**GATTU NAVANEETH RAO** ✉**:** Navaneeth.dwh@gmail.com

☏: +91- 7306389971

**5 Years’ Experience as Azure data engineer + ETL**

**Professional Summary: .**

* Over all 5 Years of IT experience on data driven application design and development with 4.4 years of relevant hands-on-experience in various Azure data services..& Talend (6.2 & 7.3 enterprise big data edition.)
* Proficient in Azure technologies such as **Azure Data Factory (ADF), Azure Data Bricks (ADB), Azure Synapse Analytics,** Azure Active Directory, Azure Storage, Azure data Lake Services (ADLS), Azure key vault, Azure SQL DB, Azure HDInsight.
* Hands on Experience developing data engineering frameworks and notebooks using Azure Data Bricks using **Spark SQL, Scala, Pyspark**
* Having extensive experience working with Unity catalog, Autoloader, Delta Live Tables(DLT).
* Created workflows for automating the tasks.
* Experience in Apache Hadoop frameworks such as **Hdfs, Map Reduce, Hive** etc.
* Experience in Microsoft Azure Cloud with Data Factory, Linked Services, HDI Cluster ,Data Lake Gen2and Data Bricks.
* Good knowledge on **Azure synapse Analytics.**
* Proficient in Big data ingestion tools like Kafka, spark streaming and Sqoop for streaming and batch data ingestion.
* Worked with Big Data distributions like **Hortonworks** (Hortonworks 2.1) with Ambari.
* Hands-on experience in application development using Java, RDBMS, and Linux shell scripting.
* Hands-on experience working with IDE tools such as Eclipse, NetBeans, and Maven.
* Worked on **Tableau and Power BI** to generate reports.
* Experience in the full-cycle software development including requirements gathering, prototyping, and proof of concept, design, documentation, implementation, testing, and maintenance.
* Good Understanding of Dimensional Modelling, Star and Snowflake Schemas
* Created Informatica mappings to load data using transformations like source qualifier, sorter, aggregator, expression, joiner, connected and unconnected lookup, filters, sequence generator, router and Update strategy.
* Experienced in doing the process of extraction, transformation and loading data from multiple sources like flat files and OLTP systems into data marts using Informatica Power Centre.
* Involved in creation of using different kinds of Sources and Destinations.
* Experience in developing all types of reports like Table, Matrix, Charts, linked reports and Sub Reports.
* Experience in Deploying SSRS Reports on Report Manager using SSDT and good understanding knowledge Report Scheduling from Report Manager.
* Good knowledge on OLAP Concepts and SQL Server.

# 

# AI & ML

* Working experience with data science python libraries like (NumPy, Pandas, Kera’s, Theano, Caffe, Open AI Gym, SciPy, Tensor flow…. etc.).
* Created models like regression, classification, LSTM, CNN (Convolutional neural network), ANN (Artificial neural network), Clustering model, N degree polynomial model…. etc.)
* Did some explorer towards problem solving using AI.
* Emotion AI: Detecting facial expressions and declaring emotion of human with CNN, ANN and ResNet
* AI in health care. (Classifier model): To localize brain tumors and giving treatment early to avoid serious injuries and helpful to save lives.
* AI in business marketing (K – means clustering) : Understanding the business in market like customer interests.
* Creative AI (Deep dream algorithm): Applying AI in culture, creating paintings, write compelling stories and compose a new music.
* Ability to adapt any new technology quickly.

**Professional Experience:**

* Worked as a Associate at “Wipro Ltd.” from Nov 2019 – Oct 2021.
* Working as a Azure data engineer at “Luxoft a DXC Technology Company” from Oct 2021 – Till Date.

**EDUCATIONAL QUALIFICATION**

* B. Tech from Jbiet.

