## **Data Engineer (Azure)**

**Name:** **Sai Teja Sri Bathina**

**Phone: (224) 286-2235**

**PROFESSIONAL SUMMARY:**

* Around 12 years of Experience as **Azure** Cloud Data Engineer in **Microsoft Azure Cloud** technologies including **Azure** Data Factory (ADF), **Azure** Data Lake Storage (ADLS), **Azure** Synapse Analytics (**SQL** Data warehouse), **Azure** **SQL** Database, **Azure** Analytical services, Polybase, **Azure** Cosmos **NoSQL** DB, **Azure** Key vaults, **Azure** **DevOps**, **Azure** HDInsight Big Data Technologies like **Hadoop**, **Apache** **Spark** and **Azure** Data bricks.
* Big Data - **Hadoop** (MapReduce & **Hive**), **Spark** (**SQL**, Streaming), **Azure** **Cosmos DB**, **SQL** Datawarehouse, **Azure** DMS, **Azure** Data Factory and SQL.
* Good experience on understanding of architecting, designing and operation of large scale data and analytics solutions on Snowﬂake Cloud Data Warehouse.
* Strong knowledge in **Spark** ecosystems such as **Spark** core, **Spark** **SQL**, **Spark** Streaming libraries.
* Very Good experience working in **Azure** Cloud, **Azure** **DevOps**, **Azure** Data Factory, **Azure** Data Lake Storage, **Azure** Synapse Analytics, **Azure** Analytical services, **Azure** Cosmos NO **SQL** DB, **Azure** HD Insight Big Data Technologies (**Hadoop** and **Apache** **Spark**) and Data bricks.
* Experience in designing **Azure** Cloud Architecture and Implementation plans for hosting complex application workloads on MS **Azure**.
* Experience working in reading Continuous **JSON** data from different source system using **Kafka** into Databricks Delta and processing the files using **Apache** Structured streaming, **PySpark** and creating the files in parquet format.
* Created manual Test Cases to check that each of the deliverables meet user's requirement.
* Good knowledge in **Apache** **Hadoop** ecosystem components **Spark**, **Cassandra**, **HDFS**, **Hive**, SQOOP, **Airflow**.
* Experienced in working with different data formats CSV, **JSON** and Parquet.
* Strong in Data Warehousing concepts, Star schema and Snowflake schema methodologies, understanding Business process/requirements.
* Expert in building hierarchical and Analytical **SQL** queries that helps in reporting.
* Expert in implementing Business Rules by creating re-usable transformations like mapplets and mappings.
* Expert in using debugger in Informatica designer tool to test and fix errors in the mappings. Supported ad-hoc reporting and analytics requests with an eye for creating scalable self-service or automated solutions.
* Later Migrated applications from Django to Flask and NoSQL (DynamoDB) to SQL(**Snowflake**)
* Developed and worked on **Machine Learning algorithms** for predictive modelling.
* Architected complete scalable data pipelines, data warehouse for optimized data ingestion.
* Collaborated with data scientists and architects on several projects to create data mart as per requirement.
* Conducted complex data analysis and report on results.
* Constructed data staging layers and fast real-time systems to feed BI applications and **Machine Learning algorithms**.
* Understanding of **Azure** webservices and at least hands on experience working in projects. Knowledge of the software development life cycle, **Agile** **methodologies**, and test-driven development.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Big Data Technologies:** | HDFS, Hive, Spark, MapReduce, YARN, Spark-Core, Spark-SQL. |
| **Programming Languages:** | .NET, C/C++, HTML, SQL, PL/SQL, and Scala. |
| **Scripting Languages:** | Shell Scripting, Bash, PowerShell, Python. |
| **Operating Systems:** | UNIX, Windows, LINUX. |
| **Web technologies:** | ASP.NET, MVC Framework |
| **Cloud Technologies:** | Azure. |
| **Azure Stack:** | Azure Data Lake, Data factory, Azure Databricks, Azure SQL database, Azure SQL Data warehouse. |
| **Databases:** | Oracle, SQL-Server, MySQL Server, MS SQL,SAP. |
| **Build Tools:** | ANT, Maven, Gradle, Docker and Jenkins |
| **IDE / Tools:** | Eclipse, IntelliJ, Spring Tool Suite (STS) |
| **Testing/Test Management / Defect Management tools:** | Selenium Web Driver/RC/IDE/Grid, HP Quick Test Pro (QTP), Load Runner, JIRA, Quality Center, ALM, Clear Quest, SOAP UI |
| **Version Control:** | Tortoise SVN, CVS and GIT |
| **Platforms:** | Windows, Mac, Linux and Unix. |
| **Methodologies:** | Agile, Waterfall, Test Driven Development |

