Using the Cloud as Platform for Data Warehousing

Colin White
President, BI Research
TDWI and Snowflake Computing Webinar
July 2015

Sponsor

Speakers

Colin White
President,
BI Research

Jon Bock
VP of Products and Marketing,
Snowflake Computing

Webinar Overview

The cloud is an ideal environment for data warehousing given the data volumes and unpredictable workloads involved

However, organizations are anxious that the cloud may just become a way of bypassing IT

Data proliferation and inconsistent analytic results are also concerns if there is no cloud *gatekeeper*

One barrier to cloud adoption is the complexity of integrating the cloud and on-premises systems

The challenge then is choosing the right projects and cloud platform for deploying DW in the cloud

Webinar topics:

- Use cases for data warehousing (DW) in the cloud
- Key requirements for cloud-based DW
- Choosing a cloud DW solution
- Integrating a cloud-based DW with existing systems

1. Understand the potential technology and business advantages of the cloud for data warehousing
2. Identify projects with pain points and needs that cloud data warehousing can address
3. Assess current cloud data warehousing technologies and services
4. Identify the cloud offering that has the best fit with project requirements and existing skills and tools
5. Assess the cost and complexity of deploying and maintaining the selected cloud data warehousing solution
6. Understand how the cloud data warehousing solution will integrate with the existing IT environment
7. Look for opportunities to use cloud data warehousing to enhance the current data warehouse environment

1. The Advantages of the Cloud for DW

Pay-as-you go pricing model reduces upfront hardware and software costs

Elastic capacity supports changing DW storage & workload requirements

Provides faster time to value for the business

Aids DW modernization because it:

- Helps reduce costs
- Increases flexibility
- Speeds up deployment

2. Identify Projects That Cloud DW Can Address

Cloud DW has been especially successful in:

- Small and medium sized companies with limited IT resources
- In business units that may already run some of their operational business processes in the cloud

Start with business areas that recognize the value of cloud computing that have DW needs that are not being addressed by IT

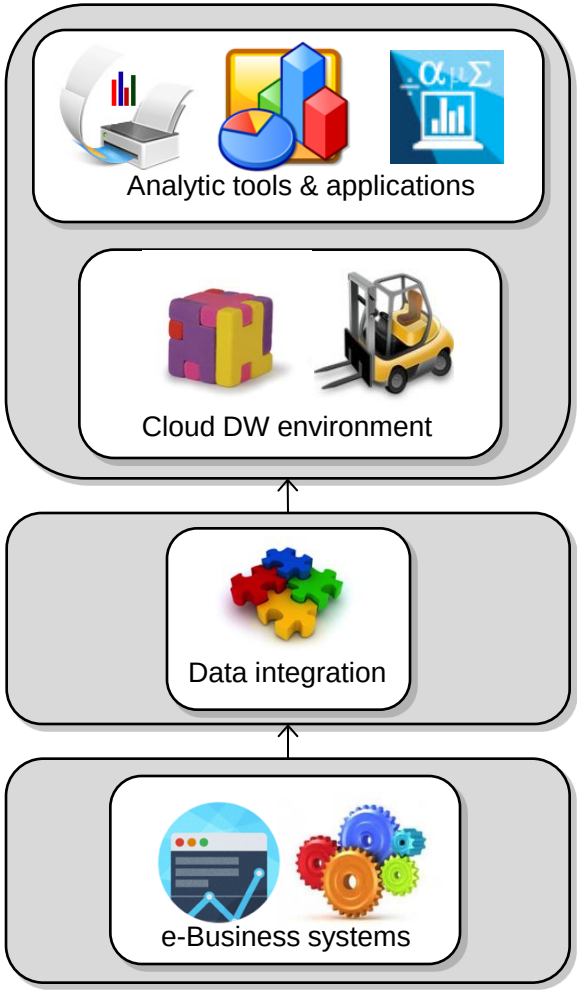
Examples of potential use cases include:

- Analysis and visualization of e-business data and processes
- Standalone analysis of Web, social media or sensor data
- Data warehouse augmentation, e.g., a data refinery and/or an investigative computing platform

Note also that the cloud can be used for prototyping, development and/or production

