**CHANDRA MOUNIKA MADDI Sr Data Engineer**

**Career Objective**

Experienced Data Engineer with 8 years of track record in Banking, Retail and insurance Domains. Proficient in Big-Data technology especially in **Hadoop, Spark, Hive, and Azure Cloud**. Ability to thrive in a fast-paced work environment. Highly analytical thinker with demonstrated talent for streamlining complex work processes. Collaborated across teams for scalable cloud-based solutions.

**Summary**

* Worked closely with stakeholders across various departments to design, build and deploy multiple initiatives within the data platform for the past 8 years.
* Vast experience in **Azure Cloud, Azure Data Factory, Azure Data Lake Storage, Azure Synapse** **Analytics, Azure Analytical services, Azure Cosmos NO SQL DB, Azure HDInsight Big Data** Technologies (**Hadoop and Apache Spark**) and **Data bricks**.
* Experience in designing Azure Cloud Architecture and Implementation plans for hosting complex application workloads on **MS Azure**
* Strong knowledge of Azure Storage Accounts, Containers, **Blob Storage, Azure Data Lake, Azure data factory, Azure SQL data warehouse**
* Migrate data to/from Microsoft Azure Cloud Platform, **Azure SQL DB, Hadoop data on the Azure Hindsight Service using ADF**
* Proficient in creating Spark applications **in DataBricks using Spark-SQL** for aggregating data and extracting data from different file formats such as Parquet and JSON
* Proficient in maintaining and implementing data warehouse business with **ETL tools and data storage technologies like Bigdata Hadoop**
* Well versed experienced in creating pipelines **in Azure Cloud ADFv2** using different activities like Move &Transform, Copy, filter, for each, Data bricks etc.
* Strong knowledge of **Data Warehousing methodologies** and including **Start Schemas, snowflake concepts**
* Maintaining **ETL, Dev, QA,** production data warehouse and BI reporting environments
* Proficiency in querying and testing the business requirements and queries in production environment using SQL
* Developed Spark jobs to transform data and applied business transformation rules to process the data across application specific layers using **Spark API`s -RDD`s, Data frames and Datasets.**
* Deployed applications using Kubernetes in GAIA customized cloud environment and Azure Cloud Services.
* Performed **ETL operations** by connecting to **Oracle DB`s using DB link**.
* Scheduled the jobs using **Control-M and Autosys tools**.
* Generated different output files and sent to **Mainframes server through SFTP**.
* Responsible for designing, developing various analytical solutions to gain insights on large data sets by ingesting and transforming these datasets in the **Big Data environment using technologies like HDFS, Spark, Sqoop, Oozie, HIVE.**
* Developed applications by implementing different optimization techniques available in **spark and hive for better performance of the application.**
* Created Tables in Hive using Partitioning and Bucketing techniques based on the business requirement and written complex queries in **Hive to extract vital information and visualized it in HUE and Impala.**
* Launched different **Oozie workflows** to export and import the data from **Oracle to HDFS and vice-**
* **versa using Sqoop.**
* Written **Unix shell scripts** to automate the jobs and to customize the input files that makes easier to process the file.
* Expertise in performance tuning of **Spark application** for setting right batch interval, memory tuning and correct level of parallelism.
* Based on business insights, we extracted information from **Hive tables and generated reports using Tableau.**
* Parallel copying of files between various clusters using **Distcp, Kafka in Hadoop**.
* Deployed applications in Cloudera Environment and having YARN as cluster manager
* Extensive knowledge of data flow diagrams, data dictionaries, database normalization procedures, entity-relationship modelling, and design methodologies
* Excellent knowledge of source control management concepts such as **Branching, Merging**, **Labeling/Tagging and Integration**, with tool like **Git, Bitbucket, SVN**
* Exposure on **ALM and JIRA** Dashboard.
* Proficient in developing clear and concise documentation of all work performed
* Strong analytical knowledge coupled with excellent communication and interpersonal skills
* Leveraged and developed multiple applications using **PySpark and Java** by increasing the efficiency by 30%.
* Expertise in **PySpark, Java and SQL.**

**Technical Skills**

* **Programming Languages:** Java, Python
* **Azure:** Azure Blob Storage, Azure Data Factory, Azure SQL, Azure Cloud, Data Lake Storage (ADLS), Databricks
* **Databases:** SQL, MySQL, Oracle
* **Data Processing and Analysis:** Apache Spark, Apache Hive, PySpark, Snowflake, Impala, HUE, DBT, Apache Pig
* **Data Transfer and Workflow:** Apache Sqoop, Apache Oozie, Apache Kafka
* **Data Storage and Management:** HDFS, NoSQL (Cassandra)
* **Cloud Computing:** Amazon Web Services, Azure
* **Hadoop Distribution:** Cloudera, Databricks
* **Agile Methodology**
* **Version Control:** SVN, Git, Bitbucket
* **Data Visualization:** Tableau
* **Scheduling Tools:** Autosys, Control-M

