**Venkata**

**PROFESSIONAL SUMMARY:**

* Around 12 years of industry experience in Data warehousing and Data Engineering with wide range of large Applications including Design, Development on end-to-end data flow for deliver business solutions using Microsoft SQL Server Integration Services and **Big Data Analytics** using **Databricks** and **Synapse Analytics** including **HIVE, Spark, Snowflake.**
* Hands - on experience in Azure Cloud Services (**PaaS & IaaS**), Azure SQL, Azure Data Factory, Azure Log Analytics, Azure Monitoring, Azure Key Vault, Azure Data Lake.
* Experience on Migrating SQL database to **Azure data Lake**, **Azure SQL Database, Azure Databricks, Delta Lake** and **Azure SQL Data warehouse** and controlling and granting database accessandMigrating On premise databases to **Azure Data Lake store** using Azure Data factory.
* Experience Life Cycle Management in Azure Data Lake and Using different Metrics and Alerts for optimizing the storage cost.
* Experience in writing medium to complex SQL queries for data analysis. Experience in Developing **Spark** applications using **Spark - SQL** in **Databricks** for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Expert in data Extraction, Transforming and Loading **(ETL)** using various tools such as **SQL Server** Integration Services **(SSIS).**
* Experience in Designing and Building the Dimensions and cubes with star schema using SQL Server Analysis Services. Developing reports in SSRS on top of Cubes and understanding of SSAS, **OLAP** cube and Architecture Experience in report writing using SQL Server Reporting Services (SSRS) and creating various types of reports.
* Excellent understanding of Relational Database Systems, Normalization, logical and physical data modeling and create automated workflows with help of triggers.
* Hands on experience and Managing **Databricks Notebooks**, Delta Lake with Python, **Delta Lake** with **Spark SQL.**
* Expert in defining and deploying Cubes using SQL Server Analysis Services (SSAS).
* Experience with Requests Prophet and Imblearn python libraries during development lifecycle.
* Experience implementation of a **CI/CD pipeline** using **Azure** DevOps (TFS) in both cloud and on-premises with GIT.
* Developed **stored procedures** and queries using **T/SQL** and proficient in databases **including ORACLE, MS SQL server, and Netezza.**
* Extensively used ETL methodology for testing and supporting data extraction, transformations and loading process, in a corporate-wide-ETL Solution **using SSIS.**
* Experienced in **NoSQL** technologies like **HBase** and relational databases like **Oracle, Netezza, a**nd **MS SQL** databases.
* Well versed with **Agile** with **SCRUM**, Waterfall Model and Test-driven development methodologies.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Versioning** | SVN, TFS |
| **Languages** | Pyspark, SQL and PL/SQL and Log Analytics Kusto |
| **Databases** | Oracle, My SQL, MS SQL, Netezza |
| **IDEs/ Tools** | PyCharm, SQL Server Integration Services, Azure Data Factory |
| **Version Control** | GitHub |
| **Big Data Tools** | Azure Synapse Analytics and Azure Databricks (Pyspark and Spark SQL), Snowflake,  Log Analytics, Azure Key vault |
| **Reporting Tools** | SQL Server Reporting Services and Power BI |

