

VidyaSagar Jonnala Mail: vsrjonnala@outlook.com Mobile: 571-597-9978

# Overview

Throughout my extensive career spanning over 12+ years, I have consistently demonstrated excellence in roles including Data Architect, Data Engineer, and Data Analyst. My primary focus has been on designing and building robust ETL integration/migration frameworks, implementing pattern-based integrations, and delivering efficient data warehousing, data management, and data visualization solutions. Notably, I have successfully led projects across a wide range of industries including retail, e-services, F&B, manufacturing, and finance.

# Skills & Capability

* Designing **conceptual, logical and physical models** for the target system is a critical component of the job.

# Expertise in designing ETL Data Migration & Integration frameworks and Proficient in building complex transformation rules & Data Cleansing rules.

* Extensive experience in design and maintain of **Enterprise Data Warehouse / data marts**.
* Solid understanding of Data Warehousing, Data marts, Dimensional Modeling, Star, Snowflake schema, OLTP and OLAP concepts.
* Drawing on Experience in all aspects of analytics/data warehousing solutions (**Data Modeling, Data mapping, ETL Development, metadata management, data migration and Analytical solutions**) I have been keying in delivering innovative database /data warehousing solutions to the clients.
* Proficient in building complex transformation rules, **Data Quality (DQ), Master Data Management, Data Profiling, Data Governance, Data Lineage, Cleansing, and Data Conversion** techniques.
* Experience in developing of **Dashboards and Reporting solutions** for the senior leadership team and the business in achieving better service quality, workforce management and business process optimizations.
* Extensive Experience in **Azure Data Platform** components (**Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, Databricks, NoSQL DB**).
* Strong Knowledge in **AWS, GCP** data components.
* Hands on Experience in various ETL tools like **DataStage, Informatica, Pentaho, Talend, SSIS, DBT and Cloud based data integration tools**.

# Hands on experience in Relational databases, Teradata, Snowflake, Vertica, No SQL database, HDFS and Cloud-based data storage.

* Experience in Data migration projects, especially data migration to CRM applications like Salesforce
* Hands on experience in various BI tools like **Tableau, QlikView, PowerBI, Pentaho BI and Business Objects.**
* Hands on Experience in **Azure Synapse**, **QLIK, DataBricks, Snowflake, Python, Git, Pyspark, Spark SQL, Unix, SaaS** and **Shell scripting.**
* Strong Knowledge on **Matillion and Fivetran**
* Excellent technical, analytical, and problem-solving skills. Strict attention to detail and ability to work independently as well as within a team environment.

# Academic and Professional Qualifications

* Master of Computer Applications (M.C.A) from Anna University (2010), India.
* Bachelors in computer science (B.sc Computer Science) from Acharya Nagarjuna University (2007), India.

**Certifications**

**Microsoft Certified: Azure Data Engineer Associate Domain Experience**

**Finance, Retail, Telecom, F&B, Manufacture, Project Management and E-Services. Experience Summary:**

**Responsibility: *Data Engineer, Walmart***

**Project: *Walmart***

**Duration: *April 2023 – till Date***

As a Data Engineer I am responsible for the data ingestion for the Customer Data Lake. (Several different sources of data) and looking at Customers outside of the Walmart + membership so the business can continue to understand how people shop, to ultimately improve customer experience.

* Design and implement data pipelines using Azure Data Factory (ADF) to extract, transform, and load (ETL) data from diverse source systems to Azure Data Storage services (ADLS, Azure Storage, Azure SQL).
* Collaborate with stakeholders to understand data requirements and design logical and physical data models.
* Ensure optimal performance and scalability of data models for Azure Data Warehouse.
* Implement data governance policies and procedures to ensure data quality, consistency, and compliance with regulations.
* Utilize Azure Integration Runtime for seamless data movement between on-premise and cloud environments.
* Work with Spark SQL for advanced data transformations and processing in Azure Databricks.
* Implement data security measures and access controls to protect sensitive customer data.
* Design and maintain data warehousing solutions in Snowflake, including the creation of entity-relationship diagrams and data schemas.
* Develop dbt (data build tool) transformations and manage dependencies between dbt models to ensure proper execution order.
* Use Snowflake for querying and reporting on processed data stored in the data warehouse.
* Collaborate with Data Engineers to ensure seamless integration of Snowflake into the overall data architecture.
* Strong knowledge in designing and developing Power BI visualization according to business requirement documents and plans for creating interactive dashboards.

