



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

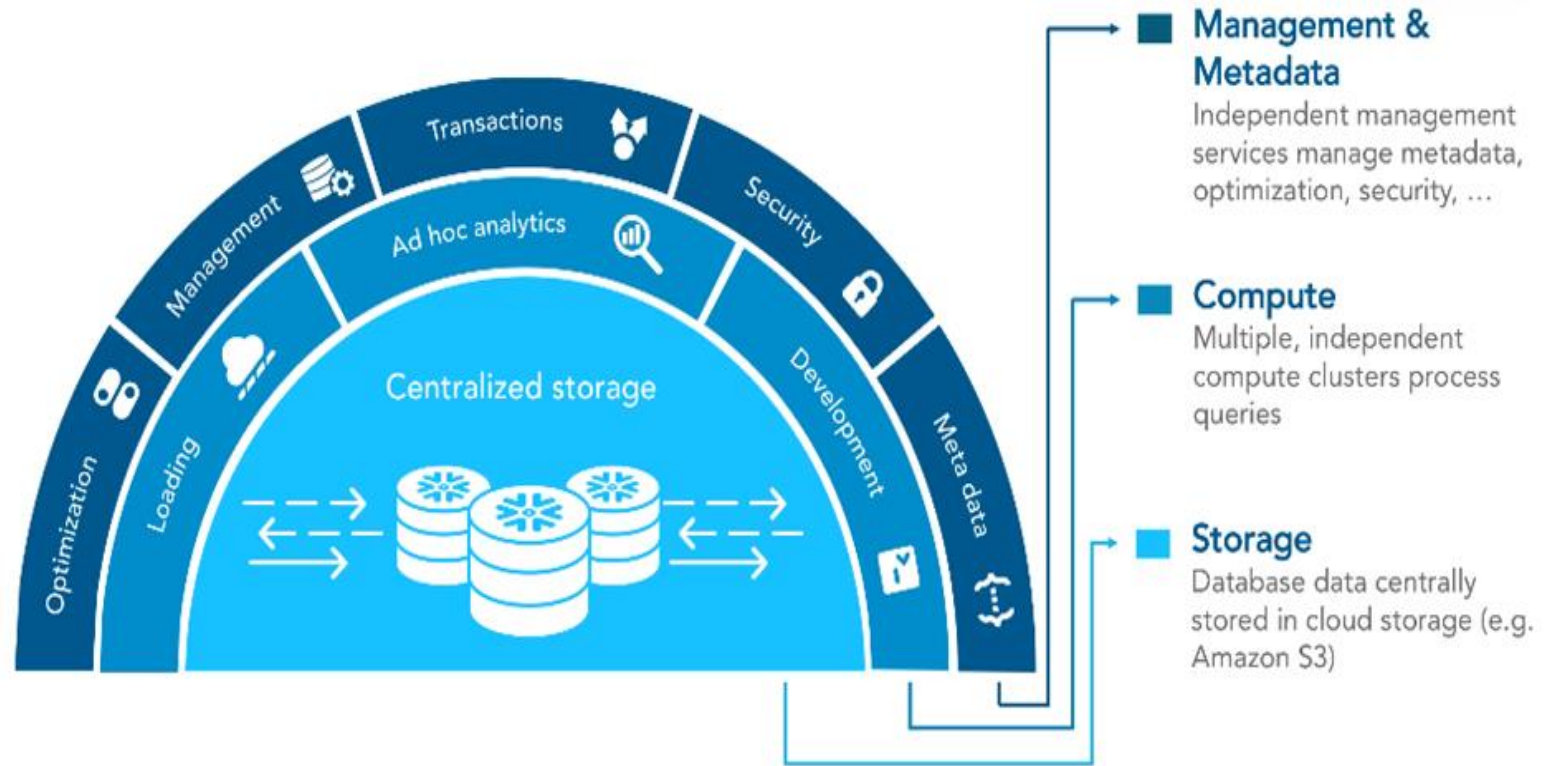
Came out of stealth mode in Oct 2014

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.

