**Lead Data Engineer**

**EDER OSVALDO VALERO ARMENDARIZ**

**Professional Summary :**

* **15+** years of experience in Analysis, Design, Development and Implementation as a Data Engineer.
* Expert in providing ETL solutions for any type of business model.
* Provided and constructed solutions for complex data issues..
* Proficient knowledge of Data Warehousing methodologies and concepts, including star schemas, snowflakes schemas, ETL processes and dimensional modelling.
* Designed and developed Power BI graphical and visualization solutions based on business requirement.
* Develop customized calculated Columns and measure using DAX Formula language in Power BI to satisfy business needs.
* Proficient in SQL, PL/SQL and Python coding.
* Experience developing On - premise and Real Time processes.
* Excellent understanding of best practices of Enterprise Data Warehouse and involved in Full life cycle development of Data Warehousing.
* Expertise in DBMS concepts.
* Leveraged big data frameworks like Apache Hive and Apache Pig for processing and analyzing large datasets, optimizing performance and efficiency.
* Demonstrated expertise in distributed computing, designing solutions that capitalize on the parallel processing capabilities of big data frameworks.
* Optimized data storage costs and performance by implementing AWS storage best practices.
* Involved in building Data Models and Dimensional Modeling with 3NF, Star and Snowflake schemas for OLAP and Operational data store (ODS) applications.
* Skilled in designing and implementing ETL Architecture for cost effective and efficient environment.
* Collaborated with business users to gather requirements and translate them into universe structures.
* Familiarity with SAP HANA data modeling concepts, such as attribute views, analytic views, and calculation views, and their role in designing efficient data structures for analytics.
* Experienced in working with Snowflake, a cloud-based data warehousing platform, for storing and analyzing structured and semi-structured data.
* Managed and mentored data engineering teams, fostering collaboration, skill development, and maintaining a positive team culture.
* Led the design and implementation of scalable and efficient data architectures, ensuring optimal performance and reliability.
* Optimized and tuned ETL processes & SQL Queries for better performance.
* Performed complex data analysis and provided critical reports to support various departments.
* Utilized AWS Glue for ETL processes, automating the extraction, transformation, and loading of data into data lakes and data warehouses.
* Work with Business Intelligence tools like Business Objects and Data Visualization tools like Power BI , Tableau and Kibana
* Acquired foundational knowledge of Kubernetes, including its architecture and core concepts.
* Acquired fundamental understanding of Ansible for configuration management and automation.
* Implemented PySpark jobs to process large-scale datasets, extract insights, and generate actionable business intelligence.
* Basic knowledge of Ansible roles, tasks, handlers, and modules for organizing and executing automation tasks.
* Basic understanding of Jenkins plugins and integrations for extending functionality and integrating with other tools.
* Familiarity with Kubernetes resources such as Pods, Deployments, Services, and Ingress controllers.
* Collaborated with data engineering teams to integrate Business Objects with data lakes or data warehouses.
* Familiar with Apache Airflow, an open-source platform for orchestrating and scheduling data workflows.
* Developed basic proficiency in Terraform for infrastructure provisioning and management.
* Extensive Shell/Python scripting experience for Scheduling and Process Automation.
* Good exposure to Development, Testing, Implementation, Documentation and Production support.
* Proficient in Spark, leveraging its distributed processing capabilities for big data analytics.
* Develop effective working relationships with client teams to understand and support requirements, develop tactical and strategic plans to implement technology solutions, and effectively manage client expectations.
* Utilized cloud platform such as AWS, to deploy and manage big data solutions in a scalable and cost-effective manner.
* Collaborated with cross-functional teams to design and implement scalable and maintainable data solutions on AWS.
* Configured AWS Identity and Access Management (IAM) policies to control access to AWS resources and data.
* Implemented ETL processes to extract, transform, and load data from source systems into Business Objects.
* Optimized data integration processes for improved performance and scalability.
* An excellent team member with an ability to perform individually, good interpersonal relations, strong communication skills, hardworking and high level of motivation.
* Knowledge of **Guidewire DataHub** for enterprise data management and integration.
* Familiar with **Guidewire InfoCenter** for reporting and analytics.
* Familiar with data migration approaches using Guidewire technology.
* Basic knowledge of the Guidewire Data Model and the ability to modify entities and attributes.
* Developed and integrated cryptographic algorithms (e.g., AES, RSA) to enhance data security in various applications.
* Experience with key management and certificate handling, strengthening access controls and safeguarding data at rest and in transit.
* Familiar with **Guidewire InsuranceSuite** (PolicyCenter, BillingCenter, ClaimCenter) for insurance lifecycle management.
* Familiar with standard C++ libraries and basic syntax, allowing for quick adaptation to project needs requiring C++ integration.

