**Suchendra**

**Sr Azure Data Engineer**

**PROFESSIONAL SUMMARY**

* **Microsoft Certified Azure Data Engineer** with **10+** years of overall technical experience in **information** **technology** as a **data** **engineer** participated in **design**, **development**, **implementation**, and enhancement of projects including requirement **gathering**, **analysis**, **process** **mapping**, **solution** **design** and **unit** **testing**.
* Experience in implementing **Azure** **Data** **Solutions**, Provisioning **storage** **account**, **Azure** **Data Factory**, **SQL** **Server**, **SQL** **Databases**, **SQL** **Data** **Warehouse**, **Azure** **Databricks**, **Azure** **Synapse** **Analytics**, and **Azure Cosmos DB.**
* Hands-on experience in developing large scale applications using **Big** **Data** **Ecosystem** technologies such as **Hadoop, Map Reduce, Pig, Hive,** and **Spark.**
* Hands-on experiences in **Hadoop** ecosystem components like **HDFS**, **Cloudera**, **YARN**, **Hive**, **HBase**, **Sqoop, Flume, Kafka, Impala** and Programming in **Spark** using **Python**.
* Experience in designing end-to-end **ETL** Strategy utilizing **SSIS** as an **ETL** tool and authored a multitude of **SSIS** packages for **data** **migration** purposes.
* Extract Transform and **Load** **data** from various source systems to **Blob** **storage**, **Cosmos** **DB**, **Azure** **SQL**, and **Azure DW (Synapse)** services using a combination of **Azure** **Data** **Factory, T-SQL, U-SQL**
* Experience in the development of **Oracle T-SQL, PL/SQL** Scripts, Stored Procedures, and Triggers for business logic implementation.
* Good experience in migrating **SQL** databases to **Azure Data Lake, Azure Data Lake Analytics**, **Azure SQL Database, Data Bricks**, and **Azure** **SQL** **Data** **Warehouse**.
* Proficient in creating, managing, and maintaining **CI/CD** pipelines to drive efficient and reliable.
* data engineering workflows and Implemented **CI/CD** **best** **practices**, security scanning/monitoring, **pipeline** **integration**.
* Strong experience in the **Analysis** **design**, **development**, **testing**, and implementation of Business Intelligence solutions using **Data** **Warehouse**/**Data** **Mart** **Design**, **ETL**, **BI**, **Client**/**Server** applications and writing **ETL** scripts using **Regular** **Expressions** and custom tools (**Informatica**, **Pentaho**, and **Sync** **Sort**) to **ETL** **data**.
* Experience in **extracting**, **transforming**, and **loading (ETL)** **data** from spreadsheets, **database** **tables** and other sources using **DataStage**, **Informatica**, **SQL Server Integration Services (SSIS)** and **SQL Server Reporting Service (SSRS)** for managers and executives.
* 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 the customer usage patterns.
* Experience with **data pipeline** building, **backend microservice** development, and **REST API** using **Python**, **Java**.
* Hands on experience in scheduling data ingestion process to data lakes using **Apache Airflow**.
* Proficient in **DevOps** practices with **Jenkins** for **CI/CD**, **Git** for source control (including **Git** **Flow**), and **Atlassian** tools **(Jira, Bitbucket, Source tree)** for **project** **management** and **version** **control**.
* Developed a **data** **pipeline** using **Kafka** stream to ingest data from their weblog server and apply.
* the transformation.
* Proficient in working on **NoSQL** technologies like **HBase**, **Cassandra**, and **MongoDB**.
* Designed and Built **ETL** pipelines using **Azure** **Databricks** and ingested data from different sources like Snowflake, **Teradata**, **Vertica**, **Oracle** **ETC** to **azure** **data** **lake** **storage** and monitored ongoing Databricks jobs.
* Experience in **Extract, Transform, and Load (ETL)** processes on **AWS**, optimizing workflows, ensuring **data** **quality**, and achieving high performance while leveraging **Azure** **Synapse** **Analytics** for efficient storage and retrieval.
* Defined **user** **stories** and driving the **agile** board in **JIRA** during **project** **execution**, participated in sprint demo and retrospective.
* Proficient in **writing**, **implementing**, and **testing** of triggers, procedures, and functions in **PL/SQL,** **Oracle**, and good command over programming languages like **Python**, **Shell** **Scripting**.
* Experienced in utilizing **Agile** **methodologies** such as **SCRUM** and **Waterfall** methodologies throughout the **Software Development Lifecycle (SDLC).**