Snowflake's dynamic, elastic cloud architecture: Multi-cluster, shared data



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

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

*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®

Disadvantages /Limitations

Slowness

- Users found that the application response lagged sometimes when there are many users

Dependent on Cloud

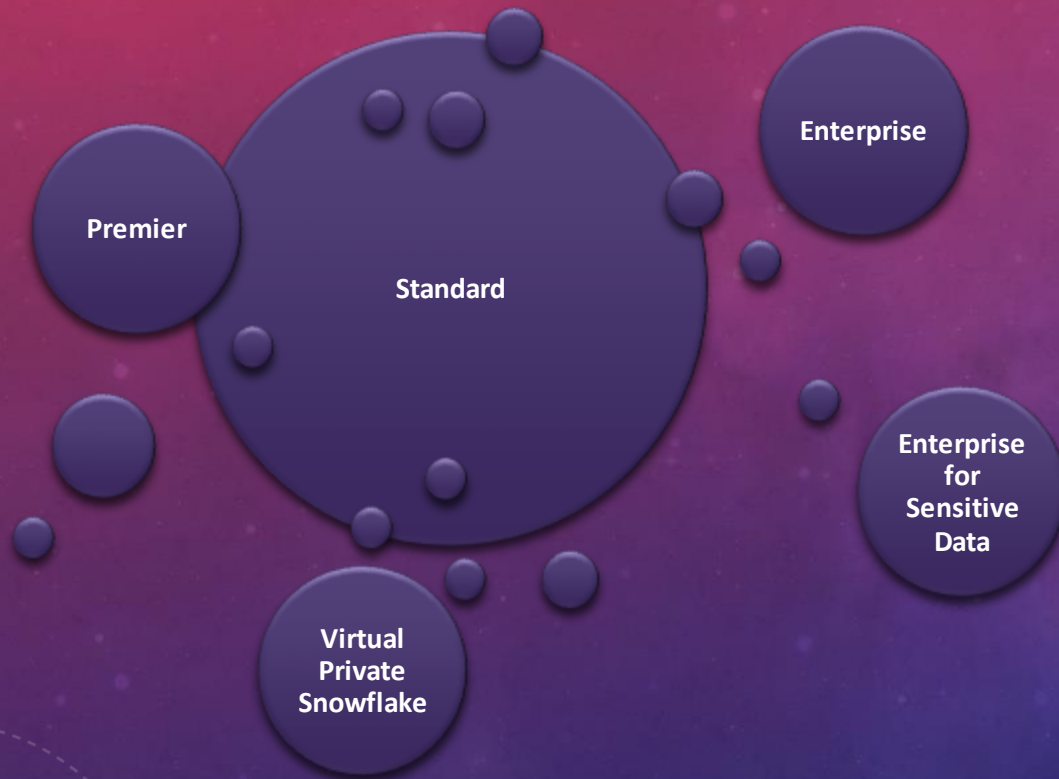
- If the cloud service provider was down, business faced interruptions

Some coding language difficulties

- Product based on standard SQL language however some statements are missing such as CREATE INDEX

