Sanjay Kumar
Senior Data Engineer

Dublin, Ireland
+353892701163 · sanjaymse@gmail.com

🔗 LinkedIn

Skills

Azure Data Factory

Azure Synapse Analytics

SQL Server

Power Bi

Dremio

Azure Data Lake

PLSQL

Github

Jira

Azure DevOps

Microsoft Fabric

Data Bricks

Pyspark , Pandas , SparkSQL

Python

PowerShell , Unix

DBT

Languages

English

Tamil

Profile

Ambitious, resolute, and collaborative Data Engineer fervent with 6.5 years of experience in Data Engineering, Big data analytics, Business Intelligence, Data Visualization, ETL and Mircrosoft Fabric, passionate about helping businesses succeed through strong technical skills rooted in substantial training. Well informed possessing master's in software engineering and professional certifications in Microsoft Azure Data Engineering, graduated Masters in Big Data Management and Analytics.

Employment History

Data Engineer, Version 1, Dublin

August 2022

Client - Sustainable Energy Authority of Ireland

- Integrating Dynamic 365 application with Azure Synapse Analytics using Power Apps.
- Deployed Azure Synapse workspace with serverless SQL pool and Integrated D365 application using Data Verse.
- Moved on-prem history data to azure synapse using synapse integration pipelines.
- Developed complex pipelines inside azure synapse for data transformation.
- Extensive experience working on Serverless SQL Pool.
- Using Polybase methodology, created external tables in synapse to view the data from csv files resides on Azure Data Lake storage. (Delta tables in serveless pool on top of data lake files)
- Created Power Bi reports and dashboards for end users based on the requirement.

Client - Renaissance Reinsurance of Europe Limited

- Worked on Dremio to develop multiple data products on finance data with the purpose of providing valuable insights, information about the business to the stakeholders.
- Created Wiki pages to present detailed information about the data products. This includes adding data lineage, metadata, sample queries, and data catalog entries to provide a holistic view of the data and its utilization.
- Developed Power BI reports on top of Dremio layer, incorporating intricate Power Queries and DAX Queries, while also integrating custom visuals to enhance the data presentation and analysis.

- Created a DBT project to facilitate the deployment of Dremio spaces to different Dremio environments. To optimize this process, a delta deployment logic has been implemented, ensuring that only the updated models are deployed to the space, thereby streamlining the deployment process..
- Developed a Power BI Paginated report with intricate logic. This report serves the purpose of printing and displaying year-wise total expenses for Jet Usage by top management. The data is sourced from an internal application API, transformed within Power BI using Power Query, and integrated into the dataset used in the Paginated report, incorporating complex expressions to deliver the required insights.
- Project operates within an agile framework, that includes daily stand-up meetings, sprint planning sessions, story refinement sessions, and retrospective meetings. All project tasks and stories are meticulously tracked and managed on a Jira Board to ensure efficient collaboration and progress monitoring
- Version control and code management for this project were conducted using GitHub to facilitate collaborative development. All code developments were pushed to the Git release branch. After receiving approval from a peer reviewer, the code was then merged into the main branch for deployment, ensuring a systematic and controlled development process.

Client - University of Limerick

.

- Created a Proof of Concept (POC) using Microsoft Fabric to demonstrate its capabilities in managing analytics workloads and showcasing its features.
- Crafted a personalized Python script to gather a set of user review data from the IMDB Open Source API and subsequently upload the information as a csv file into Azure Data Lake.
- Implemented the custom Python script within Azure Functions, ensuring automated execution for seamless extraction and uploading of files.
- Established a data pipeline within Fabric to facilitate the seamless transfer of data from Azure Data Lake Storage (ADLS) to Fabric Lake House.
- Utilized the advanced cleansing capabilities of Dataflow Gen 2 features in Fabric to process and refine the data, subsequently reloading it as a new file within the Lakehouse
- Developed a notebook and conducted Exploratory Data Analysis (EDA) using PySpark on the dataset, then stored the processed data as a delta table in the lakehouse
- Developed a notebook, loaded the data into a Pandas dataframe, and leveraged Data Wrangler in Fabric to perform additional transformations, showcasing its powerful features.
- Conducted tokenization and lemmatization on the dataset to determine sentiment scores for reviews using VADER Sentiment Analysis.
- Used Recurrent neural network deep learning model to learn the sequential relationship between each words in the reviews and for classifying the positive and negative reviews.
- Generated visual representations of sentiment scores and dataset using Matplotlib and Seaborn, with distinct axes, and subsequently saved the results as a new delta table in the Lakehouse.

- Concluded the process by crafting a Power BI dashboard for a descriptive visualization of the dataset.

Application Development Senior Analyst , Accenture, Chennai

September 2019 — January 2021

Client - Komatsu , Japan

- Experienced working with cross-functional and multi-disciplinary teams responsible for delivering project on-time.
- Integration developer for integrating different existing and new source systems with newly developed Microsoft Dynamics 365 portal.
- Used azure Boards for planning and tracking work using agile methodology.
- Extracted, Cleansed, Transformed, and loaded existing Historical ticket data to D365 portal using Azure Data Factory V2.0.
- Used Mapping dataflow and Databricks for Data Transformation.
- Used Azure logic Apps for automatic image movements from Azure blob to D365 portal in a logical flow.
- Automated auto reconciliation of text extracted from attached Images and the text filled in ticket form using Azure Computer Vision. (Azure Cognitive Services) Power BI Reports were built to visualize the Ticket data.

Software Engineer, Tata Consultancy Services, Chennai

June 2016 — August 2019

Client - HMD Global Oy, Finland

- Developed facts and dimensions in star schema model in Azure Data Warehouse
- Integrated on premises data from multiple sources with Azure data warehouse using Azure Data Factory.
- Hubble's structure multi-dimensional cube developed using SSAS (Integrated few measures in a existing cube).
- Developed SSIS Packages for data transformations and deployed in Azure SQL Server.
- Power BI Reports and Dashboards were built to visualize the Sales data of HMD Global.
- Descriptive data analytics were performed to showcase the overall sales, profit & loss of HMD Global.
- Used Azure DevOps for code check-ins and CI/CD process.
- Used Power Shell scripts to automate alerts and data reconciliation.
- Hosted Azure Virtual Machines, Encrypted and decrypted the VMs, Automated the backups of Azure resources, Automated the Data warehouse, Automated VM scale up and scale down, Automated Data Warehouse scale up and scale down.
- Enhanced the security recommendations for Azure resources and working with OMS Portal for log analytics.

Client - Prudential PLC , UK

- Developed and enhanced the existing DataMart in an OLAP Schema.
- Developed and maintained PL/SQL code for data extraction, transformation, and loading (ETL) processes.
- Conducted performance tuning and optimization of SQL queries, resulting improvement in database efficiency.

- Worked closely with QA teams to perform unit testing and resolve any identified issues promptly.
- Assisted in the creation of documentation for database processes, procedures, and standards.
- Implemented effective indexing strategies to optimize database performance.
- Implemented clustering for tables to improve data locality, minimize storage fragmentation, and enhance overall system reliability.
- Developed Unix Shell scripts for data movement from On-prem to SFTP.Utilized Jira board for Agile and Github for code management.

Education

Msc in Big Data & Analytics , Griffith College Dublin, Dublin

January 2021 — July 2022

MS in Software Engineering, VIT University , Chennai

June 2011 — May 2016

References

Jeet Jhadav from Version1
+353892070715 · jeet.jadhav@version1.com

Nimmi Narayanan from Version1
+35319019262 · nimmi.narayanan@version1.com