## Technologies: Azure Data Factory, dbt, Azure Data Lake Gen2, Azure SQL, T-SQL, DataBricks, Snowflake Pyspark, Power BI

**Responsibility: *Data Architect and Data Engineer, TERAS***

**Project: *PLUS, Unified Data Platform***

**Duration: *September 2022 to March 2023***

At TERAS, I played a crucial role as the creator, developer, and leading system integrator of Touch n Go and SmartTag, which are extensively utilized throughout Malaysia for highway, parking, public transportation payments, and access control systems. Within this position, I focused on enhancing the existing data platform and addressing data issues in pipelines, reports, and dashboards.

* Working as an Azure Data Platform Architect, collaborating with various business teams and stakeholders to achieve business results by attending workshops, gathering requirements, and recommending solutions on Azure data technologies.
* Designing, developing, and maintaining data integration pipelines using ADF and dbt to collect data from various sources, including toll booths, parking facilities, public transport systems, and payment gateways.
* Creating dbt transformations to cleanse and structure raw data for analysis and reporting within the Azure environment.
* Implementing real-time data processing solutions using Azure Stream Analytics to handle continuous data streams, such as toll booth transactions, parking spot availability updates, and public transport card usage.
* Ensuring that data is stored and processed in compliance with data security and privacy regulations within the Snowflake data warehouse.
* Monitoring and optimizing the performance of data pipelines and databases in Azure to ensure high throughput and low latency for real-time and batch processing.
* Designing and creating data models using Azure Data Platform that represent the structure and relationships of data entities.
* Collaborating with stakeholders, including business analysts and database administrators, to understand data requirements in the context of Azure, dbt, and Snowflake technologies.
* Developing data models and schemas within Snowflake to support efficient querying and reporting for toll collection, parking payments, and public transport usage.
* Collaborating with data analysts and visualization experts to create Power BI dashboards and reports that provide insights into transaction data, payment trends, and public transport usage, leveraging Azure services and dbt transformations.
* Implementing security measures and compliance standards within Azure and Snowflake to ensure the confidentiality, integrity, and availability of sensitive data, especially for financial transactions.
* Responding to data-related issues, troubleshooting problems, and ensuring the high availability and reliability of data solutions, especially for mission-critical systems like toll collection, within the Azure, dbt, and Snowflake environments.

## Technologies: Azure Data Factory, dbt, Azure Data Lake Gen2, Azure SQL, Pyspark and Power BI

**Responsibility: *Data Architect and Data Engineer for Telecom Malaysia***

**Project: *CDR***

**Duration: *January 2021 to August 2022***

Telekom Malaysia Berhad (TM) is a Malaysian telecommunications company founded in 1984. Beginning as the national [telecommunications](https://en.wikipedia.org/wiki/Telecommunications) company for fixed line, radio and television broadcasting services, it has evolved to become the country's largest provider of broadband services, data, fixed-line, pay television and network services.

* Design and implement data pipelines to ingest, process, and transform CDR data from various sources into Azure data storage solutions (Azure Data Lake, Blob Storage, Azure SQL, Azure SQL Data Warehouse).
* Design and implemented data pipelines to ingest data from various sources, including network equipment, customer databases, and external sources, into Azure data storage solutions.
* Used Azure services like **Azure Data Factory and Azure Event Hubs** to facilitate data ingestion.
* Developed ETL (Extract, Transform, Load) processes to cleanse, enrich, and transform CDR data into a usable format for analytics and reporting.
* Choose and manage appropriate data storage solutions on Azure, considering factors like cost, performance, and scalability. This may include Azure SQL Database, Azure Cosmos DB and Azure Synapse Analytics.
* Implemented data security and access controls to ensure that CDR data is protected and compliant with data privacy regulations.
* Monitor and optimize the performance of data pipelines and databases to ensure efficient data processing and reporting.
* Design and maintain data models and data warehouses using Azure Synapse Analytics.
* Created data models and schemas that support efficient querying and reporting for telecom CDR data.
* Implement data quality checks and validation processes to identify and rectify data anomalies and discrepancies.
* Implement automation and orchestration techniques to schedule, monitor, and manage data pipelines and workflows.
* Collaborated with data analysts and data scientists to create **Power BI** dashboards and reports that provide insights into CDR data.
* Collaborate with cross-functional teams, including data scientists, analysts, and business stakeholders, to understand their data requirements and deliver solutions accordingly.

