DURGAPRASAD

Phone: 972-525-1655
Email: prasadroyal6@yahoo.com

Summary

Experienced Data Engineer with over 5.5 years in data processing, analysis, and modeling. Proven ability to design and implement efficient data solutions to build data pipelines, create data warehouses, and develop insights for business operations. Proficient in Azure services, ETL processes, and data visualization tools. Seeking a Data Engineer position to further develop and optimize data-driven solutions.

Key Functional Areas

- **Proficient in Azure Data Platform services:** Data Lake, Data Factory (ADF), Synapse, Databricks, Data Lake Analytics, Stream Analytics, SQL DW, HDInsight, NoSQL databases.

- **Built and managed end-to-end data lakes:** From data ingestion, validation, and transformation to publishing.

- **Extensive experience in Azure transformation projects:** Architecting and implementing data pipelines and making informed architectural decisions.

- **Expert in ETL and data movement:** Utilizes ADF, SSIS, and Synapse for efficient data flows.

- **Developed Power BI reports and dashboards:** Effectively gathering and translating user requirements to create actionable insights.

- **Designed and implemented dashboards and visualizations:** Leveraging Microsoft tools like SSRS and Power BI for clear and actionable data presentations.

- **Extracted data from diverse sources:** SQL Server, Excel, Oracle, SQL Azure, etc., for seamless integration into Power BI reports.

- **Expertise in Databricks:** Strong understanding of its capabilities for optimal data processing.

- **Built and managed database solutions:** Azure SQL Data Warehouse and Azure SQL for scalable and secure data storage.

- **Delivered end-to-end data solutions:** Covering storage, integration, processing, and visualization tasks within the Azure ecosystem.

- **Advanced Azure Data Lake skills:** Managing ADLS and Data Lake Analytics, proficient in USQL data transformation.

- **Established database standards:** Implementing efficient operations, upgrades, migrations, and new application/customer onboarding procedures.

- **Data validation expertise:** Utilizing PySpark, Python, and SQL within Azure Databricks for high-quality data processing.

- **SQL development and maintenance:** Proficient in writing, testing, and maintaining queries, stored procedures, and triggers while adhering to established standards.

- **Automated deployment solutions:** Deployed Azure Resource Manager JSON Templates using PowerShell for streamlined infrastructure management.

- **Advanced Microsoft BI experience:** Proficient in Azure Analysis Services (SSAS) and SQL Server Reporting Services (SSRS), familiar with Excel PowerPivot, Power View, and Powerbase.

- **Azure Data Factory experience:** Worked with ADF, Integration Run Time (IR), File System Data Ingestion, and Relational Data Ingestion.

- **Data migration expertise:** Orchestrated on-premises data migration (Oracle/SQL Server/MongoDB) to Azure Data Lake Store (ADLS) using Azure Data Factory.

Technical Skills

- **Azure Services:** Azure Data Lake Storage (ADLS), Azure Data Factory (ADF), Azure Synapse Analytics, Azure Databricks, Azure SQL Database, Azure Cosmos DB (NoSQL), Azure HDInsight (Hadoop, Spark, Hive), Azure Stream Analytics, Azure Purview

- **ETL Tools:** SQL Server Integration Services (SSIS), Azure Data Factory (ADF), Apache NiFi

- **Data Visualization:** Power BI, Tableau, QlikView

- **Database Systems:** Microsoft SQL Server, Oracle Database, MySQL, MongoDB (NoSQL)

- **Big Data Technologies:** Apache Spark, Apache Hadoop, Apache Hive, Apache Kafka, dbt, Snowflake

- **Scripting and Programming Languages:** Python, SQL, PySpark

- **Version Control:** Git, GitHub, Azure DevOps

- **Data Modeling Tools:** Microsoft Visio, Erwin Data Modeler

- **Cloud Infrastructure:** Azure Storage (Blob, File, Queue), Azure Security (Azure Active Directory, Azure Key Vault)

- **Scripting and Automation:** PowerShell, Azure CLI

- **Monitoring and Logging Tools:** Azure Monitor, Azure Log Analytics, Prometheus

- **Database Administration Tools:** SQL Server Management Studio (SSMS), Azure Data Studio

Professional Work Experience

LearnBeyond, Data Engineer   **Feb 2023 – Present**   **Dallas, Texas**

- Implemented **Azure Data Factory (ADF)** extensively for ingesting data from different source systems like relational and unstructured data to meet business functional requirements.

