Sanka Weerathunga

Waterloo, ON |sankaw.com

sdweerathunga5@gmail.com, +1 778-677-8608, linkedin.com/in/sanka-weerathunga, github.com/sankaw

Dedicated, curious, and continually evolving data engineer and analytics professional with AWS certifications (Data Analytics - Specialty and Solutions Architect - Associate), seeking an opportunity to leverage expertise in data science, data-driven solutions, architecture, visualization, and data engineering to drive organizational transformation.

HIGHLIGHTS OF QUALIFICATIONS

- Over 5 years of experience in data engineering and data analytics, with a focus on building robust data pipelines.
- Skilled in delivering high-impact data visualizations, processing data in both real-time and batch modes.
- Strong understanding of modern big data architecture, data privacy, and design principles.
- Expertise in handling structured, semi-structured, and unstructured data as well as data warehousing.
- Experienced in collaborating with cross-functional teams to enhance data flow, implement governance and privacy measures to ensure data quality.

PROFESSIONAL EXPERIENCE

Senior Data Engineer, Axiata Digital Labs, Colombo, Sri Lanka   **Nov 2021 - Dec 2022**
Data Engineer, Axiata Digital Labs, Colombo, Sri Lanka   **Jan 2019 - Oct 2021**
Axiata Digital Labs, a leading technology solutions provider specializing in digital transformation services, is a subsidiary of Axiata Group, a major telecommunications conglomerate based in Malaysia. There, I performed the following tasks:
- Built flexible, automated end-to-end data pipelines on the Hadoop platform, leveraging real-time Kafka feeds and database development processes.
- Designed and implemented robust data structures to maximize the capabilities of cloud computing platforms, enriching data ingestion and integration strategies.
- Led and delivered an award-winning real-time sales operational monitoring dashboard system using Tableau, demonstrating strong analytical skills and improving decision-making efficiency.
- Drove quality management initiatives across business areas, achieving up to 99% data pipeline accuracy through daily monitoring processes while maintaining SLA.
- Performed ad hoc analyses, including sales impact assessments during disaster situations.

Data Engineer, Dialog Axiata PLC, Colombo, Sri Lanka   **Jul 2018 - Dec 2018**
Dialog Axiata PLC is the largest telecommunication provider in Sri Lanka, another subsidiary of Axiata Group. There, I accomplished the following tasks:
- Performed data preprocessing for the data modeling team, generating datasets and database views for cross-functional teams to support diverse business functions and effective data management.
- Optimized complex SQL queries by analyzing execution plans, reducing processing time from 2 hours to 5 minutes, and validated data with other sources to maintain pipeline reliability.
- Employed Agile methodologies, including JIRA, for project management and collaboration.

Software Engineer Co-op, **Undergraduate Program**   **Jul 2017 - Dec 2017**
Asoft (Pvt) Ltd, Colombo, Sri Lanka - Software Development Company
- Developed a full-stack web application for inventory management and finance system using MVC architecture, database schema for data integrity, and created dynamic Crystal Reports for insightful data visualization.
- Contributed to the entire development process, from conceptualization to deployment, collaborating with team members to ensure a user-friendly and efficient application.

E-commerce Grocery Associate   **Sep 2022 - Present**

Walmart, Waterloo, ON (The Largest Retailer in Canada)

- Collaborated with diverse teams to efficiently complete tasks, ensuring timely and accurate outcomes while demonstrating strong teamwork in a high-pressure environment.
- Communicated with customers to address inquiries and resolve issues, improving customer satisfaction and loyalty through exceptional service and problem-solving skills. Won the best employee of the month.

PROJECTS

Academic Projects

Real-Time News Data Processing Pipeline (Conestoga College, Canada)   2024

- Built a real-time pipeline using Kafka, Spark, and Spark ML for data ingestion, processing, and classification.
- Set up a Kafka producer to fetch and send news articles from NewsAPI to a Kafka topic and a Kafka consumer to save data locally then data cleaning and preprocessing, converting JSON files into CSV for analysis.
- Conducted ad-hoc analysis using Spark SQL and implemented a classification model using Spark ML to categorize articles by authorship, bypassing the need for manual labeling.

An advanced back-testing tool for Forex Traders (University of Moratuwa, Sri Lanka)   2018

- Developed an advanced back-testing tool for forex traders, integrating short-term price trend prediction, trading strategy optimization, hybrid back-testing, and anomalous time periods detection using a combination of Recurrent Neural Networks and ARIMA models.

Fit Kit CSS Injector for Android (University of Moratuwa, Sri Lanka)   2015

- Developed a mobile application that enabled users to view web applications that are not mobile responsive in a mobile-friendly format by injecting HTML and CSS to modify the DOM architecture.

