

DEMYSTIFYING DATA WAREHOUSE AS A SERVICE



My Bio

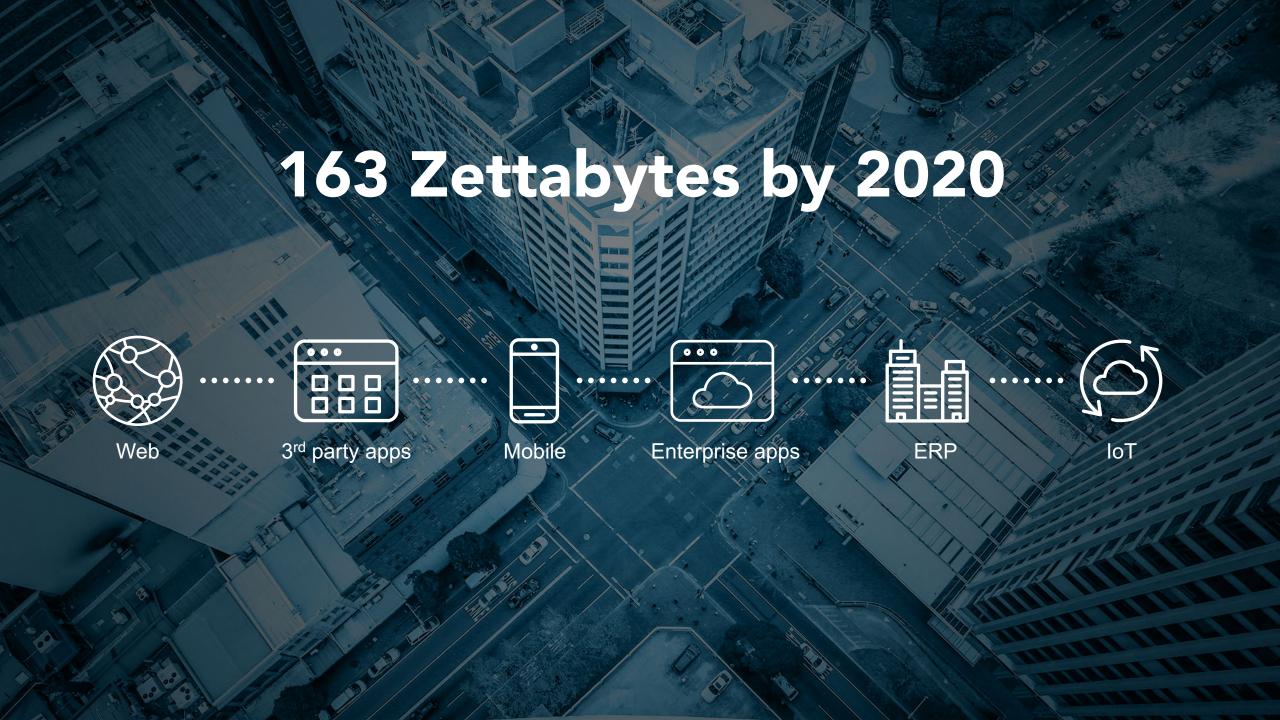
- Chief Technical Evangelist, Snowflake Computing
- Oracle ACE Director, Alumni (DW/BI)
- OakTable Network
- Blogger The Data Warrior
- Certified Data Vault Master and DV 2.0 Practitioner
- Former Member: Boulder BI Brain Trust (#BBBT)
- Member: DAMA Houston & DAMA International
- Data Architecture and Data Warehouse Specialist
 - 30+ years in IT
 - 25+ years of Oracle-related work
 - 20+ years of data warehousing experience
- Author & Co-Author of a bunch of books (Amazon)
- Past-President of ODTUG and Rocky Mountain Oracle User Group



AGENDA

- Data Challenges
- > What is a Data Warehouse as a Service?
- **►** Introducing Snowflake
- > Top 10 (or so) Cool Features of Snowflake
 - > Continuous Loading
 - > Data Sharehouse
- **➤ Agile Data Lifecycle**
- > Reference Architectures
- Comparing Options
- > Snowflake in Action

Data Challenges Today



It's not the data itself



it's how you take full advantage of the insight it provides

Most firms don't consistently turn data into action

73%

of firms aspire to be data-driven.





29%

of firms are good at turning data into action.

Source: Forrester

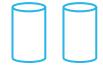


THE DATA STRUGGLE



Symptoms of fundamental challenges

Prevents Agility!











