# WHERESCAPE® & SNOWFLAKE®

DATA WAREHOUSING SOLUTIONS

# AGENDA

- History of Company
- What is it?
- Features
- Benefits
- Disadvantages/Limitations

# SNOWFLAKE®

BUILT FOR THE CLOUD

# SNOWFLAKE® - HISTORY

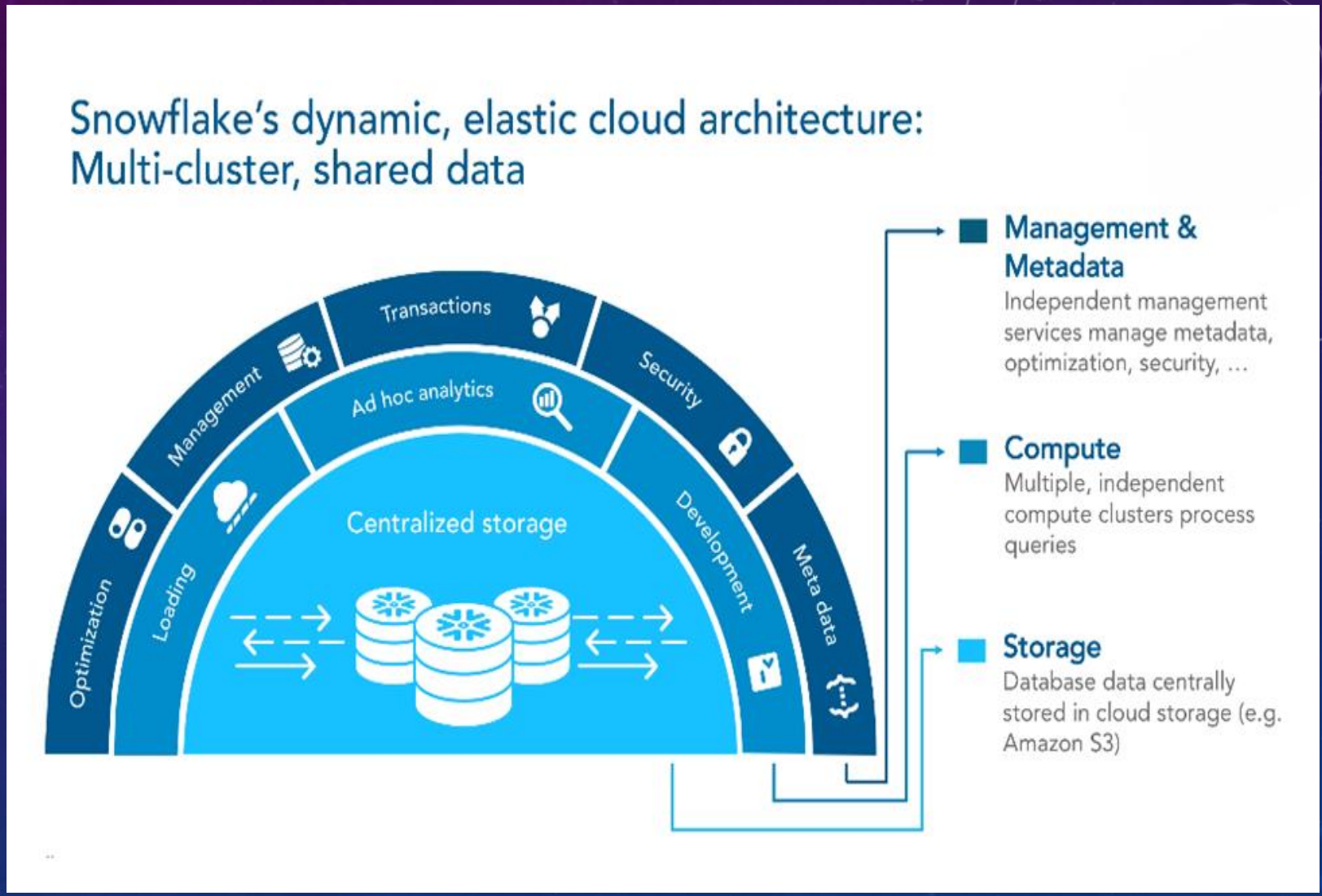
Founded in 2012 by Benoit Dageville, Thierry Cruanes and Marcin Zukowski

Came out of stealth mode in Oct 2014

Currently sells cloud-based storage and analytics service called Snowflake Elastic Data Warehouse

# WHAT IS IT?

- Cloud-based data warehouse (SaaS)
- A multi-cluster, shared data architecture that is dynamic and highly scalable thanks to enterprise class cloud-based storage systems.
- Multiple clusters all access the same underlying data, but they run independently and without contention, enabling heavy queries and operations to run simultaneously without issue.



# FEATURES

- "Time Travel"
  - Allows track changes over time: good for looking at historical data
- Cloning
  - Create an instant copy of items such as databases, schemas, tables near real-time
- Undrop
  - Useful for restoring dropped objects

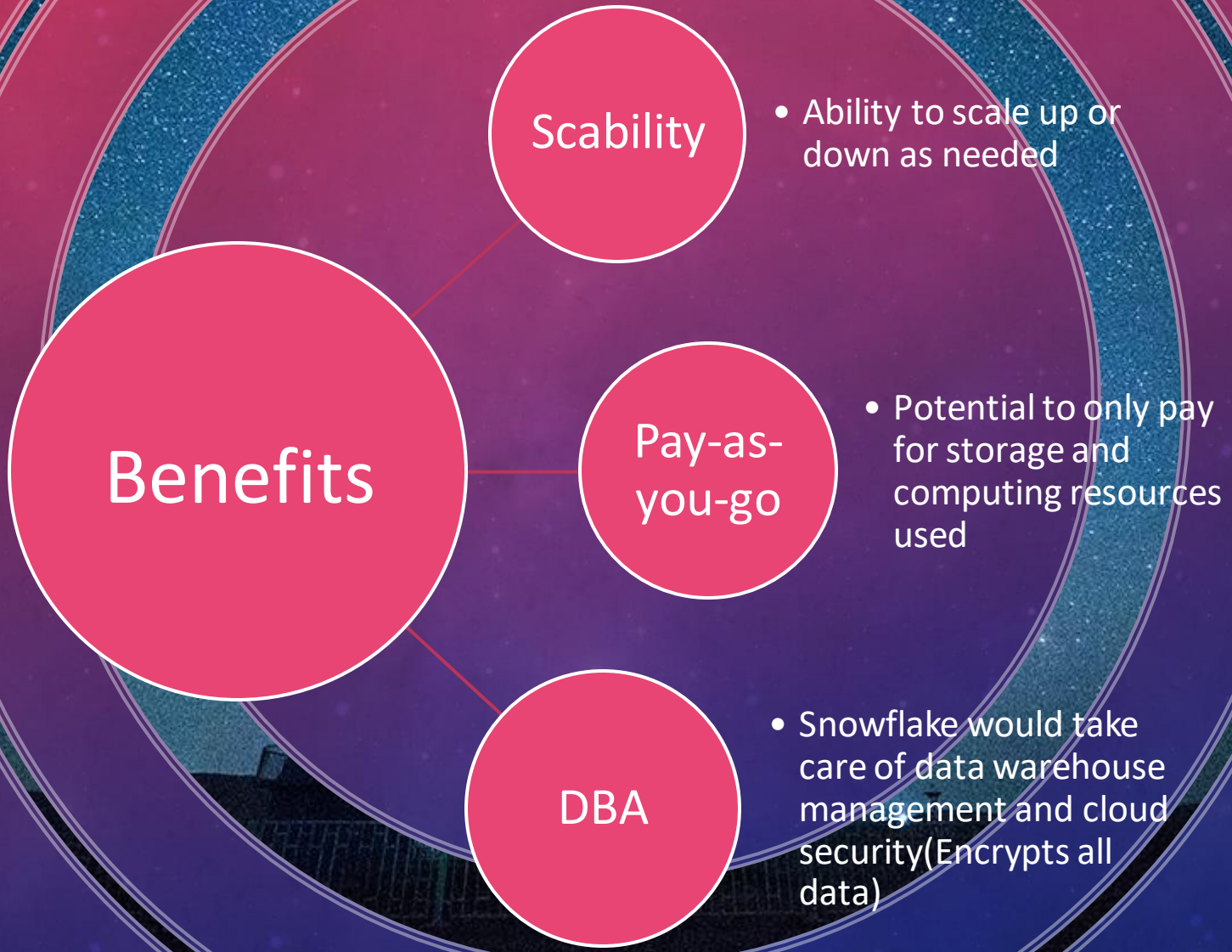
*as long as the object has not been purged from the system