**TECHNICAL SKILLS:**

* **AZURE**: Azure Data Factory (ADF), Azure Data Bricks, Unity Catalog, Delta lake, Azure Data Lake Services (ADLS), Azure Blob Services, Azure SQL DB, Azure Active Directory (AAD), Azure Devops.
* **Languages**: Scala, Core Java, Python
* **Databases**: Azure SQL, My SQL, Postgres SQL, Oracle.
* **No SQL Databases:** HBase, MongoDB.
* **Data warehouse:** *Hive, Azure Synapse, redshift, Snowflake.*
* **Big data Technologies:****A***pache Hadoop, Apache Hive, Apache Spark*
* **ETL Tools:** *Azure Data Factory (ADF V2), Talend 6.4/7.3/ Big Data Enterprise & open Source), Informatica*
* **Data ingestion***:* Sqoop, Kafka, Spark Streaming
* **Data Visualization:** Tableau, power BI

**PROJECT DETAILS:**

|  |  |
| --- | --- |
| **Project #1: EDWD** | |
| **Client: VFS Global** | |
| Environment | **PySpark, python, Unity Catalog, data bricks, ADF, Azure, Delta, Autoloader, spark Streaming** |
| Role | Azure Data Engineer |
| **Description:**  **VFS Global**, also known as "**Visa Facilitation Services Global,"** is an outsourcing and technology services company for governments and [diplomatic missions](https://en.wikipedia.org/wiki/Diplomatic_mission) worldwide. The company manages visa and passport issuance-related administrative and non-discretionary tasks for its client governments. | |
| **Roles and Responsibilities**   * Developed pipelines in azure data factory to fetch data from different sources loaded it into Azure synapse analytics. * Extensively involved in writing transforming data in PySpark and pushing into ADLS. * Experience working with Unity Catalog for data governance like Sharing Data across workspaces, providing access, masking data, storing tables, views, functions and volumns..etc * Worked on Azure Devops to setup infrastructure and to build the and deploy application. * Implemented Incremental load strategy for loading on daily basis. * Having extensive experience working with unity catalog. * Worked on Autoloader for streaming data. * Working experience with delta tables. * Automated the code by creating workflow tasks and with data factory by calling notebook activities, * Handled the releases for every deployment and the unit/integration testing. * Responsible for creating Linked services, datasets, pipelines in Azure data factory. * Created stored procedures using TSQL. * Responsible for creating copy activity, lookup activity, Metadata activity, * Monitor the pipelines are report fix the issues. * Creating data flows to transforming the data to Azure using Azure data factory. * Developed complete framework of the project end to end and met customer’s deadlines.   Proactively.   * Mapped columns dynamically with respect to input file in copy activity reading JSON code from SQL control table and reading in lookup activity. * Captured the logs and execution statistics of the pipeline (Source records count and target records count) with the help of stored procedure and passing parameters of stored procedure with respected activities. * Added triggers for the pipeline execution (Custom Events trigger, Storage Events trigger, tumbling window) and monitored triggered pipeline. * Created notebook and used notebook activity in ADF. * Mounted Azure data lake storage Zen 2 in notebook and did transformations using spark SQL data frames and loaded transformed data into Azure SQL using JDBC Connection. * Implemented the data flow logic to do transformation in azure data factory. * Implemented the dynamic data set components to work for different objects. * Trigger the pipelines based on schedule intervals. * Prepared the external tables, dimension and fact tables in azure SQL based on user requirements. * Prepared the business logic using stored procedures in azure SQL analytics based on user requirements. * Involved in bug fixing. | |