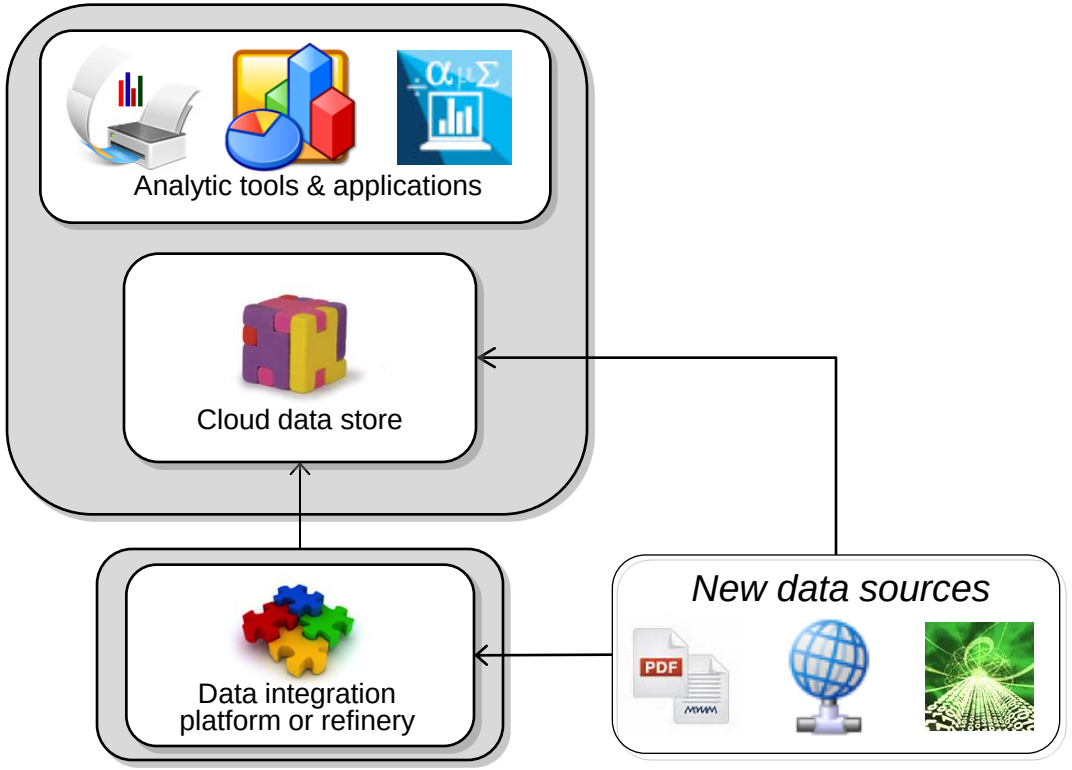
Data Warehouse for e-Business Data

Cloud-based DW is ideally suited to collecting, managing, analyzing and visualizing data that is are often deployed on hundreds of web servers that handle requests from millions of users on a variety of devices



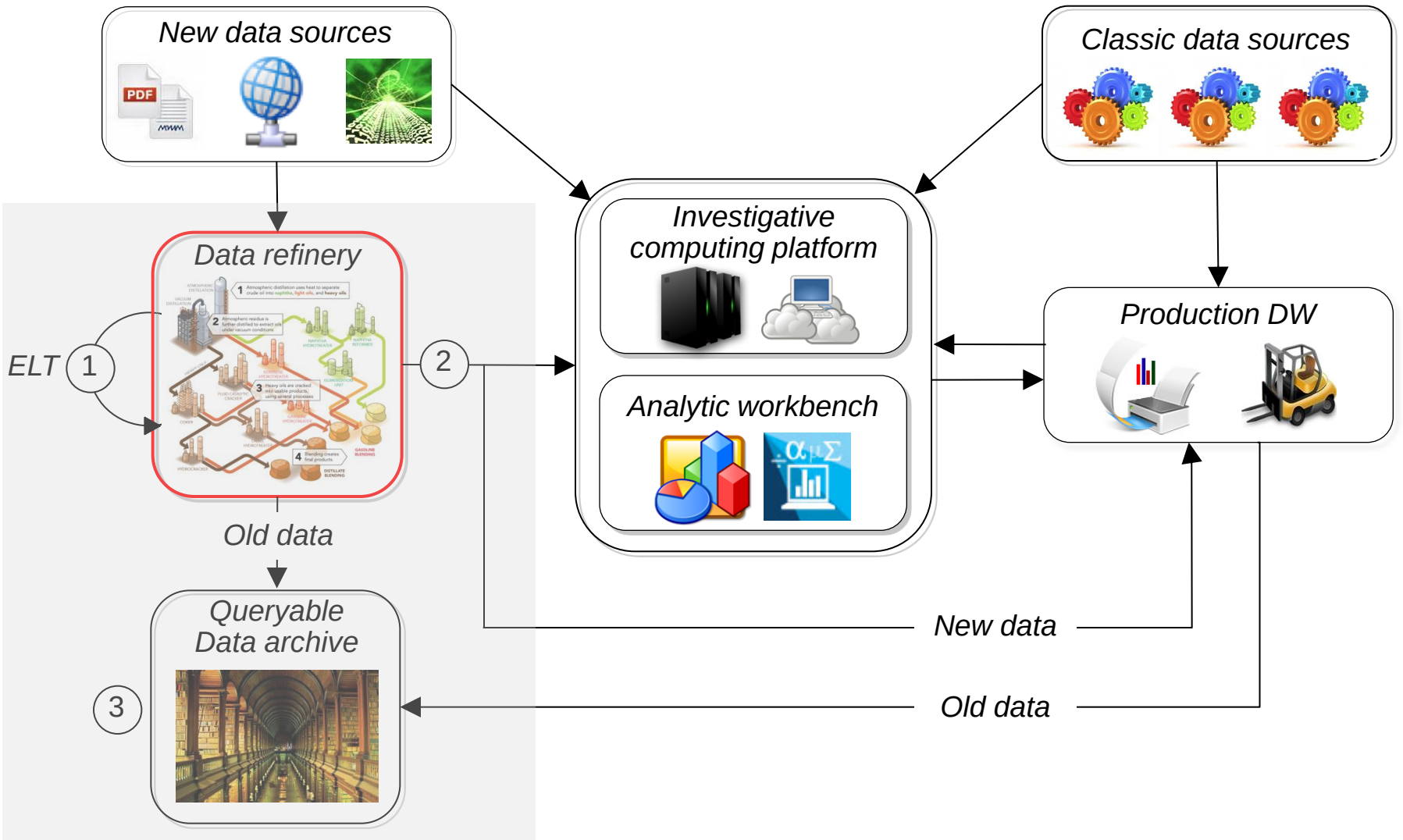
Standalone Analytic Application

Standalone processing of web, social media and/or sensor data



Data from each source is uploaded in batch or streamed directly to the cloud service for analysis

Data Refinery and Investigative Computing



3. Assess Cloud DW Technologies and Services

Organizations can optimize the value of their data by developing a comprehensive cloud-based data strategy for DW