Data silos

Data locked into separate databases, big data systems, and applications

Cost

Painful upfront costs and overprovisioned capacity

Performance

Contention for limited resources resulting in latency and delays

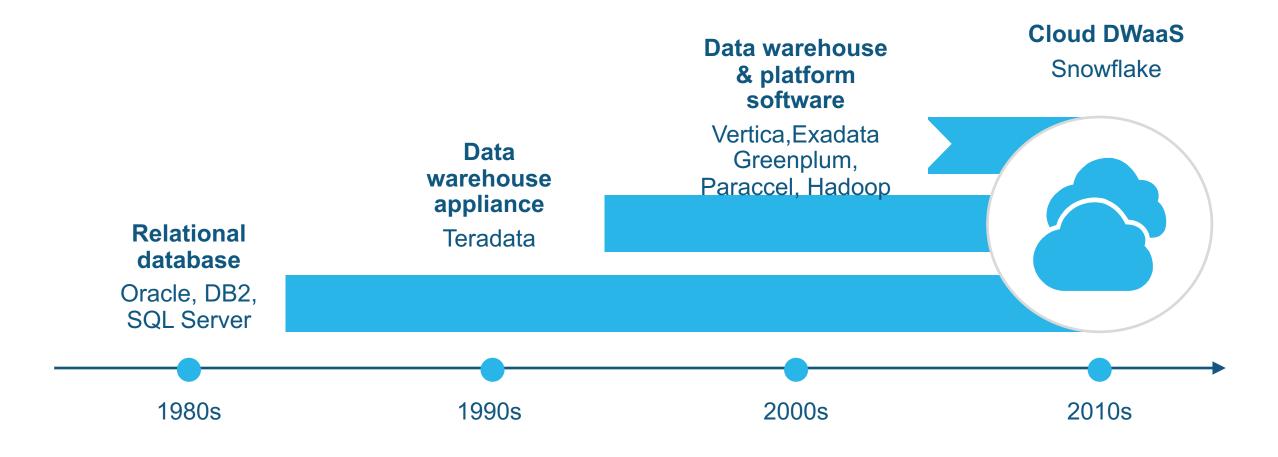
Inflexibility

Slow, cumbersome scaling and limited support for diverse data

Complexity

Multiple systems to integrate and manage requiring specialized skills and tools

The evolution of data platforms



What is a Cloud DWaaS?



DW- Data Warehouse

- Relational database
- Uses standard SQL
- Optimized for fast loads and analytic queries

aaS – As a Service

- Like SaaS (e.g. SalesForce.com)
- No infrastructure set up
- Minimal to no administration
- Managed for you by the vendor
- Pay as you go, for what you use

Goals of a Cloud DWaaS



Make your life easier

So you can load and use your data faster

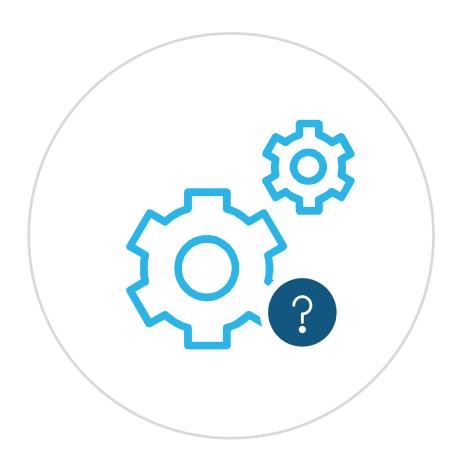
Support business

- Make data accessible to more people
- Reduce time to insights

Handle big data too!

Schema-less ingestion

What to Expect from a DWaaS



It should support standard SQL (natively)

- It should support standard ETL & BI tools
- ODBC or JDBC connectivity

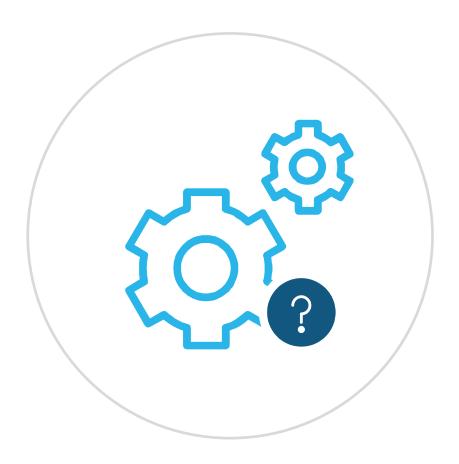
It should be infinitly scalable (cloud)

- Handle huge amounts of data
- Handle large number of concurrent queries without performance degradation

It should handle flexible schema data types

No sharding or ETL required

What to Expect from a DWaaS



It should be secure

· Built in encryption?

It shoud be stable

Resiliancy and availability should be easy to configure and manage

It should be easy to configure and manage

It should provide a lower TCO

Cloud scale pricing

Introducing Snowflake

3 years in stealth + 3 years GA

Founded 2012 by industry veterans with over 120 database patents



First customers 2014, general availability 2015





Over \$920M in venture funding from leading investors



1200+ employees Over 2000 customers today

Fun facts:

Queries processed in Snowflake per day:

100 million

Largest single table:

68 trillion rows

Largest number of tables single DB:

200,000

Single customer most data:

> 40PB

Single customer most users:

