|  |  |
| --- | --- |
| Mitra Manohar Sai Katta  Data Engineer at Maltem Australia | Email: [mitra.katta@gmail.com](#) Contact: +61 469385490  Current Location: Sydney, Australia |

# Professional Summary:

* A competent professional with 4.5+ years of experience in Admin activities, implementing Big Data Integration and Data quality business logic and business/functional requirement analysis, design, development and testing of Hadoop, Spark, Hive and Talend Big data DI ETL tools.
* Certified in **Azure Data Engineer Associate (DP-203) and** **Talend Data Integration**, **Talend DI Administration.**
* Having good hands on experience L2 support, to resolve the critical issue on time limits.

**Technical Skills:**

* ETL Tools: Talend Bigdata 6.1,6.2.1,6.4.1,7.1, Azure Data Factory
* Languages: SQL, HQL, Python, Java
* Ecosystem: Pyspark, Hadoop
* Platform: Databricks Unified Analytics Platform
* Database: Oracle 10g/11g,Sysbase,Hive
* NOSQL: Cassandra.
* Tools: TOAD, Control-M Scheduling tool
* Cloud: Azure
* Operating Systems: UNIX and Windows

# Education:

* Bachelor of Technology in Computer Science from SRM University, Chennai, Tamil Nadu (2011-2015)
* Master of Information Technology in Networking from Central Queensland University, Sydney (Nov2018/JUl2020)

# Professional Experience:

**POC Project:**

**Company:** Maltem Australia

**Duration:** Oct 2022 to Present

**Project Description:**

This POC is mainly focused on the below Objectives

* Analysis of Different types of data on basis of file format, size of the input data, real-time requirements.
* Source Data in the form of Json and xml file format (from Rest API application) are qurated in ADF. Demonstrate of Multiple ways to be able to connect to ADF and perform transformations and monitor the triggers and logs.
* Obtain tweets about a trending topic and perform analytics on streaming data to publish BI reports.

**Responsibilities:**

* Acquisition of tables from on premise databases and ingesting data into Azure data lake
* Implementation of Linked Service to gather data from multiple sources
* Performing data quration by creating external tables using polybase in Synapse.
* Orchestration of Databricks notebook activities using the Azure Data Factory (ADF)
* Implemented SCD type 1, 2, 3 and Facts and dimensions tables in ADF
* Processing real time data from Azure Event Hubs through Azure Stream Analytics

**Project No: 1**

**Project title:** CAPITAL End to End **Client**: Macquarie Bank **Company:** Maltem Australia **Duration**: Aug/2021 to Oct/2022

# Project Description:-

Macquarie bank has Banking and Financial Services Group with expertise in Wealth Management, Personal and Business banking. New UQS requirements are built as part of current state to meet APRA timeline for the minimum capital adequacy ratios in order to perform calculations of risk weight assets that includes IRB (Internal ratings-based) approach for the calculations on risk weight assets and the other approach is Standardized approach (SA).

# Responsibilities:-

* Jobs built using Spark, Hive (HQL) and TALEND ETL with Big data.
* Practicing Data Vault modeling techniques in the (Enterprise Data Warehouse) EDW agile.
* Orchestration of jobs with Airflow DAGs.
* Creation of Control-M scripts in order to trigger the Airflow DAGs.
* Using the existing bamboo CI/CD pipeline, SQL framework and Bit bucket for the build process.

# Project No: 2

**Project title:** SAHARA lulu

**Client**: HP

**Company:** Cap Gemini, Bangalore, India

**Duration**: April/2016 to Aug/2018.

# Project Description:-

This system is used for managing the incentives given to the sales representatives of HP. HP has many clients who buy products like printers, laptops and desktops. Clients are connected to HP through sales representatives. The more sales representatives sell to clients, the more will be their incentives. The system captures transactional data coming from different zones all over the world. This data is processed based on certain rules and final information is send to another system called Credit and pay. Credit and Pay engine will calculate the incentives for sales representatives.

# Responsibilities:-

* Development of pipelines, Linked services, Datasets in ADF (Azure Data Factory) to achieve business scenario.
* Configuration of Azure Cloud services that includes Azure Blob, Azure SQL Db, and Azure data lake store.
* Orchestration of jobs with ADF pipelines using tumbling window triggers.
* Creating Azure Data Bricks Pyspark notebooks to run in sequential or parallel.

# Project No: 3

**Project title:** EDH 2.0

**Client:** TE Connectivity

**Company:** Cap Gemini, Bangalore, India

**Duration**: Jul/2015 to Mar/2016

# Project Description:-

TE Connectivity Ltd, formerly Tyco Electronics Ltd, designs and manufactures connectivity and sensor solutions for a variety of industries including automotive, industrial equipment, data communication systems, aerospace, defense, medical, oil and gas, consumer electronics, energy and subsea communications. The company serves customers in more than 150 countries.

# Responsibilities:-

* Ingested data from Oracle into Cloudera distribution as Hive tables using Sqoop scripts and processing the data from Hive using Talend spark DI and converted into Visualization views which are needed to perform analytics using PowerBI.
* Jobs are scheduled and triggered using Control-M tool.
* Visualization views are created and loaded data with different segment level group access.
* Performed error handling, evaluated job execution, and also gathered job level and component level metadata using various components.
* Published, triggered and deployed jobs in TAC using Job conductor.
* Worked on Parent and child job relation to simplify the complex jobs.

# Talend Administration Activities:-

* Installation and configuration of Talend modules TAC, Command Line, Nexus, Studio, job-server and Studio from version 6.2.1 to 6.4.1.
* Creating projects, adding users to project, granting permissions.
* Kerberos Authentication set up on the Talend machine.
* Created new URL monitor configured in Sitescope for monitoring the DEV environment Talend Administration Center (TAC).
* Alert triggered on mount directory when the usage reaches 85%