**PROFESSIONAL EXPERIENCE:**

**Client:** Humana Louisville, KY

**Duration:** July 2022 – Present

**Role:** Data Engineer

**Responsibilities:**

* Actively involved in end-to-end data migration solution for migrating on premises IBM Netezza to Azure Synapse Analytics.
* Recommending SQL Pool, Serverless SQL for different use cases to client and Creating workbooks using Log Analytics for monitoring storage account usage, different metrics and performance monitoring of Dedicated Synapse SQL pool.
* Building log DataMart for cost optimal in all synapse components (**serverless SQL pool**, **dedicated SQL pool**, **spark, storage** and **pipeline** logs detail monitoring on user level with help of Microsoft Cost API)
* Creating a standard workspace and configured it as per best practices that will act as template for deploying it
* Migrating stored procedure which is written in IBM Netezza into Azure Synapse Analytics.
* Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* Creating Databricks notebooks using Spark SQL, Pyspark for transforming large JSON files.
* Created Synapse Pipelines for orchestrating data from different source systems into ADLS landing zone with 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.
* Worked with complex SQL, Stored Procedures, Triggers and packages in very large databases from various servers.
* Worked on **bicep** to automate the **Azure** Synapse Workspace deployment and deployed virtual machine scale sets in production environment.
* Worked on data processing and transformations and actions in spark by using Python (**Pyspark**) language.
* Written Templates for **Azure** Infrastructure as code using **bicep** to build staging and production environments.
* Perform Data Cleaning, features scaling, features engineering using pandas and NumPy packages in **python**.
* Created
* Create several types of data visualizations using **PowerBI**.
* Create end-to-end solution for ETL transformation jobs that involve writing Informatica workflows and mappings for one of the business use cases.
* Helped individual teams to set up their repositories in bit bucket and maintain their code and help them setting up jobs which can make use of **CI/CD** environment.
* Worked on stored procedures/views in Synapse dedicated SQL Pool.
* Primarily responsible for designing and optimizing Azure Synapse Analytics platform by monitoring and building reliable data solutions to move data from across systems from multiple sources in real as well as Batch load.
* Providing production support for large amount of Enterprise grade Data Warehouse on Azure while adhering to SLA and submitting Root Cause Analysis.

**Client:** EY, Goldman Sachs

**Duration:** Oct 2021 - June 2022

**Role:** Data Engineer

**Responsibilities:**

* Developing and managing **Azure Data Factory** pipelines responsible for end-to-end orchestration from copying data from source to target.
* Creating Data Factory pipelines for on-cloud ETL processing using **copy activity** and mapping data flow. Implementing automatic file processing using event-based trigger.
* Optimizing Azure data factory pipeline using DIU’s and Data Parallelism for large datasets.
* Creating Storage Accounts, Integration Runtimes, Linked Services.
* Integrating the end-to-end data pipeline to take data from source systems to target data repositories
* Creating azure key vault for maintaining the scope of the **secrets**.
* Processing JSON files using Azure data factory mapping data flow activities.
* Processing large datasets using Azure Databricks using Pyspark.
* Handling all data quality issues and error handling using Databricks.
* Redirect error rows in ADF using Data flow activity into Blob Storage.
* Analyze bad data and perform additional validations based on corrupted data and provide work around fix and process data and Responsible for Processed layer which has used for reporting.
* Building reusable notebooks for automating common processes using Functions.
* Debugging stored procedure and identifying data which is not coming as expected or causing failure in load.
* Involved in Sprint Backlog refinement meetings to groom the stories for next sprints.
* Involved in moving the pipelines from Dev to QA branch using ARM templates part of **CI/CD** process.
* Monitor Data Factory pipelines using **Azure Monitor** and **Log Analytics**.

**Client: Cognizant,** **The Hartford Financial Services, Hartford, Connecticut, USA**

**Nov 2017 – Oct 2021**

**Role: Data Engineer**

**Responsibilities:**

* Involved Designing and loading of large unstructured datasets (XML and JSON) in to hive tables using Python.
* Involved in loading the data from RDBMS (MySQL and Oracle) to HDFS Using SQOOP.
* And involved in PowerShell scripting to generate multiple files in different locations.
* Involved in Designing ETL packages and Implementation of SCD, Azure, CDC, ETL Optimization and Data Cleansing using SQL Server Integration Services 2014.
* Involved in creating of Data Marts and Dimensional Data Modeling, Star Schema Modeling, Snow- Flake Modeling, FACT and Dimensions Tables using SQL Server Analysis Services (SSAS).
* Involved in creating multidimensional objects (Cubes, Dimensions) using SQL server Analysis services (SSAS) and performed designing and developing extensive dashboards using Power BI and DAX.
* Involved in creating tabular model to develop data analysis prototype to create Power Map visualize reports.
* Involved in implementation of different types of graphical reports like Cross Tabs, Azure Scatter Plots, Geographic Maps, Heat Maps, Gauge, Tree Maps visualize reports using Power BI and performed fine tuning of the SQL queries and implementation of Indexes for faster performance and Implementation of Partitions, MDX for cube performance.