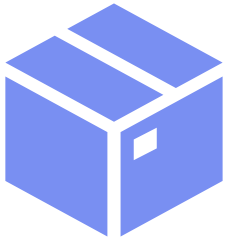
PRICING

-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

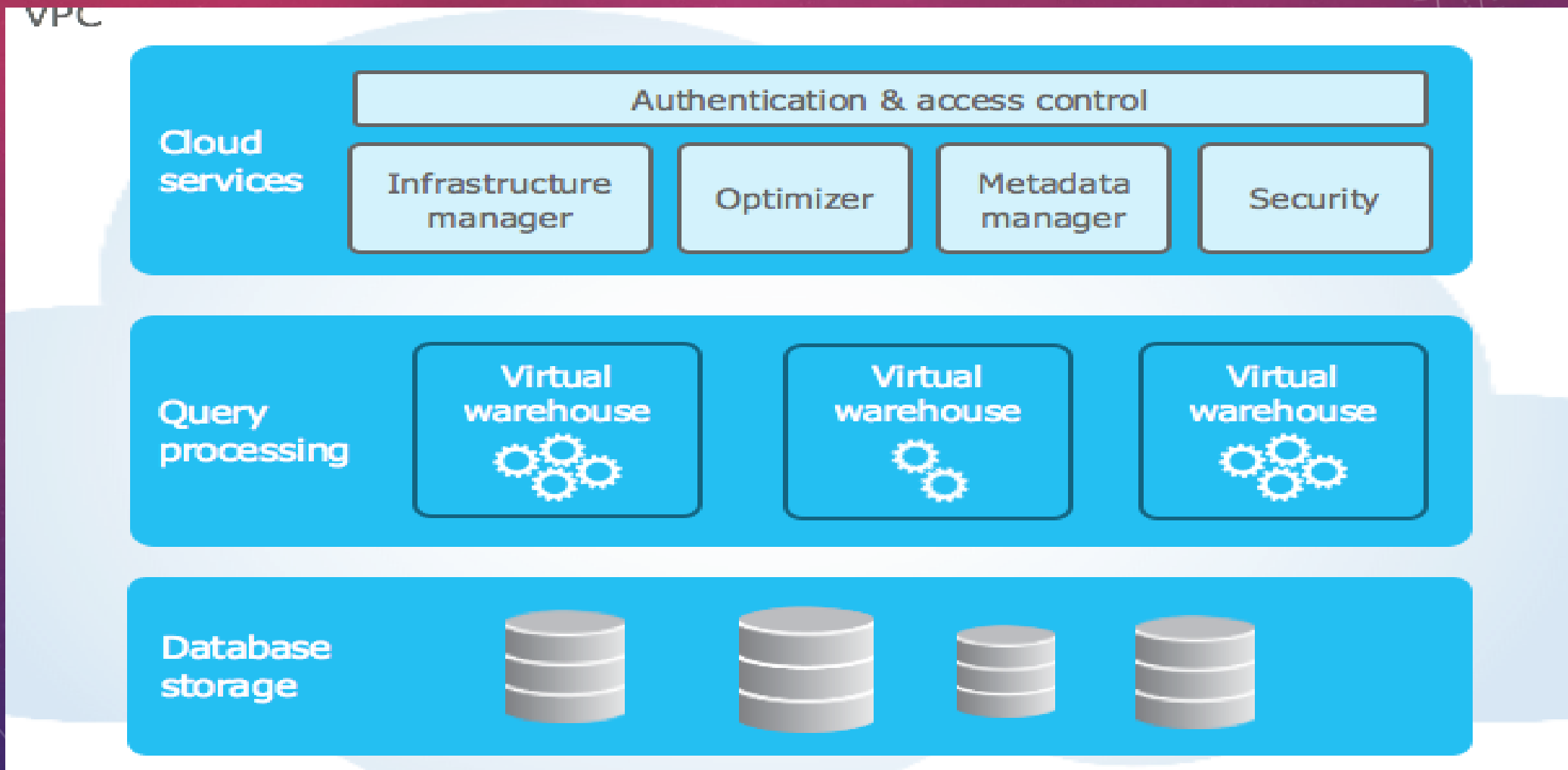
Azure – East US 2

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

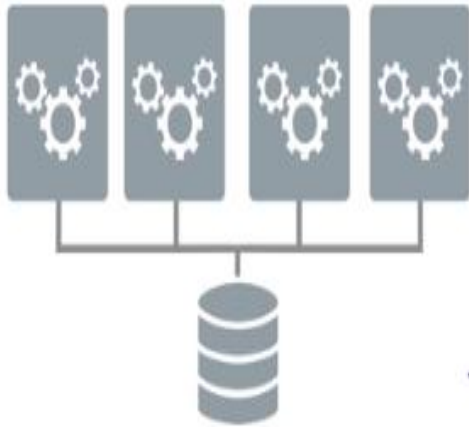
Authentication(to access the data in database), Infrastructure management, Metadata management, Query parsing and optimization

Access control

- ALLOWS TO CREATE MULTIPLE 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