Look for vendors who understand the unique requirements of data warehousing

Need to consider the complete data supply chain — from data acquisition to data delivery

- Capturing data from trusted sources
- Managing data under policy and governance guidelines
- Performing data integration, analysis, transformation and visualization

Vendor-provided *managed services* for data warehousing are also an important requirement

The Role of Managed Cloud Services for DW

Aid the implementation, administration and support of an end-to-end solution that covers the complete data lifecycle

May consist of free/fee-based consulting services and/or DW tools

These services are often technology or platform related and are not always suited to DW projects

It's important to understand exactly what a vendor means by "managed services"

- **Monitoring** – preserving the security and availability of your cloud infrastructure
- **Break-fix** – incident lifecycle management and root cause analysis for continuous improvement
- **Maintenance** – ensuring your cloud infrastructure, operating systems, middleware and applications are kept up-to-date
- **Reporting** – periodic reporting of cloud infrastructure performance against contractual Service Level Agreements
- **Optimization** – regular review of cloud service usage and capacity to aid infrastructure right-sizing
- **Provisioning** – spin-up or tear-down of infrastructure on behalf of the customer

4. Identify the Offering that Has the Best Fit - 1

Identify the best fit to hardware and software requirements

- Hardware processing and storage requirements will be dependent on the data volumes and analytics workloads associated with the project
- These requirements are often difficult to determine and this is where the elastic capacity of the cloud environment can provide benefits
- System software will depend on the DW tools used and it's important to choose a solution that supports those tools

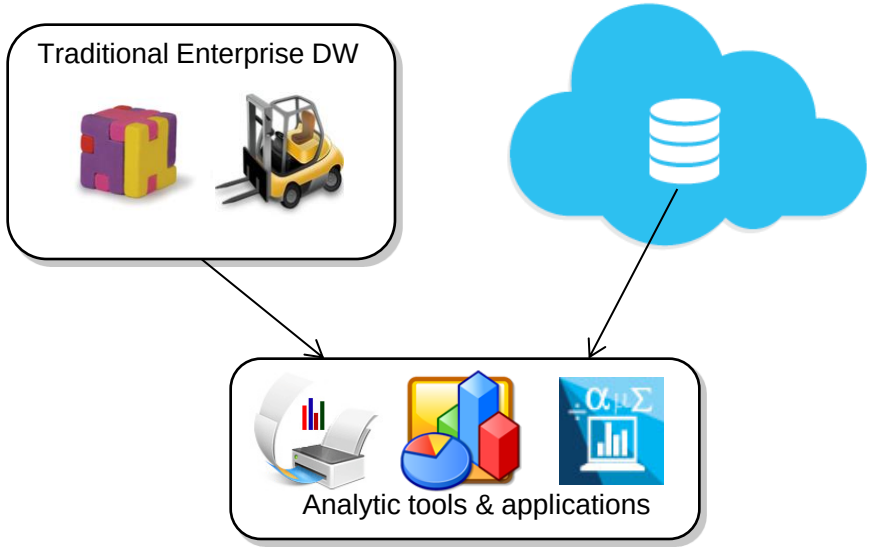
Identify the best fit to data management requirements

- Most data warehouses have been implemented with relational DBMSs
- Non-relational (e.g., Hadoop) and open source options can help improve performance and reduce software costs – but your mileage may vary

4. Identify the Offering that Has the Best Fit - 2

Identify the best fit to analytic tools and applications requirements

- The technology used in cloud-based data warehousing may require new analytic tools to be installed
- A key objective is to provide a seamless interface to data no matter where it resides



4. Identify the Offering that Has the Best Fit - 3

Identify the best fit to data integration requirements

- Modern data warehousing projects often involve the integration of both structured and/or multi-structured data
- Multi-structured data, such as web data, is more difficult to process and this makes the integration job even more difficult
- Data integration is a resource intensive DW task and the cloud solution must fit with your data integration strategy & software

Assess other differences between the cloud vendor's solution and the in-house IT environment

- Technologies used
- Data access, import and export capabilities
- Workload management, authorization and security
- Disaster recovery and help desk support

5. Assess the Cost & Complexity of Deployment - 1

Consider the complete application and data life cycle

- Total cost of ownership (TCO) is a key metric that helps assess the cost and complexity of using a given vendor's data warehousing solution
- TCO calculations must consider the complete application and data life cycle from initial design to final operation, administration and support

Importing, exporting and accessing data can be costly

- Consider the costs of importing, exporting and accessing data in a cloud service – will be dependent on where the data resides and its volume
- Potential data growth and the archiving of less active data are also important cost factors to consider

5. Assess the Cost & Complexity of Deployment - 2

Understand the managed services provided

- Many vendors provide services (consulting and/or tools) for helping companies implement projects in the cloud
- It's important to check that services support data-related tasks such as data design, acquisition, transformation, loading, exporting, archiving, etc.