**Client: Cognizant, CoreLogic**

**Duration Nov 2014 – Nov 2017**

**Role: Data Engineer**

**Responsibilities:**

* Involved Designing and loading of large unstructured datasets (XML and JSON) in to hive tables using Python.
* Involved in loading the data from RDBMS (MySQL and Oracle) to HDFS Using SQOOP.
* Involved in PowerShell scripting to generate multiple files in different locations.
* Involved in Designing ETL packages and Implementation of SCD, CDC, ETL Optimization and Data Cleansing using SQL Server Integration Services 2014.
* Involved in creating of Data Marts and Dimensional Data Modeling, Star Schema Modeling, Snow- Flake Modeling, FACT and Dimensions Tables using SQL Server Analysis Services (SSAS).
* Involved in creating multidimensional objects (Cubes, Dimensions) using SQL server Analysis services (SSAS) and performed designing and developing extensive dashboards using Power BI and DAX.
* Involved in creating tabular model to develop data analysis prototype to create Power Map visualize reports.
* Involved in implementation of different types of graphical reports like Cross Tabs, Scatter Plots, Geographic Maps, Heat Maps, Gauge, Tree Maps visualize reports using Power BI and performed fine tuning of the SQL queries and implementation of Indexes for faster performance and Implementation of Partitions, MDX for cube performance.

**Client: Birlasoft, AXA**

**Duration: July 2011 – Oct 2014**

**Role: ETL and Reporting Developer**

**Responsibilities:**

* Involved in creating functions and stored procedures
* Involved in creating packages
* Involved with designing of the reports and creation and design document.
* As per Client requirements, developed the reports in the form of Matrix, Table and Chart using SQL Server 2008 Reporting Services
* Involved assign Folder level security for each Group and Writing Expressions while calculate complex calculations.
* Involved in assign the security to each report.
* Performance tuning of sql statements and database
* Involved in writing test cases for each canned reports and Unit Testing done for the same
* Involved in Creating Report model for end customers.
* Involving in creating packages using various components of SSIS like Control Flow Tasks and DFT to implement Business functionality.
* Involved in implementing slowly changing Dimensions.
* Involving in using different tasks of control flow like sequence, For Each Loop, For Loop Container task, Script task, Sent-mail task, Execute SQL Tasks etc., and in DFT level using different Sources and Destinations and transforming the data-by-Data Merging, Data Cleansing to load the data into multiple data marts.
* Build and deployed packages to SQL server location and created configurations and breakpoints to trace errors.
* Involving in configuring packages, Logging and event handling at different levels, and involved in bottleneck process of the packages. Prepared deployment document for the DBA Team.
* Involving in deploying and scheduling SSIS packages in different environments
* Involved in the ETL Framework which consist of five stages in Every SSIS Package.
* Involved in understanding requirements from the client and design ssis packages.
* Involved in Production Support for one month and handled Some Tickets (Issues).
* Involved in Migration SSIS Packages from SQL Server 2005 to SQL server 2008.
* Scheduling And Monitoring ETL packages for Daily load in SQL Server Job Agent

**EDUCATION:**

|  |  |  |
| --- | --- | --- |
|  | **Aug 2007 - Jul 2010** | Master of Computer Applications: Computer Applications  *BVCITS Engineering College - Amalapuram, AP* |

|  |  |  |
| --- | --- | --- |
|  | **Jul 2004 - Apr 2007** | Bachelor of Science  *Abhyudaya Degree College – Srikakulam, AP* |

|  |  |  |
| --- | --- | --- |
|  | **Jun 2002 - Apr 2004** | M P C  *Sri Sai Krishna Jr College - Srikakulam, AP* |

|  |  |  |
| --- | --- | --- |
|  | **Jun 2001 - Mar 2002** | High School Diploma  *RCM St Johns High School - Srikakulam, AP* |
|  |  |

**Certifications:**

AZ - 900 Azure Fundamentals

70 - 448 SQL Server Database