DHVANI MITESH KOTHARI

New York | (716)-936-4705 | dhvanikothari03@gmail.com | linkedin.com/in/dhvani-kothari | github.com/dhvani-k

EDUCATION

University at Buffalo, The State University of New York   Buffalo, NY
Master of Science, Computer Science & Engineering (GPA: 3.75)   **August 2022-December 2023**
Coursework- Data Intensive Computing, Algorithms for Modern Computing Systems, Data Model & Query Language, Machine Learning

Yeshwantrao Chavan College of Engineering   Nagpur, Maharashtra, India
Bachelor of Engineering, Computer Technology (GPA: 4.0)   **June 2015-May 2019**

PROFESSIONAL EXPERIENCE

Solutions Architect Intern, BRAINCHIP INC., Laguna Hills, CA   **June 2023-Present**
- Empowered technical sales teams with **FastAPI** web-based GUI tool, driving **3x efficiency** and customer engagement driving upsell
- Streamlined integrated circuit scaling estimation process, reducing **30% processing time**, elevating **data accuracy** by **15%**
- Automated unit testing with **Pytest** and **Drone**, reducing time-to-market by **2 weeks**, elevating user-centric testing productivity

Data Engineer, WALMART GLOBAL TECH, Bengaluru, India   **May 2021-July 2022**
- Developed time-series ensemble forecast solution for **599** Sam's Clubs, saving **~$1.2 million** in annual operation planning costs
- Implemented real-time monitoring for critical model performance metrics, reducing model drift and data quality issues by **20%**
- Led a team of **6** to develop **ETL** mappings and **data modeling** on GCP, empowering data-driven insights for business derivatives
- Revamped **PySpark** pipelines with partitioning and query optimization, giving **20x speed** improvement, **saving $500k** annually
- Spearheaded migration of on-premises data pipeline handling **1.5** TB data to GCP using GCS Bucket, Cloud Dataflow/Dataproc, Pub/Sub, Bigquery, enabling **3x** faster data processing

Big Data Developer, CAPGEMINI, Navi Mumbai, India   **June 2019-April 2021**
- Optimized **Spark** applications, achieving **30% faster processing** through batch intervals, parallelism, and query enhancements
- Pioneered holistic automation using Azure Logic Functions and **Azure Data Factory**, delivering **100 man-hours** in quarterly savings
- Designed a comprehensive data lake spanning **11+** manufacturing sites, enabling seamless data access for over **~150** data scientists
- Engineered generic **data recovery** library, resulting in a **50% reduction** in storage & memory costs and annual savings exceeding **$100K**
- Conducted workshops for business teams, enabling them to leverage data insights, improving the **data literacy** of the organization by 25%

ACADEMIC PROJECT/EXPERIENCE

NYC-311 Service Insights (Python, Google Cloud Platform, Terraform, Docker, Looker)   **March 2023-May 2023**
- Constructed NYC 311 data pipeline, cutting manual data processing by **80%**, enabling analysis of **13** years of service requests data
- Enhanced **BigQuery** performance by **40%** using strategic table partitioning and clustering for faster data retrieval and analytics

CineETL: Movie Insights Data Pipeline (AWS S3, Redshift, Airflow, Spark)   **December 2022-February 2023**
- Crafted ETL pipeline for **26M** user ratings and **45K** movies with a data ingestion rate of **10K** records/minutes into AWS **Redshift**
- Normalized data model, automated data quality checks, orchestrated using **Airflow** and achieved **99%** daily ETL cycle success rate

YouTube Trending Video Analysis (AWS Services - Athena, Glue, S3, QuickSight)   **October 2022-December 2022**
- Architected scalable data pipeline with AWS **Glue** ETL to transform over **100k** records, stored securely in an S3 bucket
- Employed **Lambda** functions to clean and preprocess JSON data, converting it into optimized parquet format, and automating the process through a file load trigger, saving **10%** of manual work
- Used Utilized **QuickSight** to analyze engagement metrics to identify top-performing videos and improve marketing strategies

Netflix Navigator: A Data-Driven Content Optimization Dashboard (Tableau, Excel)   **September 2022-October 2022**
- Engineered a scalable Netflix dashboard on **Tableau**, enabling streamlined data integration and analysis for **50k+** records
- Leveraged advanced **data analytics** techniques to uncover insights and trends, driving a **2x** ease in understanding user preference

SKILLS

- **Programming Languages:** Python, C++, C, SQL, Java, HTML/CSS, JavaScript, Scala, R, Bash
- **Database:** SQL Server, Oracle, MySQL, MongoDB, Cassandra(NoSQL) , SSIS, HBase
- **ETL/Big Data:** Apache Kafka, Talend, Informatica, Apache Hadoop, MapReduce, Databricks, ElasticSearch
- **Cloud Technologies/DevOps Framework:** Linux, Jenkins, Docker, Kubernetes, Google Cloud Platform, AWS, Microsoft Azure
- **Miscellaneous:** Power BI, Tableau, Excel, Rest API, Git, Agile,

LEADERSHIP EXPERIENCE AND ACTIVITIES

- Bravo Award: Recognition for consistent demonstration of outstanding and exemplary performance by Walmart Global Tech
- Chaired as YCCE ACM Student chapter President to work towards goals promoting community service, academics, and unity
- Published paper at International Conference on Advancements in Computing & Management, Jaipur, India

CERTIFICATIONS AND TRAINING

- Microsoft Certified: Azure Fundamentals (AZ-900)
- Oracle Cloud Infrastructure Foundations 2022 Certified Associate