

DHARMIK NAIK

Arlington, TX • dharmikjanakkumarnaik@gmail.com • (213) 646 2420 • [LinkedIn](#) • [GitHub](#)

(Open to Relocate)

Experienced **Data Engineer** with **5 years** of expertise in enhancing business performance through automated reporting processes, ETL pipeline creation, and expert guidance in reporting initiatives. Demonstrated success in streamlining multiple business processes for diverse clients. **Actively seeking full-time opportunities** to contribute and work in dynamic environments

AWARDS AND CERTIFICATIONS

- [Oracle Certified Professional, Java SE6 Programmer](#)
- [Oracle Certified Expert, Java EE6 Web Component Developer](#)
- University Topper - Indus University, B.Tech. - Computer Engineering, 2020
- Inducted to MS in CS Honors Program at the University of Southern California
- Team Contributor Award at Synoptek LLC - 2021

PROFESSIONAL EXPERIENCE

UNIVERSITY OF SOUTHERN CALIFORNIA

Los Angeles, CA

Teaching Assistant

Aug 2023 - May 2024

- Course Teaching Assistant for CSCI 401 - Capstone: Design and Construction of Large Software Systems: Responsible for managing and instructing senior Computer Science student groups in real industry projects
- Course Teaching Assistant for CSCI 103 - Introduction to Programming: Design and conduct labs, assignments, and exams covering the basics of OOPS, debugging, programming practices, and more concepts in C++ along with GNU Debugger, Valgrind, **LINUX/UNIX**, to name a few
- Skills: Public speaking, Leadership, Teaching, Software Architectural Design, C++, Debugging, Object Oriented Design, Software Project Management, Conflict Resolution

SYNOPTEK INC.

Ahmedabad, India

Associate Engineer (Data Engineer)

Dec 2019 - May 2022

Client: Large Tourism and Sightseeing Company based in Southern CA

Addressed critical integration challenges within the finance department, where data discrepancies and lack of a unified source hindered business operations

- Served as a Subject Matter Expert (SME) for reporting and data-sharing initiatives while engaging with business users for over 50+ clients to understand and translate reporting requirements
- Designed and implemented scalable data solutions using **Snowflake** and **Azure SQL Database** to support business intelligence and analytics
- Developed and automated **data pipelines** using **Python**, ensuring efficient and reliable data workflows.
- Utilized **Tableau** for data visualization and business intelligence, developing comprehensive dashboards and reports to enable effective data analysis and decision-making
- Successfully integrated and consolidated data from over 20 disparate systems and manual uploads into a cohesive **data warehouse**, eliminating redundant data points to streamline data management
- Transitioned client reporting from Excel to **PowerBI** to overcome scalability limitations, enhancing the analytical capabilities and report generation speed
- Designed and developed Power BI reports leveraging Drill-through, Bookmarks, Row-level Security, and Conditional Formatting to present multi-client detailed-level analysis
- Developed specialized data cubes in **Azure Analysis Services** for power users to facilitate deeper, scalable, and performant data analysis

- Addressed challenges associated with legacy systems that were prone to errors and lacked integration; enhanced data reliability through rigorous data cleaning and validation processes
- Achieved a **70%** increase in operational efficiency post-implementation, earning client commendation and securing additional contracts for ongoing work and maintenance
- Employed **Ralph Kimball's** dimensional modeling techniques to address complex modeling challenges
- Expertly crafted bus and stakeholder matrices to align data architecture with business objectives
- Developed and deployed 10+ changes building **SQL** objects (CTEs, joins, views, stored procedures, functions) with robust documentation
- Optimized query performance by reducing processing time by **66%** through execution plan analysis and implementing indexes replacing sub-queries with joins
- Implemented cost-optimization workflows for Azure services, significantly reducing client expenditures through automated management of resource utilization
- Developed a VBA script to facilitate on-demand refresh of Azure Analysis Services data cube
- Promoted to **Team Leader**, responsible for overseeing and mentoring a team of three junior data analysts
- Implemented robust logging practices to expedite problem resolution and enhance system reliability
- Applied **Agile methodology**

Client: Large Manufacturing Company with 5 Business Segments

Greenfield project. Integrated sales and inventory data from all business segments

- Developed integrated reporting solutions tailored for executive decision-making, enabling more strategic management of business operations
- Conducted comprehensive data discovery sessions with key stakeholders to identify and understand specific data requirements and objectives
- Built and implemented a robust data infrastructure from scratch, designed specifically to support the company's diverse reporting and analytical needs
- Constructed a new data warehouse that consolidated data inputs effectively, serving as the backbone for all reporting activities
- Facilitated knowledge transfer sessions for both in-house staff and contractor employees, ensuring they were equipped with the necessary skills to maintain and utilize the new data systems effectively
- Managed and participated in the complete **Software Development Life Cycle (SDLC)** process, ensuring the successful planning, development, testing, and deployment of data projects
- Conducted rigorous testing of jobs to guarantee the accuracy and completeness of data, meeting quality standards
- Provided end-user support, including user documentation and effective communication, ensuring users could leverage systems efficiently

Client: Large E-commerce Auction Company

Modernized analytics platform

- Designed and built a comprehensive **data warehouse (DW)** tailored to the client's needs, facilitating advanced data analysis and decision-making
- Executed one-time data migration from the legacy system and created **ETL pipelines** to ingest changes from the new system with a high **CDC volume**
- Responded to client requirements by streamlining and reducing the complexity of the ETL pipelines, enhancing performance, and reducing operational costs
- Managed overall infrastructure, maintaining high standards of reliability and scalability to support the client's growing business needs
- Authored **CI/CD pipelines** using **Azure DevOps** and **ARM** templates, automating deployment processes and ensuring consistent and error-free releases
- Managed and participated in the complete Software Development Life Cycle (SDLC) process, ensuring the successful planning, development, testing, and deployment of data projects

