**Omprakash**

vinod@atvsllc.com 832-862-3709 Ext: 112

Data Engineer with 8 years of experience in implementing Big Data, Spark, Azure components and Data Lake solutions to

solve strong analytical problems and aid in business decision making. Experienced working in highly reputable organizations including Retail, insurance, and finance industries

# Professional Summary:

* Data Engineer with 8 years of experience specializing in Azure-related technologies, big data, SQL Server, and ETL processes
* Proficient in Azure services such as Azure Data Factory, Azure SQL Database, Azure Data Lake, and Azure Databricks, Azure Synapse Analytics
* Extensive knowledge and hands-on experience in designing and implementing efficient data processing pipelines using Hadoop, Spark, and Kafka
* Expertise in ETL processes, ensuring data accuracy, availability, and security throughout the data lifecycle.
* Strong proficiency in working with SQL Server, designing, and managing complex database systems, and developing optimized queries
* Skilled in collaborating with cross-functional teams to translate business requirements into technical solutions
* Proven track record of successfully tackling data engineering challenges and delivering robust data solutions
* Experienced in migrating on-premises data infrastructure to Azure cloud, optimizing efficiency, and reducing costs
* Solid understanding of cloud architecture and its application in data engineering projects
* Continuously up-to-date with the latest industry trends and emerging technologies in data engineering and Azure ecosystem
* Strong analytical mindset and problem-solving skills, enabling effective data analysis and troubleshooting
* Ability to ensure data integrity and quality through meticulous data governance practices
* Excellent communication and interpersonal skills, fostering effective collaboration and teamwork
* Proactive and adaptable, capable of thriving in dynamic and fast-paced environments
* Committed to driving data-driven decision-making and unlocking the full potential of data assets for organizations
* Build reporting dashboards in Microsoft Power-BI for different stakeholders, to analyze and review modeling data, identify historical data trends and compare model outputs
* Comprehensive experience in ETL implementation, Big Data Analytics, orchestrated data pipelines and implemented cloud-based data stacks and big data solutions
* Experienced with version control systems like Git, GitHub, to keep the versions and configuration of the code organized
* Experience with various file formats such as Avro, Parquet, ORC, Json and XML
* Implemented and Utilized Kubernetes and Docker for the runtime environment for the Continuous Integration, CI/CD system to build, test, and deploy

# EDUCATION:

* Bachelors in Electronics and Communication Engineering from Osmania University
* Master’s in Business Analytics and Information Systems from University of South Florida

# Technical Skills:

|  |  |
| --- | --- |
| **Programming** | SQL, Python, Pyspark, Unix Shell Script, Spark,Scala |
| **Databases** | MySQL, SQL, Oracle, Teradata, Cosmos DB |

|  |  |
| --- | --- |
| **Cloud Technology/Technologies** | Azure Databricks, Azure Data Factory, Data Lake Service, Data Lakehouse, Snowflake, Amazon S3, Kuberentes, Github, Kafka |
| **Big Data Frameworks** | Hadoop, Hive, Spark |
| **Analysis and reporting services** | MS Excel, Microsoft Power BI |
| **Architecture** | Relational DBMS, OLTP, OLAP |
| **Development Methodologies** | Agile/Scrum |
| **Software** | Anaconda, Postman-API, Jupiter Notebooks |

**Professional Experience**

**Organization: JP Chase June 2023 – Present**

**Role: Big Data Engineer**

**Summary: -** Established a versatile infrastructure for data migration from on-prem to Hadoop or AWS S3, supporting multiple LOBs. Responsible for end-to-end implementation, maintenance, and resolving support tickets. Proficiently manage the entire infrastructure, addressing issues at the code level and optimizing for efficiency. Integrated AWS S3, Hadoop, Kafka, Kubernetes, Spark, GitHub, and Jira for a comprehensive solution. Ensured scalability, real-time streaming, and efficient container orchestration. Dedicated to code-level issue resolution and proactive code optimization. Delivered a unified solution for smooth data migration and sustained infrastructure efficiency.

