeh

778-232-9525 | homayoonbanazadeh@gmail.com | linkedin.com/in/homayounbanazadeh | github.com/homayoo-n

Summary

- Proficient in designing and implementing robust data pipelines using Azure Data Factory, Databricks, and Apache Spark, with expertise in Scala and SQL.
- Experienced in migrating and optimizing legacy warehouse pipelines to modern cloud-based solutions to enhance performance improvements.
- Skilled in data modeling and schema design for effective storage and retrieval of structured and unstructured data in Azure SQL Database, Cosmos DB, and Hadoop.
- Capable of leveraging advanced data processing techniques such as partitioning strategies, indexing, and windowing functions to optimize query performance and enhance data processing efficiency.

Education

Simon Fraser University Sep. 2019 – Dec. 2023

Bachelor of Science in Data Science, Minor in Computing Science *Burnaby, BC*

- Dean's Honor List, SFU Transfer Scholarship, Peer Education Volunteering Group Manager

Certification: Microsoft Azure Data Engineering Oct. 2023 – Dec. 2023

Experience

Freelance Data Engineer Aug 2022 – Present

DrJobCanada *Vancouver, BC*

- Implemented efficient file formats (Parquet, JSON) and access tiers (Cool, Hot) in Blob Storage Gen2, resulting in a 20% reduction in storage costs for the client.
- Designed star schema structures in Azure Synapse with effective partitioning strategies, achieving a 30% improvement in query performance through optimization.
- Implemented robust authentication mechanisms using Active Directory and encryption techniques, ensuring compliance with industry security standards and safeguarding sensitive data.

Software Analyst Sept. 2021 – April 2022

Western Stevedoring *North Vancouver, BC*

- Conducted software research and gathered stakeholder requirements over interviews, which led to producing a report that directed the company's decision to invest $10,000 annually.
- Developed a Python-based automated data pipeline tailored to handle monthly financial Excel files, which were often irregularly formatted and required thorough scripting due to loose filling guidelines.
- Implemented an automated approval system for vehicle permits using Power Automate. Leveraging Microsoft Forms, designed a process for truckers to submit information and certificates.
- Led training sessions and authored a 20-page documentation covering maintenance, risks, and challenges.

Data Engineer Sep. 2020 – May 2021

Insurance Corporation of British Columbia(ICBC) *North Vancouver, BC*

- Migrated legacy warehouse pipelines in IBM Data Stage to a modern Hadoop implementation using Spark and Scala for a 60% improvement in the speed of data memory access.
- Authored 10 data mapping documents pivotal for the seamless data migration from legacy systems to Hadoop, enabling the engineering team to grasp complex data transformations without delving into code analysis.
- Processed extensive SQL logic containing hundreds of lines to develop new transformations in Spark and translated the SQL scripts into various dialects.

Application Developer Jan. 2017 – Jan. 2018

Snapp *Tajrish, Tehran*

- Developed and maintained a database infrastructure using MS SQL Server, integrating it with MVC frameworks for efficient data management in web applications, while ensuring version control with Git.
- Implemented user interfaces with HTML5, CSS3, and JavaScript, ensuring optimal experience across devices, and leveraging Bootstrap for responsive design.
- Utilized React.js to create UI components, enhancing code maintainability and scalability in web applications, while adhering to RESTful API standards for communication with back-end systems.

Projects

LSTM-Based Stream-flow Forecaster for Drought Prediction | *PyTorch, Pandas* March 2024 – Present

- Developed a notebook to explore linear regression and LSTM models for stream-flow prediction, investigating various modeling approaches, including predicting the next day's data and computing the monthly index monthly versus prediction of the daily index
- Provides training procedures with the Adam optimizer and evaluates metrics such as Nash-Sutcliffe Efficiency.
- Utilizes the CAMELS dataset spanning 35 years, incorporating daily precipitation and temperature.

Emotional Voice Recognizer | *HCI, Feature extraction* Feb. 2022 – Apr. 2022

- Developed a method with 74% accuracy in sentiment by utilizing YouTube video data for specific phrases.
- Curated a dataset of 211 audio files sourced from YouTube, manually annotated to ensure accuracy, and applied noise reduction techniques for improved data quality.
- Tested SVM and Decision Tree classifiers for sentiment prediction, alongside investigating the effectiveness of MFCC features and pitch, resulting in finding the best model.

Package Manager | *Java, Client/Server model* Aug. 2022 – Oct. 2022

- Built An inventory management application to add, save, load, and remove various package types.
- UI was made using Java Swing with sections for adding/removing packages and viewing overdue packages.
- Employs Factory and Adapter design patterns for handling packages efficiently to ensure that packages are created and managed consistently while preventing unnecessary resource duplication.
- Data is saved on server side as a JSON list. Server side was made with Spring Boot framework. Client code sends and receives data using HTTPS requests with CURL commands and API calls.

Yearly Canadian Temperature Visualizer Shiny App | *Shiny, R* Nov. 2019 – Jan. 2020

- Developed a Shiny app that visualizes and examines the relationships between Canadian temperature, CO2, and Wind Speed data over last 2 decades.
- Includes a time-series analysis of temperature data, and regression analysis between CO2 and temperature with the capability to visualize multiple aspects of these relationships.

Twitter Sentiment Analyzer | *NLP, API Access* Feb. 2019 – April 2019

- Utilized R programming language and twitteR, tidytext, and ggplot2 packages to extract real-time Twitter data.
- Interacted with the Twitter API to access Twitter data for analysis.
- Conducted sentiment analysis on tweets using natural language processing techniques.
- Employed visualization methods like word clouds to analyze tweet content and sentiment trends.
- Conducted sentiment predictions on individual tweets, contributing to comprehensive data analysis.

Technical Skills

Languages: Java, Python, C/C++, SQL (Postgres), JavaScript, HTML/CSS, R
Frameworks: React, Node.js, Flask, JUnit, WordPress, Material-UI, FastAPI
Developer Tools: Microsoft Power BI, Git, IBM Data Stage, Visual Studio Code, Microsoft SQL Server Management Studio, Azure Data Factory, Databricks, Spark, Scala, React.js
Libraries: Power BI query editor, DAX, Python, Pandas, NumPy, PyTorch, Keras