* The redesign of Snowflake data warehouse system was to adapt to the change that public cloud platforms offer unlimited compute and storage resources on demand, and the emerging of SaaS model.
* The advantages of Snowflake data warehouse system include: Pure SaaS experience, relational (supporting ANSI SQL and ACID transactions), semi-structured (offering built-in functions and SQL extensions), elastic, highly available, durable, cost-efficient, and secure.
* Elasticity is the ability to scale independently and seamlessly. As data and platform change overtime, it’s important to process large amount of live data without impacting on data availability or performance of concurrent queries. Snowflake employs Amazon S3 and Virtual Warehouse to enable this elasticity, which allows users to dynamically match their compute resources to usage demands, independent of the data volume.
* Snowflake separates storage and compute. Compute is provided through Snowflake’s proprietary shared-nothing engine for its scalability and commodity hardware, while storage is provided through Amazon S3, to reduce network traffic between compute nodes and storage nodes.
* When a new query is submitted, each worker node in the respective VW spawns a new worker process. Each worker process lives only for the duration of its query, and never causes externally visible effects. Therefore, Snowflake can easily contain and routinely resolve worker failures by retries. MapReduce, on the other hand, would have the master reschedules the tasks handled by the worker if a worker node fails.