VIGNESH KUMAR SATHYA MURTHY

0466903192 • ess.vigneshkr@gmail.com • Linkedin • Portfolio • Melbourne

SKILLS

Programming Languages: Python, PySpark, SQL (PostgreSQL, Oracle), KQL, R, Shell Scripting.

ETL & Data Engineering: Data Pipelines, Data Ingestion, Data Warehousing, ETL Processes, Data Modeling, Business Intelligence, Automation.

Data Processing: Apache Spark, HDFS, Stream Processing, Batch Processing, Data Transformation, Parallel Processing.

Cloud Platforms: Microsoft Azure, Azure Databricks, Microsoft Fabric, Azure Synapse Analytics, Azure Data Factory (ADF), Informatica PowerCenter, AWS.

Data Visualization & BI: Microsoft Power BI, Tableau, Reporting Dashboards.

Development Tools: Databricks Notebooks, GitHub, Jira.

EXPERIENCE

VisionVerse Interactive Pty Ltd | Link — Melbourne, Australia

Data and AI Intern — 07/2024 - 12/2024

Technologies Used: Python, OpenAI, RAG, PostgreSQL, Next.js, GraphQL, Jira, GitHub

- Developed an AI-driven solution automating personalized webpagesfor restaurants, generating it within 2 minutes.
- Implemented a universal scraping module using OpenAI's LLM, Python, and Selenium to extract restaurant data (e.g., menus, logos), overcoming layout inconsistencies, with a 95% success rate.
- Utilized RAG techniques to combine retrieved restaurant data with AI generation for contextually accurate webpage content.
- Designed a dynamic frontend in Next.js, enabling real-time webpage generation from user-inputted URLs.
- Designed and managed PostgreSQL database for storing structured restaurant data, ensuring efficient querying and integration with GraphQL APIs via PostGraphile.
- Led end-to-end project phases, collaborating in a five-member Agile team.

Tata Consultancy Services Ltd | Link — TamilNadu, India

Data Engineer — 01/2021 - 01/2023

Technologies Used: Azure Data Factory, Azure Synapse Analytics, SQL, Azure Databricks, Power BI, Azure SQL Database

- Developed and maintained robust data pipelines using Azure Data Factory for a large-scale banking project, ensuring seamless data flow and processing.
- Leveraged Azure Databricks for complex data transformations and processing, implementing PySpark jobs to handle large-scale banking datasets efficiently.
- Implemented medallion architecture (Bronze, Silver, Gold layers) to optimize data processing workflows, improving data quality and accessibility.
- Optimized data pipelines and ETL processes, reducing processing time by 50% and enhancing overall system performance.
- Enhanced database performance by implementing indexing strategies, partitioning, views, and CTEs, reducing query response time by 25%.
- Designed comprehensive Power BI dashboards for banking stakeholders, enabling real-time data visualization and faster decision-making.
- Delivered 98% of sprint goals on time in Agile teams, ensuring consistent project delivery and stakeholder satisfaction.

EDUCATION

RMIT University — Melbourne, Australia

Masters of Data Science — 02/2023 - 12/2024

Sona College of Technology — TamilNadu, India

Bachelors of Computer Science and Engineering — 08/2016 - 09/2020

PERSONAL PROJECTS

Azure Healthcare Revenue Cycle Management (RCM) Data Engineering Solution | GitHub Link
Technologies Used: Azure Data Factory, Azure Databricks, Delta Lake, Azure SQL Database, Azure Data Lake Storage Gen2, Azure Key Vault, Power BI, Python, SQL.

- Designed and implemented end-to-end healthcare data pipeline processing daily EMR data from multiple hospital systems using medallion architecture (Bronze-Silver-Gold) with automated data quality checks, SCD Type 2 implementation, integrated external APIs (NPI, ICD codes), and Power BI dashboards for comprehensive healthcare analytics and revenue cycle optimization.

Real-time Taxi Data Stream Processing | Github Link
Technologies Used: Apache Kafka, Apache Spark Streaming, HDFS, Scala.

- Processed NYC taxi trip data in real-time using Apache Kafka and Spark Streaming to compute ride metrics, store data in HDFS (Parquet), and ensure system reliability with checkpointing and state management.

Cloud-Based Music Subscription Application | Github Link
Technologies Used: AWS EC2, S3, API Gateway, Lambda, DynamoDB, Python

- Designed a music app leveraging AWS services like EC2 for hosting, S3 for scalable storage, Lambda for serverless operations, and DynamoDB for data management. Enabled seamless user interactions, ensuring high availability, fast data retrieval, and a robust backend infrastructure.

CERTIFICATION

Fabric Data Engineer Associate
Microsoft

Fabric Analytics Engineer Associate
Microsoft

Azure Data Fundamentals
Microsoft

Azure Fundamentals
Microsoft

Industry Experience Program
RMIT

Global Leader Experience
RMIT