JAS H VAGH AS I Y A

437-436-3011| [jashmukeshvaghaisya@gmail.com](mailto:jashmukeshvaghaisya@gmail.com)

# E X P E R I E N C E

**NextDayAI – SpicyChatAI Canada**

*Data Engineer/Analyst August 2024 – Present*

*Data Developer – Co-op May 2024 – August 2024*

* Developed and maintained data pipelines processing 850K daily records using Azure Data Factory, PySpark on Databricks, and Azure Synapse Analytics with comprehensive workflow orchestration.
* Optimized ETL performance reducing critical queries from 45 minutes to 12 minutes (73% improvement) and processing time from 8 hours to 2.5 hours (~69% improvement) through strategic partitioning, materialized views, and intelligent batching across 70GB+ data bi-weekly.
* Built real-time streaming architecture using Confluent Kafka and Spark Structured Streaming to process session completion events, enabling immediate data analysis and performance monitoring with variable data sizes.
* Implemented data lakehouse solutions using Unity Catalog, Azure Purview, and Delta Lake with automated compliance workflows, PII detection via ML models, and multi-layered validation while reducing manual review workload.
* Integrated AI models with data pipelines including automated training data collection, validation metrics collection, and real-time model validation dashboards in Power BI for comprehensive lifecycle management.
* Designed hybrid real-time/batch processing solutions reducing critical processing time during peak periods while optimizing infrastructure costs by 35% through resource allocation and automated retry mechanisms.
* Provided debugging and troubleshooting support for data pipelines, implemented monitoring solutions, and delivered executive dashboards improving load times.
* Participated in CI/CD practices using Azure DevOps pipelines for automated deployments and maintained version control with Git for collaborative development.

# Dione Apps India

*Data Engineer January 2021 – November 2022*

* Developed scalable data migration strategies using Azure Data Factory, PySpark, and Synapse Analytics, enabling the seamless transfer of terabytes of data with zero data loss and minimal downtime.
* Designed and implemented end-to-end ETL pipelines integrating data from sources, ensuring data quality, business logic alignment, and high-performance transformations in Azure Databricks and SQL-based warehouses.
* Optimized data processing workflows handling 100GB+ of loads by fine-tuning PySpark transformations, reducing execution time and improving job performance and reliability.
* Implemented Slowly Changing Dimension (SCD) logic in cloud-based databases, ensuring accurate historical tracking and enhancing BI reporting reliability.
* Proactively monitored and resolved data pipeline failures using Azure Monitor, Log Analytics, and automated alerting, leading to an 80% reduction in operational issues.
* Automated DDL deployments and data ingestion workflows using CI/CD pipelines and orchestration tools, reducing manual interventions by 60% and ensuring compliance with governance standards.
* Designed custom ETL utilities to streamline data migrations between on-premise and cloud platforms, improving efficiency and standardization across projects.
* Partnered with data architects and business stakeholders to gather requirements, document processes, and ensure alignment with strategic data initiatives, driving 100% project success rates.

# C E R T I F I C A T I O N S

[**Microsoft Certified:**](https://learn.microsoft.com/en-us/users/jashvaghasiya-4649/credentials/d944bfaa682ef208) **Azure Data Engineer Associate** *July 2024*

[**Databricks Certified:**](https://www.credential.net/71959e09-697b-4f64-8a15-ae7213dbf55c#gs.dwgbfp) **Associate Developer for Apache Spark 3.0 – Python** *August 2024*

[**Databricks Certified:**](https://www.credential.net/1001cd75-9e1d-4340-ba0c-78332c105021#gs.dwgbkx) **Data Engineer Professional** *August 2024*

# T E C H N I C A L S K I L L S

* **Databases & SQL:** Microsoft SQL Server, PostgreSQL, MySQL, NoSQL (MongoDB, CosmosDB)
* **Cloud & Data Platforms:** Azure, AWS, Databricks, Microsoft Fabric
* **Programming & Scripting:** Python, SQL, T-SQL, Scala, Bash, PowerShell
* **Big Data Tools:** PySpark, Hadoop, Hive, Kafka, Azure Data Lake, Delta Lake
* **ETL & Data Integration:** Azure Data Factory (ADF), SSIS, Azure Synapse Pipelines
* **Workflow Orchestration:** Azure Data Factory, Apache Airflow
* **Data Lakehouse & Warehousing:** Delta Lake, Azure Synapse, Microsoft Fabric Lakehouses, Data Modeling
* **Testing & Quality Assurance:** pytest, Unit Testing, Integration Testing, Data Quality Validation
* **DevOps & CI/CD:** Azure DevOps, Git, Docker, Kubernetes, Terraform
* **Monitoring & Performance:** Azure Monitor, Azure Log Analytics, Performance Optimization, Troubleshooting
* **Security & Compliance:** Role-Based Access Control (RBAC), Unity Catalog, Data Encryption, Azure Purview
* Project **Management:** Agile, JIRA, Code Reviews, Project Planning

# E D U C A T I O N

**Artificial Intelligence and Machine Learning |** Lambton College in Toronto *August 2024*

**Bachelor’s Degree in Computer Applications |** Veer Narmad South Gujarat University *June 2021*