**Data warehouse notes**

**0 Introduction:**

**Data warehouse:**

* central repository which has data from disparate sources, consistent and comprehensive data analysis
* current and historical data in one place which means we keep historical data for analyzing by time for different use cases
* separation from operations: data in data warehouse separate from operation databases because it helps improve performance, ensure data consistency, and support better decision-making across the business.
* separate compute and storage computational resources (like processing power and memory) are kept independent from the storage resources (where data is kept). This separation provides greater flexibility, scalability, and cost-efficiency
* scalable: horizontally, vertically
* support bi, analysis, reporting

**Data warehouse lifecycle:**

**01 Course structure: stack technology**

**02 Setup snowflake snowsql:**

Path for config file in snowflake: /Users/rajneet/.snowsql/config/

Account name

ev60976.eu-north-1.aws

Terminal/Git commands:

|  |  |
| --- | --- |
| cd | Go back to home directory |
| ls | List files and direcxtories in the current directory |
| pwd | Print current directory |
| cd .. | Go up one directory level |
| cd ~ | Go to the home directory |
| cmd + k | clear |
| cd <DIRECTORY> | Change directory |
| . | Current directory |
| .. | Home directory |
| git clone | create a copy of a specific repository from a remote source, such as GitHub |
| mkdir | make directory |
| rmdir | Remove directory |
| ls -a | Get hidden files also |
| .git | Git can Version control |
| .gitignore | Because we choose python in version control |
| cat | Is used for read .gitignore file |
| code . | Open visual studio |
| ls -al | Authentication(read, write or read/write) |
| git status | used to display the current state of a Git repository. |
| git add . | used to stage changes in your working directory for the next commit in a Git repository. |
| git push | this command will push your changes to the corresponding branch on the default remote repository |
| git commit -m”message” | command is used to save your changes to the local repository in Git |
| git pull |  |

**Glossary:**

|  |  |
| --- | --- |
| Data Warehouse | digital library where all the information a company collects is organized and stored |
| Upstream | The steps involved in collecting, generating, or processing data before it is stored in the data warehouse |
| Downstream | The steps that happen after the data is stored—accessing, analyzing, reporting, distributing, and using the data to make decisions. |
| OLTP | Online transactional processing |
| OLAP | Online analytical processing |
| Cloud computing | It is a specific type of computing that leverages remote, internet-based resources managed by a cloud provider, offering greater flexibility, scalability, and reduced management overhead. |
| Computing | It is the broader term, referring to using computers and other devices to process data and perform tasks, usually on physical hardware you manage yourself. |
| Storage | the technology and methods used to store digital data |
| Scalable | the ability of a system, process, or technology to handle increased workload or demand without compromising performance or efficiency. |
| Vertical scalability (Scaling Up) | This involves adding more power to an existing system, such as upgrading the CPU, adding more RAM, or increasing storage capacity in a single server. |
| Horizontal scalability (Scaling out) | This involves adding more machines or nodes to a system, such as adding more servers to a network or cloud environment |
| dlt | Data load |
| dbt | Data build tool |
| IDE | Integrated development environment |
| LLM | Large language model |
| CLI | Command line interface |