## Technologies: Azure Data Factory, SSIS, Azure Synapse Analytics, Data Lake Gen2, Cosmos DB, Pyspark, Power BI, SQL Server, Oracle

**Responsibility: *Data Architect and Data Engineer for JKR***

**Project: *J10***

**Duration: *September 2019 to December 2020***

The Malaysian Public Works Department is the federal government department in Malaysia under Ministry of Works Malaysia which is responsible for construction and maintenance of public infrastructure in Malaysia. BPI focuses on establishing and proposing to -be business process for JKR to enhance its service delivery with focus on customer as focal point, speed, compression, service quality. This covers the project management core function of JKR namely planning design, procurement, construction, and handover. And type of implementation/projects which are conventional inhouse and conventional consultant for both assets building and infrastructure (road, bridge, slope etc.,).

* Collaborated with end users to identify analytical KPIs and documented wire-framed dashboards based on their requirements.
* Utilized **Pentaho Data Integration and Talend** to migrate data from legacy systems (Postgres) to the target database (Oracle) and designed a data mart for data visualization.
* Developed data models and centralized storage to facilitate efficient data retrieval and analysis.
* Led the migration of Historical data and unstructured data, such as design documents, notes, construction videos, and pictures, into **HDFS (Private Cloud)** for seamless integration into the analytical system.
* Used Pentaho Data Integration and Talend for migrating data from legacy systems (Postgres) to target databases (Oracle) and designing data marts.
* Experienced in working with unstructured data migration, including design documents, notes, construction videos, and pictures into HDFS.
* Design and implementation of a new data warehouse to fulfil business requirements and support analytical needs.
* Used **Tableau** for developing dashboards and visualizations, incorporating analytical KPIs and wire- framed dashboards based on end-user requirements. Proficient in integrating Google Maps for consolidated views based on geographical locations.
* Used **Pentaho Report designer** for real time Reports for providing stakeholders with valuable insights and facilitating informed decision-making.
* Worked both **hybrid cloud and on-premises environments**, ensuring effective data management and security.

## Technologies: Pentaho Data Integration, HDFS(Cloud), Data Lake, Oracle, PostgreSQL, Tableau and Pentaho Report Designer.

**Responsibility: *Data Engineer for JLM***

**Project: *MMDIS***

**Duration: *October 2018 to August 2019***

JLM is a department under [the Ministry of Water, Land and Natural Resources Malaysia](https://ms.wikipedia.org/wiki/Kementerian_Air%2C_Tanah_dan_Sumber_Asli_Malaysia) . This department is responsible for monitoring the safety and environment of the gazette [marine parks](https://ms.wikipedia.org/wiki/Taman_laut) in the waters [of the Malacca](https://ms.wikipedia.org/wiki/Selat_Melaka) [Strait](https://ms.wikipedia.org/wiki/Selat_Melaka) and [the South China Sea](https://ms.wikipedia.org/wiki/Laut_China_Selatan) , [the Sulawesi](https://ms.wikipedia.org/wiki/Laut_Sulawesi) Sea and [the Sulu Sea which are included in](https://ms.wikipedia.org/wiki/Laut_Sulu) [Malaysian](https://ms.wikipedia.org/wiki/Laut_Sulu) waters.

* Played a key role in implementing JLM, an ERP system focused on registering ships in Malaysian Ports, encompassing modules such as Ship Registration, Ship Clearance, Target Positions, Document Management, and Human Resource.
* Actively participated in client meetings to gather requirements, ensuring a thorough understanding of their needs and expectations.
* Analyzed requirements and formulated Key Performance Indicators (KPIs) for data migration and visualization, aligning them with business objectives.
* Prepared mapping documents for data migration, ensuring accurate and seamless transfer of data from legacy systems (MySQL and MSSQL) to the target database (MySQL) using Pentaho.
* Collaborated with cross-functional teams to gather data requirements and designed data models to meet business needs.
* Ensured data quality and integrity through data validation techniques and data cleansing procedures.
* Developed migration scripts and designed the DataMart once data migration was successfully completed, enabling efficient data storage and retrieval.
* Utilized Pentaho Business Intelligence to create interactive dashboards and generate reports, providing stakeholders with valuable insights and facilitating informed decision-making.