**PROFESSIONAL EXPERIENCE:**

**Client: Kennametal INC (Remote, USA) Oct 2023-Till Date**

**Role: Senior Cloud Data Engineer (Migration Project)**

**Responsibilities:**

* Migrated on-premises environment in **Azure Synapse.**
* Worked on Creating Models in Cloud environment with SAP source data.
* Used **Azure** Data Factory extensively for ingesting data from disparate source systems.
* Used **Azure** Data Factory as an orchestration tool for integrating data from upstream to downstream systems.
* Automated jobs using different triggers (Event, Scheduled and Tumbling) in ADF.
* Used **Cosmos DB** for storing catalog data and for event sourcing in order processing pipelines.
* Designed and developed user defined functions, stored procedures, triggers for **Cosmos DB**
* Analyzed the data flow from different sources to target to provide the corresponding design Architecture in **Azure** environment.
* Take initiative and ownership to provide business solutions on time.
* Created High level technical design documents and Application design documents as per the requirements and delivered clear, well-communicated and complete design documents.
* Created DA specs and Mapping Data flow and provided the details to developer along with HLDs.
* Created Build definition and Release definition for Continuous Integration and Continuous Deployment.
* Designed **Snowflake** Schema for Data Warehouse, ODS architecture by using tools like Data Model, Erwin.
* Developed data models and data migration strategies utilizing concepts of the **snowflake** schema.
* Created Application Interface Document for the downstream to create new interface to transfer and receive the files through **Azure** Data Share.
* Creating pipelines, data flows and complex data transformations and manipulations using ADF and **PySpark** with Databricks
* Ingested data in mini-batches and performs RDD transformations on those mini-batches of data by using **Spark** Streaming to perform streaming analytics in Data bricks.
* Created, provisioned different Databricks clusters needed for batch and continuous streaming data processing and installed the required libraries for the clusters.
* Integrated **Azure** Active Directory authentication to every **Cosmos DB** request sent and demoed feature to Stakeholders.
* Experience with AZURE cloud platforms and their data services.
* Improved performance by optimizing computing time to process the streaming data and saved cost to company by optimizing the cluster run time.
* Perform ongoing monitoring, automation and refinement of data engineering solutions prepare complex **SQL** views, stored procs in **Azure** **SQL** DW and Hyperscale
* Designed and developed a new solution to process the NRT data by using **Azure** stream analytics, **Azure** Event Hub and Service Bus Queue.
* Created Linked service to land the data from SFTP location to **Azure** Data Lake.
* Created numerous pipelines in **Azure** using **Azure** Data Factory v2 to get the data from disparate source systems by using different **Azure** Activities like Move &Transform, Copy, filter, for each, Databricks etc.
* Created several Databricks **Spark** jobs with **PySpark** to perform several tables to table operations.
* Extensively used **SQL Server** Import and Export Data tool.
* Created database users, logins and permissions to setup.
* Working with complex **SQL**, Stored Procedures, Triggers, and packages in large databases from various servers.
* Helping team members to resolve any technical issue, Troubleshooting, Project Risk & Issue identification and management.
* Addressing resource issues, Monthly one on one, Weekly meeting.

**Environment**: Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL Server, Teradata Utilities, Windows remote desktop, UNIX Shell Scripting, AZURE PowerShell, Data bricks, Python, Erwin Data Modelling Tool, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Azure Machine Learning., SAP Native HANA, SAP BW and SAP