**Technical Skills:**

* **BigData Eco Systems**: Hadoop
* **Programming**: Python
* **Data Warehousing**: Informatica Power Center 9.x/8.x/7.x, Informatica Cloud, Snowflake and Bigquery
* **ETL Tools** : SSIS
* **Databases**: MySQL, SQL, Mongo DB
* **BI Tools**: Power BI, Tableau 9.1, Kibana
* **Query Languages**: SQL, PL/SQL, T-SQL
* **Scripting Languages**: Unix, Python, Windows PowerShell
* **RDBMS Utility**: Toad, SQL Plus, SQL Loader
* **Scheduling Tools**: ESP Job Scheduler, Autosys, Windows scheduler, Apache Airflow
* **Version Control Systems :** Github and Gitlab

**Education Details** :

Bachelor of Engineering from University of Mechanical and Electrical Engineering (FIME UANL)

**Professional Experience:**

**Lead Data Engineer, Pleasant Hills, CA NTT DATA-Citi Bank August 2022-Present**

Responsibilities:

* Exploratory Data Analysis (EDA) Jupyter Notebook.
* Large-scale Data Processing using Python Frameworks (Pandas, NumPy, Matplotlib, Seaborn, DASC, SPARK, VAEX).
* Designed universes with a focus on performance optimization and data integrity.
* Managed databases, ensuring optimal performance, security, and scalability in environments that may include SQL and NoSQL databases.
* ETL Integration service with SSIS.
* Implemented complex joins, contexts, and aliases to ensure accurate data retrieval.
* Proficient in Spark, leveraging its distributed processing capabilities for big data analytics.
* Development of projects in python for customer segmentation and clustering.
* Customer segmentation for marketing campaigns with a greater possibility of purchasing a certain product, based on their consumption habits.
* Model development for hyper-segmented campaigns using Python.
* Designed and implemented scalable big data processing pipelines using technologies such as Apache Hadoop and Apache Spark.
* Optimized universe design to align with best practices and business needs.
* Gained knowledge of SAP HANA architecture, including its in-memory database, columnar storage, and parallel processing capabilities.
* Conducted universe testing and validation to ensure data consistency across reports.
* Gained knowledge of techniques for optimizing SAP HANA performance, including data partitioning, indexing, and aggregation, to ensure fast query processing and analytics.
* Logical design for databases and data models (Diagrams)
* Design, development, creation, and implementation of databases, as well as updating and maintenance of the same (SQL structured language).
* Implemented CI/CD pipelines for Kafka infrastructure, stream processing applications, and related services, enabling automated testing, deployment, and monitoring of changes to ensure reliability and agility in data delivery.
* Led the design and implementation of high-performance Kafka-based streaming architectures to support real-time data processing and analytics for enterprise-scale applications.
* Implemented data governance practices to maintain data integrity and security within Business Objects.
* Designed and developed Business Objects reports based on user requirements.
* Developed and maintained Business Objects universes to provide a semantic layer for reporting purposes.
* Versioning and documentation of projects in GITLAB.
* Skilled in using PySpark, the Python API for Spark, to develop data processing pipelines and analytics solutions.
* Developed basic proficiency in Terraform for infrastructure provisioning and management.
* Demonstrated ability to deploy and manage containerized applications using Kubernetes.
* Integration of services from the AWS cloud (Athena) to Datawarhouse Snowflake.
* Designed ETL pipelines in Amazon EMR using Pyspark to process the raw data from Amazon S3 and copy the data into Amazon Redshift and created views to enable fast access to the data & improved view performance by using sortkeys
* Implemented and maintained scalable data solutions using AWS services such as Amazon S3, Amazon Redshift, and Amazon RDS.
* Implemented advanced report functionalities such as drill-down, cross-tabulation, and conditional formatting.
* Gained introductory knowledge of Jenkins for implementing CI/CD pipelines.
* Expertise in snowflake to create and Maintain Tables and views.
* Developed Python Scripts, UDFs using both Data frames/SQL and RDD in Spark for data aggregation, queries and writing data back onto HDFS.
* Developed a Collaborative Recommendation Engine to recommend most suitable events, sessions & various products.
* Built Data Preprocessing Pipeline and performed data cleaning, features scaling, features extraction using Pandas and NumPy; used Matplotlib, Seaborn & Plotly in Python to visualize the data and performed Featuring Engineering such as detecting outliers, missing value and interpreting variables.
* Designed & created SQL Server and MYSQL Workbench database from scratch in order to send recommendations and receive feedback data from the frontend.
* Utilized Apache Spark with Python to develop and execute Big Data Analytics.
* Involved in converting Hive/SQL queries into transformations using Python
* Performed complex joins on tables in SQL with various optimization techniques
* Work extensively with flat files. Loading them into on-premise applications and retrieve data from applications to files.
* Prepare TDDs, Test Case documents after each process has been developed. Identify and validate data between source and target applications.
* Verify data consistency between systems.
* Leveraged AWS Lambda functions to trigger automated data processing tasks based on event-driven architecture.
* Worked with deployments from Dev to UAT, and then to Prod.

**Technologies Used** Hadoop, HDFS, Cloudera, MapReduce, Python, SQL, Power BI, Tableau, Kibana, SSIS, Unix scripts, FlatFiles, XML files,Kafka, Parquet Files.

