Name: Sachin Pralhad Magadum

Happiest Minds Email id: sachin.pm@happiestminds.com

Primary Skills: Python, PySpark, SQL, Azure Databricks, Azure Data Factory, ADLS, Airflow

Experience Summary: A highly skilled **Azure Data Engineer** with over 3+ years of experience in designing, developing, and optimizing scalable data solutions on **Microsoft Azure**. Expertise includes building and managing **ETL pipelines**, implementing efficient data storage strategies, and performing advanced data transformations to support high-performance analytics. Demonstrated success in utilizing key **Azure cloud services** such as **Azure Data Lake**, **Azure Data Factory**, **Azure SQL Database**, **Azure Databricks**, and **Azure Synapse Analytics** to drive data-driven decision-making and deliver actionable business insights. Proven ability to ensure **data security**, **compliance**, and **data integrity** while optimizing performance and maintaining reliability across large-scale data environments.

Key Skills & Expertise:

- **Cloud & Big Data Solutions:** Extensive experience in building and deploying scalable cloud-based data solutions using **Microsoft Azure** services, including **Azure Data Factory**, **Azure Databricks**, and **Azure Datalake** for large-scale data processing, real-time analytics, and data warehousing.
- **ETL Pipeline Development:** Proficient in designing, developing, and maintaining **ETL** pipelines using tools like **Azure Data Factory**, **Apache Airflow**, and **Python** to automate the extraction, transformation, and loading of both structured and unstructured data from diverse systems to cloud environments.
- **Data Transformation & Modeling:** Advanced skills in data transformation and modeling using **PySpark**, **Apache Spark**, **SQL**, and **Python** for complex data processing tasks, including data aggregation, cleansing, and optimization in large-scale environments.
- **Data Integration & Orchestration:** In-depth knowledge of integrating data from various sources (cloud, on-premises, APIs, third-party systems) through **REST APIs**, **Azure Logic Apps**, and **SQL Server Integration Services (SSIS)**, ensuring cohesive, end-to-end data solutions that enhance reporting and analytics.
- **Performance Optimization:** Expertise in optimizing data processing workflows and queries, leveraging **SQL Server**, **Azure Databricks**, and **Apache Spark** to improve performance, reduce processing times, and maintain data quality and integrity for accurate business insights.

Total Industry Experience: 7 Years & 2 Months

  ➢ **Project 1 – Duration (From July 2023 to June 2024 (12 months)**
    1. **Description of the Project: EPR (Enterprise Party Repository)** is a centralized master data management solution used across the ABN AMRO Group to capture, store, and maintain party data. It integrates data from over 30 sources, ensuring a unified, accurate view of client and party information.

EPR supports regulatory compliance, client due diligence, and internal analytics by providing a single source of truth for data. Its robust data governance framework ensures data quality, while its scalability meets evolving business and regulatory needs.

2. **Responsibilities:** I designed and developed scalable data pipelines using Azure tools, ensuring efficient data flow and processing across multiple platforms. I implemented end-to-end ETL workflows to extract, transform, and load data from various internal and external sources, including systems like SFR, Alfam, KVK, and DNB. This involved transforming raw data into standardized formats suitable for analysis and reporting and loading it into centralized databases for easy access. Additionally, I contributed to module design, participated in code reviews, and regularly attended project status meetings to ensure alignment with project goals. My role also involved optimizing data pipelines for performance and reliability, ensuring the smooth integration of diverse data sources into the overall system architecture. Through these efforts, I helped streamline data processing workflows, enabling more efficient and actionable insights.

3. **Technical tools used:** Python, PySpark, Azure Databricks, Data Lake, Airflow, Azure Data Factory

➢ **Project 2 - Duration (From May 2020 & To Mar 2022 (23 months)**
   1. **Description of the Project: Edge by Ascential** is a powerful platform designed to provide sales-driving insights for global brands, helping them optimize their e-commerce strategies, particularly across Amazon's marketplace. The project focuses on processing and analyzing large volumes of data generated from Amazon properties, including product listings, sales data, customer reviews, and competitive intelligence.
   The initiative involves building and maintaining complex data engineering pipelines to handle the vast and dynamic data streams coming from Amazon's ecosystem. These pipelines aggregate, clean, and transform raw data into actionable insights that enable brands to make data-driven decisions, optimize pricing strategies, track product performance, and improve visibility on Amazon's search results.
   2. **Responsibilities:** Developed ETL workflows to process and ingest data. Built dashboards using Looker for data visualization. Wrote unit and functional tests in Python. Participated in agile development processes, code reviews, and project planning.
   3. **Technical tools used:** Python, PySpark, Azure Databricks, Airflow, Bitbucket, JIRA, Pytest

➢ **Project 3 - Duration (From August 2019 & To May 2020 (8 months)**
   1. **Description of the Project**: **AIT** is a cloud-based analytics framework designed to generate detailed audience insights and reports based on user-selected traits. The project focused on developing a serverless infrastructure to enable efficient data ingestion, processing, and real-time reporting. Using scalable cloud services, AIT processes large volumes of data, transforming raw information into actionable insights tailored to specific audience segments. The system leverages advanced data pipelines to automate data flow, ensuring high availability and seamless performance. By providing

customizable reporting features, AIT empowers users to make data-driven decisions and enhance audience targeting strategies. This solution is highly adaptable, supporting rapid scaling as data requirements grow.

2. **Responsibilities:** Designed and implemented cloud-based infrastructure using AWS services. Automated data workflows using Glue jobs, Lambda, and Athena. Developed a serverless web application using API Gateway, Lambda, and S3. Participated in module design, requirement analysis, and performance optimization.

3. **Technical tools used:** Python, PySpark, AWS Lambda, Athena, EC2, S3, SQS, PostgreSQL, GIT, JIRA

Technical Skills:

- **Database Systems:** Azure SQL server, PostgreSQL
- **Subject Area:** Big Data Engineering, Azure Data Engineering
- **Big Data Platforms:**
  - ➢ Spark
- **Language/Platforms:**
  - ➢ Python, PySpark, SQL, Azure Databricks, Airflow, Azure Data Factory, ADLS

Certificate and Trainings:

- **Certificate of Microsoft Azure Data Engineer Associate (DP-203)**
- **Python — HackerRank, June 2023**