- Fail-Safe
  - Ensures historical data is protected in case of disk failure or any other hardware failures

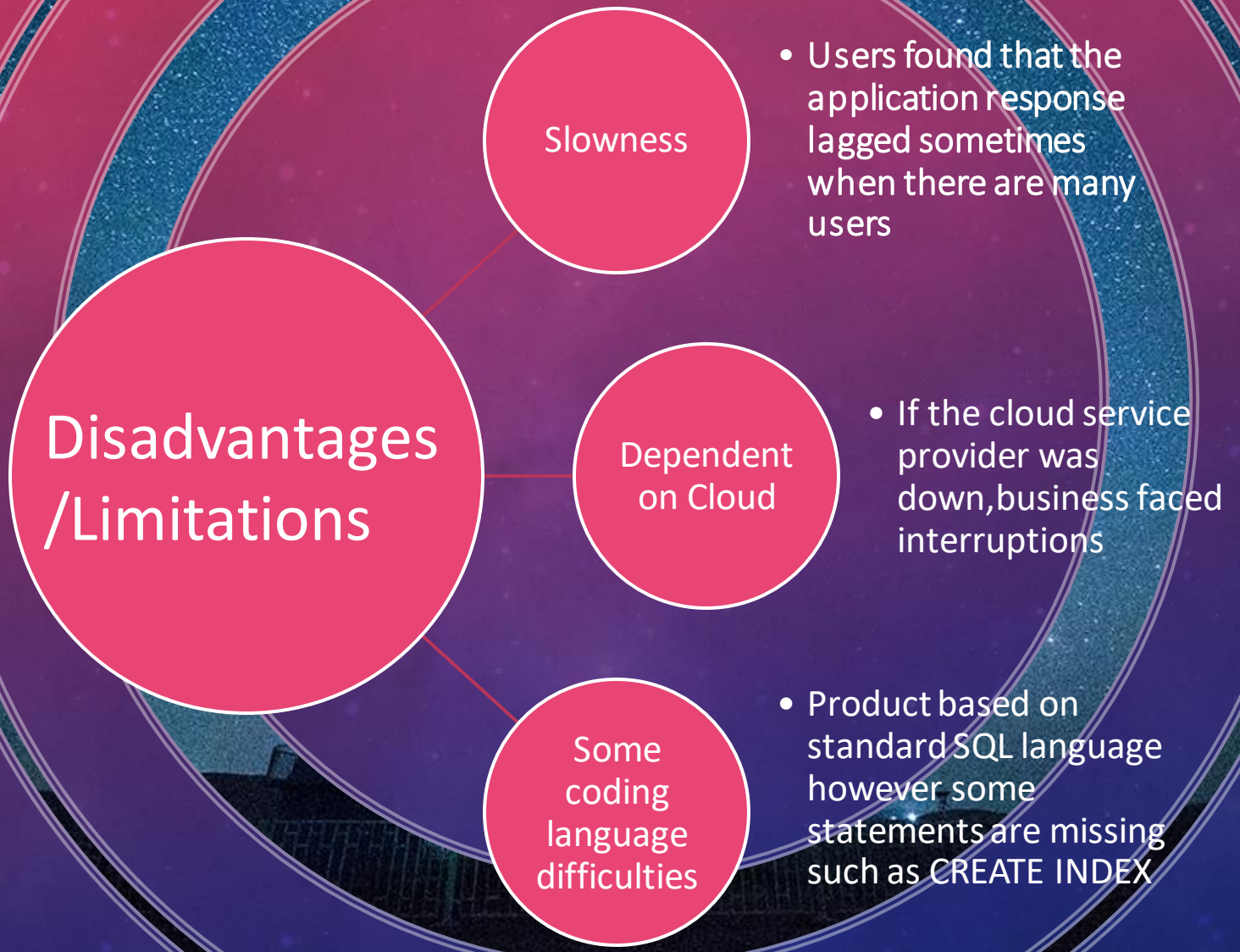
*up to 7 days worth but only recoverable by Snowflake and must be classified as catastrophic

# HOW IT WORKS

# SNOWFLAKE®

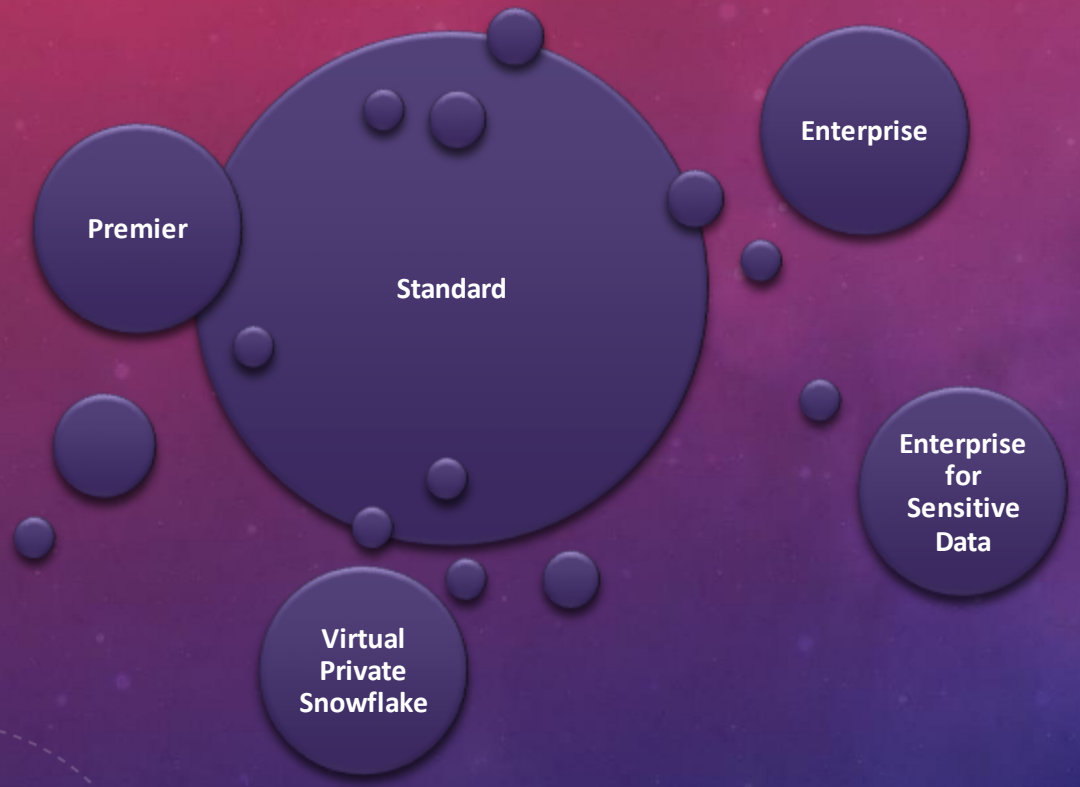

Benefits
Scability
Ability to scale up or down as needed
Pay-as-you-go
Potential to only pay for storage and computing resources used
DBA
Snowflake would take care of data warehouse management and cloud security(Encrypts all data)

