

# Sunny Kumar

Mail: sunnykrr3377@gmail.com

Contact: +91- 8092429920

## **Carrer Objective:**

Dedicated Professional with 4 years of IT experience with 3 years' experience as a Azure Data Engineer of expertise designing, implementing, and optimizing data solutions on the Azure platform. Proven ability to drive actionable insights through data migration, cleansing, transformation, and integration across diverse industries. Skilled in Python, PySpark, Spark SQL, and Azure Data Services (ADLS, Databricks, Data Factory, etc.). overall, 4 years' experience in TCS.

## **PROFILE SUMMARY**

- Having 3 years of hands-on experience in Cloud computing technologies (**Azure Data Factory, Azure Databricks and Azure Data Lake**).
- Experience in ETL processes. Worked extensively on Data Extraction, Transformation and loading from different data sources using **ADF**.
- Monitor **performance of pipelines, troubleshoot incidents**, and provide support to end-users.
- Good understanding on Application Data models and data modelling activities.
- Designing ETL/ELT processes from the source systems to Azure cloud.
- Knowledge in **Azure Data bricks** and notebooks development
- Having good experience **Azure Databricks Delta Lake**.
- Designed various ingestion and processing patterns based on use cases in **Delta Lake**.
- Experience in Managing and storing confidential credential in **Azure Key vault**.
- Orchestrated the end-to-end data integration pipelines using **Azure Data Factory**.
- Hands-on experience in python and Spark components like **Spark-core and Spark-SQL**
- Worked on creating the RDDs, DFs for the required input data and performed the data transformations using Spark-core.
- Hands on experience on **Power BI Desktop, Power Query and Power BI Service, DAX, Azure**
- Experience in creating Charts like **Pie Charts/Donut/Word/charts/cards/gauges/slicers** etc. and working with custom visualizations using Power BI
- Good at analyzing, gathering, managing, and documenting business requirements.
- Extensively used transformations like **Merge Queries, Append Queries, Pivoting the Column, Un Pivoting Columns, Adding Custom Columns, Splitting the Columns, Merging the Columns** etc.
- I worked extensively on Enhancing the Data Model by Creating New Columns, New Measure and New Tables using **DAX Expressions**.
- Hands on Experience with **Visual level, Page level, Report level and Drill Through filters** for filtering the data in a Report.
- Experience on Migrating SQL database to **Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse** and controlling and granting **database access and migrating on premise databases to Azure Data Lake store using Azure Data factory**.
- Experienced with the Spark improving the performance and optimization of the existing algorithms in **Hadoop using Spark Context, Spark-SQL, Data frame API, Spark Streaming, Pair RDD's and worked explicitly on PySpark**.
- Worked on Performance Tuning with the help Indexing and SQL Profiler for reducing the Stored procedures run time and to have faster generation of reports.
- Self-directed, organized, capable of multi-tasking and willing to shift priorities to meet project needs and expectations.

- Expertise in OLTP/OLAP System Study, Analysis and E-R modelling, developing Database Schemas like Star schema and Snowflake schema used in relational, dimensional, and multidimensional modelling.
- Experience of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance. Experience with different file formats like Avro, parquet, ORC, Json, XML and compressions like snappy & zip.

### **WORK EXPERIENCE:**

- Currently Working as Data Engineer in TCS from June-2021 to till now.

### **TECHNICAL SKILLS:**

- **Core Technologies:** Azure Data Services (ADLS, Databricks, Data Factory, SQL Database), Python, PySpark, Spark SQL, SQL
- **Data Management & Analytics:** Data Migration & Integration, Data Cleaning & Transformation, Data Modelling, Data Visualization (Power Bi)
- **Additional Tools & Technologies:** Azure Key Vault, Jupyter Notebook, VS Code, MongoDB Compass, SQL Workbench.

### **EDUCATION DETAILS:**

- Masters in Computer Application (MCA) from Vellore Institute of Technology, Chennai, India – (2019-2021)
- Bachelor's In Computer Application (BCA) from GIIT College, Jamshedpur, Jharkhand - (2016-2019)

### **Project 1: Building a Scalable Data Pipeline for Marks and Spencer (M&S)**

**Project Description:** Develop a robust and automated data pipeline to ingest, process, and transform large volumes of customer and sales data from various M&S sources.

#### **Roles and Responsibilities:**

- Collaborated with business stakeholders to identify all relevant customer and sales data sources within M&S.
- Developed existing tools (APIs, data scraping, Azure Data Factory - ADF) to extract data from each source in a reliable and efficient manner.
- Designed and implemented data transformation processes using Databricks and PySpark to clean, standardize, and enrich the extracted data (address formatting, missing values, data aggregation).
- Chosen and configured a suitable data warehouse solution (Azure Synapse) to store the transformed data for easy access and analysis.
- Implemented an orchestration tool (ADF, Databricks) to automate the entire data pipeline, ensuring data is ingested, processed, and loaded into the data warehouse on a scheduled basis.
- Continuously monitor the data pipeline for errors, performance bottlenecks, and data quality issues. Implement mechanisms for automated alerts and troubleshooting procedures using ADF and Databricks.

#### **Benefits:**

- Consistent and reliable data ensures accurate reporting and analysis.
- Automated data pipelines enable faster access to valuable insights for business decisions.
- Manual data processing tasks are eliminated, freeing up resources for other data-driven initiatives.
- The pipeline can be easily scaled to accommodate future data volumes and changing data processing needs.

## **Project 2: Customer Segmentation & Targeting Dashboard**

### **Description:**

This project aims to leverage the power of Power BI to segment Verizon's customer base and develop targeted marketing strategies for each segment. By understanding customer behavior and preferences, we can personalize marketing campaigns, improve customer engagement, and ultimately drive revenue growth.

### **Responsibilities:**

- Connect to different Data sources from Power BI like Database, flat files, Excel.
- Gathered data using direct query mode.
- Implemented DAX (Data Analysis Expressions) functions for calculated columns and measures.
- Experience in Custom visuals and Groups creation.
- Reported EDA analysis in Power BI and RLS and worked on Power BI services.
- Handled End to end process to deliver the project.
- Publish Power BI Reports and create scheduling the reports using Data Gateway.
- Experience in creating different visualizations like Line, Bar, Histograms, Scatter, Water, Bullet, Heat maps, Tree maps, KPIs.
- Develop various types of complex reports like Drill Down, Drill through Reports.
- Responsible for providing security for reports/dashboards by assigning roles.
- Experience in creating Joins, Sub-Queries, sub queries, Stored Procedures and Views.
- Interact with client directly to get requirements.
- Collaborated with Azure data engineers and operation team to implement the ETL process using Azure Data Factory wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
- Performed data cleaning including transforming variables and dealing with missing value and ensured data quality, consistency, integrity using ADF.
- Develop and maintain standardized reports, dashboards and portals utilizing the business intelligence platform's software developer tools.

### **DECLARATION:**

- I hereby declare that the above-mentioned information is correct, and I bear the responsibility of the correctness of the above-mentioned.

**Sunny Kumar**