> 10,000



Snowflake: a team of data experts



Benoit Dageville CTO

Lead architect of Oracle parallel execution and a key manageability architect



Marcin Zukowski Founder & VP of **Engineering**

Inventor of vectorized query execution in databases



Thierry Cruanes Founder Architect

Leading expert in query optimization and parallel execution at Oracle

Team















Microsoft









Investors

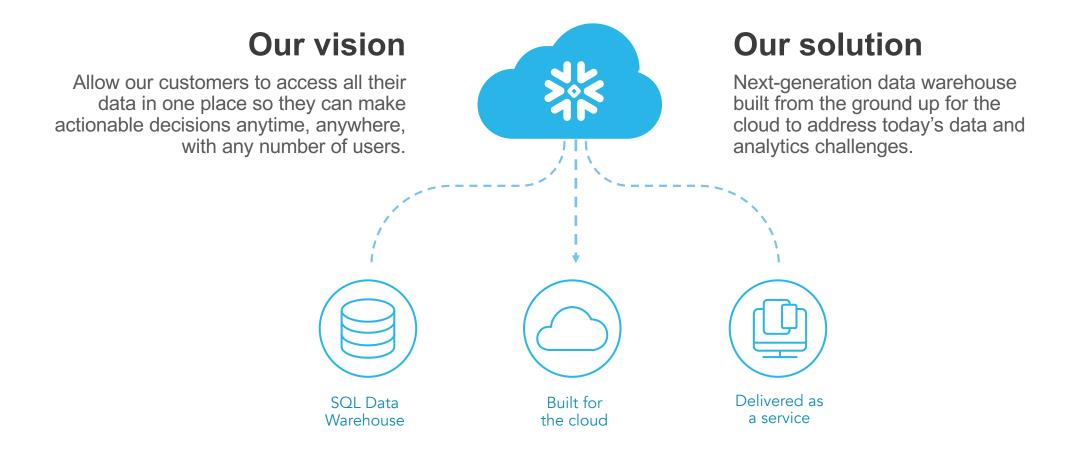






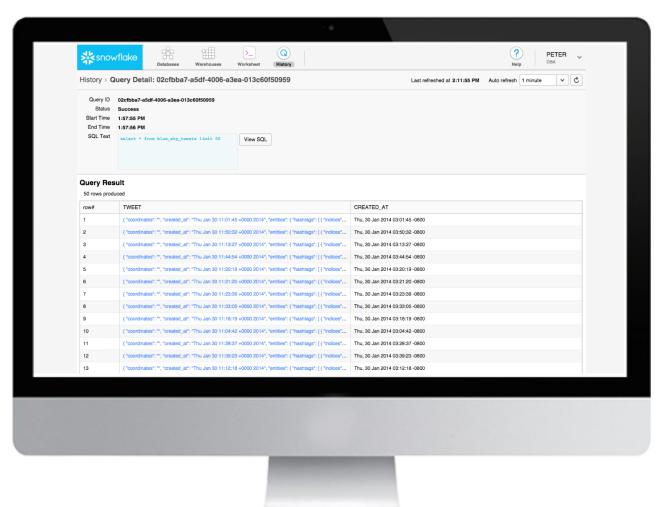


SNOWFLAKE: A FULL DATA WAREHOUSE, BUILT FOR THE CLOUD



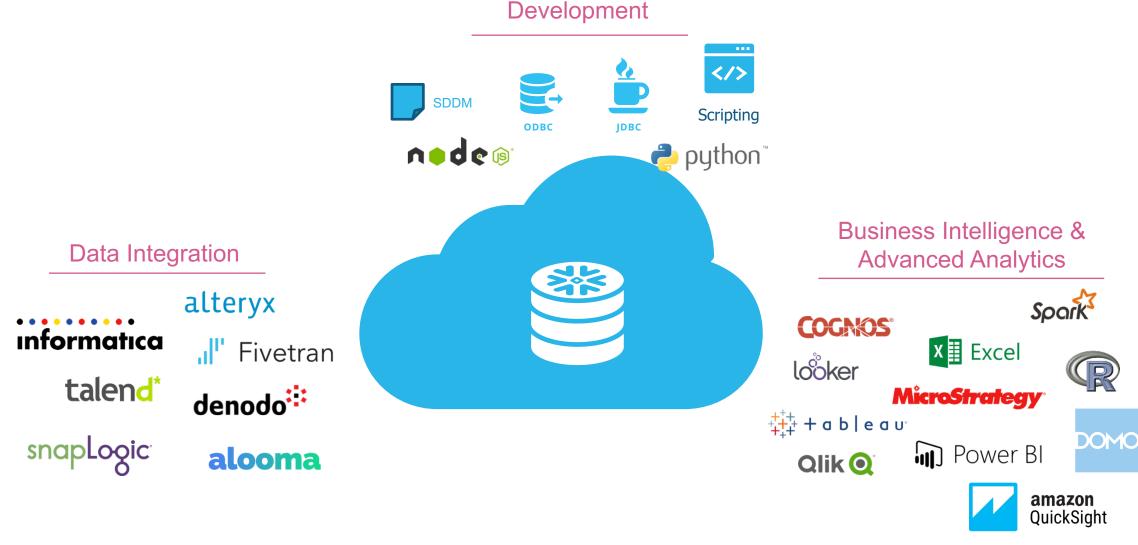
The Data Warrior's
Top 10+ Cool Things About
Snowflake
(A Data Geeks Guide to DWaaS)

#10 – Persistent Result Sets



- No setup
- In Query History
 - By Query ID
- 24 Hours
- No re-execution
- No Cost for Compute