# SNOWFLAKE®




- Users found that the application response lagged sometimes when there are many users
- If the cloud service provider was down, business faced interruptions
- Product based on standard SQL language however some statements are missing such as CREATE INDEX

# PRICING
## -CHOICE BETWEEN AWS & AZURE




- The Snowflake structure separates data warehousing into two distinct functions: storage and virtual warehouses
- Pricing is based on the volume of data stored in Snowflake and the compute time used
- Payment Options: On Demand & Pre-Purchased Capacity

# PRICING

## Storage

Customers charged monthly fee for data stored in Snowflake

Cost is measured using the avg storage used per month after compression

## Virtual Warehouse

Paid for using Snowflake Credits

Snowflake Credits: a unit of measure that is used to pay for the processing time of a Virtual Warehouse

Snowflake Credits are only consumed when a Virtual Warehouse is running and is consumed at different rates based on the size of the running warehouse

# PRICING EXAMPLE

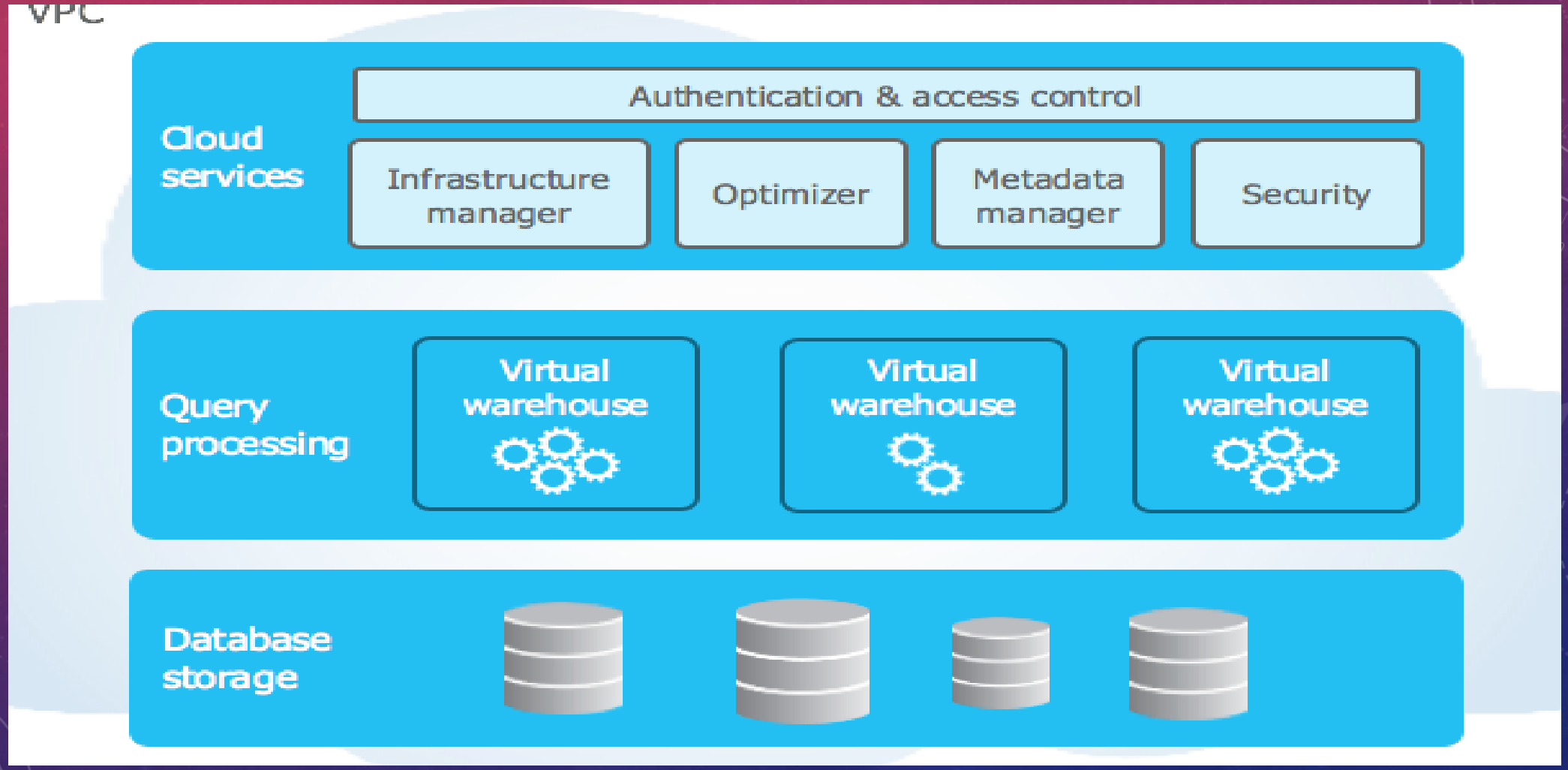
## Azure – East US 2

| **Plans** | |
|---|---|
| Standard | 2.00 **compute cost per credit** |
| Premier | 2.25 **compute cost per credit** |
| Enterprise | 3.00 **compute cost per credit** |
| Enterprise for Sensitive Data | 4.00 **compute cost per credit** |
| On-Demand Storage | **40.00 per TB/month** |
| Capacity Storage | **23.00 per TB/month** |

## AWS - Canada

| **Plans** | |
|---|---|
| Standard | 2.25 **compute cost per credit** |
| Premier | 2.50 **compute cost per credit** |
| Enterprise | 3.50 **compute cost per credit** |
| Enterprise for Sensitive Data | 4.50 **compute cost per credit** |
| On-Demand Storage | **46.00 per TB/month** |
| Capacity Storage | **25.00 per TB/month** |

# ARCHITECTURE



# ARCHITECTURE

Database Storage

Snowflake reorganizes the data into its internal optimized, compressed, columnar format. It stores this optimized data in cloud storage.

Query Processing

Snowflake processes queries using “virtual warehouses”. Each virtual warehouse is an MPP compute cluster composed of multiple compute nodes allocated by Snowflake from a cloud provider.