**Professional Experience**

AllPro Consulting LLC July 2023 – Present Azure Data Engineer

* Established a Cloud infrastructure from scratch and created various pipelines to bring the data from other sources to **Azure SQL server, Data Lake, and Storage using AzureDataFactory, Azure Data Bricks.**
* Collaborated cross-functionally to conceive and implement data initiatives utilizing **Hadoop, Spark**, **and Hive**, yielding a remarkable 25% efficiency boost.
* Created pipelines, data flows, and complex data transformations and manipulations **using ADF and PySpark with Data bricks.**
* Worked on transformation, Data Ingestion, and extraction using the **ETL** processes developed using Azure Event Hub
* Worked on developing data-driven workflows using **Spark, Cloud, Hive, Python**, Impala for further ingesting the data
* Developed Batch and Near **Real-Time data processing using ADF and Azure Databricks**
* Designed **ETL** data pipeline with **PySpark and Spark** to automate the process of collecting data and storing data in **SQL database**
* Responsible for building and deploying the pipelines from Dev to **UAT and Prod using Azure DevOps**
* Improved Query Efficiency **of Azure Synapse tables** using Ordered **Clustered** Column Stored Index
* Developed a detailed project plan and helped to manage the data conversion migration from the legacy system to the target **Snowflake database**
* Leveraged Azure Data Lake Services to streamline data storage; **Azure Databricks expedited ETL** operations efficiently.
* Developed Spark applications using **PySpark and Spark-SQL** for Data aggregation and transformation from different file formats for analysing and transforming the data
* Application deployment adeptly using **Kubernetes in the customized GIAI cloud**, ensuring precision
* Worked on **Azure Synapse** to stage data and added external tables in Azure Synapse loaded data from Azure Data Lake Storage
* Monitored the **SQL** scripts and modified them for improved performance using **PySpark SQL** and managed relational database service in which **Azure SQL** manages scalability, stability, and maintenance.
* Worked on complex **SQL Queries and converted them to ETL tasks**
* Responsible for transforming and loading large sets of structured, semi-structured, and unstructured data
* Engineered transformative data sets with **Spark,** optimizing processing efficiency and significantly elevating application performance.
* Incremental Load approach **via Delta Lake, curtailing data processing** time.
* Played a pivotal role in enhancing data quality by conducting comprehensive **data validation** processes, contributing to a 15% reduction in data errors.

Mphasis Sep 2021 – Jul 2023 Sr Data Engineer

Client: JPMorgan Chase & Co

* Orchestrated Spark sub-jobs for **Active data transformation, delivering a substantial 25% efficiency enhancement through strategic optimization.**
* As part of the implementation, we created multiple spark sub jobs where each job has different **business functionality** and **dependency** on one another.
* Efficiently employed DB links for seamless **Data Lake and Oracle DB integration**, expediting data extraction and loading.
* **Synthesized essential reports** from queried tables, profoundly impacting strategic decision-making processes.
* Reports generated are sent to **Mainframe server** to validate the **business data** by adding specific indicators which plays a vital role to contact the customers.
* Spearheaded application deployment adeptly using Kubernetes in the customized **GIAI cloud**, ensuring precision and efficiency.
* Applied **Spark Core and Spark SQL** adeptly, resulting in a noteworthy application performance enhancement of 20%.
* Migrated Ab-initio jobs to **Spark jobs** for better performance and to increase the efficiency
* Prepared detailed documentation for the jobs and shared in the resource available here
* Monitored and guided few of the teammates and worked together under stressful work environment and delivered the project on time.
* Worked with QA closely to solve and fix the defects immediately and to share the knowledge on the jobs which helps them to prepare the test cases.

Tata Consultancy and Services June 2016 – Sep 2021 Data Analyst / Data Engineer

Client: Bank of America

* Identified customer needs and engaged with customer daily to understand the business requirements for big data analytical solutions and break down large scale requirements into detailed system specifications.
* Implemented the flow using **Spark API`s** in JAVA.
* Deployed the application in **Cloudera Environment** using **YARN a**s cluster manager.
* Scheduled the **spark jobs using Autosys** and monitored the performance of the jobs.
* Optimized the Spark jobs for the best performance and increased the efficiency by 30%
* Queried tables and analysed the **data using Impala**.
* Prepared documents by mentioning about each **sub flow**, details of the files to be generated and **Hive tables** that need to be created.
* Performed **script validations** based on business requirements.
* Participated in the Analysis phase and gathered all the requirements for the flow.
* Prepared different pipelines to extract and to process **different source of data, based on few indicators** which differs for each sub flow of specific job.
* Generated few reports and different delimited files as part of the process which are sent to downstream via **NDM and to mainframes server through SFTP**.
* 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.
* Performed unit testing in **Dev and SIT environments** and provided support to QA Team.
* Created Oozie workflows to get the data from **Oracle to HDFS using Sqoop.**
* Produced different reports using Tableau based on the business insights.
* **Engineered Spark flows** proficiently using Core Java, optimizing data processing and reinforcing strategic data management by 30%.
* Orchestrated precise **Spark job scheduling via Autosys**, realizing a remarkable 30% efficiency elevation in data operations.
* Leveraged Impala for comprehensive data analysis, shaping invaluable insights for stakeholders' informed decision-making.
* Formulated, honed, and streamlined **Oozie workflows** meticulous for efficient data migration between Oracle and **HDFS through Sqoop**.

DPI Technology India Nov 2015 to May 2016

Intern Data Analyst

* Maintained **Data Models, Metadata and data dictionaries**, defined source fields and their definitions: collaborated with **Data Governance Teams**
* Developed **UNIX shell scripts** for file manipulation and transfer
* Efficiently defined scope, gathered business requirements, performed **GAP analysis** for the Project
* Executed Various **Data Mapping** tasks from Source Systems
* Worked on Data Analysis on reverse engineering models and deleted pointless tables and columns on **SQL server database**

**Education**

* Bachelor of Technology in Computer Science and Engineering Oct 2012 – Apr 2016 Gudlavalleru Engineering College, Kakinada, India

**Certifications**

* Certified **PYTHON 101 for Data Science** - Cognitive.AI (IBM's Cognitive Artificial Intelligence Program)
* **Data Visualization in Tableau by Great Learning**
* Hands on Essentials – **Data Warehouse issued by Snowflake**

**Key Achievements**

* Acknowledged as the Quarterly Top Performer for consistently surpassing project objectives.
* Received the Innovation Excellence Award for devising an optimized data processing workflow, resulting in a remarkable 30% reduction in processing time.