## Technologies: Pentaho Data Integration, Pentaho business intelligence, MySQL, SQL Server

**Responsibility: *Data Engineer for MYEG Services***

**Project: *MYEG Data Analysis***

**Duration: *December 2016 to September 2018***

MYEG is a concessionaire for Malaysian Electronic-Government (“E-Government) MSC Flagship Application. MYEG builds, operates, and owns the electronic channel to deliver services from various Government agencies to Malaysia citizens and businesses.

The business is growing bigger as more consumers registered and purchased at least one of all services. Marketing team needs the detailed report on analyzing and comparing how consumers used MYEG services and which service they purchased.

* Collaborated closely with cross-functional teams, including IT, business analysts, and management, to understand data requirements and ensure the accuracy and relevance of the implemented solution.
* Implemented a star-schema data warehousing approach to support comprehensive data analysis across all MYEG services, customers, and transactions.
* Developed a robust data integration solution using Pentaho to efficiently extract, transform, and load data from multiple production databases into the data mart.
* Conducted thorough testing and performance optimization to ensure the efficient and reliable functioning of the data warehousing system.
* Utilized identifier keys within the data warehousing system to optimize data storage and retrieval, minimizing storage space requirements.
* Leveraged QlikView and Crystal reports to create interactive and visually appealing dashboards, reports enabling stakeholders to easily explore and analyze the data.

## Technologies: Pentaho Data Integration, SAP HANA, PostgreSQL, QlikView, and Crystal Reports

**Responsibility: *Data Analyst for Coles Myer***

**Project: *Sales Information System***

**Duration: *September 2015 to November 2016***

Coles Myer is Australia's largest retailer with more than 1,900 stores throughout Australia and New Zealand serving more than 4.5 million customers every week. Coles operates fully serviced supermarkets in more than 460 stores throughout Australia. Today Coles is a leader in Australian food retailing and Sales lead information provides the capturing and maintaining information capabilities on all sales.

* Use SQL Server Integration Services (SSIS) to design and develop data integration processes that extract, transform, and load (ETL) sales data into a SQL Server database.
* Maintain and optimize the SQL Server database to ensure data integrity, performance, and availability.
* Implement an Enterprise Data Warehouse to centralize and store sales data for decision support purposes. Ensure that the EDW supports data quality and consistency.
* Set up processes to provide data access and updates, enabling stakeholders to make decisions based on the most current information.
* Implement data quality checks and validation procedures to minimize errors and ensure data accuracy and reliability.
* Created visually engaging data visualizations that help users quickly understand trends, patterns, and insights from sales data.
* Utilize tools such as QlikView and Crystal Reports to design and create interactive and visually appealing dashboards and reports for sales and business stakeholders.
* Enable stakeholders to explore and analyze sales data through user-friendly dashboards, allowing them to make data-driven decisions.

## Technologies: SSIS, SQL Server and QlikView.

**Responsibility: *Migration Consultant for Schneider Electrics***

**Project: *SENA***

**Duration: *April 2014 to May 2015***

Schneider Electric is the world leader in electrical distribution, industrial control and automation products, systems and services, and the only electrical manufacturer dedicated solely to the distribution and control of electricity. We are in data migration from their existing applications Data to SAP, under current phase we are migrating data to a Staging location with Data Transformation, Cleansing with various Business validation rules.

* Design, develop, and maintain ETL (Extract, Transform, Load) processes using Informatica PowerCenter or other related tools.
* Create and optimize data integration workflows and mappings.
* Collaborate with business analysts and data analysts to understand data requirements.
* Debug and troubleshoot ETL processes to ensure data accuracy and reliability. Perform performance tuning and optimization of ETL processes.
* Have imported data from applications such as NetSuite to build data warehouse to serve end reporting needs used for tracking subscription sales, invoices, subscription expiries, renewals and open/closed opportunities.
* Conducted testing of data load results and error log status in RDBMS Oracle using SQL queries, ensuring data accuracy and integrity.
* Utilized Informatics power center to extract data from source transactional databases, perform transformations, and load it into the target system.
* Played a role in performance tuning of Informatics jobs, identifying and resolving performance issues to optimize job execution.
* Conducted Unit, Systems Integration testing, and User Acceptance Testing (UAT) to ensure the successful implementation of migration projects.