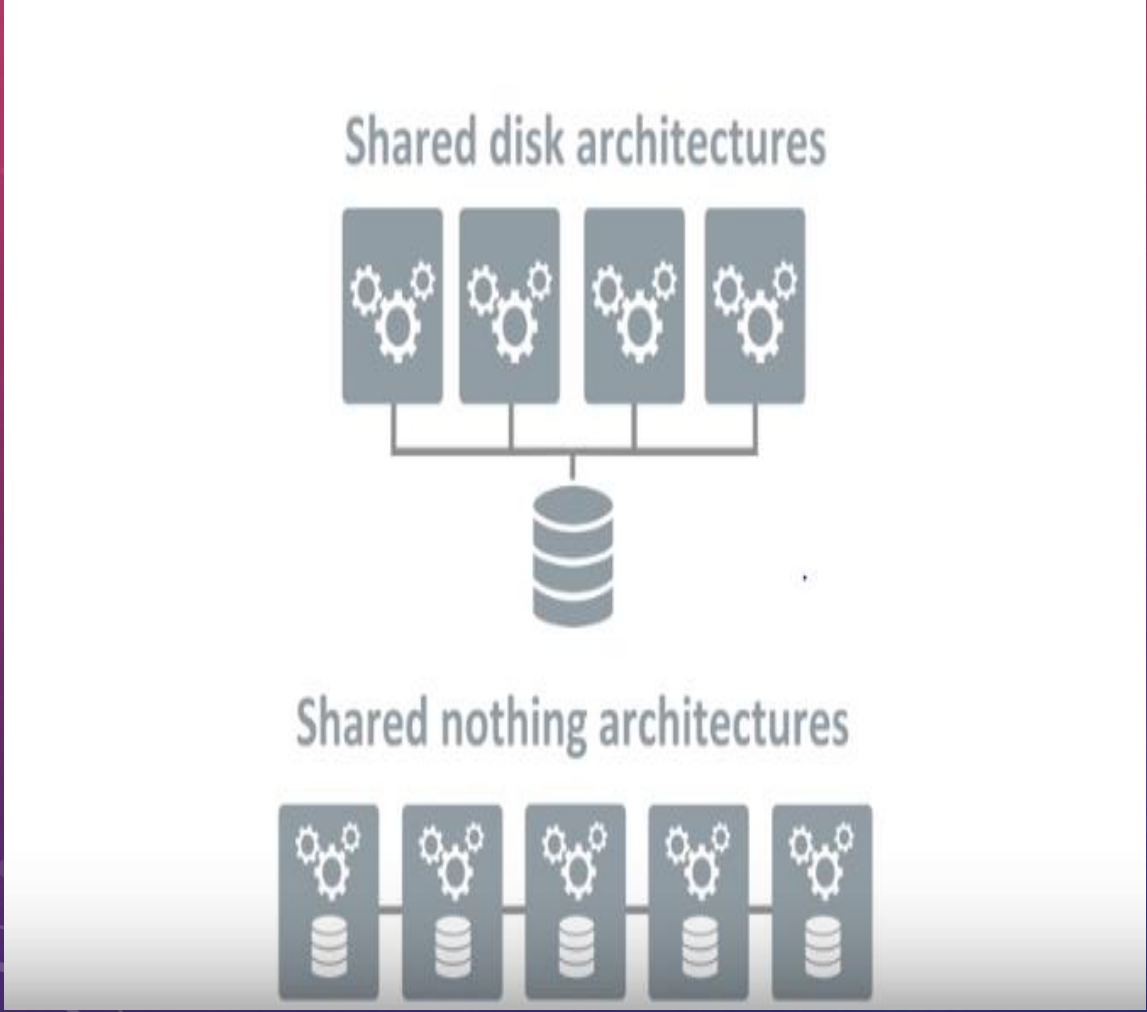
Cloud Services

Authentication(to access the data in database ), Infrastructure management, Metadata management, Query parsing and optimization

Access control

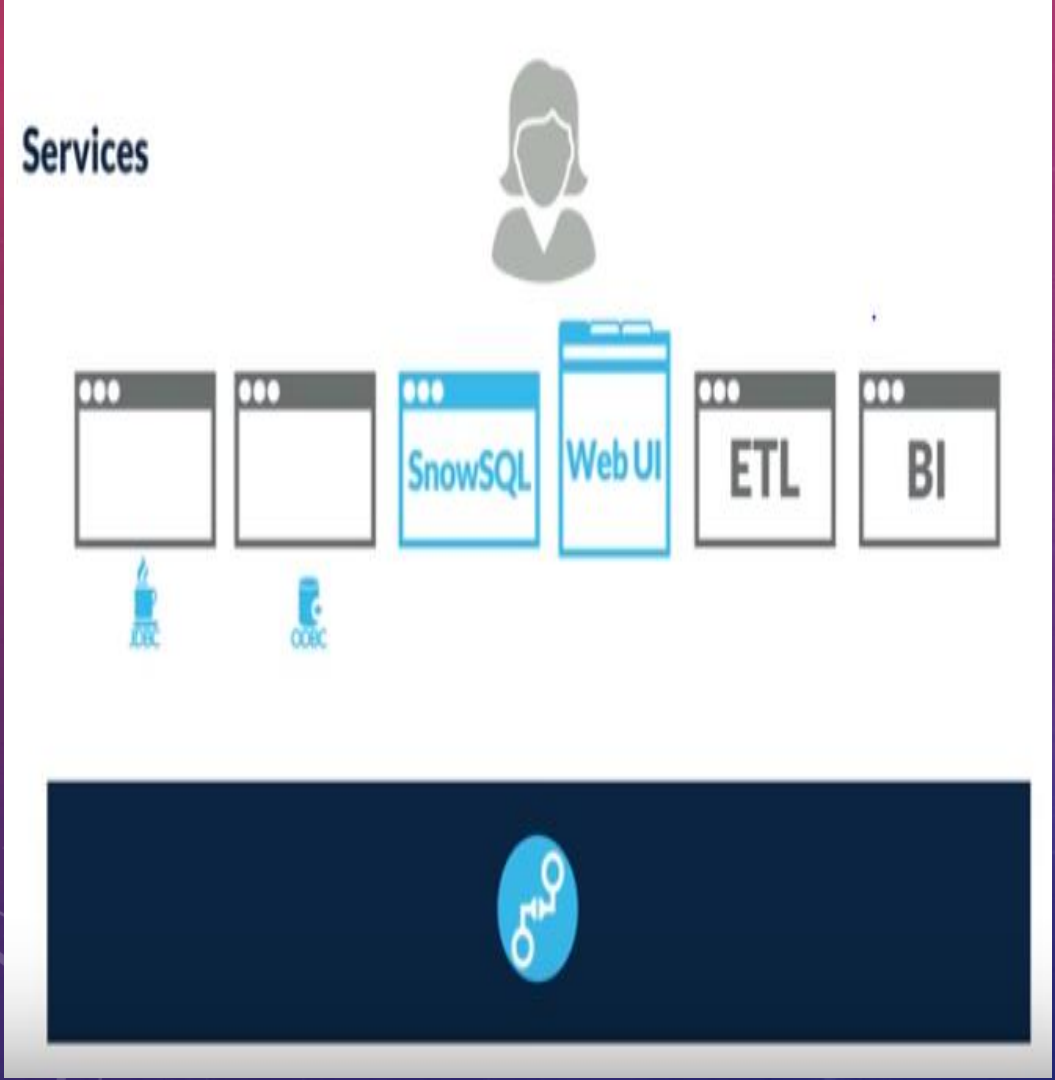
- ALLOWS TO CREATE MUL TIPLE COMPUTE INDEPENDENT CLUSTERS called VIRTUAL dw THAT ALL ACCESSES THE SAME DATA STORAGE LAYER without performance issue
- STORE SAME DATA LAYER

# DIFFERENT FROM TRADITIONAL ARCHITECTURE




- Hybrid of Shared and shared nothing architecture
- Shared - Uses central data repository for data that is accessible from all compute nodes
- Shared-nothing architectures - Snowflake processes queries using MPP (massively parallel processing) compute clusters where each node in the cluster stores a portion of the entire data set locally.