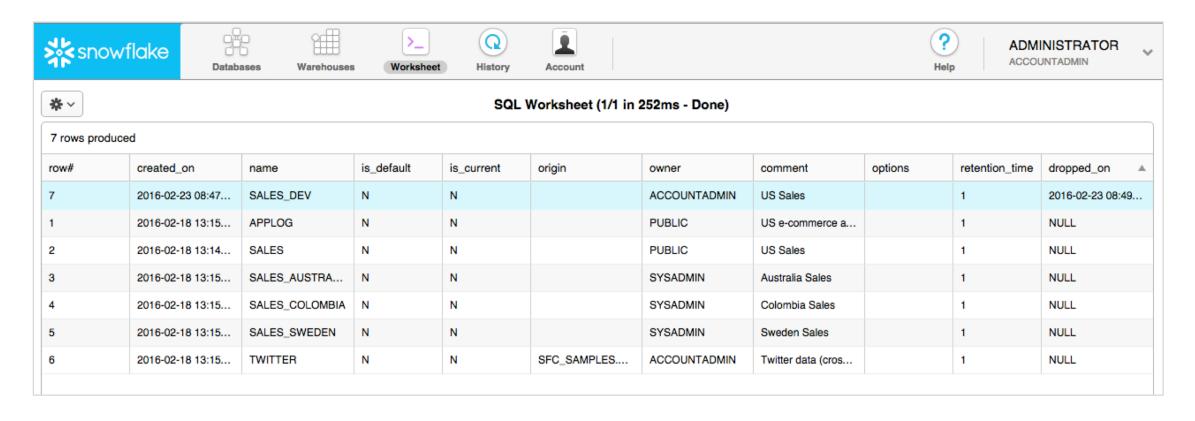
#9 Works with the tools and skills you already have.



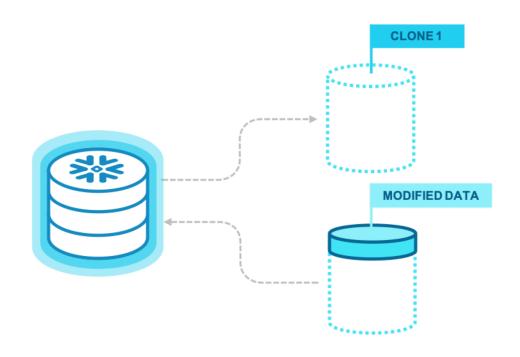
#8 - UNDROP

UNDROP TABLE UNDROP SCHEMA <schema name> UNDROP DATABASE <db name>

Part of Time Travel feature: AWESOME!



#7 - ZERO-COPY DATA CLONING



Instant data cloning operations

Databases, schema, tables, etc

Metadata-only operation

Modified data stored as new blocks

Unmodified data stored only once

No data copying required, no cost!

Instant test/dev environments

Test code on your entire production dataset

Swap tables into production when ready

Simply SQL to Clone – Fast!

Instant copy of table, schema, or database:

CREATE OR REPLACE TABLE MyTable_V2
CLONE MyTable;

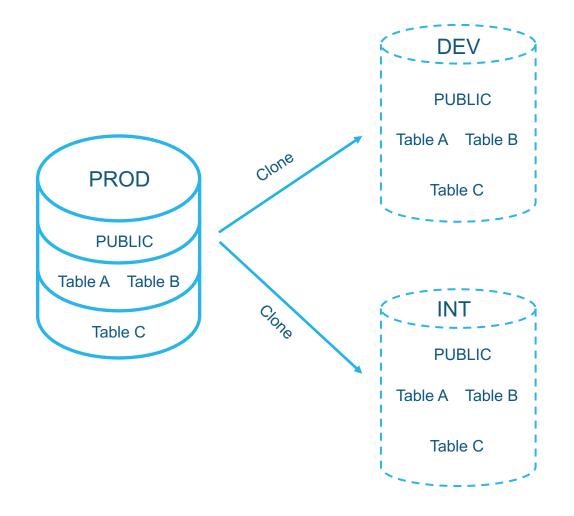
With Time Travel:

CREATE SCHEMA mytestschema_clone_restore

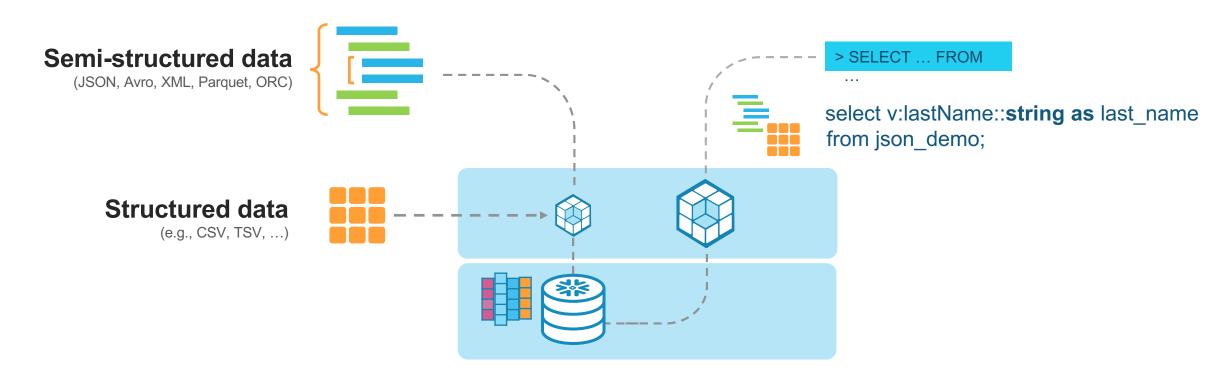
CLONE testschema

BEFORE (TIMESTAMP =>

TO_TIMESTAMP(40*365*86400));



#6 – JSON Support with SQL!