**EDUCATION**

* **Bachelor of Electronics & Communication Engineering** from Sri Chandrasekarendra Saraswathi Viswa Maha Vidyalaya.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **ETL Tools & Technologies** | Azure Data Factory (ADF), SSIS, DataStage, Talend, Informatica, Pentaho, Sync Sort, Spark-SQL, U-SQL, SQL Server Reporting Services (SSRS), Snowflake, Vertica, Oracle, DB2 |
| **Azure Data Solutions** | Azure Data Factory (ADF), Azure Synapse Analytics, Azure Databricks, Azure SQL Database, Azure Data Lake, Azure SQL Data Warehouse (DW), Azure Cosmos DB, Azure Blob Storage, Azure Event Hub, Azure Stream Analytics, Azure Key Vault, Azure Active Directory, Azure Monitoring, Azure Search, Azure Analysis Services, Azure Logic Apps |
| **Big Data & Cloud Technologies** | Hadoop Ecosystem (HDFS, YARN, Hive, HBase, Sqoop, Flume, Kafka, Impala), Apache Spark (PySpark, Spark-SQL), Cloudera, HDInsight, Azure Databricks, AWS S3, AWS Lambda, AWS Data Transfer |
| **Programming & Scripting Languages** | Python, T-SQL, PL/SQL, Shell Scripting, SQL |
| **NoSQL Databases** | MongoDB, Cassandra, HBase |
| **CI/CD & DevOps Tools** | Jenkins, Git, GitFlow, Atlassian Jira, Bitbucket, SourceTree, Azure DevOps, TFS |
| **Data Modeling & Data Migration** | Snowflake Schema, Data Modeling, Source to Target Mappings, ETL Strategy, Data Migration (Lift and Shift), Data Lake Ingestion |
| **Database Management** | Azure SQL Database, Azure SQL Synapse Analytics (Data Warehouse), Oracle, SQL Server, DB2, Cosmos DB |
| **Data Pipelines & Integration** | REST APIs, Flask, JSON, Spark Applications, Kafka Streams, Data Migration Frameworks, Data Transformation, Data Aggregation, Data Ingestion, Batch/Real-time Data Processing |
| **Business Intelligence & Reporting** | Power BI, SSRS, Self-service BI Tools, Data Visualization, Business Intelligence Solutions, Data Quality Analysis |
| **Version Control & Project Management** | GIT, TFS, Azure DevOps, Atlassian Jira, Bitbucket, SourceTree |

**PROFESSIONAL WORK EXPERIENCE**

**Client: Citizens Bank, Johnston, Rhode Island Feb 2020 – Present**

**Role: Sr Azure Data Engineer**

**Responsibilities:**

* Extract **Transform** and **Load** **data** from Sources 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 SQL, Azure DW**) and processing the data in **Azure** **Databricks**.
* Design and Implementing the **Data** **Migration** framework with appropriate data load process and sequencing (one-time and incremental loads) using **Azure Data Factory**, **Azure** **Blob** Storage and orchestrate data from On-Premises systems using **Azure** **Data** **Factory** by creating **Azure** **Data** Pipelines and **Data** **Flows**.
* Have good experience working with **Azure** **Blob** and **Azure** **data** **lake** storage and loading **data** into **Azure SQL Synapse Analytics (DW).**
* Developed **data** **pipelines** connecting on-premises and cloud sources using **SSIS** and **Azure** **Data** **Factory**.
* Used **PySpark** and **Spark**-**SQL** to clean, transform and aggregate data with proper file and compression types as per requirement before writing data to azure data lake storage.
* Implemented a **CI/CD** framework for data pipelines using **Azure DevOps**, enabling efficient automation and deployment.
* Developed an executable application that securely transfers files and creates folders in **AWS S3** and created **Lambda** functions in **AWS** for Application development to manage **S3** security.
* Created **Restful** **API** using **Flask** to integrate functionalities & communicate with other applications.
* Worked on different data sources such as **Teradata**, **Oracle**, **Flat** **files**, etc.
* Developing **JSON** Scripts for deploying the Pipeline in **ADF** that processes the data.
* Responsible for estimating the **cluster** **size,** monitoring, and troubleshooting of the **Spark** data bricks **cluster**.
* Analyzed existing systems and propose improvements in processes and systems for usage of modern scheduling tools like Airflow and migrating the legacy systems into an enterprise data lake on Azure Cloud.
* Worked on **Microsoft Azure services** like **HDInsight Clusters**, **BLOB**, **Data Factory** and **Logic** **Apps** and did **POC** on **Azure** **Data** **Bricks**.
* Written templates for Azure Infrastructure as Code (IaC) using Terraform to build staging and production environments. Integrated Azure Log Analytics with Azure VMs for monitoring the log files, store them and tract metrics and used terraform as a tool.
* Worked on **Snowflake** **Schema**, **Data** **Modeling** and elements and Source to **Target** **Mappings**, Interface **Matrix**, and **design** **elements**.
* Performed **data** **quality** issue analysis using **Snow** **SQL** by building analytical warehouses on **Snowflake**.