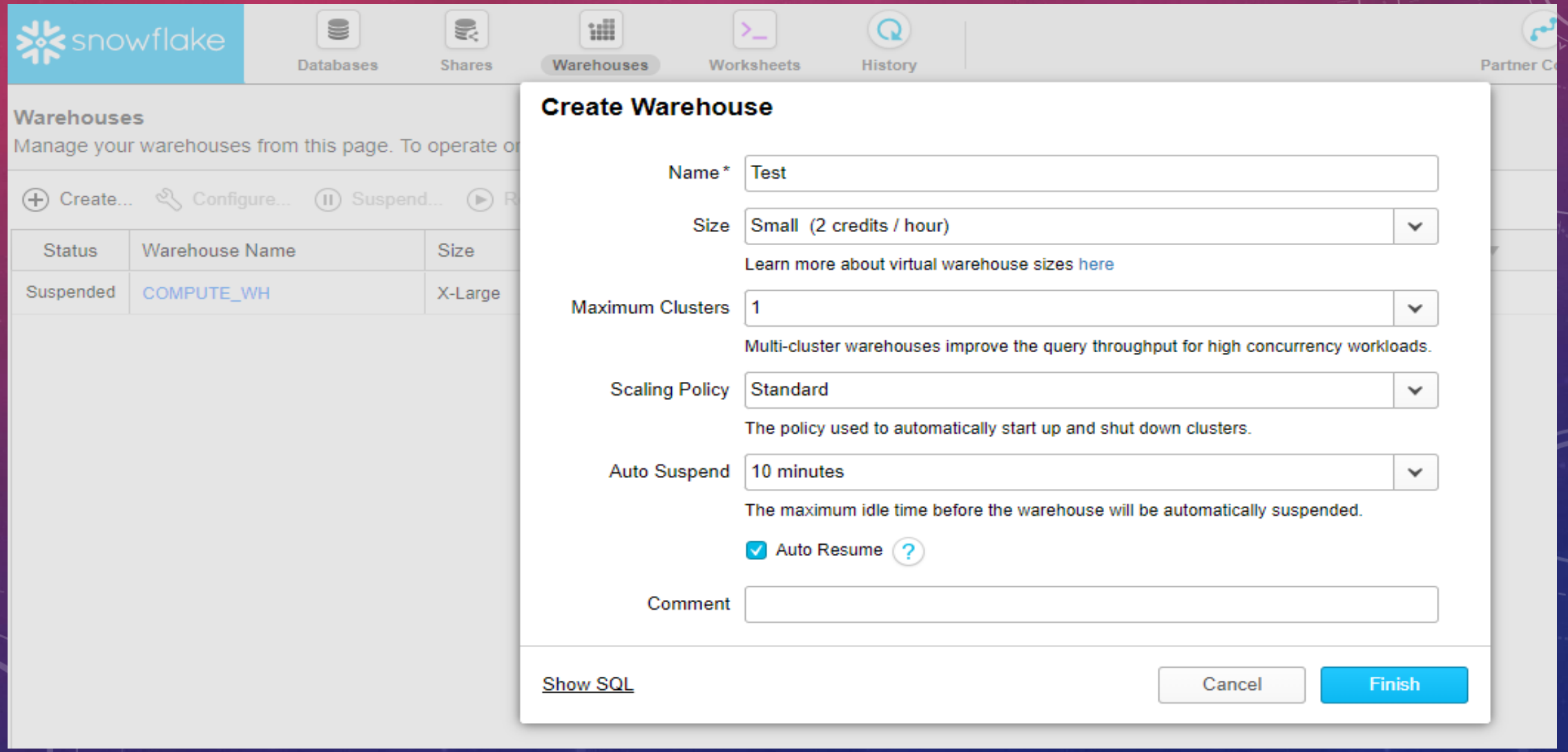
# CONNECTING TO SNOWFLAKE



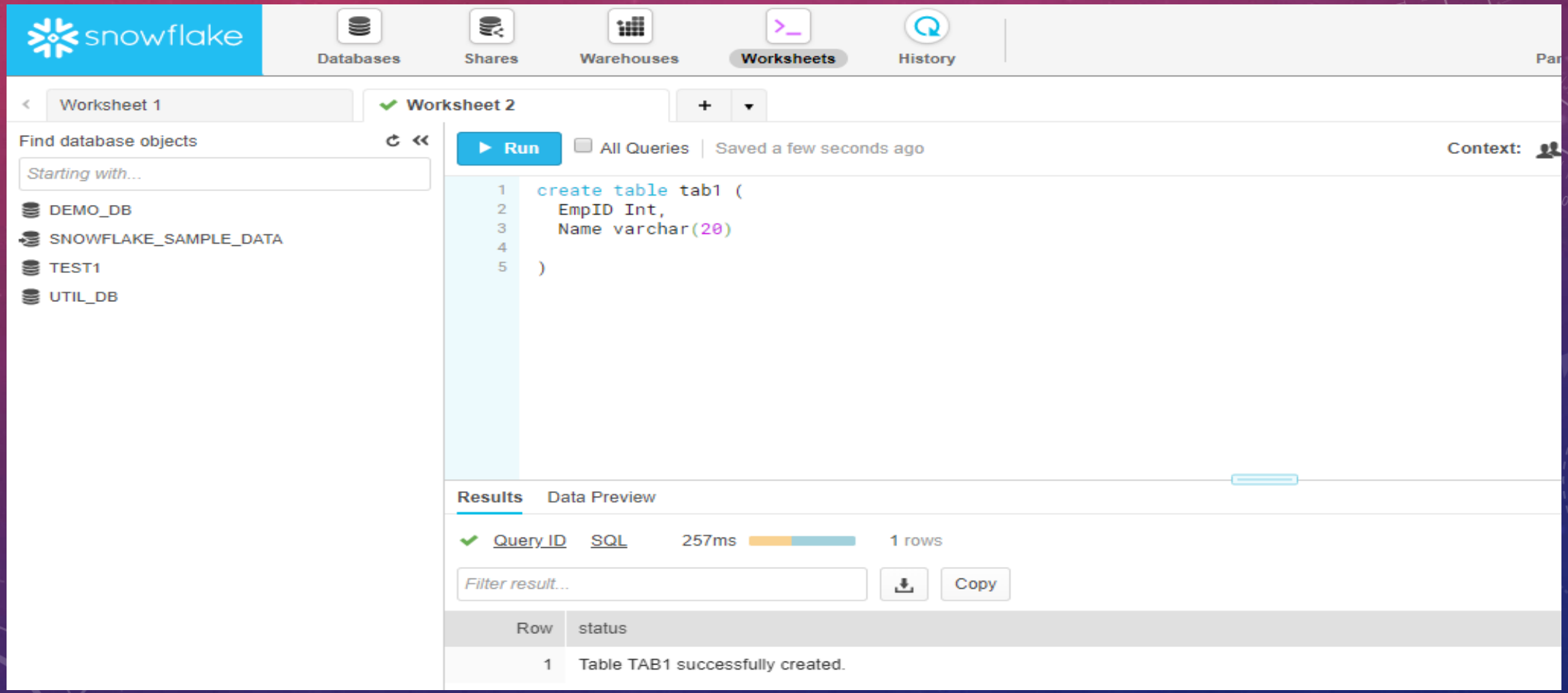

Multiple ways of connecting to the service:

- A web-based user interface from which all aspects of managing and using Snowflake can be accessed.
- Command line clients (e.g. SnowSQL) which can also access all aspects of managing and using Snowflake.
- ODBC and JDBC drivers that can be used by other applications (e.g. Tableau) to connect to Snowflake.
- Native connectors (e.g. Python) that can be used to develop applications for connecting to Snowflake.
- Third-party connectors that can be used to connect applications such as ETL tools (e.g. Informatica) and BI tools to Snowflake.