Storage optimization

Transparent discovery and storage optimization of repeated elements



Query optimization

Full database optimization for queries on semi-structured data



© 2019 Snowflake Inc. All Rights Reserved 26

#5 – Standard SQL w/Analytic Functions

Partner post: https://sonra.io/2018/02/04/create-custom-aggregate-udaf-window-functions-snowflake/

```
select Nation, Customer, Total
from (select
      n.n name Nation,
      c.c name Customer,
      sum(o.o totalprice) Total,
      rank() over (partition by n.n name
      order by sum(o.o totalprice) desc)
           customer rank
  from orders o,
  customer c.
  nation n
  where o.o custkey = c.c custkey
  and c.c nationkey = n.n nationkey
  group by 1, 2)
where customer rank <= 3
order by 1, customer rank
```



Complete SQL database

- Data definition language (DDLs)
- Query (SELECT)
- Updates, inserts and deletes (DML)
- Role based security
- Multi-statement transactions







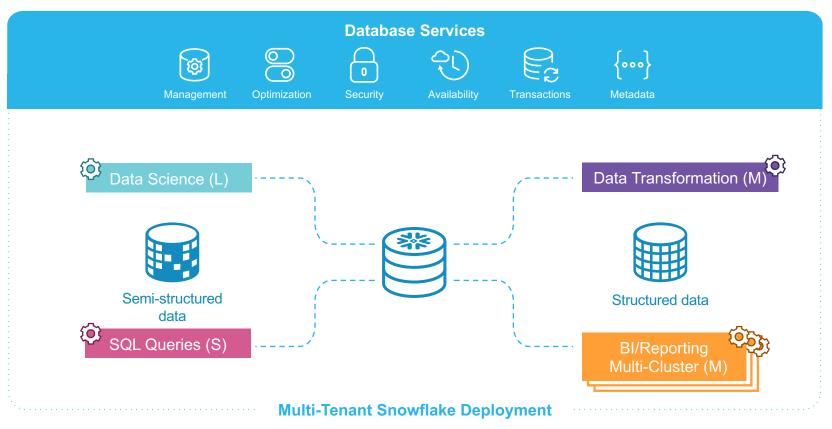








#4 – Separation of Storage & Compute New multi-cluster, shared data architecture

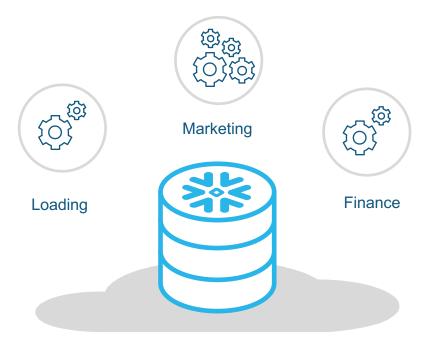


Flexible Cloud Storage For All Kinds of Data
Unlimited Compute Clusters to Serve Every Use
Easy-to-use Service with No Management

© 2019 Snowflake Inc. All Rights Reserved

#3 – Support Multiple Workloads

Deliver faster analytics at any scale



Accelerate the data pipeline

Run loading & analytics at any time, concurrently, to get data to users faster

Scale compute to support any workload

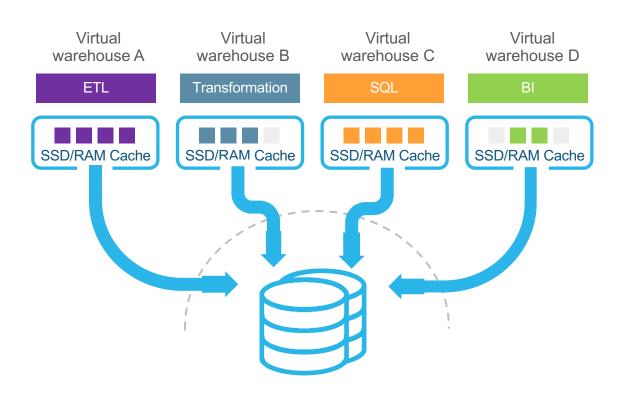
Scale processing horsepower up and down on-the-fly, with zero downtime or disruption

Scale concurrency without performance impact

Multi-cluster "virtual warehouse" architecture scales concurrent users & workloads without contention

DATA PROCESSING

How to allow concurrent workloads to run without impacting each other?