**Data Engineer ATOS-Mexico April 2022-Aug 2022**

**Responsibilities:**

* Acted as SME for Data Warehouse related processes.
* Performed Data analysis for building Reporting Data Mart.
* Conducted regular performance tuning and optimization of AWS resources to improve data processing efficiency.
* Worked with Reporting developers to oversee implementation of report/universe designs.
* Tuned performance of Informatica mappings and sessions for improving the process and making it efficient after eliminating bottlenecks.
* Worked on complex SQL Queries, PL/SQL procedures and convert them to ETL tasks
* Worked with PowerShell and UNIX scripts for file transfer, emailing and other file related tasks.
* Worked with deployments from Dev to UAT, and then to Prod.
* Performed exploratory data analysis like statistical calculation, data cleaning and data visualizations using Numpy, Pandas and Matplotlib; Created interactive Dashboards on desktop platform to visualize the data by using Power BI in MS Excel and Tableau
* Established data storage solutions in AWS, managing data with Amazon S3 for object storage and Amazon DynamoDB for NoSQL databases.
* Designed and developed Power BI graphical and visualization solutions with business requirement documents and plans for creating interactive dashboards.
* Worked with Reporting developers to oversee the implementation of reports/dashboard designs in Tableau.
* Assisted users in creating/modifying worksheets and data visualization dashboards in Tableau.
* Tuned and performed optimization techniques for improving report/dashboard performance.
* Implemented encryption protocols to secure sensitive data, ensuring compliance with industry standards for data protection.
* Assisted report developers with writing required logic and achieve desired goals.
* Met End Users for gathering and analyzing the requirements.
* Worked with Business users to identify root causes for any data gaps and developing corrective actions accordingly.
* Created Ad hoc data reports for presenting and discussing the data issues with Business.
* Performed gap analysis after reviewing requirements.
* Identified data issues within DWH dimension and fact tables like missing keys, joins, etc.
* Wrote SQL queries to identify and validate data inconsistencies in data warehouse against source system.
* Validated reporting numbers between source and target systems.
* Finding a technical solution and business logic for fixing any missing or incorrect data issues identified
* Coordinating and providing technical details to reporting developers
* Working with data, using visualization tools and predictive modeling, analyzing past data to inform new data and solve business problems, working with stakeholders for make the business more efficient and make a solution more accurate.
* Collaborated with cross-functional teams to apply cryptographic practices in system design, emphasizing secure data handling and storage.
* Management and development of projects for large amounts of data, pipeline design as well as its ingestion either by batch or streaming through PUB/SUB, ETL and pipeline design through Dataflow, Processing with Dataproc and Workflow Orchestration with Cloud Composer, data storage for short- long term in data lakes in Cloud Storage or Datawarehouses in Bigquery for analysis through BI tools (Tableau, Kibana, Google Studio and Google looker). Use of tools such as Python and libraries for data analysis and processing, as well as SQL and BSON-JSON language management for non-relational databases, Dataprep with repositories for manual transformations without code line.

**Technologies Used**: SQL Server 2005, Tableau 9.1, GCP, Power BI, SSIS, RightNow, PowerShell, Unix,

**Data Engineer AXTEL SA DE CV-Mexico Jan 2009-March 2022**

**Responsibilities:**

* Implemented reporting Data Warehouse with online transaction system data.
* Developed and maintained data warehouse for PSN project.
* Provided reports and publications to Third Parties for Royalty payments.
* Managed user account, groups and workspace creation for different users in Powercenter.
* Wrote complex UNIX/windows scripts for file transfers, emailing tasks from FTP/SFTP.
* Worked with PL/SQL procedures and used them in Stored Procedure Transformations.
* Extensively worked on oracle and SQL server. Wrote complex sql queries to query ERP system for data analysis purpose
* Worked on most critical Finance projects and had been the go to person for any data related issues for team members.
* Migrated ETL code from Talend to Informatica. Involved in development, testing and post production for the entire migration project.
* Documented the code.
* Tuned ETL jobs in the new environment after fully understanding the existing code.
* Maintained Talend admin console and provided quick assistance on production jobs.
* Involve in designing Business Objects universes and creating reports.
* Built adhoc reports using stand-alone tables.
* Involved in creating and modifying new and existing Web Intelligence reports.
* Created Publications which split into various reports based on specific vendor.
* Wrote Custom SQL for some complex reports.
* Worked with business partners internal and external during requirement gathering.
* Worked closely with Business Analyst and report developers in writing the source to target specifications for Data
* warehouse tables based on the business requirement needs.
* Exported data into excel for business meetings which made the discussions easier while looking at the data.
* Performed analysis after requirements gathering and walked team through major impacts.
* Provided and debugged crucial reports for finance teams during month end period.
* Addressed issue reported by Business Users in standard reports by identifying the root cause.
* Get the reporting issues resolved by identifying whether it is report related issue or source related issue.
* Creating Ad hoc reports as per users needs.
* Investigating and Analysing any discrepancy found in data and then resolving it.
* Leveraged C++ for prototyping certain algorithmic solutions, gaining foundational knowledge of object-oriented programming in the language.
* Assisted in basic C++ module development for data processing and analysis tasks, contributing to performance improvements.

**Technologies Used**: SQL Server 2005, UNIX, Windows Scripting, JIRA