Client: Paper Manufacturing Company

Optimized ETL workloads

- Collaborated with stakeholders to understand existing **ETL** practices, **data models**, and data inputs
- Enhanced **Power Query ETL** processes in **Power BI**, achieving significant efficiency improvements, reduced resource consumption, and elevated system performance

Ad Hoc Requests and Projects

- Provided **ETL** support, meticulously mapping data, and crafting **DDL** queries for dimension, fact, and aggregate tables.
- Leveraged advanced **SQL** techniques, including window functions and Common Table Expressions (CTEs), to perform intricate data transformations, aggregations, and analysis within **SQL Server**
- Utilized SQL to create optimized views, stored procedures, scripts, and summary records, ensuring efficient data manipulation and query performance optimization
- Leveraged advanced Power BI features, such as slicers, filters, **DAX** measures, and columns, to enhance dashboard functionality and usability
- Fixed bugs in ETL jobs
- Demonstrated strong ability to plan and prioritize tasks while working on multiple projects simultaneously with minimal supervision
- Served as the primary escalation contact for Tier 1 production issues, swiftly addressing and resolving critical incidents
- Recognized as one of the highest billable resources within the department, reflecting substantial contributions to project revenue and client satisfaction
- Contributed to scoping, proof of concept (POC), and resource allocation for projects
- Integrated data from on-prem client services billing system and resource projection web app used by department leads and projects managers using **SSIS** and built Power BI dashboards to enable executives and managers with advanced resource-centric analytics
- Designed and implemented end-to-end scalable **ETL/ELT** pipelines that integrated data from a diverse range of **APIs**, on-premise, and cloud-based source systems using **Azure Data Factory** and SQL
- Collaborated with multiple stakeholders and cross-functional teams to build end-to-end data warehouses and analytic platforms to enable clients to make data-driven decisions, gain a comprehensive understanding of their business operations, and improve their productivity
- Identified and addressed performance bottlenecks to enhance the efficiency of data workloads and storage systems, ensuring faster data retrieval and optimized storage using query optimization, indexes, and caching of data
- Enhanced observability in data pipelines by instituting comprehensive logging practices, capturing key metrics and events, and facilitating a rapid diagnosis and resolution of issues
- Demonstrated proficiency in problem-solving by addressing multiple production issues and supporting ad-hoc analytical queries on time
- Migrated **SSIS ETL** packages to the cloud and optimized data transformations to reduce the duration of ETL jobs by **35%**
- Technologies: **SQL**, **Python**, **C#**, **Azure Data Factory**, SQL Server Integration Services, **Azure Data Lake**, **Azure Blob Storage**, PowerBI, Azure Analysis Service, **Azure DevOps**, Azure Functions, ARM templates, Azure Key Vault, Git, **Linux**, Agile, **CI/CD**

PROJECTS

Finding Similar Businesses on Yelp for Recommendations

- Developed a content-based recommendation system leveraging Min-hashing and approximate near-neighbor search using Locality Sensitive Hashing (LSH) to efficiently identify similar businesses on Yelp based on user ratings for user retention and satisfaction
- Achieved precision ≥ 0.99 and recall ≥ 0.97 on 500k reviews using limited computing resources.
- Skills: Python, **PySpark**, Distributed Computing, **Git**, **Docker**, Locality Sensitive Hashing, Min-hashing, Recommendation Systems

Discovery of Frequent Itemsets in Ta-Feng Grocery Transactions

- Developed a scalable batch-processing data pipeline to pre-process the Ta-Feng grocery transactions, extracting relevant features and transforming customer transactions into analyzable data points
- Applied and implemented the Savasere, Omiecinski, and Navathe (SON) distributed algorithm to discover the frequent itemsets in the transactions efficiently, addressing challenges such as the inability to fit all transactions in the main memory of a single machine
- Utilized the A-Priori Algorithm to process segmented chunks of data in parallel on worker nodes
- Skills: **Python**, **PySpark**, Distributed Computing, Git, Docker

Named Entity Recognition

- Experimented with a bi-directional LSTM and LSTM-CNN model using GLOVE embeddings for representation of text
- Used CNN to capture character-level information
- Achieved F1-score of 90.1% on CoNLL-2003 corpus
- Technologies: Python, **PyTorch**, **LSTM**, **CNN**, Neural Networks, Word Embeddings

SKILLS

- **Databases:** Microsoft SQL Server, MySQL, Azure Analysis Service
- **Programming Languages:** Python (NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch), R (Shiny, dplyr, ggplot2), Power Query, DAX
- **Tools & Libraries:** Excel, PowerPoint, JIRA
- **Data Visualization:** PowerBI, Tableau, Tableau Prep Builder, Alteryx
- **Machine Learning:** Supervised ML (Linear and Logistic Regression, Decision Trees, RF, Xgboost), Unsupervised ML (K-means, PCA), NLP, Classification techniques (binary classification, multi-class classification), A/B testing

EDUCATION

UNIVERSITY OF SOUTHERN CALIFORNIA

Los Angeles, CA

Master of Science in Computer Science (Honors Fellow), GPA: 4.00/4.00

Aug 2022 - May 2024

Courses: Algorithms, Operating Systems, Data Mining, Machine Learning, Natural Language Processing, Database Systems

INDUS UNIVERSITY

Ahmedabad, India

Bachelor of Technology in Computer Science, GPA: 9.92/10.00, Rank: 1/60

Aug 2016 - May 2020