## Technologies: Informatica Power Center, Oracle, PL/SQL, SAP ECC

**Responsibility: *Migration Consultant for Target***

**Project: *TGT$100 and ISPA***

**Duration: *January 2012- April 2014***

TGT$100 is a transformation program for integration of the different legacy systems to **SAP ECC** (SAP FICO). **Data Stage SAP R/3 PACK** and **SAP BW PACK** are being used to integrate with the ECC systems.TGT$100 is the business and technology effort to transform financial applications and related processes with a Finance Integrated Network that will support Target as a $100 + Billion company.

Target is a $64.9 billion general merchandise retailing business that is one of the most admired companies in the United States. Located in Bangalore, Target India fulfills a unique role in ensuring the success of this widely respected retailer that ranks #28 in the Fortune 500 List–2009.

* Analyzed existing jobs and successfully migrated them to the AIX server, adhering to development standards and best practices.
* Developed Server and Parallel jobs utilizing Oracle, ODBC, FTP, Pivot, Hashed file, Aggregator, STP, Filter, Funnel, Copy, Change Capture, Merge, Lookup, Join, Sort, and other stages within IBM Infosphere DataStage.
* Employed Quality Stage stages such as Standardize, Match, Investigate, and Survive to ensure data validation in accordance with business requirements.
* Conducted thorough unit testing and optimized performance by fine-tuning data warehouse tables, utilizing IBM Infosphere DataStage Director for job monitoring and troubleshooting.
* Created the data warehouse repository using IBM Infosphere DataStage Manager, importing source and target database schemas.
* Implemented shared containers, enabling their reuse in multiple jobs to promote efficiency and consistency.
* Utilized IBM Infosphere DataStage Job Sequencer extensively for scheduling jobs and ensuring proper job sequencing.
* Conducted unit testing and facilitated deployments to ensure successful project implementations.
* Collaborated with cross-functional teams to identify and resolve script-related issues, ensuring smooth operation of scheduled jobs and monitoring processes.
* Assisted in issue tracking, monitoring, and reporting throughout the migration process.

## Technologies: DataStage, Oracle, PL/SQL, DB2, SAP ECC, SAP BO, Linux, Shell Scripting

**Responsibility: *Migration Consultant for Target***

**Project: *AIX Migration***

**Duration: *December 2010- December 2011***

AIX migration project is meant for migrate existing applications from Sun Solaris System to AIX for better performance also provides flexibility for future applications. Analyze existing jobs and migrate to AIX server with development standards.

Analyzed existing jobs and successfully migrated them to the AIX server, adhering to development standards and best practices.

* Developed Server and Parallel jobs utilizing Oracle, ODBC, FTP, Pivot, Hashed file, Aggregator, STP, Filter, Funnel, Copy, Change Capture, Merge, Lookup, Join, Sort, and other stages within IBM Infosphere DataStage.
* Employed Quality Stage stages such as Standardize, Match, Investigate, and Survive to ensure data validation in accordance with business requirements.
* Conducted thorough unit testing and optimized performance by fine-tuning data warehouse

tables, utilizing IBM Infosphere DataStage Director for job monitoring and troubleshooting.

* Created the data warehouse repository using IBM Infosphere DataStage Manager, importing source and target database schemas.
* Implemented shared containers, enabling their reuse in multiple jobs to promote efficiency and consistency.
* Utilized IBM Infosphere DataStage Job Sequencer extensively for scheduling jobs and ensuring proper job sequencing.
* Conducted unit testing and facilitated deployments to ensure successful project implementations.
* Assisted in issue tracking, monitoring, and reporting throughout the migration process.
* Implemented and maintained shell scripts for job scheduling and automation, ensuring timely execution of critical tasks and reducing manual intervention.

***Technologies: DataStage, Oracle, DB2, Linux, PL/SQL, Shell Scripting***