Virtual warehouses

A warehouse is one or more MPP compute cluster Use multiple warehouses to segregate workload

- e.g., ETL warehouse versus query warehouse Resizable on the fly
 - Adjust cluster size (up/down) based on data size and/or query complexity
 - Automatically add/remove clusters as level of concurrency varies

Able to access data in any database

Transparently caches data accessed by queries

Transaction manager synchronizes data access

Automatic suspend/resume when needed

Cloud-based architecture for big data analytics

Data Warehousing Cloud Service

Data storage is separate from compute

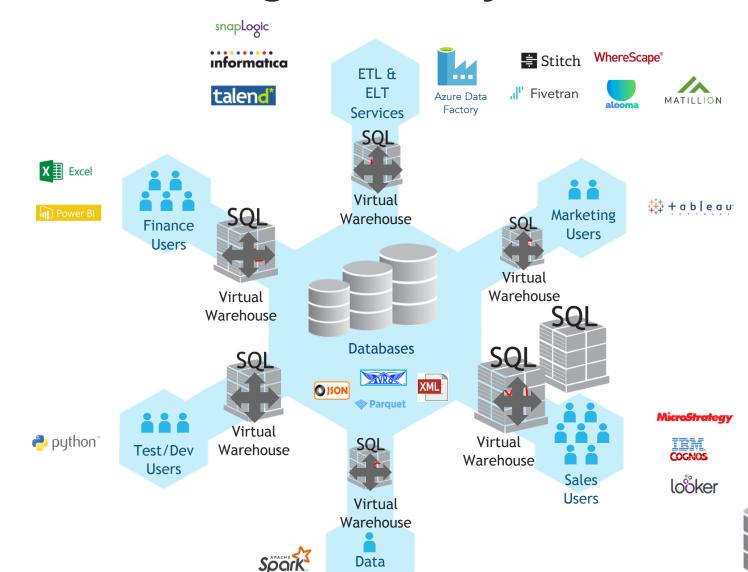
One Virtual Warehouse, multiple Databases

One Database, multiple Virtual Warehouses

Virtual Warehouses scale independently

Multi-Cluster warehouses support high concurrency

Workloads don't compete with each other



Scientist

Metadata

Instant, unlimited scalability

- Elastic scaling for storage
 Low-cost cloud storage, fully
 replicated and resilient
- Elastic scaling for compute
 Virtual warehouses scale up &
 down on the fly to support
 workload needs
- Elastic scaling for concurrency
 Automatically scale concurrency using multi-cluster virtual warehouses



Pay for what you actually use...down to the second



Data Scientist

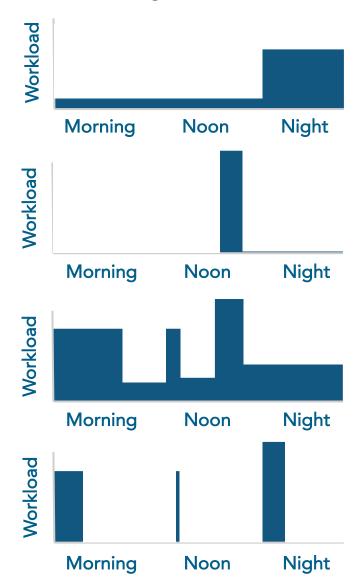


Reporting



Ad-hoc analytics





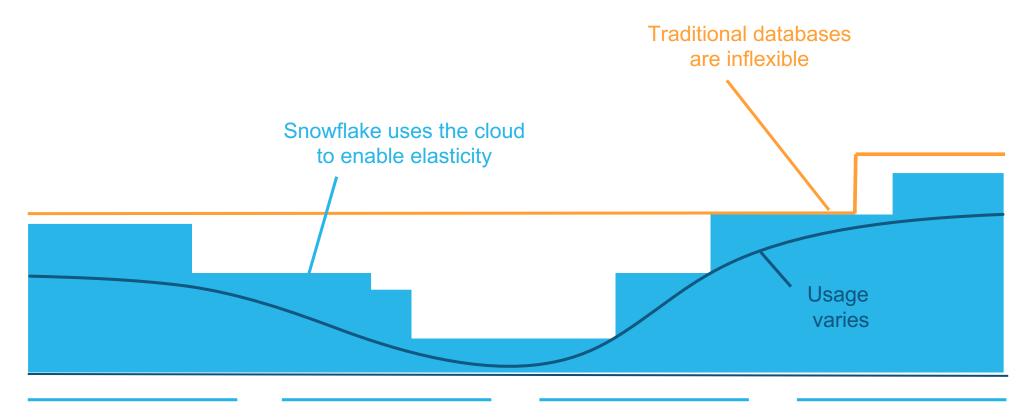












Pay for only what you use with no overprovisioning

Eliminate overbuy

Scale compute up and down, transparently and automatically No need for capacity planning, make capacity decisions on the fly

© 2019 Snowflake Inc. All Rights Reserved.

