PAVAN **A**

Dallas, TX 75201 | (475)-837-3173 | Pavan.kumar4c2@gmail.com

Summary

Over five years of expertise in data engineering, focusing on sophisticated data warehousing and ETL processes using Informatica Power Center. Advanced proficiency in backend programming, including schema design, development of stored procedures, triggers, views, and indexes. Demonstrated skill in data quality management and profiling to ensure data accuracy and integrity. Strong expertise in Relational database systems like Oracle, MS SQL Server, Teradata, MS Access, DB2 design, and database development using SQL, PL/SQL, SQL PLUS, TOAD, SQL - LOADER. Highly proficient in writing, testing, and implementation of triggers, stored procedures, functions, packages, and Cursors using PL/SQL. Mastered advanced data visualization techniques, developing custom Tableau portals integrated with PostgreSQL for enhanced analytics. Led complex data mapping and model maintenance, covering both logical and physical aspects of large datasets. Directed agile projects, closely collaborating with product owners and end-users to deliver bespoke data solutions. Expertise in OLTP/OLAP System Study, Analysis and E-R modeling, developing Database Schemas like Star schema and Snowflake schemas. Spearheaded major data migration projects to Snowflake and AWS, optimizing systems with Unix-based scripts and scheduled jobs. Developed end-to-end automated data workflows using bash scripts and Control-M, enhancing operational efficiency. Created API services and Spark applications using Python (PySpark), handling intricate data processing tasks. Managed installations and configurations of the Hadoop ecosystem, including Cloudera Hadoop, HDFS, MapReduce, Pig, Hive, and Sqoop. Utilized Avro for JSON data serialization and handled various data formats like JSON, AVRO, Parquet with compression techniques. Implemented Oozie for job scheduling and workflow automation in Hadoop clusters, monitored using Ganglia and Nagios. Designed secure, efficient data pipelines to Snowflake and established Azure Data Lakes, optimizing data storage and analytics. Developed comprehensive cloud data pipelines using Azure Data Factory, orchestrating data services across multiple platforms. Proficient in applying Systems Development Life Cycle (SDLC) methodologies and delivering solutions using the Software as a Service (SaaS) model. Transitioned ETL processes to more efficient scripts and platforms, demonstrating expertise in data transformation and storage optimization.

Skills

- SQL
- Software Development Life Cycle (SDLC)
- Data Warehousing
- Business Intelligence (BI)
- Extract, Transform, Load (ETL)
- Informatica
- PL/SQL
- Microsoft SQL Server
- Hive
- Data Warehousing
- Apache Spark
- Amazon Redshift
- Excellent Communication
- Key Performance Indicators
- RDMS Design
- PostgreSQL
- Hive
- Apache Kafka
- Machine Learning
- Tableau
- Big Data
- Database structures

Experience

Data Engineer **09/2022 to Current**
CVS Healthcare

- Reliably provided recommendations as a part of the DRB review sessions, to observe if the attributes at the physical and logical level abide by the naming standards
- Strong understanding of the principles of Data warehousing, Fact Tables, Dimension Tables, Star and Snowflake Schema Modelling