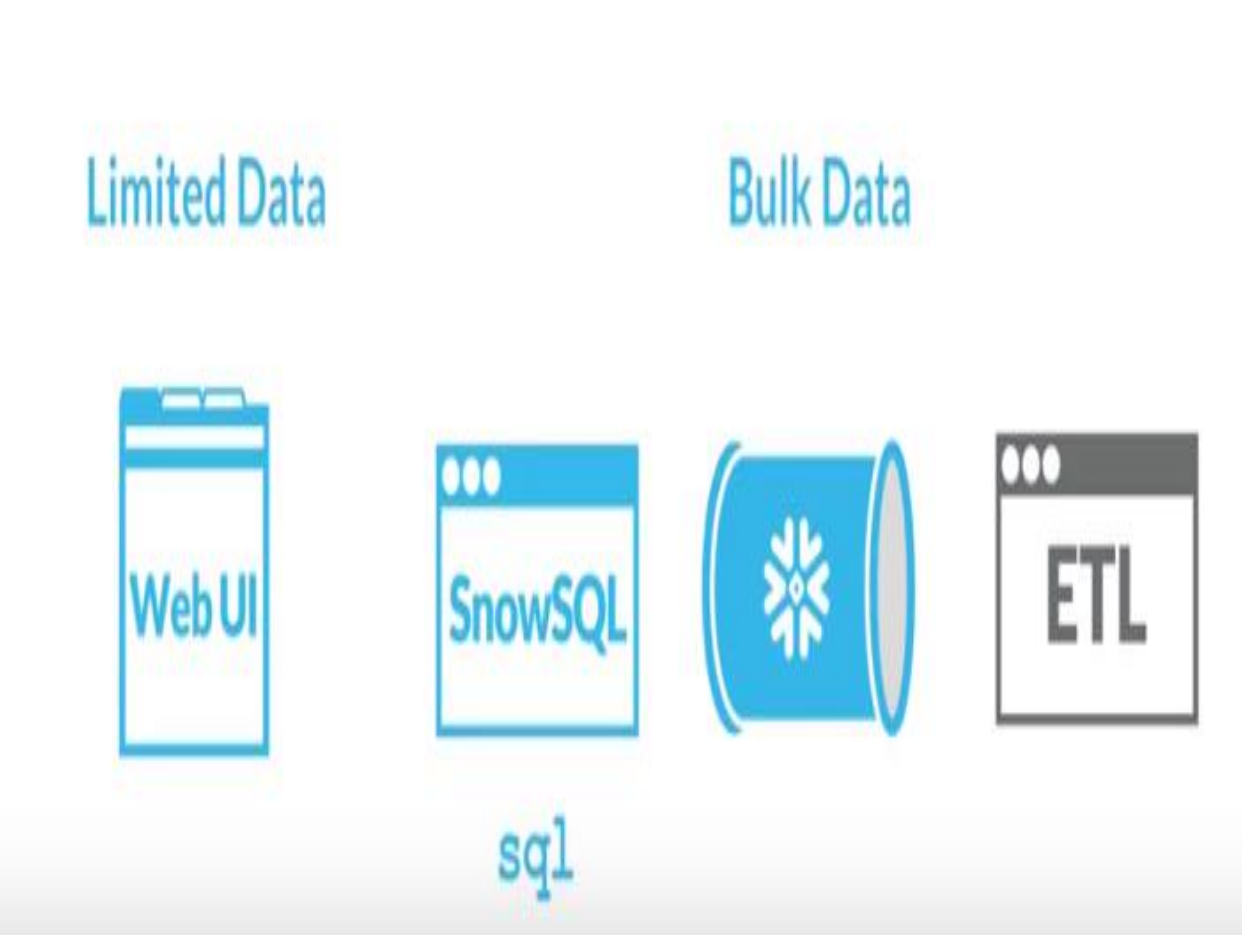
# CREATING A WAREHOUSE



# WORKSHEET AND QUERIES

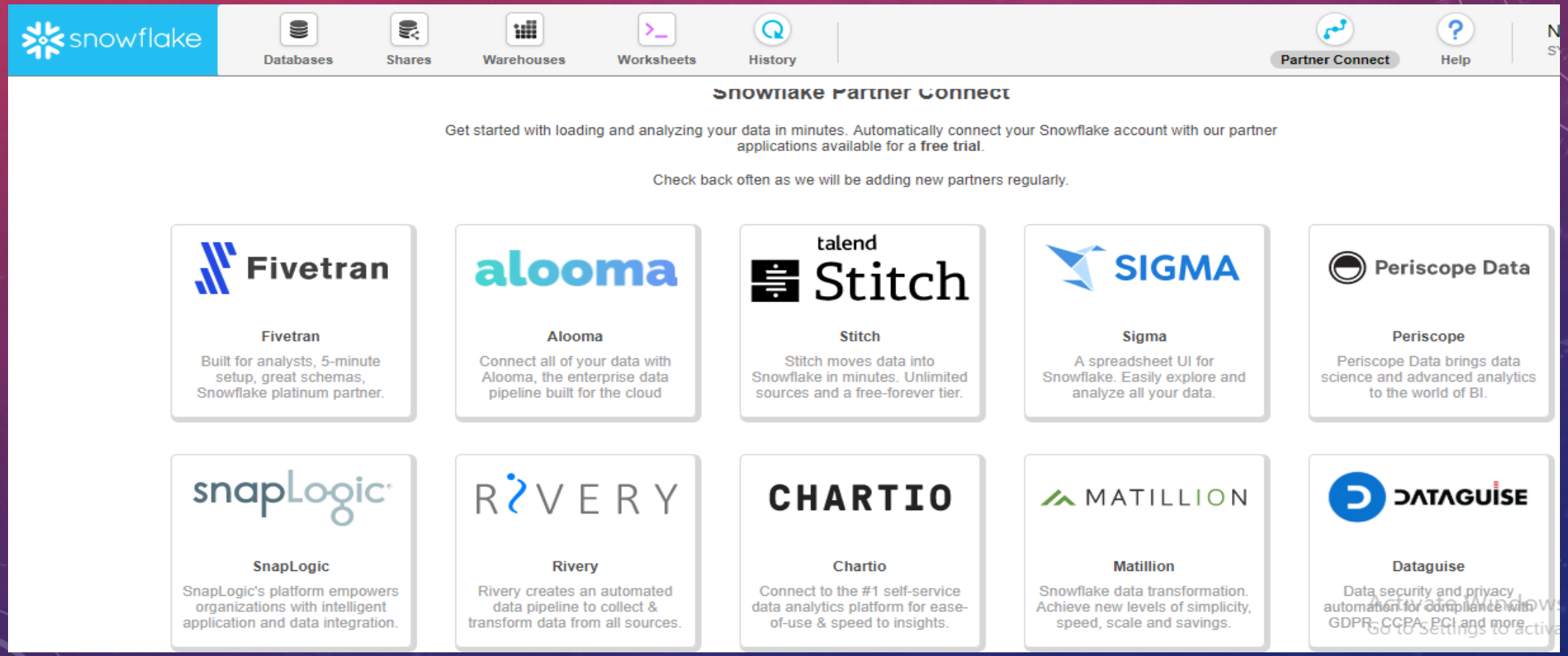


# DATA LOADING OPTIONS




- Bulk load is done in 2 phases. Stage the data first and then load into DW.
- Web interface using Load button (structured and semi structured)
- Snowpipe uses the copy command along with other features to load data in DW
- ETL process is used

# SNOWFLAKE PARTNER



# WHERESCAPE®

# WHERESCAPE® - HISTORY

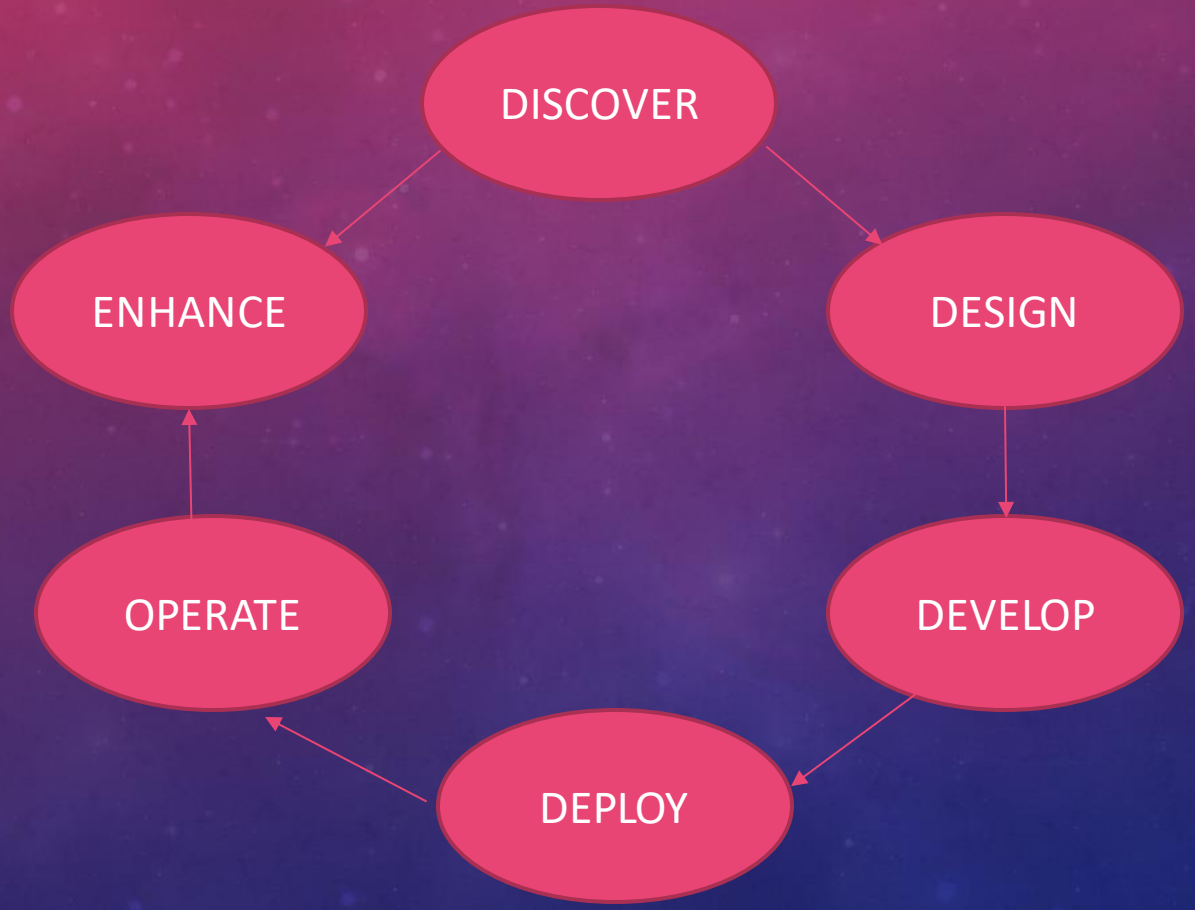
- Founded in 1997 by co-founders Michael Whitehead and Wayne Richmond and formerly known as Profit Management Systems
- Currently offers three products
  - WhereScape® 3D
  - WhereScape® RED
  - WhereScape® Data Vault Express