#2 - SECURE BY DESIGN

Authentication



- Embedded multifactor authentication
- Federated

 authentication
 available

Access control



- Role-based access control model
- Granular privileges on all objects & actions

Data encryption



- All data encrypted, always, end-to-end
- Encryption keys managed automatically

External validation









 Certified against enterprise-class requirements

#1 – Automatic Query Optimization



Load data and run queries, we do all the rest

Zero infrastructure and admin costs

Secure and highly available

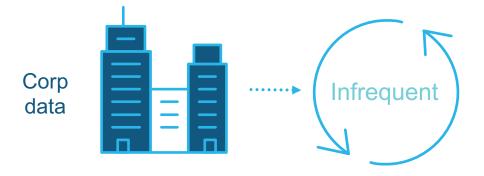
Fully managed with no knobs or tuning required

No indexes, distribution keys, partitioning, or vacuuming

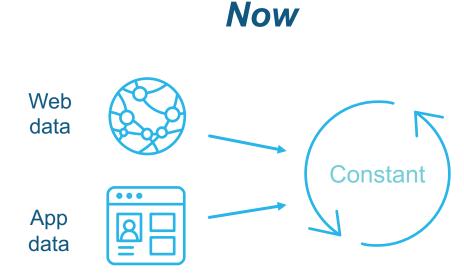
Continuous Data Loading with Snowpipe

Data is Being Generated Faster than Ever

Before



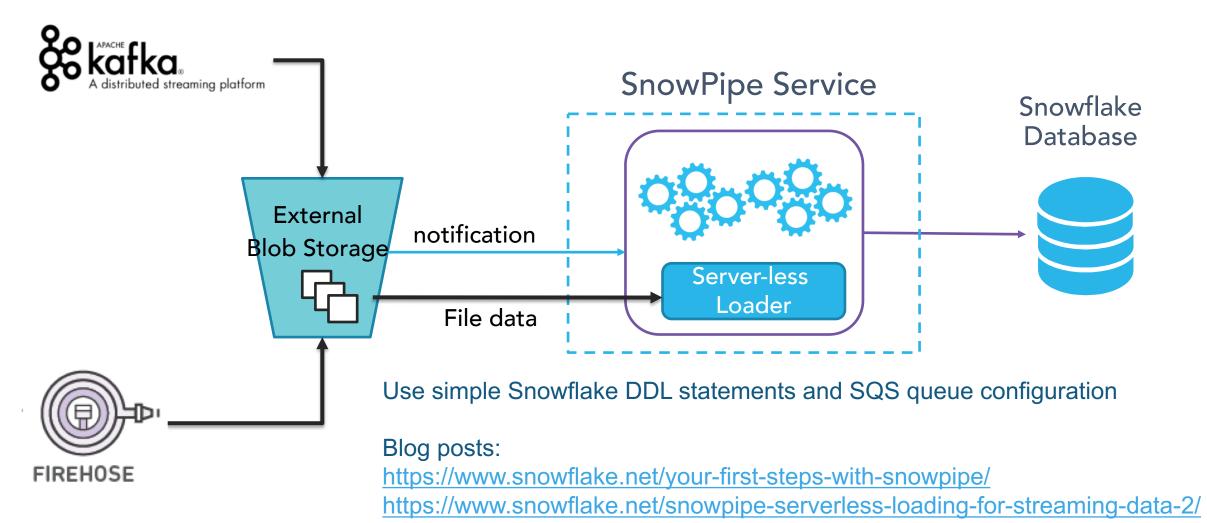
Data accumulated over time.



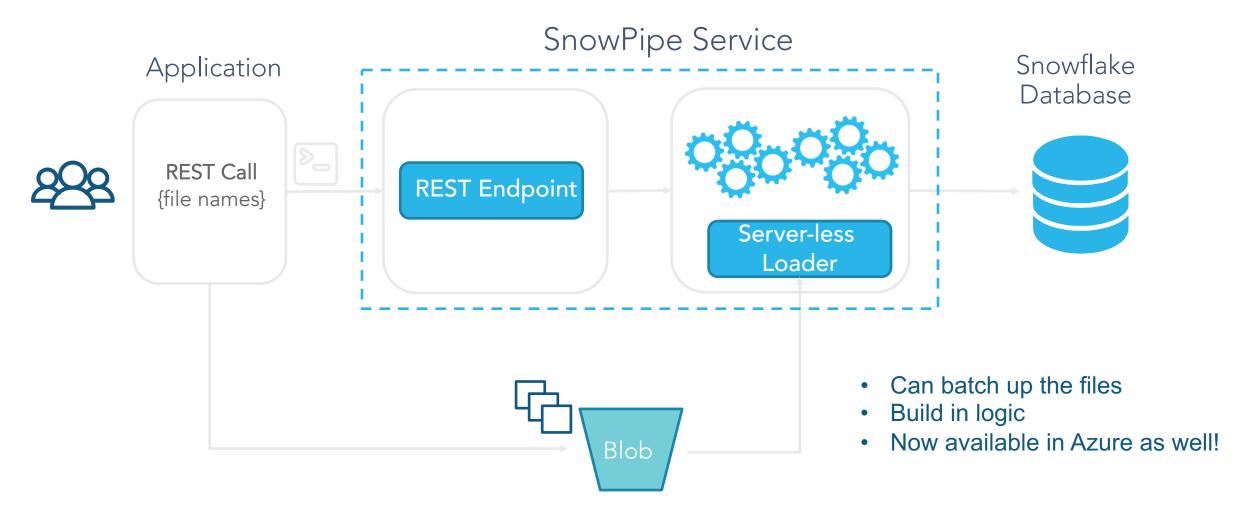
Data is created constantly.