Shared disk architectures

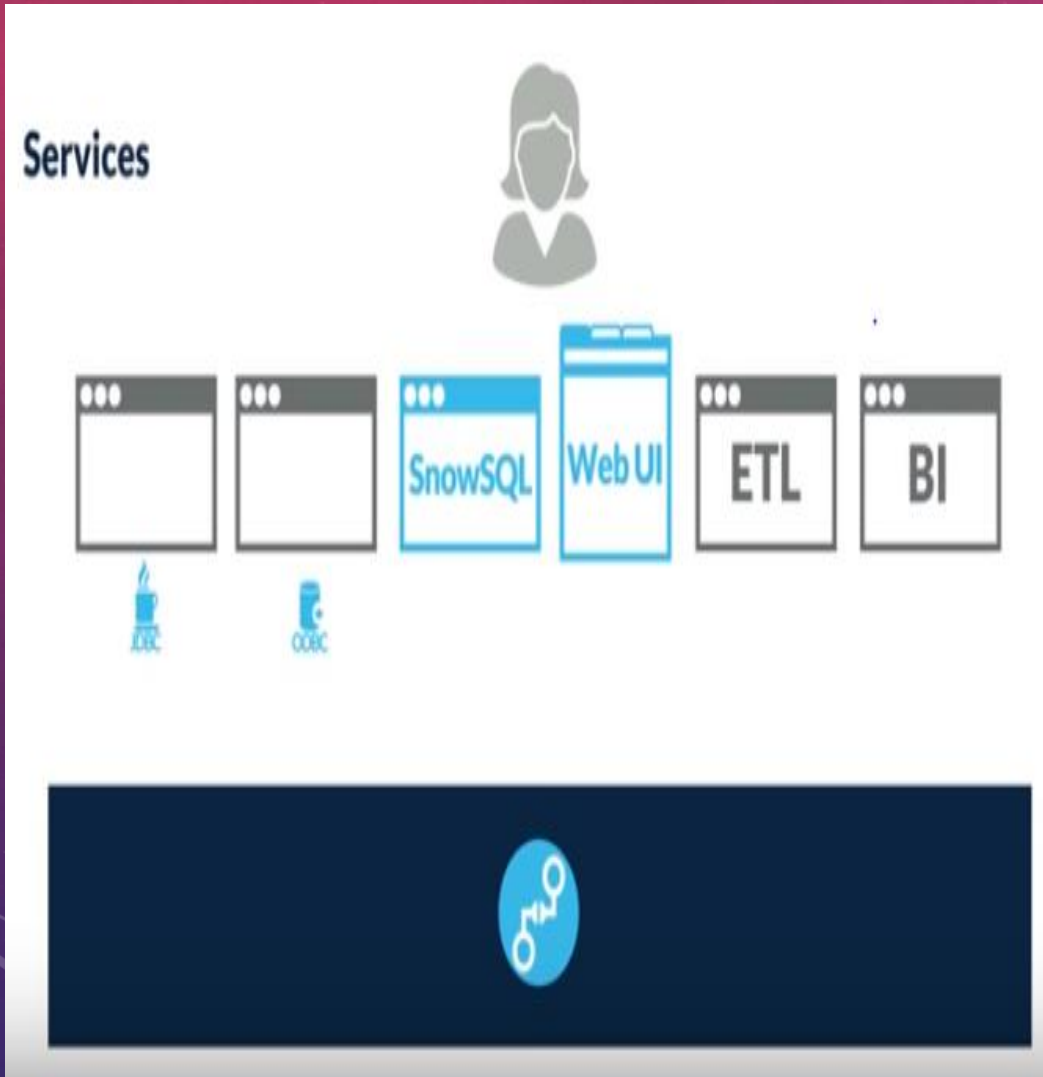


Shared nothing architectures



- 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

Warehouses
Manage your warehouses from this page. To operate on a warehouse, click the icon next to its name.

[+ Create...](#) [🔗 Configure...](#) [⏸ Suspend...](#) [▶ Resume...](#)

Status	Warehouse Name	Size
Suspended	COMPUTE_WH	X-Large

Create Warehouse

Name *

Size ▼
[Learn more about virtual warehouse sizes here](#)

Maximum Clusters ▼
Multi-cluster warehouses improve the query throughput for high concurrency workloads.

Scaling Policy ▼
The policy used to automatically start up and shut down clusters.

Auto Suspend ▼
The maximum idle time before the warehouse will be automatically suspended.