# WHAT IS IT?

- Automation software that allows companies to access information and insight in an agile, scalable and extensible way by automating the data management lifecycle.
- Eliminates the manual effort required to design, deploy and operate a data warehouse

# WHERESCAPE AUTOMATION LIFECYCLE

"WhereScape helps IT organizations of all sizes leverage automation to design, develop, deploy, and operate data infrastructure faster".




SOURCE: www.wherescape.com

DISCOVER

- Connection to a wide range of source systems.
- Integrates data from multiple sources.
- Applies various versions of SQL to query data.

DESIGN

- Creation of entity relationship diagrams.
- Data profiling.
- Modeling star schema, DVs, DLs, DMs.
- Model conversions (example 3nf to DV, 3nf to star schema).
- Tag columns as business keys in the source model (Volvo).

SOURCE: www.wherescape.com

## DEVELOP

- Generation of consistent, robust and stable codes with automation optimized for target data platforms.
- Guidance with wizards and templates.

## DEPLOY

- Leverage schema management and risk reduction with governed secured and audited approach to deployment.

SOURCE: www.wherescape.com

## OPERATE

- Integrated scheduling.
- Dependency management.
- Central logging and auditing.
- Performance metrics.
- Generates profiling documentation.

## ENHANCE

- Be ready for change, easily identify impacts and quicker to iterate.

SOURCE: www.wherescape.com

## BENEFITS

AGILE : Uses a metadata driven approach to generate schema and manage workflows.

LINEAGE : Shows where data comes from and what rules have been applied.

CLOUD READY : Templates for all major PaaS vendors (Snowflake, Azure SQL Database etc)

## CHALLENGES

OLD HABITS : Companies still rely on traditional, hand-coded approach to building warehouses.

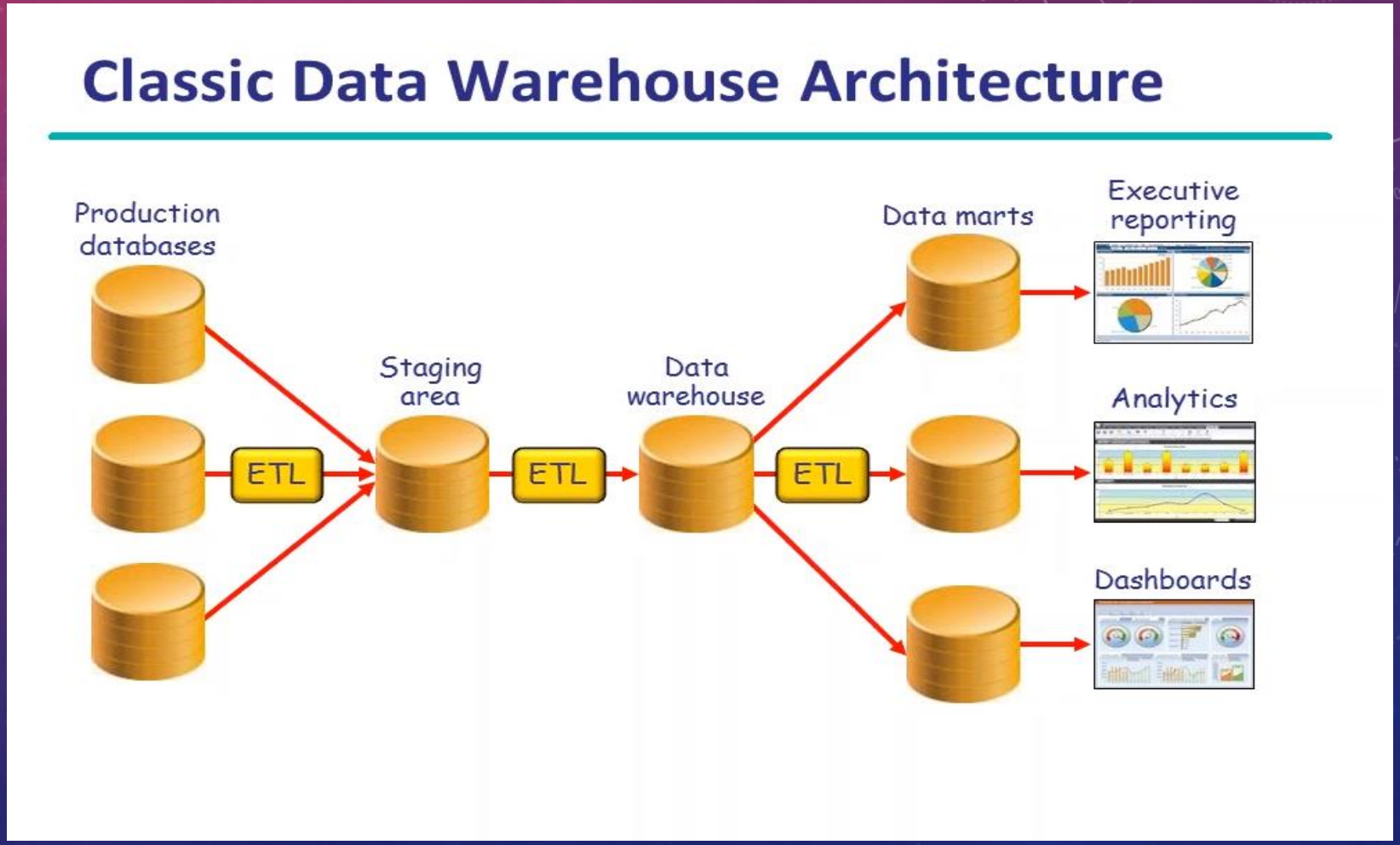
COST : DWA tools are not inexpensive. License can cost ~ $50,000.