**Client: Kennametal Apr 2020 – Sep 2023**

**Industry: Sales and Manufacturing.**

**Role: Sr Data Engineer**

**Responsibilities:**

* Migrated data warehouses to Azure Data warehouse.
* Used Data bricks to build Pyspark jobs for ETL transformations on the source data. Used spark’s narrow and wide transformations to transform data as per business requirements.
* Design, develop, and implement ETL (Extract, Transform, Load) processes using Azure Data Factory.
* Built Data mart on Data bricks by creating internal tables. Built partitioned external tables on the data available in Blob storage using mount point.
* Built Pyspark jobs on Data bricks to check for data quality using open-source python package GREAT\_EXPECTATIONS and deployed using Azure DevOps.
* Built Pyspark Streaming jobs using Azure Event hubs and Kafka using Data bricks and loaded the streaming data into Data bricks Delta tables.
* Migrated Hive database to Azure cloud and hosted it on Data bricks. Used both internal and external data bricks tables to mimic the current architecture of hive megastore.
* Used Data bricks Job clusters on ADF for scheduling purposes. Used Azure self-hosted integration runtime to create connection object between Data bricks and Azure Data factory.
* Loaded Coupon information from Azure Blob to Snowflake internal stage using External storage integration. Configured external storage integration on Snowflake to connect with Azure blob using STOARGE\_PROVIDER, AZURE\_TENANT\_ID, and STORAGE\_ALLOWED\_LOCATION parameters.
* Created Notification integration to automate the Snow pipe triggering mechanism. Integrated the Azure Event Grid to send a notification to Snow pipe automatically to trigger the Snow pipe data load process.
* Built Snowflake stage over Azure blob and later mounted Snowflake External table using the stage. Proficient in table design strategy using Data clustering keys.
* Built complex SQL joins on Snowflake to derive data for dash boarding from the DataMart’s. Used COPY INTO command for loading data from local machine (XLS files) into target tables.
* Used the PUT command to load local system files to the Snowflake table stage and built stage tables to join with other staged tables and loaded the final data set into the target table using COPY INTO command.
* Used Snowflake tasks to automate the tasks for data loading, stored procedure runs, and data transformations.
* Performed data migrations from on-prem to Azure Data Factory and Azure Data Lake. Used Kafka and Spark Streaming for data ingestion and cluster handling in real-time processing.
* Built dataflow pipeline on top of Azure Synapse for generic ingestion pattern (CSV files) which are later used by other teams as templates.
* Developed data pipelines and ETL workflows in Azure Synapse Analytics using Azure Data Factory, Azure Data bricks, and Azure Synapse Studio.
* Designed and implemented data models in Azure Synapse Analytics using Azure Synapse SQL and Spark pools.
* Worked with Azure Synapse Analytics components such as Azure Synapse Workspace, Azure Synapse Analytics Server less SQL Pool, and Azure Synapse Analytics Dedicated SQL Pool.
* Built and deployed Azure Synapse Analytics pipelines and workflows using Azure DevOps and Azure Resource Manager templates.
* Integrated Azure Synapse Analytics with other Azure services such as Azure Blob Storage, Azure Event Hubs, and Azure Key Vault.
* Used Azure Event Grid to capture events on Sales from kiosks at stores and processed the Azure Stream Analytics. Azure functions are used to capture the event notification and launch the Data bricks job to process the data.
* Used Azure IOT hub is used to capture the TERMINAL IOT events, the IOT Hub triggers Azure functions. Built Azure function to connect with Cosmos DB and store pre-processed telemetry data. Proficient in tuning the db. for consistency levels.
* Configured Azure Container Registry for building and publishing Docker container images and deployed them into Azure Kubernetes Service (AKS).
* Deﬁned virtual warehouse sizing for Snowﬂake for diﬀerent type of workloads.
* Created tables in snowﬂake DB, loading and analyzing data using Spark-Scala scripts.
* Developed ETL pipelines in and out of data warehouse using combination of Python and azure SQL.
* Written POCs in Python to analyze the data quickly before applying big data solutions to process at a scale.

**Environment**: Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL Server, Teradata Utilities, Windows remote desktop, Data bricks, Python, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Azure Machine Learning.

**Client: Kennametal**

**Industry: Sales and Manufacturing. April 2016 – Mar 2020**

**Role: Data Engineer**

**Responsibilities:**

* Designed and developed a real-time matching solution for customer data
* Design end-to-end data integration solutions based on business needs.
* Work with different data sources and destinations, ensuring efficient data movement and transformation.
* Troubleshoot and optimize data pipelines for performance and reliability.
* Manage and optimize data integration runtimes.
* Investigate and resolve issues related to data pipelines and data integration processes.
* Analyze and troubleshoot errors in data movement and transformation.
* Collaborate with support teams and vendors to resolve technical issues.
* Administer and monitor Azure Data Factory instances.
* Collaborate with infrastructure teams to ensure proper resource allocation.
* Created Different tableau reports using Live and Extract Connection.
* Involved in developing weekly, daily, monthly, yearly Tableau reports and sending the reports according to the business requested format.
* Specify permissions for Tableau workbooks, this includes setting up who can see or modify data, dashboards, and reports. You can set these permissions at the workbook.
* Proficiency in rapidly creating precise ad hoc reports.
* Participated in code reviews and version control processes to maintain code quality and foster collaboration among team members.
* Designed and implemented efficient database structures to enhance data organization and retrieval.
* Created Multiple SQL complex queries, stored procedures and optimizing database performance.
* Proficient in End-to-End design of Business Intelligence solutions to meet global and regional reporting needs.
* Developed and implemented indexing strategies to enhance query performance for fast data retrieval.
* Collaborated with data suppliers to establish data quality standards and processes, ensuring the accuracy and reliability of incoming data.
* Interacted with consumers to gather requirements and feedback, customizing data views to meet their specific needs and enhance data utilization.
* Created and maintained database schemas to ensure data integrity and compatibility with application requirements.
* Preserved robust backup and recovery measures to secure data and ensure business continuity in case of failures.
* Involved in UAT testing of report code to ensure fulfilment of requirements and accuracy.
* Played a key role in supporting divestitures and acquisitions by facilitating the smooth transfer of data, both outbound and inbound to ensure business continuity and data integrity.
* Extensive experience in implementing Complex queries using SQL server and create SSRS reports for data utilization to the business to make data easier to understand.
* Ensured strict adherence to audit and compliance requirements, maintaining data security, integrity, and regulatory compliance at all times.
* Drove the standardization of data processes and automation of routine tasks, leading to increased efficiency, reduced errors, and enhanced data consistency.
* Experience in Performance tuning scripts in Stored Procedures to extract large datasets.
* Participated in on-call support rotation to address and resolve data-related issues outside of regular working hours, ensuring uninterrupted data operations.
* Creating the change request (CHG) in service-now for any code deployment to production servers.
* Maintained reports and dashboards in tableau and power BI based on business needs.
* Understand and meet business needs and extract the reports from SQL Server management using Custom SQL queries or views and send to the business.