# Responsibilities:

* Proficient in AWS S3, providing scalable and durable storage solutions for effective data migration projects
* Deployed and managed Hadoop clusters, emphasizing the Hadoop Distributed File System (HDFS) for optimal storage and processing of large datasets
* Implemented real-time data streaming using Kafka, ensuring seamless integration with diverse sources during on-premises to Hadoop/S3 migration
* Utilized Kubernetes for container orchestration, optimizing deployment and scaling of applications within the data migration pipeline
* Applied Apache Spark for distributed data processing and ETL tasks, ensuring high performance and rapid data transformations
* Managed migration scripts and documentation on GitHub, utilizing version control for collaborative and efficient development workflows
* Leveraged Jira for project management, tracking tasks, managing milestones, and addressing issues throughout the data migration lifecycle
* Integrated ETL tools seamlessly with AWS S3, Hadoop, Kafka, and Spark, optimizing data movement and transformation processes
* Implemented Kubernetes for scalability, ensuring efficient resource management during the execution of data migration tasks
* Conducted rigorous testing and validation at each stage, ensuring data accuracy, consistency, and reliability in the target Hadoop or AWS S3 environment

**Environment**: AWS s3, Spark, Hdfs, Kafka, Kubernetes, Jira, Github

**Organization: Tiger Analytics March 2022 – May 2023**

# Client: PepsiCo Role: Data Engineer

**Summary: -** Managing the Hadoop and Spark ecosystems to ensure efficient data processing and analysis. I optimize workflows to handle large-scale datasets effectively, leveraging my expertise in these technologies. Leading migration projects is a key aspect of my role, where I oversee the seamless transition of on-premises infrastructure to the Microsoft Azure cloud platform. Utilizing Python for automation and data manipulation tasks, I enhance productivity and streamline processes. Additionally, I collaborate closely with cross-functional teams to deliver data-driven solutions that align with the organization's business objectives.

# Responsibilities:

* Proficient in Hadoop ecosystem technologies like HDFS, MapReduce, Hive, enabling scalable and fault-tolerant data processing
* Extensive experience in Apache Spark for real-time and batch data processing, utilizing RDDs, DataFrames, and Spark SQL
* Skilled in managing large-scale datasets and complex data structures, leveraging Big Data technologies to derive actionable insights
* Led successful migration projects, seamlessly transitioning on-premises Hadoop and Spark clusters to the Microsoft Azure cloud platform.
* Collaborated with cross-functional teams, including database administrators and developers, to ensure smooth migration and minimize disruptions to business operations
* Implemented efficient data replication, transformation, and validation processes during migration to ensure minimal downtime and data integrity
* Proficient in deploying and managing Azure data services such as HDInsight, Databricks, and Data Lake Storage for scalable and cost-effective solutions
* Employed Delta Lake tables for reliable data versioning and ACID transactions, enhancing data quality and availability in the silver and gold layers
* Optimized data migration processes through performance tuning and query optimization, ensuring efficient data retrieval and analysis
* Provided training and guidance to junior team members on data migration best practices and technologies
* Designed, deployed, and automated data pipelines using Azure DevOps, ensuring smooth data movement, and processing from source to destination.
* Collaborated with cross-functional teams to define requirements, implement version control, and monitor pipeline performance
* Implemented Infrastructure as Code using Azure DevOps, creating ARM templates and scripts to automate Azure resource provisioning for data projects
* Improved scalability and reduced manual configuration efforts while maintaining version-controlled infrastructure
* Applied Agile and DevOps methodologies in data projects, managing tasks and sprints through Azure DevOps boards
* Introduced automated testing practices to ensure data quality, fostering continuous improvement and enhancing pipeline reliability

**Environment**: Hadoop, Spark, Azure Data Factory, Azure Databricks, Python, Delta lake Tables, Azure Blob Storage, Logic Apps, Synapse, Azure Devops

**Organization: Assurant Inc June 2021 – Dec 2021**

**Role: Data Engineer Intern**

**Summary:** Data Engineer with expertise in optimizing Databricks notebooks and implementing Delta Lake tables. Skilled in designing and implementing efficient data processing pipelines using Apache Spark, Databricks, and Delta Lake. Proficient in performance tuning, query optimization, and data governance. Committed to delivering scalable and reliable data solutions to enable organizations to derive valuable insights from their data.