Data warehouse specific services are a distinguishing factor

- Support for DW specific tasks is a key distinguishing feature between cloud offerings
- A managed cloud service for DW helps improve time-to-value when implementing a new project in the cloud
- The managed service model is especially attractive to companies with limited IT resources

6. Understand Cloud DW Integration - 1

An integrated solution for DW simplifies implementation

- Cloud users have the choice of integrating various cloud products and services themselves or using an integrated end-to-end solution
- An integrated end-to-end cloud solution for data warehousing offers similar benefits to an appliance approach

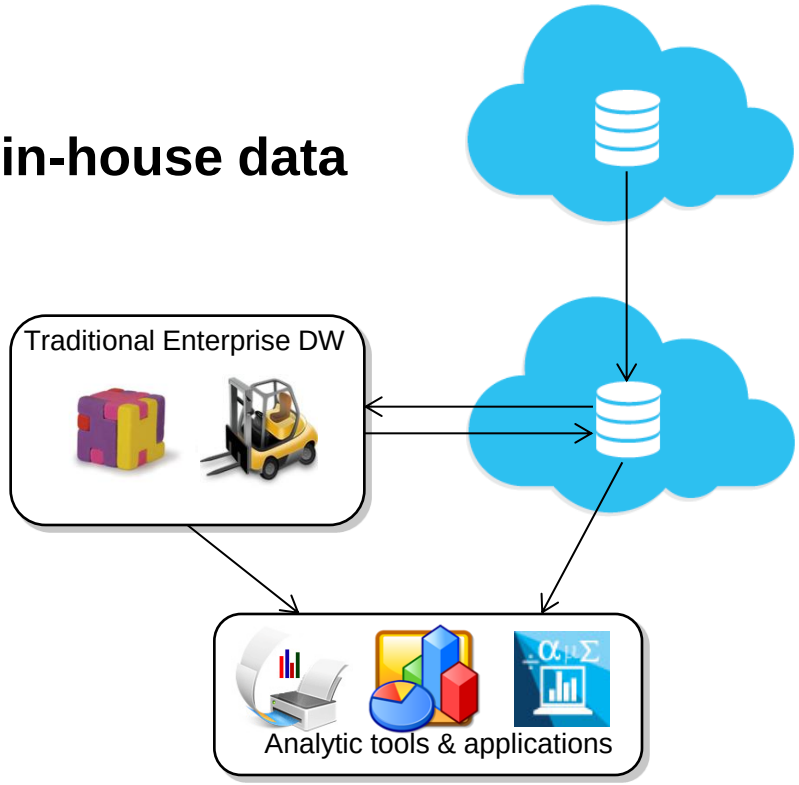
Data integration & movement can become a barrier to success

- Data integration and movement can become a barrier to successful cloud implementation
- When possible, the data should be processed where it resides
- When the source data already resides in the cloud it may still have to be moved to a different cloud system for processing

6. Understand Cloud DW Integration - 2

Users need to access both cloud and in-house data

- A DW project may involve a mixture of cloud data and in-house data
- In-house data may be accessed using data virtualization or copied from the in-house system to the cloud
- This is the same as in an in-house system where DW projects often use data from a variety of sources



- It is important to realize however that data movement in cloud environment occurs across a public Internet connection

7. Look for Other Opportunities - 1

Overcome in-house performance and cost issues

- Cloud computing can be used to reduce the costs and/or improve the performance of existing DW operations
- The use of cloud DW can even make possible what cannot be achieved using an in-house system
- Increasing data volumes and software costs and the loss of skills can affect the ability to maintain existing DW service levels – outsourcing resource constrained projects to the cloud can alleviate cost and service issues

7. Look for Other Opportunities - 2

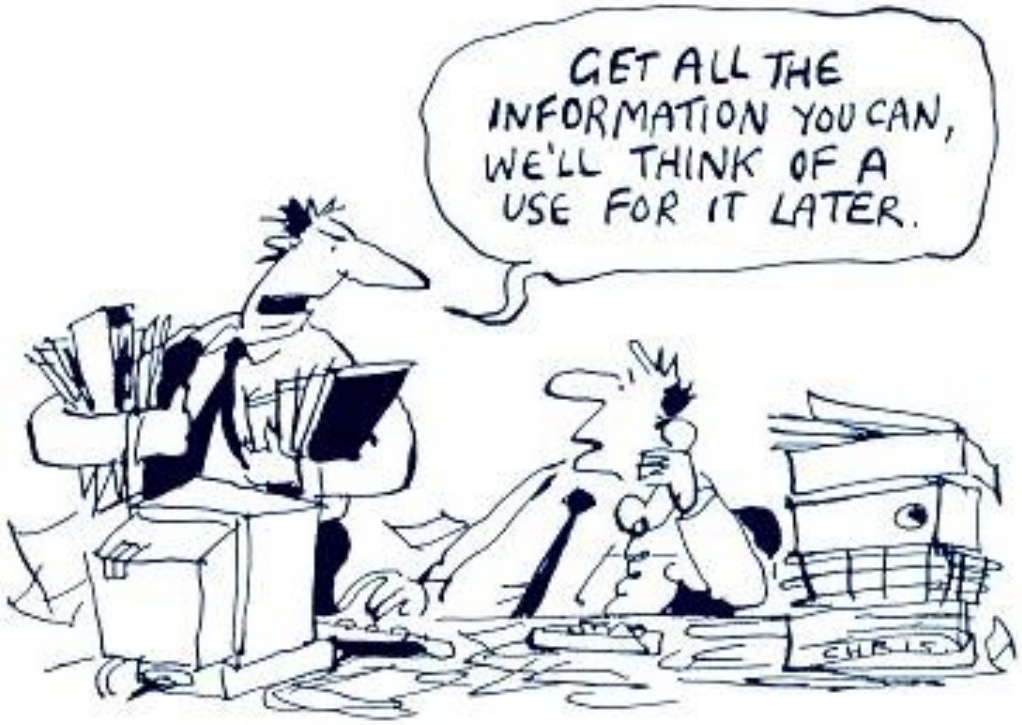
Enhance the existing data warehouse

- Data warehouse modernization and cloud computing can also be used to enhance the current data warehouse environment
- New types of data can be used in existing applications to broaden and deepen an organization's understanding of the business
- New analytic processes enable business analysts and managers to move beyond basic reporting and descriptive (i.e., diagnostic) analytics