**Environment:**  MS Access, SQL Server, SSIS, SSRS, Azure, Tableau, Excel,SAP.

**Client: Automotive Manufacturer Pvt Ltd. Apr 2014 – Mar 2016**

**Role: Data Engineer**

**Responsibilities:**

* Performed migration of Reports (Crystal Reports, and **Excel**) from one domain to another domain using Import/Export Wizard.
* Wrote complex **SQL**, PL/**SQL**, Procedures, Functions, and Packages to validate data and testing process.
* Used advanced Excel formulas and functions like Pivot Tables, Lookup, If with and/index, match for data cleaning.
* Redesigned some of the previous models by adding some new entities and attributes as per the business requirements.
* Reviewed Stored Procedures for reports and wrote test queries against the source system (**SQL Server**) to match the results with the actual report against the Data mart (**Oracle**).
* Involved with data profiling for multiple sources and answered complex business questions by providing data to business users.
* Performed **SQL** validation to verify the data extracts integrity and record counts in the database tables.
* Created Schema objects like Indexes, Views, and Sequences, triggers, grants, roles, Snapshots.
* Effectively used data blending feature in **Tableau** to connect different databases like **Oracle**, MS **SQL Server**.
* Transferred data with **SAS/Access** from the databases **MS Access**, **Oracle** into SAS data sets on Windows and **UNIX**.
* Provided guidance and insight on data visualization and dashboard design best practices in Tableau.
* Performed Verification, Validation and Transformations on the Input data (Text files) before loading into target database.
* Executed data extraction programs/data profiling and analyzing data for accuracy and quality.
* Wrote complex **SQL** queries for validating the data against different kinds of reports generated by Business Objects.
* Documented designs and Transformation Rules engine for use of all the designers across the project.
* Designed and implemented basic **SQL** queries for testing and report/data validation.
* Used ad hoc queries for querying and analyzing the data.
* Performed Gap Analysis to check the compatibility of the existing system infrastructure with the new business requirements.

**Environment:** SQL, PL/SQL, Oracle9i, SAS, Business Objects, Tableau, Crystal Reports, T-SQL, SAS, UNIX, MS Access 2010

**Client: Reliance Retail. India July 2012 – Mar 2014**

**Role: Data Engineer**

**Responsibilities:**

* Developed Mappings to extract data from Oracle database and transferred aggregated data into summary tables using ETL Tool.
* Implement complex database queries using strong and advanced SQL skills.
* Obtained initial requirements from the users of various departments and corresponded with them throughout the process.
* Created DDL scripts to create database schema and database objects like tables, stored procedures, views using T-SQL.
* Involved in creating ETL Flows using multiple transformations to load data into the staging & data warehousing layer.
* Scheduling email alerts at each stage of ETL process as part of monitoring.
* Responsible to monitor jobs and take actions as per severity.
* Responsible for develop, deploy and monitoring SSIS packages for ETL Processes and created scheduled jobs through SQL server job agent as per the requirement.
* Extensively used Extract Transform Loading (ETL) tool of SQL server to populate data from various data sources.
* Experience in loading files to SQL server using transformations in SSIS.
* Developed and maintained ETL packages to extract the data from various sources and Responsible for debugging and upgrading of several ETL structures as per the requirements.
* Experience in managing and automating Control flow, Data flow, Events using SSIS packages.
* Used SSIS and T-SQL stored procedures to transfer data from OLTP databases to staging area and finally transfer into data warehouse.
* Used various SSIS task such as conditional split, derived column, lookup which were used for data scrubbing, data validation checks during staging, before loading the data into the data warehouse.

**Environment:**SQL server, Oracle DB, CSV files.