Industry Projects

An Operational Dashboard for One-wallet   2022

- Provided data reporting and visualizing solution for the project One-wallet (which is a digital wallet that unites two money reloading options for retailers). Retailers use this wallet to recharge mobile plans & bill payments.
- Developed a dashboard for One-wallet to visualize sales, balances, targets, and past achievements of 30,000+ retailers and 1,500+ sales managers with hierarchical filtering.
- Enhanced performance with advanced query optimization for high-volume data.

Sales Operations Center   2020

- Developed a real-time sales monitoring system for 60,000+ retail outlets, enabling efficient sales tracking and target management by hierarchy and locality.
- Used technologies inducing Tableau, PySpark, Impala, Kafka, Kudu Tables, PostgreSQL, and shell scripting.

Other Reporting Dashboards   2019 - 2022

- Created Tableau dashboards for customer journey analysis, fraud detection, and app performance, managing ETL pipelines and providing insights for senior management.

General Projects

AI Customer Support Agent with Amazon Bedrock   2024

- Deployed a serverless AI customer support agent using Amazon Bedrock, integrating a GenAI LLM model (anthropic.claude-3-haiku) and AWS Lambda for real-time query resolution.

Unlocking the Numbers of International Students in Canada   2024

- Developed a Tableau dashboard to visualize trends of international students in Canada by analyzing open data.
- Built a serverless data pipeline on AWS using Lambda and S3 for data processing, with monthly updates via EventBridge.

- Implemented a chatbot with LangChain and OpenAI API for user inquiries about the visualization.

Toronto Auto Theft Dashboard   **2024**
- Developed an interactive Tableau tool that visualizes auto theft incidents using open data from the Toronto Police Service.
- It highlights vulnerable locations and trends over time, with filtering options for year, month, and location type.

Data-Driven Approach to Understanding Child Abuse in Sri Lanka: **2024**
- Developed a Tableau dashboard to visualize child abuse data in Sri Lanka from government reports.
- Used image processing and PDF extraction, aiming to raise awareness and enable discussions by allowing users to filter data by district and type of abuse.

TECHNICAL SKILLS

Programming Languages: SQL, Python, PySpark, R, Scala, Java, Impala, JavaScript, PHP, Bash/Shell Scripting
Data Analytics: Spark, Databricks, Hadoop, AWS EMR, Scikit-Learn, Data Modeling, Pandas, NumPy, Athena
Cloud Services: AWS S3, IAM, Lambda, DynamoDB, AWS Data Catalog, EC2, AWS RDS, Kinesis, Glue
Data Visualization: Tableau, PowerBI, MS Excel, Python (Matplotlib, Seaborn), R
Data Warehouse: Snowflake, AWS Redshift, Star Schema, Snowflake Schema, Delta Lake, S3
Workflow Orchestration/ ETL Tools: Apache Airflow, Cron Jobs, AWS Glue
Generative AI Tools: OpenAI API, AWS Bedrock, LangChain, Amazon Q, ChatGPT, Perplexity AI
Databases: Oracle, MS SQL Server, MySQL, KUDU, MongoDB, PostgreSQL, Hive, Amazon S3, HDFS, Neo4j
Data Extraction Tools: Kafka, Web Scraping (Selenium, Beautiful Soup)
DevOps and CI/CD Tools: Terraform, Circleci, Docker
Software Testing and Automation: Postman, Selenium, Manual Testing, Automated Testing, REST APIs
Data File Formats: CSV, JSON, Parquet, AVRO, ORC
Version Control & Agile Tools: Git, Bitbucket, JIRA

CERTIFICATIONS

AWS Certified Data Analytics - Specialty   **Aug 2023**
AWS Certified Solutions Architect – Associate   **May 2023**
Distributed Computing with Spark SQL   **Feb 2023**
Advanced Snowflake: Deep Dive Cloud Data Warehousing and Analytics   **Sep 2024**

Over 30 IT mini certifications, including nine in Cloud and Distributed Computing, nine in Data Engineering and Data Science, and five in Visualization and Analytics, among others.

EDUCATION

Post Graduate Certificate, Conestoga College, Kitchener, ON   **Sep 2023 - Aug 2024**
Big Data Solution Architecture - One Year (GPA - 3.87/4.00)
> **Coursework:** Relational Database Design, NoSQL Database Implementation, Programming for Big Data, Data Analysis Mathematics, Algorithms and Modeling, Big Data Integration and Storage, Web Analytics and Business Intelligence Tools, Ethics and Security, Software Quality.

Bachelor of Science (Hon), University of Moratuwa, Sri Lanka   **Mar 2014 - Aug 2018**
Information Technology – Four years including 6 months Co-op (GPA - 3.29/4.20)
> **Coursework:** Machine Learning (ML) and Pattern Recognition, Big Data Analytics, Data mining, Artificial Neural Networks, Artificial Intelligence (AI), Computational Mathematics, Operational Research, Data Structures and Algorithms, Object-oriented Programming, Web Technologies, Enterprise Application Development, Advanced Database Management Systems, Project Management.