Allows the organization to focus on the data rather than the technology

- Cloud computing eliminates the obstacles and pains of managing infrastructure, enabling the organization to focus on using its data

But … Data is the WHAT, Cloud is the HOW and Insight is the WHY!



snowflake

Reinventing the Data Warehouse for the Cloud

Jon Bock
VP of Products & Marketing
jon.bock@snowflake.net

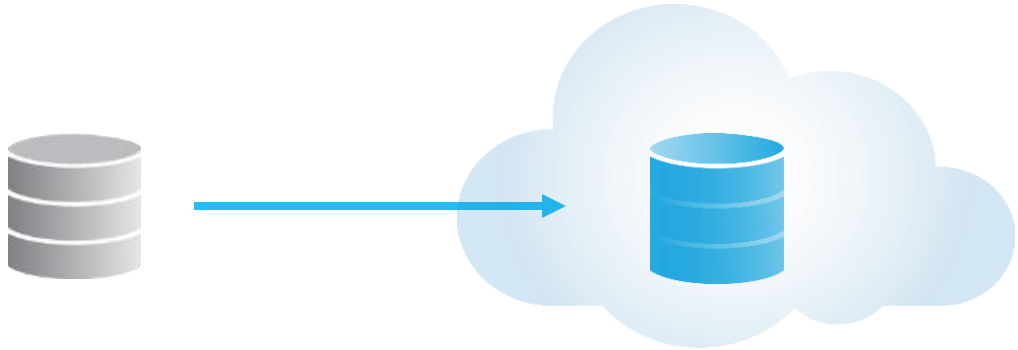
About Snowflake

A team of data experts

snowflake

About Snowflake

Who reinvented the data warehouse for the cloud



The Snowflake Elastic Data Warehouse™

Enterprise-class database
Completely new architecture and SQL engine

The Snowflake Elastic Data Warehouse™

Designed as a service
Self-service for data users--security, availability, optimization built in

Enterprise-class database
Completely new architecture and SQL engine

snowflake

The Snowflake Elastic Data Warehouse™

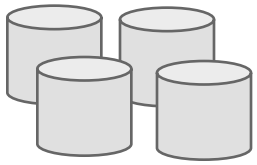
Supporting diverse data & analytics
Native support & optimization for diverse data and data usage

Designed as a service
Self-service for data users--security, availability, optimization built in

Enterprise-class database
Completely new architecture and SQL engine

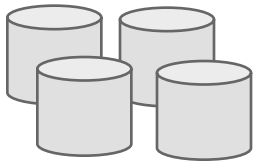
snowflake

Principles of on-premises data warehousing

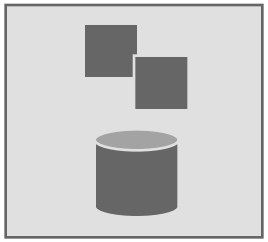


Scalable, enterprise-class external storage is expensive

Principles of on-premises data warehousing

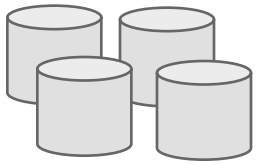


Scalable, enterprise-class external storage is expensive

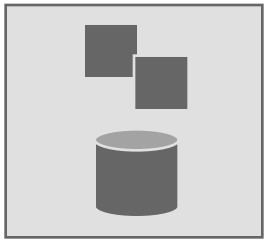


Resources are limited and hard to change

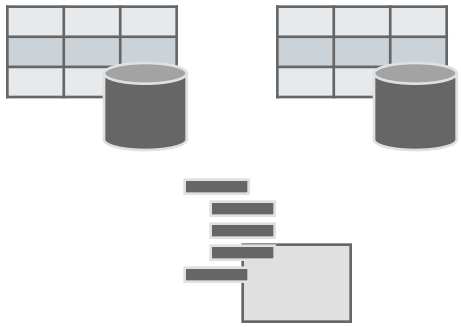
Principles of on-premises data warehousing



Scalable, enterprise-class external storage is expensive

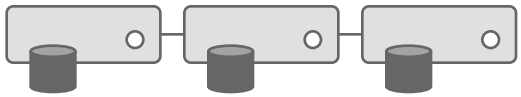


Resources are limited and hard to change



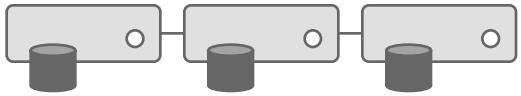
Different data and uses require different silos