☒ Auto Resume [?](#)

Comment

[Show SQL](#) [Cancel](#) [Finish](#)

WORKSHEET AND QUERIES

The screenshot displays the Snowflake web interface. At the top, the navigation bar includes icons for Databases, Shares, Warehouses, Worksheets (which is the active tab), and History. Below the navigation bar, the interface is divided into several sections. On the left, there is a sidebar with the text 'Find database objects' and a search input field containing 'Starting with...'. Below this, a list of database objects is shown: DEMO_DB, SNOWFLAKE_SAMPLE_DATA, TEST1, and UTIL_DB. The main area of the interface is titled 'Worksheet 2' and contains a SQL query editor. The query is as follows:

```
1 create table tab1 (  
2     EmpID Int,  
3     Name varchar(20)  
4 )  
5
```

Below the query editor, there is a 'Run' button and a status bar indicating 'All Queries' and 'Saved a few seconds ago'. The 'Results' section is visible, showing a 'Data Preview' of the query results. The results are displayed in a table with two columns: 'Row' and 'status'. The first row shows the status 'Table TAB1 successfully created.'.

Context:

Results Data Preview

✓ Query ID SQL 257ms 1 rows

Filter result...

Row	status
1	Table TAB1 successfully created.

DATA LOADING OPTIONS

Limited Data



Bulk Data





sql





- 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


snowflake


Databases


Shares

Warehouses

Worksheets

History

Partner Connect

Help

**Fivetran**
Fivetran
Built for analysts, 5-minute setup, great schemas, Snowflake platinum partner.

**alooma**
Alooma
Connect all of your data with Alooma, the enterprise data pipeline built for the cloud

**talend**
Stitch
Stitch
Stitch moves data into Snowflake in minutes. Unlimited sources and a free-forever tier.

**SIGMA**
Sigma
A spreadsheet UI for Snowflake. Easily explore and analyze all your data.

**Periscope Data**
Periscope
Periscope Data brings data science and advanced analytics to the world of BI.

**snapLogic**
SnapLogic
SnapLogic's platform empowers organizations with intelligent application and data integration.

**R?VERY**
Rivery
Rivery creates an automated data pipeline to collect & transform data from all sources.

**CHARTIO**
Chartio
Connect to the #1 self-service data analytics platform for ease-of-use & speed to insights.

**MATILLION**
Matillion
Snowflake data transformation. Achieve new levels of simplicity, speed, scale and savings.

**DATAGUISE**
Dataguiise
Data security and privacy automation for compliance with GDPR, CCPA, PCI and more.

WHEREscape®

WHERESCAPE® - HISTORY

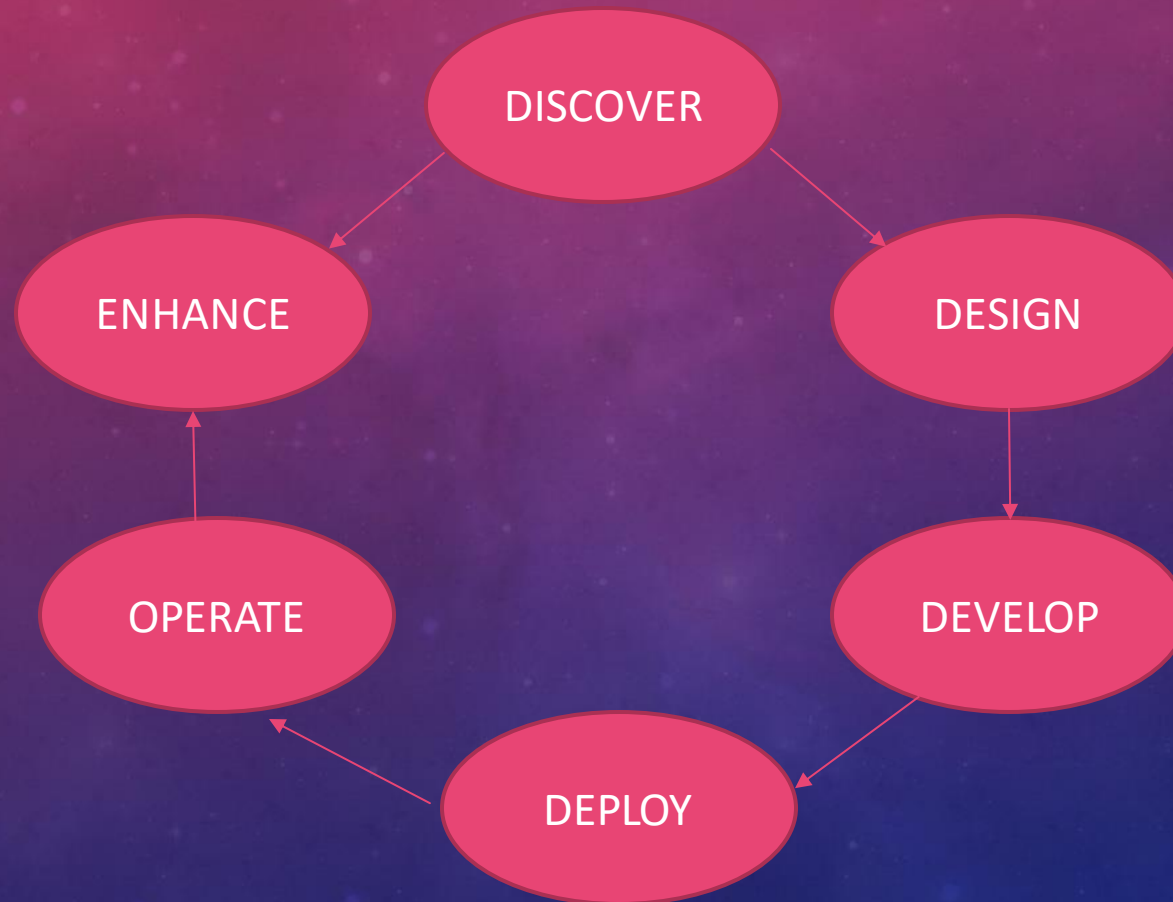
- Founded in 1997 by co-founders Michael Whitehead and Wayne Richmond and formerly known as Profit Management Systems
- Currently offers three products
 - WhereScape® 3D
 - WhereScape® RED
 - WhereScape® Data Vault Express

