Dheeraj K

Toronto, ON, Canada | dheeraj1208.k@gmail.com | LinkedIn: Dheeraj-K Website:

Professional Summary

Results-oriented Data Engineer with 3 years of experience designing and deploying scalable data pipelines, cloud architecture, and analytics solutions across enterprise and non-profit environments. Proficient in Azure (ADF, Databricks, Synapse), GCP, Spark, Kafka, SQL, and Python, with a proven record of building efficient ETL workflows and enabling self-service analytics through Power BI and Looker Studio. Strengthened analytical and modeling skills through data science consulting roles, delivering predictive models, interactive dashboards, and end-to-end cloud solutions that transformed complex data into actionable insights. Passionate about leveraging engineering, analytics, and cloud technologies to drive data-driven decision-making and organizational impact.

Technical Skills

Data Engineering & Processing: Azure Data Factory (ADF), Databricks, Apache Spark, Apache Kafka, Delta Lake, Azure Data Lake Gen2, Azure Synapse Analytics, Hive

Cloud Platforms: Microsoft Azure, Google Cloud Platform (GCP), Amazon Web Services (AWS), Google Cloud Storage

Databases: MS SQL Server, MySQL, PostgreSQL

Data Analytics & Visualization: Power BI, Tableau, Looker Studio

Programming & Scripting: Python, SQL, Scala, R

Machine Learning & AI: Scikit-learn, TensorFlow, XGBoost, Pandas, NumPy, Matplotlib, CNN, ANN, NLP, Exploratory Data Analysis (EDA)

DevOps & Version Control: Jenkins, Git/GitHub, Azure CI/CD

Methodologies & Frameworks: Agile, Scrum, Value Stream Mapping, Statistical Data Analysis

Operating Systems: Windows, Linux

Other Core Areas: Data Engineering, Distributed Systems, Data Structures & Algorithms (DSA), Object-Oriented Programming (OOP)

Certifications & Accreditations

- Databricks Accredited Databricks Fundaments
- Databricks Accredited Databricks Azure Analyst

Professional Experience

DATA SCIENCE & ANALYTICS CONSULTANT

Younge Street Mission (YSM) – Toronto, Canada   Jan 2025 – Apr 2025

Led a data-driven initiative to support YSM's mission of combating poverty by analyzing participant pathways and building an end-to-end analytics solution integrating predictive modeling, interactive dashboards, and cloud architecture.

- Conducted exploratory data analysis (EDA) across multiple datasets (demographics, service delivery, assessments) to uncover participant trends and service effectiveness.
- Engineered and cleaned a multi-source dataset by merging records from program initiations, service usage, and TIMES assessments, with custom Python-based imputation logic.
- Built predictive models using XGBoost and Gradient Boosting, achieving R^2 scores of ~0.57 to forecast future participant outcomes based on prior assessments, service usage, and demographics.
- Created a series of interactive Power BI dashboards with custom KPIs, participant journey flows, demographic trends, and service impact visualizations to drive programmatic decision-making.

- Designed and deployed a fully functional end-to-end cloud pipeline in Google Cloud Platform (GCP):
    1. Stored and managed data via Google Cloud Storage.
    2. Configured a Compute Engine VM for data preprocessing and model training.
    3. Proposed a scalable workflow for real-time processing and Looker Studio integration.
- Used R and Python for statistical analysis and visual comparison of historical assessment patterns and modeling effectiveness.
- Delivered key insights on service impact, demographic disparities, and high-value indicators (e.g., Sense of Power, Self-Awareness), supporting YSM's strategic planning and targeted interventions.

DIGITAL ANALYTICS CONSULTANT

RCIScience – Toronto, Canada   Apr 2024 – Jul 2024

Led the end-to-end modernization of RCIScience's web analytics infrastructure to support more accurate, actionable insights and long-term measurement strategy.

- Successfully migrated from Universal Analytics (UA) to Google Analytics 4 (GA4), ensuring data continuity and alignment with evolving tracking standards.
- Configured custom events and conversions in GA4 to match updated engagement goals and KPIs.
- Built a Looker Studio dashboard to visualize user behavior, engagement, and conversion trends, providing stakeholders with real-time insights.
- Conducted historical data analysis using R and Python, identifying trends, comparing legacy metrics, and guiding GA4 implementation decisions.
- Improved data quality and reporting accuracy by streamlining event tagging and refining data layers.
- Delivered clear documentation and walkthroughs to enable internal teams to confidently use GA4 and Looker Studio for ongoing analysis.

DATA ENGINEER

Tata Consultancy Services (TCS) – Telangana, India   Jan 2021 – Sep 2023

As a Data Engineer with a focus on analytics, I design and develop scalable data solutions across hybrid environments (on-prem and cloud). My work spans data ingestion, transformation, and visualization, enabling business stakeholders to make informed decisions.

- Built ETL pipelines using Azure Data Factory for seamless data movement between HDFS/Hive (on-prem) and Azure cloud.
- Developed batch and streaming data workflows using Apache Spark, python, Scala, and Kafka, supporting real-time analytics.
- Prototyped transformations and pipelines in Azure Databricks, leveraging Spark for large-scale data processing.
- Automated validation and reporting tasks using Python, improving ETL accuracy and efficiency.
- Designed data warehouse solutions using Azure Synapse and Azure SQL Database.
- Created complex SQL stored procedures and implemented business logic for analytics and reporting use cases.
- Built interactive Power BI dashboards with dynamic filters, KPIs, and hierarchies, enabling self-service analytics.
- Established CI/CD pipelines (Jenkins) to automate deployment and testing of ETL workflows.
- Monitored and supported critical data pipelines, resolving L1/L2 issues and ensuring high availability.
- Configured Azure Blob Storage for secure, scalable data storage.
- Collaborated with cross-functional teams to align architecture, risks, and deliverables.

- Documented ETL workflows, DR plans, and operational procedures for ongoing support and knowledge transfer.

Education

Master of Professional Studies, Analytics Sep 2023 – Apr 2025
Northeastern University - Toronto, ON, Canada

Bachelor of Technology, Computer Science, Aug 2016 – Aug 2020
Anurag Group of Institutions - Hyderabad, India