Conventional on-premises architecture

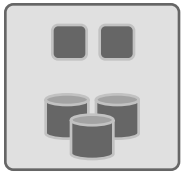


Bring data to compute

Conventional on-premises architecture

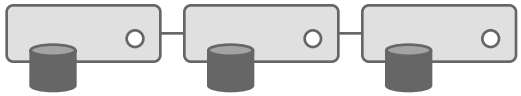


Bring data to compute

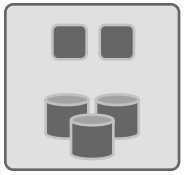


Carefully plan for and balance capacity

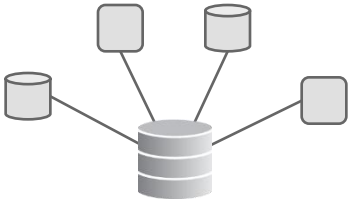
Conventional on-premises architecture



Bring data to compute

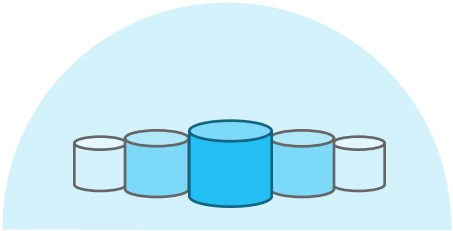


Carefully plan for and balance capacity



Different systems and marts for different uses

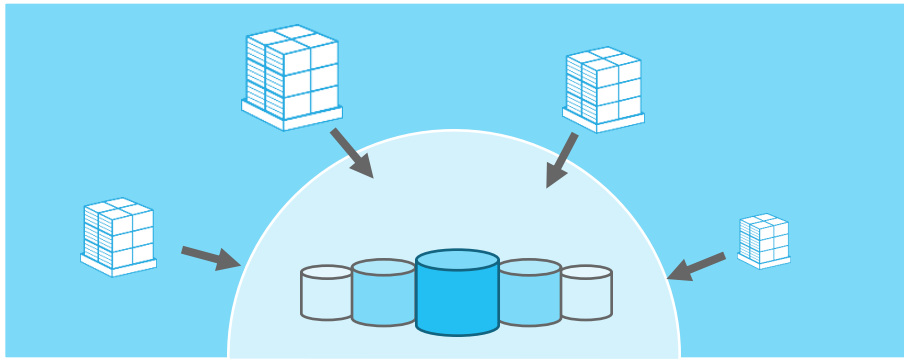
Snowflake's new architecture: multi-cluster, shared-data



Database storage

- **Data centralized in enterprise-class cloud storage**

Snowflake's new architecture: multi-cluster, shared-data

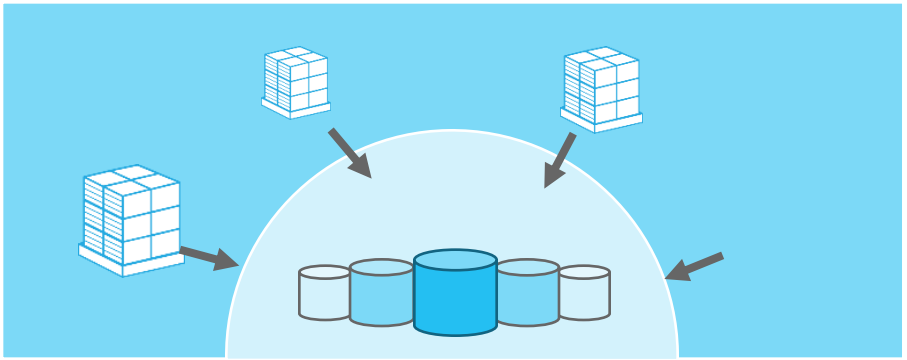


Processing

Database storage

- Data centralized in enterprise-class cloud storage
- **Independent processing clusters access centralized data**

Snowflake's new architecture: multi-cluster, shared-data

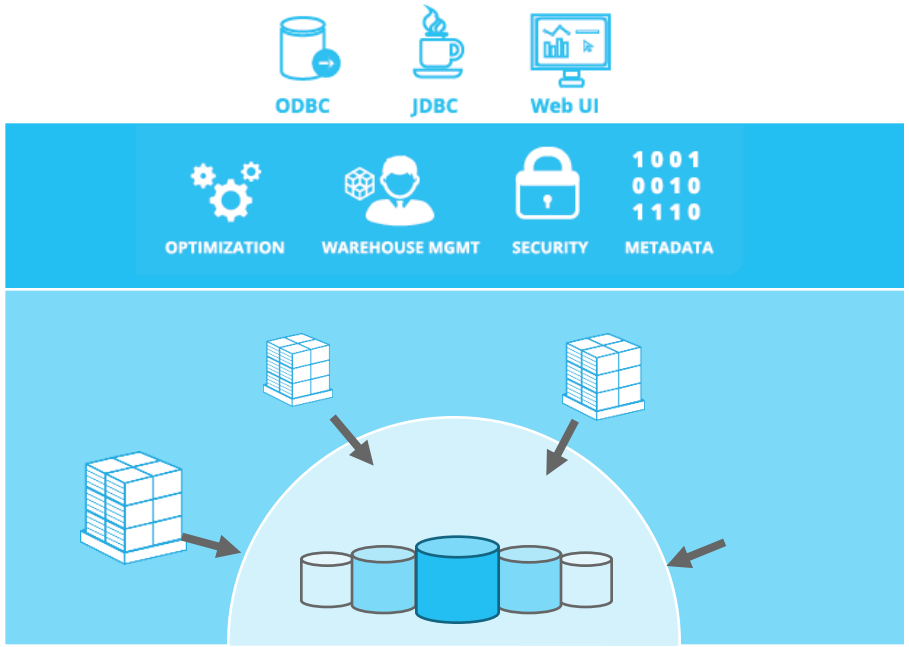


Processing

Database storage

- Data centralized in enterprise-class cloud storage
- Independent compute clusters access centralized data
  - **Scale up and down on-demand**