- Experience in backend programming including schema and table design, stored procedures, triggers, views, and indexes
- Using Informatica Power Center for extraction, transformation, and load (ETL) of data in the data warehouse
- Analyzing the data sources and targets using the Informatica data profiling option
- Work closely with data scientists and engineers to design and maintain scalable data models and pipelines
- Build complex tableau dashboards with advanced functions to provide insightful analysis, identify growth opportunities, and maximize profits
- Create a custom Tableau Report Portal with PostgreSQL to show real-time usage and provide easier navigation
- Prepared scripts to ensure proper data access, manipulation, and reporting functions with R programming languages
- Formulated procedures for integration of R programming plans with data sources and delivery systems
- Skillfully defined the list codes and code conversions between the Source Systems and Data Mart as well as actively involved in extensive data analysis on Teradata and Oracle Systems querying and writing in SQL and Load
- Developed mapping spreadsheets; provided the Data Warehouse Development (ETL) team with source to target Data Mapping; developed logical and physical data models that documented source data from multiple sources
- Work directly with Product Owners and end-users to develop solutions in a highly collaborative and agile environment
- Creation of Mapping document for Source System (SQL Server) to Target System (Snowflake)
- Prepared technical design and database design document
- Migrated data from Oracle, and MySQL databases to AWS using Unix-based File Watcher too
- Scheduled jobs to run promptly using Control-M
- Integrated the ingestion tool to be able to communicate with any external data sources like RDBMS databases like Oracle, MySQL, etc
- Designed and developed an end-to-end data ingestion framework
- Automated workflows using bash scripts and Control-M jobs to pull data from various databases and migrate it to AWS and Snowflake
- Creation of datasets and data-wise rules to load data from SQL databases to S3 buckets and Snowflake
- Used the Arrow scheduler system to automate the pipeline workflow
- Actively participated in software development lifecycle (scope, design, implement, deploy, test), including design and code reviews
- Developed Spark Application by using Python (PySpark)
- Developed and implemented API services using Python in spark
- Worked with different file formats like JSON, AVRO, and parquet and compression techniques like snappy
- Implement (code and test) to defined set of requirements to improve the existing functionality
- Configured Load Balancing features to different applications depending on the load and visibility of the application to the enterprise.

Data Warehouse Engineer **01/2020 to 07/2021**
Mphasis

- Documented the entire lifecycle of source-to-target mappings, data dictionaries, transformations, and data quality controls within the first 19 days of employment, establishing a strong foundation for ongoing data management processes.
- Collaborated effectively with over 30 developers and 120 business users to enhance proficiency in Cognos reporting methodologies and data warehousing practices, significantly improving data utilization across the organization.
- Successfully resolved a 17% variance between the company's sales reports and general ledger by

performing a detailed forensic analysis of SSIS packages, thereby restoring accuracy and trust in financial reporting.
- Led requirement gatherings, data mappings, design, and maintenance of enterprise applications for more than 10 companies in the first quarter of 2022, demonstrating strong project management and technical skills in high-pressure environments.
- Refined master data and implemented automated clean-up processes, significantly enhancing data quality across more than 10 key business processes and internal reporting systems, ensuring higher accuracy and reliability of data.
- Developed tailored reporting solutions for Polyhire by identifying over 20 innovative approaches to enhance sales campaigns, directly contributing to improved sales strategies and outcomes.
- Engineered an end-to-end solution for ETL process monitoring using IBM Business Monitor V6.1, which improved the team's response times by 84%, greatly enhancing operational efficiency and system reliability.

Jr.Data Engineer **05/2019 to 12/2019**
Capgemini **Bangalore, India**

- Automated and optimized a complex data pipeline handling millions of rows daily, reducing manual workload by 33% and ensuring high availability using Python and SQL
- Designed and deployed scalable multi-tier applications across core AWS services (S3, RDS, DynamoDB, Redshift, Lambda), focusing on enhancing data availability and system resilience
- Maintained a data pipeline uptime of 99.9% by managing continuous data streams from third-party sources, utilizing a combination of AWS tools and Python scripting
- Reduced data synchronization time by 50% through the implementation of Stitch, an ETL service that automated data flows between DynamoDB and AWS Redshift
- Developed SQL procedures and views to dynamically support complex business requirements, significantly aiding the customer support team in mobile and web application environments
- Spearheaded the creation of interactive dashboards using AWS QuickSight, which transformed well-defined business requirements into actionable insights, thereby facilitating strategic decision-making.

Education and Training

Masters in Technology Management
Lindsey Wilson College Colombia, Kentucky

Bachelors in Electronic Engineering
NRI Institute of Technology Pothavarappadu