|  |  |
| --- | --- |
| **Project #2: HPI- Tera Restatement** | |
| **Client: HPI** | |
| Environment | **Azure data factory, Azure data bricks, Pyspark, Unity Catalog, Spark Streaming, Delta Live Tables(DLT), Azure SQL DB, ADLS 2, Talend (7.3), Hive, Hadoop, Linux,** |
| Role | Azure Data Engineer |
| **Description:**  The Business process involves adding three more restated fields and loading of data from Ingestion Hive to Distillation Hive table. There after data is exported in the form of Parquet file and pushed to Azure Data Lake Store Gen2 – Landing container. Parquet file is then loaded into ADW and ADB layer which is consumed by Power BI tool to generate dashboards which are used by end users for analysis and decision making. | |
| **Roles and Responsibilities**   * Involved in ETL workflow using Azure ADF, Databricks with PySpark as per business requirements, which includes extraction of data from Relation Database and loading it to Azure Sql DB. * Experience working with Unity Catalog for data governance like Sharing Data across workspaces, providing access, masking data, storing tables, views, functions and volumns..etc * Extensively involved in writing transforming data in PySpark and pushing into ADLS Gen2. * Automated the code by creating workflow tasks and with data factory by calling notebook activities, * Worked on Autoloader for streaming data. * Storing the resulted data into Azure SQL DB for PowerBI and Sportfire Consumption. * Experienced in data migration from On-Prem to Azure Cloud with Databricks using Spark api. * Attend daily scrum calls and update the ADO user stories on daily basis. * Monitor the pipeline jobs and report/fix them if any issues. * Parameterized the hive scripts and passed from sql server to talend using context variable to run the hive scripts in hive. * Extracted the data from hive using tHiveinput and did transformations and loaded into hive. * Altered 3 columns for 10 tables. * Implemented performance tuning logic on targets, source to provide maximum efficiency and performance. * Deployed and run tac jobs for Ingestion and distillation. * Loaded data from file to table using tfileInputExcel and tHiveLoad. * Used various components like tFileList, thdfsFilelist, thdfsFilecopy, tjavaFlex, tFlowtoIterate, tFileInput/Output, tOracleInput/ Outputs etc. * Performed to shift the partitions in hive from one table to another table using thdfslist, tjavaflex, thiveRow. * Used TAC to schedule the Talend jobs and monitored daily Loads. | |

|  |  |
| --- | --- |
| **Project #3: SDM** | |
| Role | Data Engineer. |
| Environment | Talend 7.2 Big Data (HDFS, Hive), UNIX. |
| **DESCRIPTION: -**  what they want most. And their real- time approval technology helps them do it all on the spot.  Segmentation data mart (SDM) delivered core capabilities to support the Email and Direct mail marketing campaigns. Core capabilities includes Universe of marketable contacts, Account association, Job title categorization, Sales rep assignments, Suppressions, 3rd party list loads and integration to Eloqua. | |
| **Roles and Responsibilities**     * Creating the Jobs to extract the data from source to target using Talend tool. * Extracted data from Oracle database and Flat files and staged into a single place and applied business logic to load them in the Hive. * Implemented performance tuning logic on targets, source to provide maximum efficiency and performance. * Parameterized the Talend Jobs and Increased the re-usability. * Applied slowly changing dimensions like Type 1 effectively to handle the Delta Loads. * Developed Talend jobs and published into TAC. | |

|  |  |
| --- | --- |
| **Project #: EPICX** | |
| Role | Data Engineer. |
| Environment | Talend 7.3, Unix, Aws, Sftp, Snowflake, Redshift, Athena, S3, Sap |

|  |
| --- |
| **DESCRIPTION: -**  Integrating the data from different source systems to SFTP location and making available this data to all business users across the world and respected stake holders for their business decisions. |
| **Roles and Responsibilities**     * Prepared the jobs based on business requirements using different components like tfineinoutdelimited, tfileist, tlogcather, trunjo,tdir, tflowtoiterate and so on. * Mapped columns dynamically with respect to input file in tfineinoutdelimited using dynamic schema. * Captured the logs and execution statistics of the pipeline (Source records count and target records count) with the help of logcather and tscatcather. * Added triggers for the job execution in TMC and monitored triggered jobs. * Extracted from Azure data lake storage Zen 2 into talend and did transformations using components like tmap, tjava, tjavarow loaded transformed data into oracle. * Implemented the dynamic data set components to work for different objects. * Trigger the jobs based on schedule intervals. * Prepared the external tables, dimension and fact tables in azure SQL based on user requirements. * Involved in bug fixing. |

---------Thanks.,

**G Navaneeth Rao**