Snowflake's new architecture: multi-cluster, shared-data



Interfaces

Cloud services

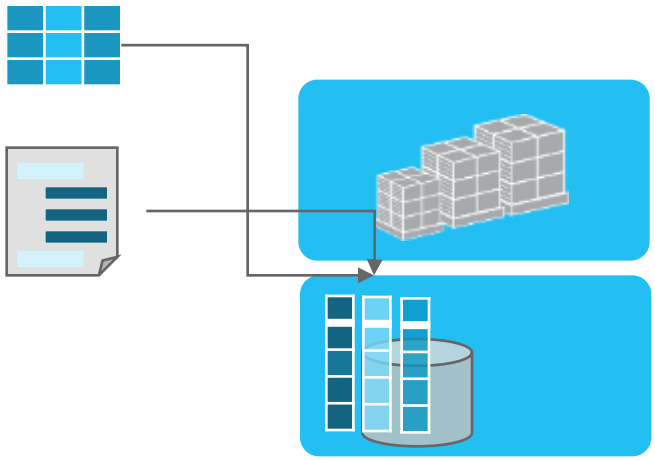
Processing

Database storage

- Data centralized in enterprise-class cloud storage
- Independent compute clusters access centralized data
  - Scale up and down on-demand
- **Single service accessible by any SQL tool**

snowflake

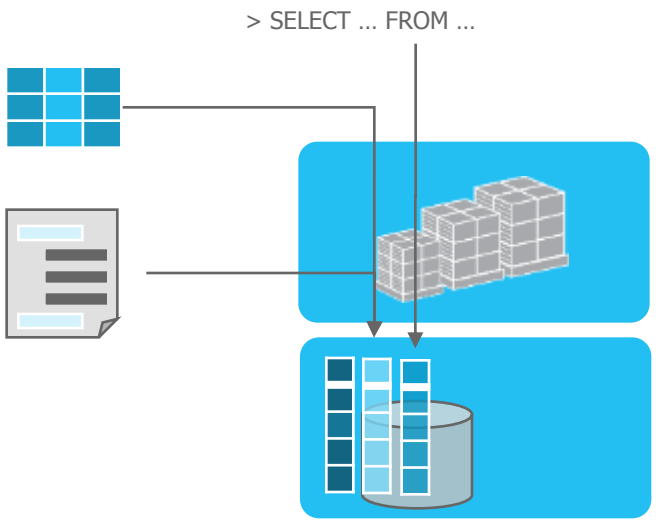
Diverse data in one system



Storage optimization

- Transparently optimize storage of structured and semi-structured data

Diverse data in one system



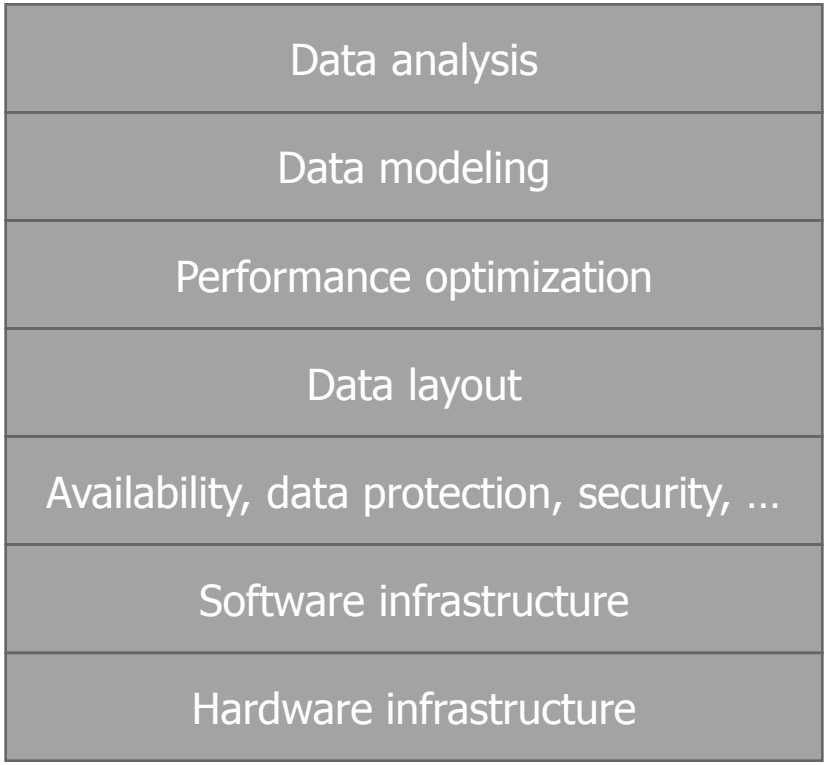
Storage optimization

- Transparent discovery and storage optimization of repeated elements

Query optimization

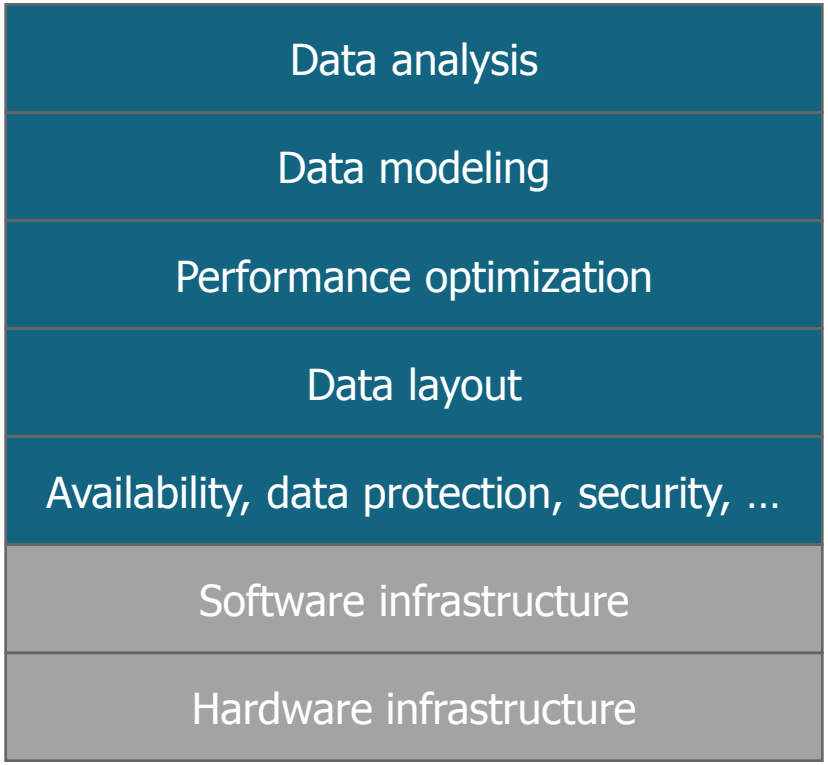
- Full database optimization for queries on all data

