Last updated on Aug 22, 2023

What's your data warehouse documentation process?

Powered by AI and the LinkedIn community

Data warehouse documentation is the process of creating and maintaining clear, accurate, and consistent information about the design, structure, and functionality of your data warehouse. Documentation helps you and your team understand the data warehouse better, troubleshoot issues, and communicate with stakeholders. But how do you document your data warehouse effectively? Here are some tips and best practices to follow.

1 Choose a documentation tool

When deciding on a tool for documenting your data warehouse, there are many options to choose from, ranging from simple text editors and spreadsheets to more specialized software and platforms. You should consider how easy it is to create, update, and access the documentation, as well as how well it integrates with your data warehouse and other tools. Additionally, you may want to look into the flexibility and customizability of the tool for different formats and styles, as well as its security and reliability for storing and sharing the documentation.

2 Define a documentation standard

The next step is to define a standard for what, how, and when to document your data warehouse. This will help guarantee consistency, quality, and completeness of the documentation across your team and projects. When deciding on a standard, consider what types of documentation you need (e.g. data models, data dictionaries, ETL processes, business rules, and user guides), how it should be formatted and structured (e.g. using headings, tables, diagrams, and naming conventions), and when it should be created, updated, and reviewed (e.g. during development, testing, deployment, and maintenance).

3 Document the data warehouse design

The design documentation for a data warehouse describes the conceptual and logical aspects, such as data sources, data models, data flows, and data quality. This documentation helps plan, implement, and validate the architecture and components. For instance, a data source inventory is a list of the data sources that feed into the warehouse, including their names, types, locations, and owners. Additionally, a data model diagram is a visual representation of the schema with tables, columns, keys, and relationships. Furthermore, a data flow diagram displays the ETL process with its transformations, validations, and loads. Lastly, a data quality plan defines the criteria, metrics, and checks for the warehouse. Overall, design documentation assists in creating a successful data warehouse.

4 Document the data warehouse content

Content documentation is essential to understanding, managing, and optimizing your data warehouse data and processes. It includes data dictionaries that define the data elements such as names, descriptions, data types, formats, and constraints; data lineage that traces the origins, transformations, and destinations of the data; a data usage report that summarizes the consumption and utilization of the data; and a data performance report that measures the efficiency and effectiveness of the data. All of these elements are necessary for effectively managing your data warehouse.

5 Document the data warehouse maintenance

The maintenance documentation describes the ongoing and periodic tasks and procedures for keeping your data warehouse functional and up-to-date, such as backup, recovery, security, and testing. This documentation helps to prevent, resolve, and improve data warehouse issues and performance. Examples of maintenance documentation include a backup and recovery plan that outlines the backup frequency, location, and method for your data warehouse, as well as the recovery steps in case of data loss or corruption. Additionally, a security policy can define the data access and protection rules and roles for your data warehouse. Lastly, a testing plan should include the testing scope, strategy, test cases, test data, and test results.

6 Document the data warehouse feedback

The feedback documentation describes the user and stakeholder feedback and suggestions for your data warehouse, such as the requirements, expectations, and satisfaction. This documentation helps you align, adapt, and enhance your data warehouse to meet the business needs and goals. For example, you can create a requirements document that captures the user and stakeholder requirements for your data warehouse, like data scope, quality, and analysis. Additionally, an expectations document can clarify the user and stakeholder expectations for your data warehouse, like data availability, accuracy, and insights. Lastly, a satisfaction survey collects the user and stakeholder satisfaction for your data warehouse such as relevance, usability, and value.

7 Here's what else to consider

This is a space to share examples, stories, or insights that don't fit into any of the previous sections. What else would you like to add?

More relevant reading

Data Architecture
How can you create a data warehousing project plan?

Data Engineering
How can you enforce data warehouse standards?

Data Architecture
How can you implement data warehousing best practices across domains and platforms?

Data Warehousing
What's the best way to evaluate data warehouse tools?