- Designed and developed **batch processing and real-time processing solutions** using ADF, Databricks Clusters, and Stream Analytics.

- Created numerous **pipelines** in Azure using ADF v2 to get the data from disparate source systems by using different Azure activities like Move & Transform, copy, filter, for each, Databricks etc. Maintain and provide support for optimal pipelines, data flows, and complex data transformations and manipulations using ADF and PySpark with Databricks.

- Automated jobs using different **triggers** like events, schedules, and tumbling in ADF.

- Created, provisioned different **Databricks clusters**, notebooks, jobs, and autoscaling.

- Performed **data flow transformation** using data flow activity.

- Used **PolyBase** to load tables in Azure Synapse.

- Implemented **Azure** self-hosted integration runtime in ADF.

- Improved performance by optimizing computing time to process the streaming data by optimizing the cluster runtime.

- Perform ongoing monitoring, automation, and refinement of data engineering solutions.

- Scheduled, automated business processes and workflows using **Azure Logic Apps**.

- Designed and developed a new solution to process the NRT data using **Azure Stream Analytics, Azure Event Hub**, and service bus queue.

- Created **linked services** to connect external resources to ADF.

- Worked with complex **SQL views, stored procedures, triggers**, and packages in large databases from various servers.

- Used **Azure DevOps** & Jenkins pipelines to build pipelines and deploy different resources (code and infrastructure) in Azure.

- Ensured the developed solutions are formally documented and signed off by business.

- Worked with team members to resolve any technical issues, troubleshooting, project risk, and issue identification and management.

- Experience managing **Azure Data Lake (ADLS)** and Data Lake Analytics and understanding how to integrate with other Azure services.

- Migration of on-premises data (Oracle/Teradata) to **Azure Data Lake Store (ADLS)** using ADF (ADF V1/V2).

- Worked closely across teams (support, solution architecture) and peers to establish and follow best practices while solving customer problems.

- Created infrastructure for optimal extraction, transformation, and loading of data from a wide variety of data sources.

- Designed and created optimal pipeline architecture on the Azure platform.

- Created linked services to land the data from different sources to ADF.

- Created, provisioned different Databricks clusters needed for batch and continuous data processing and installed the required libraries for the clusters.

Infosys, Data Engineer   **Sep 2017 − Aug 2021**   **Hyderabad, India**

- Analyzed, designed, and built modern **data solutions** using Azure PaaS service to support visualization of data. Understand current production state of the application and determine the impact of new implementation on existing business processes.

- Created objects, pipelines, notebooks in **Azure Synapse**.

- Extracted, transformed, and loaded data from source systems to **Azure Data Storage** services using a combination of **Azure Data Factory, T-SQL, Spark SQL**, and U-SQL Azure Data Lake Analytics. Data ingestion to one or more Azure services - (Azure Data Lake, Azure Storage, Azure Synapse, Azure SQL, Azure DW) and processed the data in **Azure Databricks**.

- Created pipelines in ADF using Linked Services/Datasets/Pipeline to extract, transform, and load data from different sources like **Azure SQL, Blob Storage, Azure SQL Data Warehouse**, write-back tool, and backwards.

- Developed **Spark applications** using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into customer usage patterns.

- Responsible for estimating the cluster size, monitoring, and troubleshooting of the **Spark Databricks cluster**.

- Experienced in **performance tuning** of Spark applications for setting the right batch interval time, correct level of parallelism, and memory tuning.

- Developed **JSON scripts** for deploying the pipeline in ADF that processes the data using the SQL Activity.

- Hands-on experience in developing **SQL scripts** for automation purposes.

- Created build and release for multiple projects (modules) in the production environment using **Visual Studio Team Services (VSTS)**.

Certifications

- Certification of Appreciation from Wipro for Cloud, Infrastructure, and Security Department in 2020