Delivered as a software service



Customer

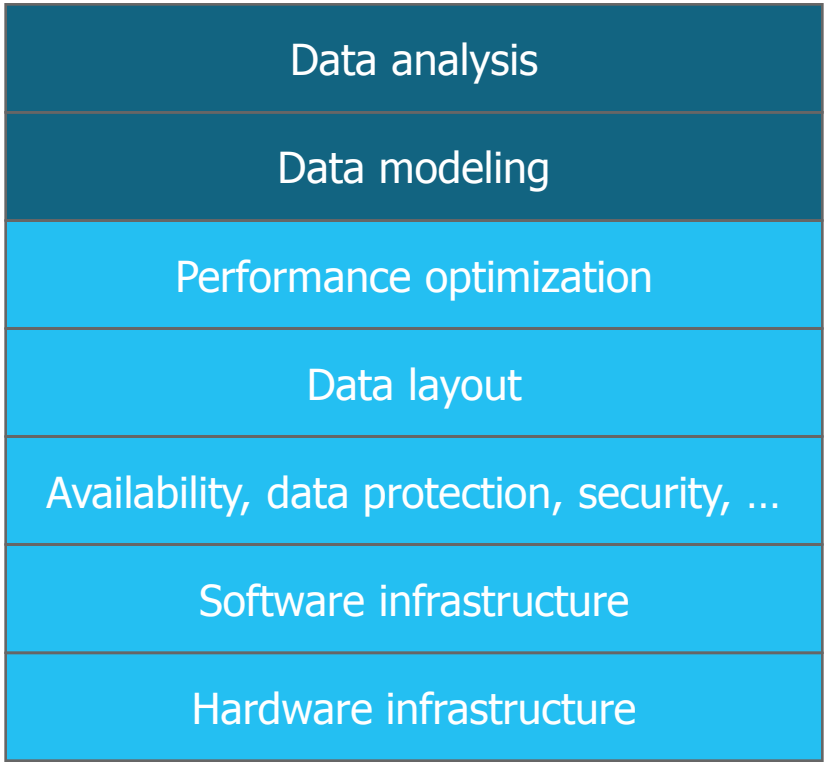
Delivered as a software service



Customer

Cloud vendor

Delivered as a software service



Customer

snowflake

Eliminate knobs and tuning

Data distribution

Data dynamically distributed and redistributed based on query patterns

Eliminate knobs and tuning

Data distribution

Data dynamically distributed and redistributed based on query patterns

Metadata management

Statistics automatically created and updated by Snowflake

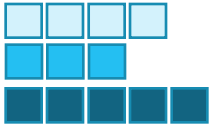
Eliminate knobs and tuning

Data distribution

Data dynamically distributed and redistributed based on query patterns

Metadata management

Statistics automatically created and updated by Snowflake



Concurrency

Parallelism and query dispatch automatically determined based on matching resource needs with resource availability

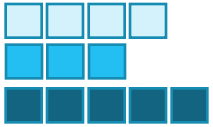
Eliminate knobs and tuning

Data distribution

Data dynamically distributed and redistributed based on query patterns

Metadata management

Statistics automatically created and updated by Snowflake



Concurrency

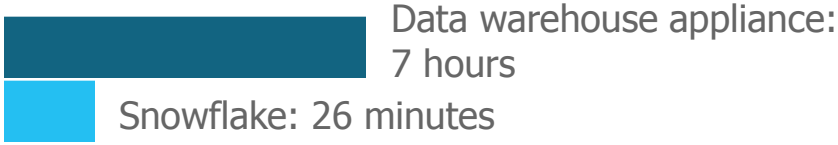
Parallelism and query dispatch automatically determined based on matching resource needs with resource availability

Availability and resiliency

Distributed and resilient across components and datacenters

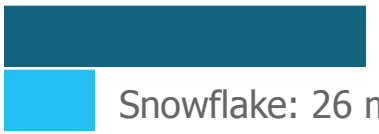
Customer results

Market research company
Replace on-premises data warehouse with Snowflake for analytics workload



Customer results

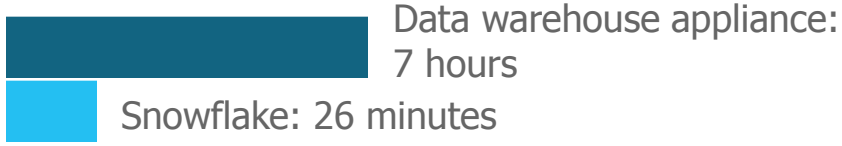
Market research company
Replace on-premises data warehouse with Snowflake for analytics workload



Online retailer
Eliminate need to grow staff to support additional analytics projects

Customer results

Market research company
Replace on-premises data warehouse with Snowflake for analytics workload



Online retailer
Eliminate need to grow staff to support additional analytics projects

Gaming company
Replace noSQL data store with Snowflake for storing & transforming event data

Questions?

Contact Information

If you have further questions or comments:

Colin White, BI Research

info@bi-research.com

Jon Bock, Snowflake Computing

jonb@snowflakecomputing.com