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).

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.

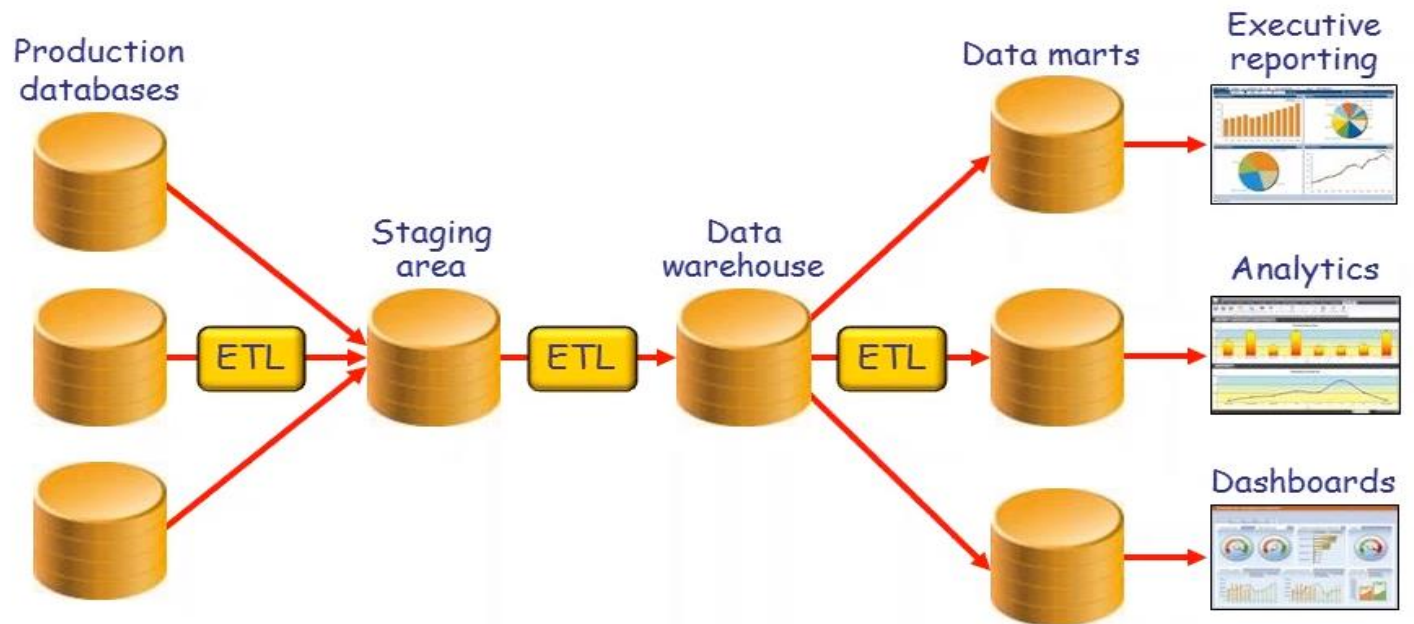
COST : DWA tools are not inexpensive. License can cost ~ \$50,000.

