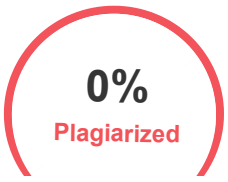
May 07, 2024

Plagiarism Scan Report



Characters:7029	Words:979
Sentences:50	Speak Time: 8 Min

Excluded URL None

Content Checked for Plagiarism

Week I Learning relevant tools and technology This week, I'll dive into mastering the essential tools and technologies crucial for our cloud data engineering journey. Starting with AWS, I'll explore its key services and functionalities, gaining a solid understanding of cloud infrastructure management. Additionally, I'll delve into Terraform, learning how to provision and manage cloud resources efficiently through infrastructure as code. To streamline our development workflow, I'll also familiarize myself with GitHub Actions, enabling seamless automation and integration within our projects. This week sets the foundation for our cloud data engineering endeavors, equipping me with the skills needed to tackle upcoming challenges effectively. Week II Started on new project ACTEN3 as data engineer. As a data engineer, I've embarked on a new project, ACTEN3, diving into its intricacies. This week, I'm immersing myself in understanding the project flow, conducting background research, and familiarizing myself with its components. I've begun by setting up the project locally, ensuring smooth functionality, and rigorously testing all its functions. Moreover, I'm actively connecting with new team members, fostering collaboration and knowledge exchange. This phase marks the initial steps toward effectively contributing to ACTEN3, laying the groundwork for successful project implementation and integration. Week III Getting familiar with the MW This week, I'm immersing myself in understanding the middleware (MW) component of our project. I'm delving deep into the codebase, dissecting its functionality, and gaining insights into its operations. Additionally, I'm focusing on learning about pre-processing files, understanding how data is transformed and prepared for further analysis or processing. By familiarizing myself with MW and its intricacies, I'm equipping myself with the knowledge and skills necessary to effectively contribute to our project's development and optimization. Week IV Learning new software Gathr. This week, I'm dedicating time to learning a new software called Gathr. I'm diving into its documentation, absorbing its features and functionalities. Alongside, I'm actively testing the software to understand its potential applications and capabilities. In this process, I'm also identifying any limitations of the software and devising strategies to overcome them by leveraging existing functions or alternative approaches. By gaining proficiency in Gathr and addressing its limitations, I aim to enhance our toolkit and streamline our data engineering processes effectively. Week V Working on Gathr This week, my primary focus is on Gathr, where I'm dedicated to optimizing its functionalities and addressing any outstanding tasks. Simultaneously, I'm tasked with developing a series of ETL (Extract, Transform, Load) processes tailored to meet specific business requirements. These include ETLs for managing Business Data, Business Relation Data, Driver Data, Truck Data, Trailer Data, Stop Data, Location Data, and Assignment

Data. Following development, I meticulously test each ETL to ensure accuracy, reliability, and alignment with business objectives. This week's efforts aim to streamline data management processes and enhance the effectiveness of our data workflows within Gathr. Week VI Resolving bugs and daily ticket. This week is dedicated to maintaining project stability by resolving bugs and addressing daily tickets promptly. Additionally, I'm implementing an automated process to trigger ETLs upon file uploads through the user interface, enhancing efficiency and reducing manual intervention. I'm also adapting ETLs to accommodate changes in business requirements, ensuring alignment with evolving needs. Furthermore, I'm creating comprehensive workflows for all ETL processes, optimizing data management procedures. Lastly, I'm integrating TigerGraph with the ETLs to leverage its capabilities for enhanced data analysis and insights. This week's focus is on enhancing productivity, adaptability, and connectivity within our data ecosystem. Week VII Migration from gathr to python driven ETL This week, we initiated the development of a configuration-driven ETL (Extract, Transform, Load) process focused on the "stop.csv" file. The primary steps included creating a dedicated configuration file to streamline and customize the ETL operations for "stop.csv". Following this, extensive testing was conducted to ensure the ETL process operates as intended, accurately handling data extraction, transformation, and loading phases. A significant enhancement was the integration of a Pydantic model, which was implemented to validate the data integrity and structure, ensuring that the information processed adheres to predefined schemas and standards. This addition not only improved the reliability of the data handling process but also enhanced the overall robustness of our ETL solution. Next Steps: Further refine the ETL process by incorporating advanced data validation techniques and explore optimization opportunities to enhance performance and scalability. Week VIII Migration from gathr to python driven ETL This week marked a significant expansion in our ETL (Extract, Transform, Load) processes, extending our operational scope to include a variety of crucial data files. The focus was on developing and implementing configuration-driven ETL processes for the following datasets: * Business Data: Established a foundational structure for handling core business metrics and information. * Load Data: Developed processes to manage data related to cargo and freight loads. * Load_BusinessRelation Data: Created a specialized ETL to handle the nuanced relationships between loads and business entities. * Driver Data: Focused on aggregating and transforming data concerning our fleet drivers. * Truck Data: Developed a configuration to process data related to the trucks in our fleet, focusing on operational metrics. * Trailer Data: Implemented an ETL process for managing data associated with our trailers, crucial for logistics and planning. * Location Data: Created a process to handle geospatial and location-based data, essential for route planning and optimization. * Assignment Data: Developed a system to manage and transform data related to assignments, including driver, truck, and load assignments. Each of these data files now has a corresponding configuration file designed to tailor the ETL process to the specific needs and structures of the data. This approach ensures flexibility, scalability, and precision in handling diverse datasets. Testing Phase: Following the development of these ETL processes, extensive testing was carried out. This phase aimed to validate the efficacy, accuracy, and reliability of our ETL operations, ensuring that data transformation and loading meet our stringent standards for data integrity and quality.

Sources