# Responsibilities:

* Optimized Databricks notebooks, implementing performance tuning techniques and best practices to enhance query execution time and overall job efficiency
* Utilized Spark UI and other monitoring tools to identify and resolve bottlenecks in Databricks clusters, improving overall data processing performance
* Implemented Delta Lake tables for reliable and scalable data storage, enabling ACID transactions and versioning for data integrity and efficient data querying
* Designed and developed data processing pipelines using Apache Spark and Databricks, ensuring the smooth and efficient extraction, transformation, and loading of data from various sources
* Conducted data quality assurance checks to ensure accuracy, consistency, and completeness of data processed by Databricks notebooks
* Developed efficient ADF pipelines, optimizing data processing and reducing query response time by 10%

**Environment**: Azure Databricks, Python, Delta lake Tables, Azure Data Factory, Azure Devops, CI/CD

# Organization: Wipro Tech Jan 2014 – May 2019

**Client: Microsoft Teams Role: Data Engineer**

**Summary:** Expertise in Microsoft Teams data leveraging ADLS, ADF, Blob Storage, Azure Data Lake Analytics, U-SQL, and SQL Server. Skilled in designing and implementing scalable data solutions for Microsoft Teams analytics and reporting. Proficient in data ingestion, transformation, and analysis using Azure data technologies. Committed to delivering high- quality data solutions to drive data-driven decision-making and enhance collaboration.

# Responsibilities:

* Designed and implemented data ingestion pipelines to capture Microsoft Teams data using Azure Data Factory (ADF) and Blob Storage, ensuring reliable and efficient data ingestion
* Leveraged Azure Data Lake Storage (ADLS) for storing and organizing Microsoft Teams data, ensuring scalability and high-performance data storage
* Utilized Azure Data Lake Analytics and U-SQL for data transformation and analysis, enabling comprehensive insights into Microsoft Teams data
* Developed complex U-SQL scripts to process and transform large volumes of data, optimizing query performance and ensuring efficient data processing
* Collaborated with stakeholders to define data requirements and create data models and schemas in SQL Server, facilitating accurate data analysis and reporting
* Conducted performance optimization and tuning of data processing jobs to improve processing efficiency and reduce execution time
* Implemented data governance and compliance measures, ensuring data security, privacy, and adherence to regulatory standards
* Collaborated with cross-functional teams, including data scientists and analysts, to understand data requirements and deliver data engineering solutions aligned with business needs
* Worked closely with Microsoft Teams product teams to understand data structures and capture relevant data for analytics purposes
* Collaborated with stakeholders to validate data quality, ensuring data integrity throughout the data engineering process

**Environment**: Azure Data Factory, Azure Blob Storage, ADLS Gen1/2, Azure Data lake Analytics, U-SQL, MS SQL, Cosmos DB

# Client: British Petroleum Role: SQL Developer

**Summary:** Dynamic and results-driven Data Engineer with expertise in SQL Server and data management at British Petroleum (BP). Proficient in designing and implementing data solutions, creating, and altering tables, developing stored procedures, and generating reports on transactional data. Skilled in ad-hoc query development and optimization. Committed to delivering accurate and timely data-driven insights to support informed decision-making at BP.

# Responsibilities:

* Designed and implemented SQL Server database solutions to support data management and reporting requirements at BP
* Created and altered tables in SQL Server, ensuring efficient data storage and retrieval for transactional data
* Developed stored procedures to automate data processing and enhance data consistency and integrity
* Generated reports on transactional data, providing insights into key business metrics and supporting decision- making processes
* Proficiently developed end-to-end ETL solutions using Informatica PowerCenter, orchestrating seamless data extraction, transformation, and loading processes from diverse sources to target systems
* Demonstrated expertise in performance optimization by fine-tuning SQL queries, session configurations, and partitioning techniques, resulting in notable improvements in ETL job efficiency and overall processing times.
* Conducted performance tuning and query optimization to enhance overall system performance and reduce query execution time
* Worked closely with data architects and developers to design and optimize databases, ensuring data modeling best practices and efficient data storage and retrieval
* Implemented data governance and compliance measures, ensuring data security, privacy, and adherence to regulatory standards
* Collaborated with cross-functional teams, including data analysts and business users, to understand data requirements and deliver data engineering solutions aligned with business needs

**Environment**: SQL, Microsoft Excel, ETL