**Environment**: AWS, S3, Azure Data Factory, T-SQL, Spark SQL, Oracle, Snowflake, Azure Data Bricks, DataStage, Spark, Pyspark, Spark-SQL, Azure SQL Synapse Analytics, Hive, CI/CD, ADF, JSON, BLOB.

**Client: Allstate Insurance Company, Bangalore, India Sep 2018 - Jan 2020**

**Role: Data Engineer**

**Responsibilities:**

* Design and implement database solutions in **Azure** **SQL** **Data** **Warehouse**, **Azure** **SQL**.
* Architect & Implement medium to large scale **BI** solutions on **Azure** **using** **Azure** **Data** **Platform** **services** **(Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).**
* Design and implement migration strategies for traditional systems on **Azure (Lift** and **Shift**/**Azure** **Migrate**, other third-party tools).
* Engage with business users to gather requirements. **Design** **visualizations** and provide training to use self-service **BI** tools.
* Used various sources to pull **data** into **power** **BI** such as **SQL** **Server**, **Excel**, **Oracle**, **SQL** **Azure** etc.
* Propose architectures considering **cost/spend** in **Azure** and develop recommendations to right size **data** **infrastructure**.
* Develop conceptual solutions & create **proof**-**of**-concepts to demonstrate viability of solutions.
* Technically guide projects through to completions within target timeframes.
* Collaborate with application architects and **DevOps**.
* Identify and implement best practices, **tools**, and standards.
* Design, Setup, maintain, and administer the **Azure** **SQL** **Database**, **Azure** **Analysis** **Service**, **Azure** **SQL** **DW, Azure Data Factory.**
* Build complex distributed systems involving huge amounts of **data** **handling**, **collecting** **metrics**, building **data pipelines, and Analytics.**

**Environment**: Azure Data Factory, Azure Data Bricks, Azure Data Lake

**Client: Max Life Insurance, Bangalore, India Jul 2014 - Aug 2018**

**Role: Software Engineer**

**Responsibilities:**

* Involved in **Extraction**, **Transformation**, and **Loading** of data.
* Understanding **project** **requirements** by actively participating in regular meetings with customers and internal teams.
* Worked on **ETL** operations in Talend using **Database** **components**, **Processing** Components, File components, Misc components like **tDBConnection, tDBInput, tDBOutput, tFilterRow, tSortRow, tUniqueRow, tAggregateRow, tMap, tNormalize, tDenormalize, tExtractDelimitedFields, tRunJob, tJavaRow, tFileInput&Output, tFilelist, tReplicate, tFileCopy, tDie, tFileCopy, tSendmail, tPrejob, tPostjob, etc.**
* Experienced in working with context variables and **Hash** and **Buffer** Components.
* Worked on error handling techniques by using **tLogCatcher** Component.
* Running sub jobs in parallel using the **tParallelize** component.
* Implemented Slowly Changing **Dimensions** **Type1** and **Type2**.
* Worked on Flat files like **.txt** and **.csv files** and done transformation by proper components like **tFileList,** **tFileInputDelimited** and **tFileOutputDelimited**, **tFileOutputPositional** Components.
* Worked on Exporting and Importing of jobs and involved in Performance improvement techniques in the **ETL** flow.
* Experienced in working with various transformations like **Filtering**, **joining**, **aggregate** operations, update **strategy**, **sequence** **generator** etc.
* Worked on different phases of the **software** **development** **lifecycle** from **development**, **testing**, and deployment of the project.

**Environment**: Talend Cloud 7.2.1, MS SQL Server, Oracle

**Certification**

**Microsoft Certified: Azure Data Engineer Associate (DP-203)**