RELATIONAL FOCUSED : Still focused on relational data warehousing

SOURCE: www.wherescape.com

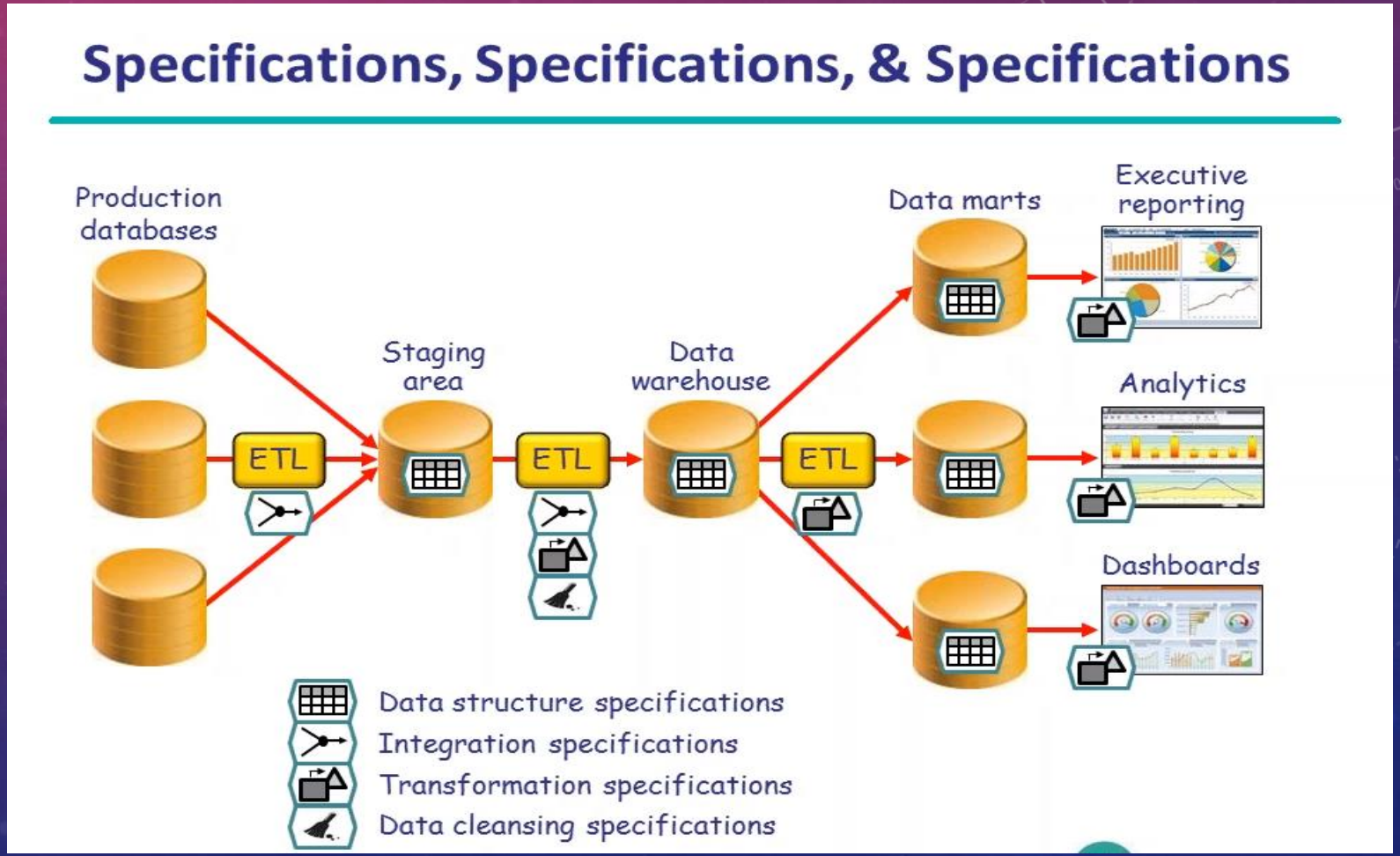
# WHERESCAPE FEATURES

Comparison with Traditional DataWarehouse Architecture



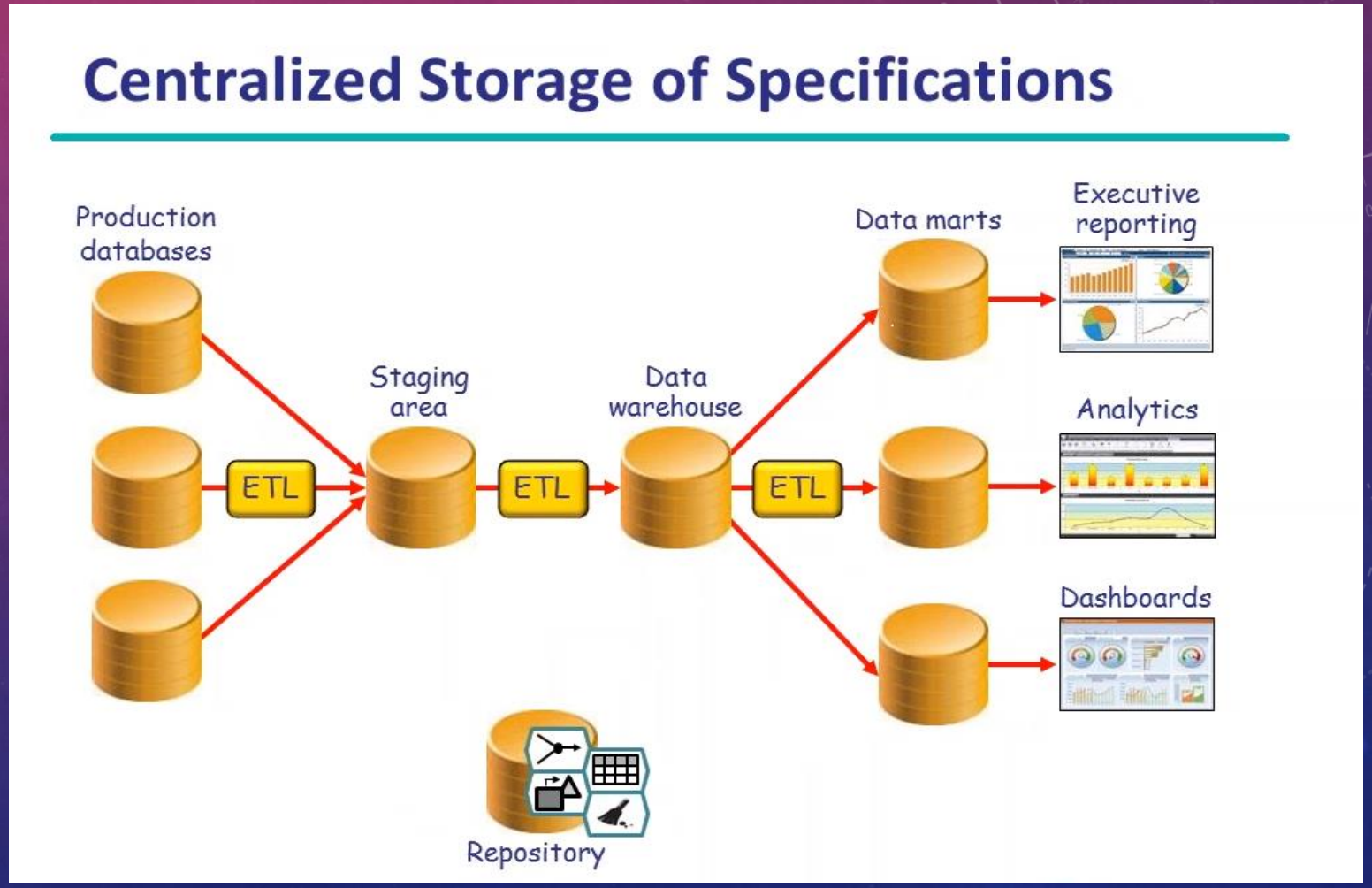
# WHERESCAPE FEATURES

Issues with Traditional Datawarehouse Architecture



# WHERESCAPE FEATURES

Wherescape's meta-data driven approach.



# CONCLUSION

- Automation is critical to the success of data warehousing projects.
- It eliminates the manual efforts required to design, deploy and operate a datawarehouse
- Snowflake allows businesses the flexibility to innovate, scale and grow at own pace
- Combining WhereScape’s data warehouse automation tools with Snowflake's zero administration service is the key to agile data warehousing

# QUESTIONS?

# THANK YOU!