**Maddi Chandra Mounika**

Sr Data Engineer; [chandramounika.maddi94@gmail.com;](mailto:chandramounika.maddi94@gmail.com;%20) +1 214-919-3885; [LinkedIn](https://www.linkedin.com/in/chandra-mounika-maddi-b65990b8/)

**Profession Summary**

* Experienced **Senior Data Engineer** with around **10 years** of expertise in developing and deploying robust and scalable data solutions.
* **Certified** in **Microsoft Azure Data Fundamentals (DP-900)** and **Databricks Associate Developer for Apache Spark3.0 (Python)** and proficient in other Azure Cloud Services.
* Possesses advanced problem-solving skills and deep proficiency in database technologies. Skilled in business analysis, Agile methodologies, architecture design, and team leadership, ensuring successful project outcomes aligned with strategic objectives.
* Experienced in crafting applications utilizing Azure services including **Azure Data Factory, Blob Storage, Azure Databricks, Azure SQL, Azure Data Lake Storage Gen2 and Azure Key Vault.**
* Demonstrated proficiency in crafting and deploying applications using Kubernetes within the **GAIA** **customized cloud environment,** **Azure**, and the **Cloudera Platform** leveraging **YARN** as the cluster manager in On-Premises Environment.
* Showcased expertise in implementing Spark jobs to revolutionize data, applying business transformation rules across application-specific layers using various **Spark APIs** such as **RDDs**, **Dataframes**, and **Datasets**.
* Accountable for designing and developing numerous analytical solutions on vast datasets by ingesting and transforming data within the **Bigdata** environment leveraging **Hadoop** Components such as **HDFS, Spark, Sqoop, Oozie, HIVE.**
* Crafted optimized tables in **Hive** using advanced techniques like **Partitioning** and **Bucketing** to meet the business requirements and better performance, adeptly extracting crucial insights through complex queries visualized in **HUE** and **Impala**.
* Leveraged **Oozie** **workflows** to seamlessly export and import data between **Oracle** and **HDFS** and vice-versa orchestrating efficient data transfers using **Sqoop**.
* Implemented **Unix shell scripts** to automate tasks and customize input files, streamlining processes and enhancing overall efficiency.
* Exhibited expertise in **fine-tuning Spark applications, memory utilization, and parallelism** to ensure optimal performance.
* Acquired excellent proficiency in **Snowflake**, harnessing its capabilities to drive data-driven insights and analytics. Skilled in **Snowflake (SnowPipe, SnowSQL, Stored Procedures, Cloning, COPY, Time Travel Concept, Fail Safe, Data Sampling, Task Scheduling, Access Control, Performance Tuning, Virtual Warehouse, Snowpark).**
* Expertise in **Microsoft Fabric** and strong knowledge on **Medallion Architecture Design, One Lake Architecture** and other Fabric Services.
* Adept at driving business insights through data visualization with **Tableau** and **Power BI**.
* Showcased extensive knowledge of **source control managem**ent concepts including **Branching**, Merging, **Labeling/Tagging, and Integration** and utilizing tools like **Git**, **Bitbucket**, and **SVN**.
* Frequently utilized **ALM** and **JIRA** Dashboard as part of daily development activities for project management and tracking progress.
* Expertise in NoSQL DB like Cassandra and good knowledge in MongoDB and PostgreSQL.
* Expertise in Database development and RDBMS in real-time.
* Expertise in **PySpark**, **Python**, **Core Java** and **SQL**.

**Technical Skillset**

|  |  |
| --- | --- |
| Programming Languages | Java, Python |
| Bigdata /Hadoop | Hadoop distributed file system, Spark (Spark Core, Spark SQL), Hive, Sqoop, Oozie, MapReduce, YARN, Flume |
| Databases | SQL, SQL Server, Oracle, MySQL |
| Data Processing, Data Analysis and Data warehouse | Spark, Impala, HUE, Hive, PySpark, **Snowflake** (SnowPipe, SnowSQL, Stored Procedures, Cloning, COPY, Time Travel Concept, Fail Safe, Data Sampling, Task Scheduling, Access Control,Performance Tuning, Virtual Warehouse, Snowpark), Microsoft Fabric |
| Data Transfer and Workflow | Apache Sqoop, Apache Oozie, Apache Kafka |
| Cloud Platforms | **Azure** (Azure Data Factory, Blob Storage, Azure Databricks, Datalake, Delta Lake, Azure SQL, Azure Data Lake Storage Gen2 and Azure Key Vault) |
| Version Control | SVN, Git, Bitbucket |
| Data Visualization | Tableau, PowerBI, Microsoft Fabric |
| Scheduling Tools | Autosys, Control-M |
| Project Management Methodologies | Waterfall, Agile Methodology |
| IDE`s & Tools | Eclipse, Spring Tool Suite, IntelliJ, PyCharm, Jupyter Notebook, TOAD, Splunk |
| OS | Linux, Windows |

# Projects

***Client: Staples  
Azure Data Engineer 11/2022 – Present***

* Engaged with retail Business stakeholders to gather and analyze critical Business Requirements.
* Established distinct Databricks mount points for each stage of data storage, ensuring efficient data management.
* Ingested raw data into designated mount points and processed it using Azure Databricks, adhering to business rules.
* Applied various transformations and actions to refine the data, moving it to the processed folder for further processing.
* Processed diverse source files including delimited, JSON, and Nested JSON formats, utilizing PySpark APIs to execute complex transformations with precision and scalability.
* Orchestrated data movement and transformation using Azure Data Factory, streamlining pipeline creation and management.
* Created corresponding tables and stored final data in Parquet format, ensuring optimal storage and query performance.
* Implemented incremental load strategies based on daily upstream data updates, ensuring data freshness and accuracy.
* Leveraged PySpark APIs for efficient data transformations, writing complex code in notebooks to handle diverse requirements.
* Utilized Azure Databricks to transform and to load data to ADLS, facilitating seamless data flow.
* Scheduled data processing tasks using Azure Data Factory pipelines, automating job execution for improved efficiency.
* Utilized Azure Data Lake Services for secure and scalable data storage, ensuring reliability and accessibility for data processing tasks.
* Connected to PowerBI to generate comprehensive reports, visualizing data insights for stakeholders and facilitating actionable decision-making.