RELATIONAL FOCUSED : Still focused on relational data warehousing

WHERESCAPE FEATURES

Comparison with
Traditional
Data Warehouse
Architecture

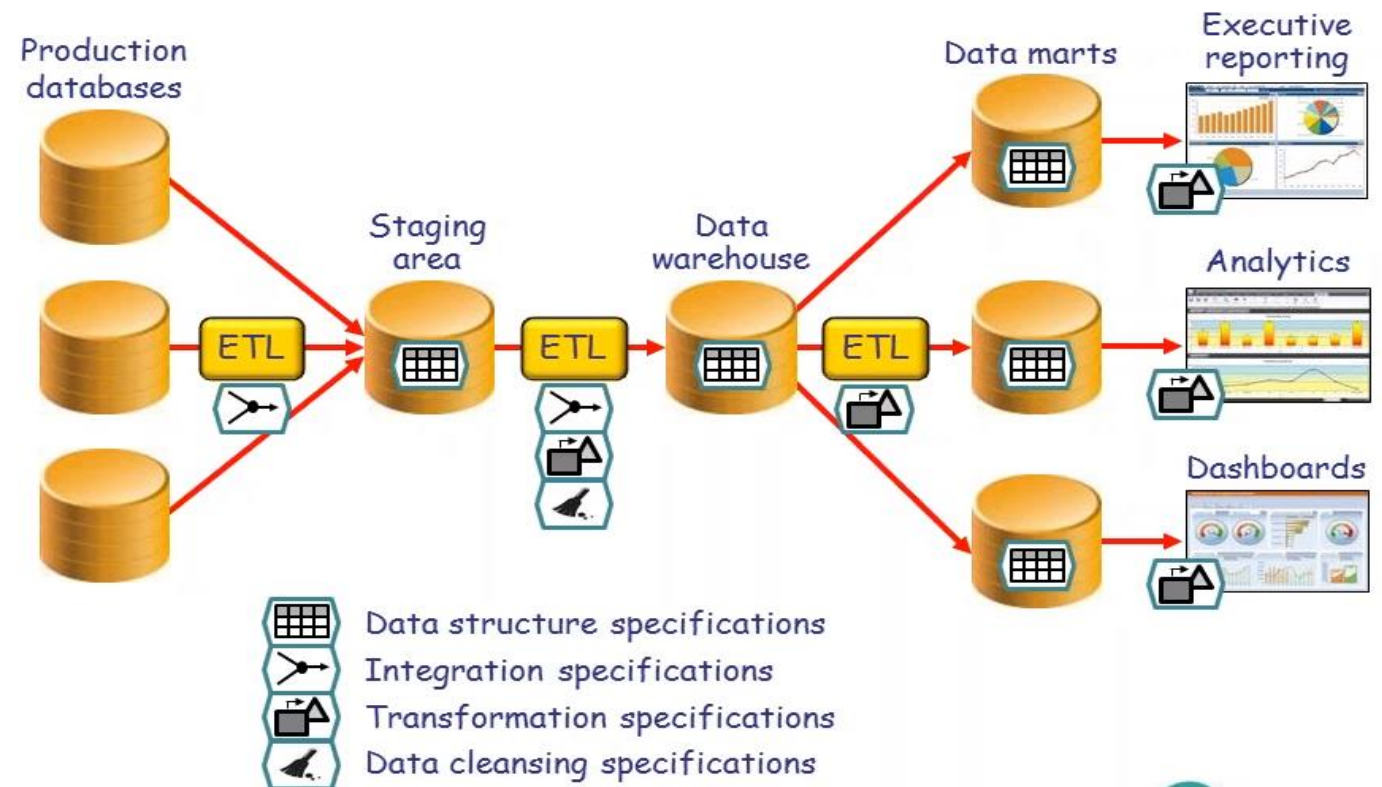
Classic Data Warehouse Architecture



WHERESCAPE FEATURES

Issues with Traditional
Datawarehouse
Architecture

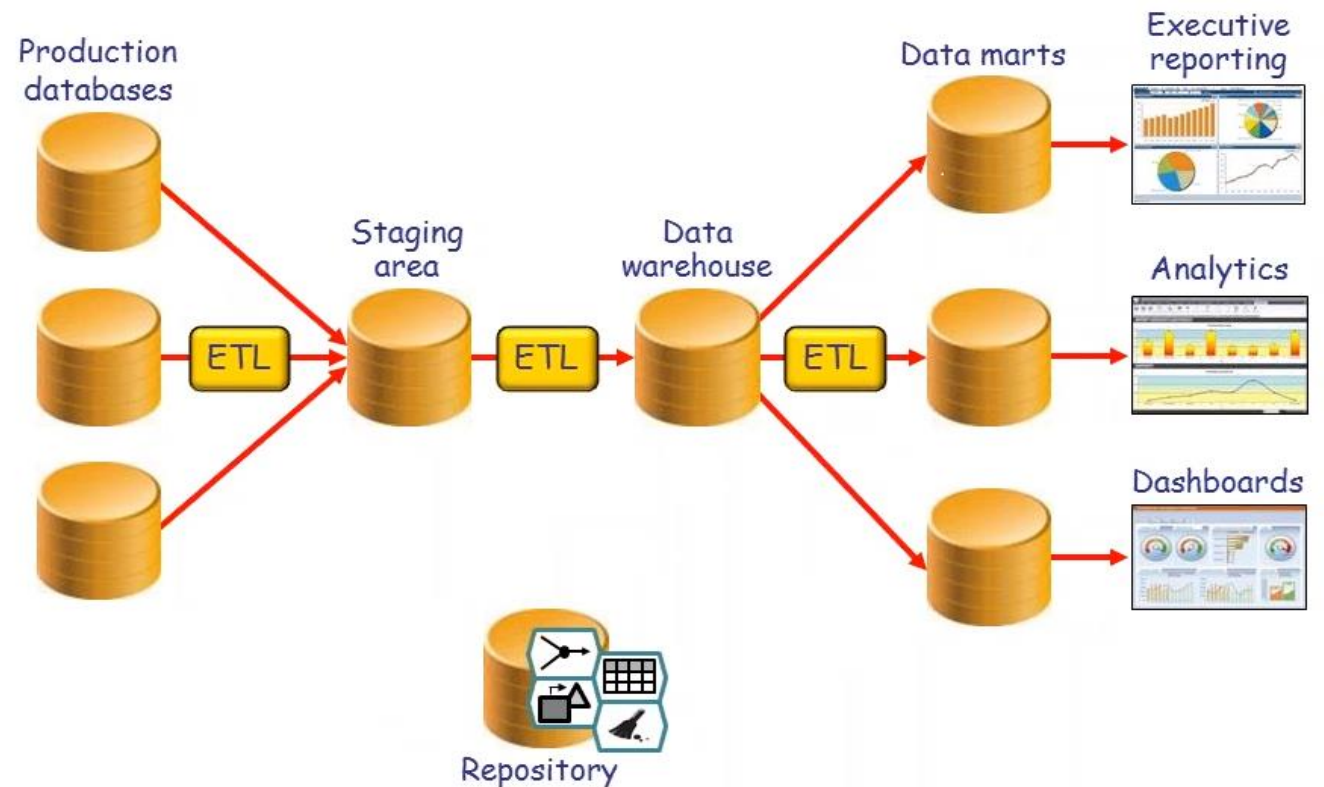
Specifications, Specifications, & Specifications



WHERESCAPE FEATURES

Wherescape's meta-data driven approach.

Centralized Storage of Specifications



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

The background is a dark blue field filled with numerous out-of-focus light circles in shades of blue, cyan, and purple, creating a bokeh effect. Overlaid on this are several thin, white or light blue circular lines and arcs, some of which form complete circles. Some of these circles have small arrows indicating a direction of movement. Faint, white numerical values are scattered across the background, particularly along the curved lines, including 40, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, and 260. A large, dark blue rectangular box with a thin white border is centered horizontally and vertically, containing the word "QUESTIONS?" in white, bold, sans-serif capital letters.

QUESTIONS?



THANK YOU!