*A blue and white logo

Description automatically generated*A logo for a company

Description automatically generated

Manideep D| *+1-940-331-4847 |* [*04manideep@Gmail.com*](mailto:04manideep@Gmail.com)

*LinkedIn:* [*linkedin.com/in/manideep-d-a74a5b12b*](https://www.linkedin.com/in/manideep-d-a74a5b12b/)

**PROFESSIONAL SUMMARY:**

Results-oriented Data engineer with extensive 5 years of experience in designing, implementing, and optimizing end-to-end data solutions within the Azure ecosystem. Skilled in leveraging Azure cloud services such as Azure Data Lake Storage, HDInsight, Databricks, and Azure SQL Database to develop efficient data pipelines for processing and analytics. Proficient in Python, Spark, Airflow, and Tableau for data manipulation, ETL, and visualization. Experienced in Snowflake, Terraform, Informatica, Kafka, and other data integration and management technologies. Proficient in Agile methodology, actively participating in sprint planning, daily stand-ups retrospective meetings.

**TECHNICAL SKILLS:**

**Data Engineering:** ETL Development, Data Modeling, Data Pipelines, Dataflows.

**Tools**: Power BI, Tableau,

**Cloud**: Azure (Data Factory, Databricks, Azure datalake Storage(ADLS gen2), Azure blob, Azure Synapse), SSIS, SSRS, Ms.Excel.

**Languages**: Python, T-SQL,SQL, Scala, Java, Pyspark.

**Technologies:**  Bigdata, Hadoop, Map Reduce, Spark, Kafka, Hadoop, Snowflake, Informatica

­­­­­­­­**EDUCATION:**

**Masters :** Business Analytics, Texas A&M University, Commerce, USA, May-2023, GPA – 3.88

**Bachelors :** Textile Technology, Osmania University, Hyderabad, India, May 2017, GPA – 7.4

**CERTIFICATIONS***:*

**Microsoft Certified:** Azure Data Engineer Associate DP-203

Azure Fundamentals DP-900

**WORK EXPERIENCE**

**Data Engineer** *|* ***Advent Health|*** *July 2022 –Present*

* Designed and implemented a real-time data processing and analytics platform on **Azure** using **the Medallion Architecture**.
* Utilized **Azure Data Lake Storage (ADLSGEN2)** for efficient storage and retrieval of streaming data.
* Implemented **Apache** **Kafka** on **Azure** **HDInsight** for real-time data ingestion and buffering.
* Developed stream processing and transformation tasks using **Azure Databricks** to ensure high throughput and low latency. Skilled in **HIPAA**-compliant healthcare data engineering, and **HL7** integration
* Employed **DBT** (Data Build Tool) for modeling and transforming raw streaming data into structured datasets.
* Configured **Apache Airflow** on **Azure Kubernetes Service (AKS)** for **data quality checks** and monitoring.
* Stored processed data in **Azure Synapse Analytics** and **Azure SQL Database** for fast querying and analytics.
* Configured **Snowpipe** for seamless data integration into **Snowflake**, ensuring data consistency and reliability.
* Integrated **Snowflake** with **ETL** pipelines to efficiently load transformed data into **Snowflake** **tables**.
* Managed data security and access control using **Azure Key** Vault. Integrated **Event Hubs** with diverse data sources.
* Utilized **Azure Databricks** for distributed **ML model** training, optimizing performance on large datasets.
* Implemented data preprocessing pipelines on Azure for **ML model** training using **Azure Data Lake Storage** and **Azure Data Factory.** Expertise in translating raw data into actionable insights through **SSRS**.
* Documented architecture diagrams, data flow diagrams, and technical specifications for project deliverables
* Proficient in writing complex **T-SQL** queries for **JSON** format etc data extraction, manipulation, and analysis.
* Experienced in database design, normalization, and optimization using **T**-**SQL**.
* Familiarity with, **CDC**, **MDX** (Multidimensional Expressions) and **DAX** for querying and manipulating SSAS data.
* Experienced in optimizing **DAX** calculations for performance and efficiency in **Power BI** reports and dashboards.

**Big Data Engineer** *|* ***Infosys| Westpac Financial Banking Services Feb 2020 – Dec 2021***

* Collaborated with cross-functional teams to gather requirements and design data solutions.
* Designed and implemented end-to-end data pipelines within the **Azure ecosystem**, utilizing, **Azure Data Factory (ADF), Blob storage**, **Azure Data Lake Storage (ADLS GEN2),** **Synapse Analytics, HDInsight**, **Databricks,** and **Azure SQL Database**.
* Led the design and deployment of **big data** solutions on the Azure cloud platform.
* Integrated **Kafka** for real-time data streaming and event processing, enhancing data processing capabilities.
* Utilized **Terraform** for infrastructure as code (**IaC**) to provision and manage Azure resources, ensuring scalability and reliability. Designed and executed ETL processes on Azure using **Informatica** for seamless data movement between systems. Leveraged **Informatica's** scalability to process large volumes of data efficiently on **Azure**.
* Managed **Delta Tables** in Azure Databricks for efficient data ingestion, processing, and analytics.
* Integrated **Delta Lake** with **Azure services** for end-to-end data pipelines and integrated **Azure Event Hubs** for processing.
* Developed **ETL** pipelines in **Spark** using **Python** and **Airflow** for data extraction, transformation, and loading.
* Conducted data cleansing and applied transformations using **Databricks** and **Spark** to ensure data quality and integrity.
* Developed interactive dashboards and reports in **Tableau** and **Power** **BI** for visualizing key performance metrics and trends. utilized **SSIS** for optimizing data loading speeds. Designed and deployed **ETL** jobs with **DataStage** and **Talend**
* Implemented error handling and monitoring mechanisms within **ETL** pipelines to ensure data accuracy and completeness.

**Associate Software Developer** *|* ***DTCC*** *Feb 2019 – Feb 2020*

* Imported data from various sources **(Microsoft SQL Server, MySQL, Teradata)** into **HDFS** using **Sqoop**.
* Automated data loading tasks using **Oozie** workflows on **EMR**.
* Analyzed partitioned and bucketed data in **Hive** to compute reporting metrics.
* Developed and optimized **Hive** queries for data processing and analysis.
* Utilized **Spark** for transformations and aggregations before storing data in **HDFS**.
* Developed and optimized **Apache Pig** scripts for efficient **ETL** of large datasets in Hadoop.
* Built **RESTful** **APIs** for seamless data exchange between systems.
* Processed ingested raw data using **Python**.Utilized **MapReduce** for parallel processing and analysis of large datasets.
* Managed **Hadoop** **clusters** and monitored performance using **Cloudera Manager.**
* Converted **Hive/SQL** queries into **Spark** transformations using **DataFrames**.
* Created **Tableau** dashboards for reporting analyzed data and Managed and reviewed Hadoop log files for optimization.
* Utilized **GitHub** for code repository management and **Docker** and **Jenkins** for continuous integration**(CI/CD)**.
* Skilled in building **RESTful** services in **.NET, Java and** .Experienced in **DevOps** for data engineering, optimizing workflows and Proficient in Microservices architecture for scalable applications.

**ACADEMIC PROJECTS:**

* Led SQL dashboard & database dev using SSIS & SSMS. Managed ETL, enhancing BI ops. Implemented SCD strategies for accurate historical data tracking.