**Tools & Technology Used**: PySpark, Azure Data Lake Storage Gen2, Azure Data Factory, Azure Databricks, Databricks Mounts, Azure Key Vault, SQL, PowerBI, Databricks Notebooks, Apache Spark, Python.

***Mphasis Limited Client - JP Morgan Chase & Co Azure Data Engineer 09/2021 – 09/2022***

**PreChargeoff flow**: analyzing credit card customer data to identify delinquent accounts, utilizing systems including EPAY, XLOB, ODM, and Business Action, among others, to facilitate efficient repayment processes.

* Each stage has a specific functionality and processed different formats of batch data using Spark API`s like Spark Core and Spark SQL.
* Orchestrated Spark jobs to actively transform data, achieving a significant 25% efficiency boost through strategic optimization.
* Seamlessly integrated Data Lake and Oracle DB using efficient DB links, accelerating data extraction and loading processes.
* Led precise and efficient application deployment using Kubernetes in a customized GIAI cloud environment.
* Demonstrated proficiency in Spark Core and Spark SQL, resulting in a notable 20% improvement in application performance.
* Created table in Oracle database in which data gets loaded through spark jobs
* Jobs are scheduled using Control- M and they get triggered at specified intervals.
* Utilized tables to query and extract vital information and generated insightful reports on top of them.
* Generated reports have specific indicators which helps the business to reach out to the credit card customers for their next payments.
* Transferred output files to different servers using SFTP for efficient data distribution.

**Tools & Technology Used**: Java, Apache Spark, Oracle DB, Control -M, SQL, Kubernetes, Cloud, IntelliJ IDE

***Tata Consultancy and Services Client: Bank of America***

***Big Data Developer 05/2014 – 09/2021***

* Designed and implemented solutions to process customer information from various Lines of Business (LOBs) of Bank using big data components like Hadoop Framework, Spark, Oozie, Hive, Sqoop and Kafka.
* Engaged with customers daily to understand business requirements for big data analytical solutions, translating large-scale needs into detailed system specifications
* Utilized different big data components at various data stages within the Data Application to meet business requirements effectively.
* Employed Hadoop Framework for distributed storage and processing of large datasets.
* Leveraged Spark for fast and flexible data processing and analytics.
* Orchestrated workflow management with Oozie to ensure smooth data processing pipelines.
* Utilized Hive for data warehousing and querying capabilities.
* Implemented Hive partitioning and bucketing techniques for better performance of the jobs.
* Integrated Sqoop for efficient data transfer between Hadoop and relational databases.
* Implemented data processing using Spark APIs (Spark Core and Spark SQL) in Java and optimized Spark jobs for an increase in efficiency by 30%.
* Deployed applications in CDH (Cloudera) Environment using YARN as cluster manager.
* Scheduling Spark jobs with Autosys and monitored the performance of the jobs at different intervals.
* Analyzed data using Impala and HUE.
* Written Complex SQL Queries to query different tables by which we can derive actionable insights of customer transactions.
* Prepared comprehensive documentation outlining job flows, file details, and corresponding Hive tables
* Validated scripts based on business requirements and participated in the analysis phase to gather requirements
* Developed pipelines to extract and process diverse data sources, tailoring processes based on specific indicators
* Generated reports and delimited files as output files and sent them to downstream via NDM and SFTP to mainframe servers.
* Written scripts to purge the data in few tables and implemented a process to store the data for N no of days required and delete the older Nth partition data by which we save the storage by 10% in the cluster and cleanup the unwanted data.
* Mounted the data from LFS to HDFS locations and processed the data using Spark
* Conducted unit testing in Dev and SIT environments. Provided support to the QA team
* Performed script validations based on business requirements
* Created Oozie workflows to move data from Oracle to HDFS and vice-versa using Sqoop
* Produced business reports using Tableau to visualize insights derived from data analysis.
* Deployed the code in Bit Bucket using GIT
* Mentored new team members on work environment and targets, teamwork dynamics.

**Tools & Technologies used**: Apache Spark (Spark Core and Spark SQL), Hive, Impala, UNIX, Core Java, Autosys, YARN, CDH, Sqoop, GIT, Scala, Oozie, Hadoop, HDFS, Splunk, WinSCP, TOAD.

**Certifications**

* Microsoft Certified -Azure Data Fundamentals by Microsoft (DP 900)
* Databricks Certified Associate Developer for Apache Spark 3.0 by Databricks
* Certified in Advanced SQL by Hacker Rank.
* Hands on Essentials – Data Warehouse issued by Snowflake
* PYTHON 101for Data Science - Cognitive.AI powered by IBM Developer Skills Network-IBM & Cognitive AI
* Data Visualization in Tableau issued by Great Learning

**Education**

* Bachelor of Technology in Computer Science & Engineering, JNTU Kakinada - India -**2014**