There's an enormous opportunity to use continuously generated data in analysis.

Snowpipe Auto-Ingest: Fully Automatic Loading from Blob Storage



Snowpipe – Expert Mode



Key Snowpipe Benefits



Continuously generated data is available for analysis in seconds



Avoid repeated manual COPY commands



Full support for semi-structured data on load



You only pay for the compute time you use to load data



0 management. No indexing, tuning, partitioning or vacuuming on load



Server-less: No servers to manage or concurrency to worry about.

Monetizing Your Data with The Data ShareHouse

Business Case for Snowflake Data Sharing



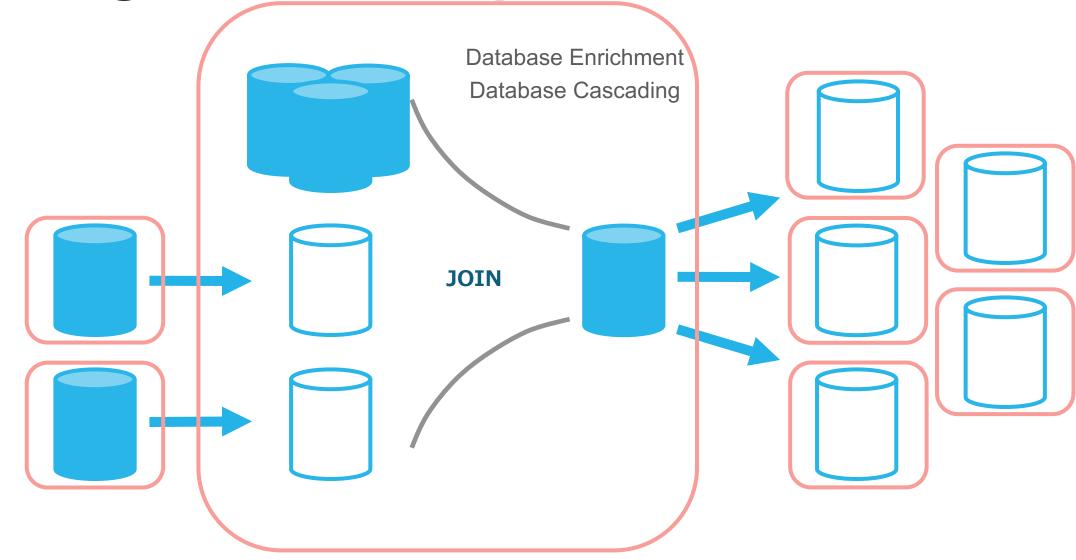
- Access to deliver, query-able data as a service
 - Digital Marketing
 - Internet of Things
 - Healthcare
 - Financial services
- Collaboration
 - Partners
 - Vendors
 - Customers
- Connecting enterprise data silos
 - Connecting multiple enterprise accounts



Live Data Sharing

Share Live Data Instantly NO FTP or EDI **Data Provider Snowflake Account Grant Secure Row Level Access NO Duplicate Files or Storage Use One Share for Many NO Additional Charge** Data to Share Consumer 2 Consumer 1 **Snowflake Account Snowflake Account** Data Secure Data Share Analysts Science WH Wants provider data for SQL Wants provider data for data scientists analysts

Enabling Database Pipelines



Snowflake Data Sharing

Business benefits

- Easy access to more complete data
- Better and faster decision making
- New monetizing opportunities

Better access to data

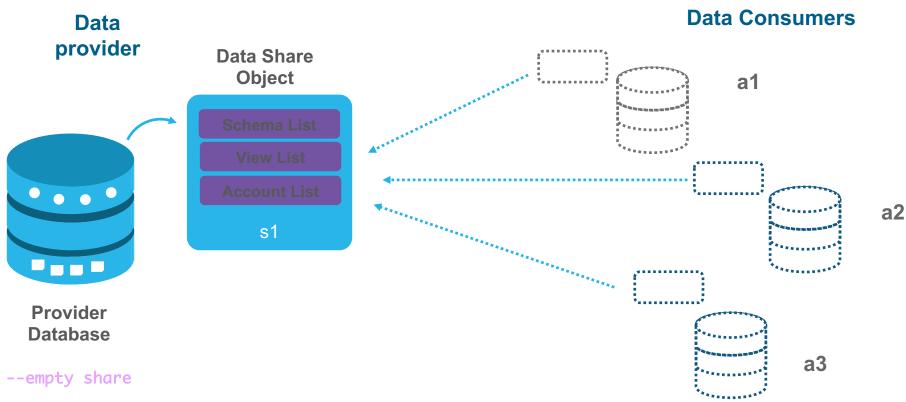
- Removes the pain point of data movement
- Provides access to live data with fine granularity
- Enables easy access to queryable data

Built into the Snowflake architecture

- Separation of Compute and Storage
- Global Metadata Store
- Cloud Connectivity



Data Sharing - Configuring access



create share s1; --empty share

grant usage on database sales to share s1; -- add database grant usage on schema sales.east to share s1; -- add schema grant usage on view sales.east.accts to share s1; -- add view

alter share s1 add accounts=a1, a2, a3; -- add accounts

create database sales from share p1.s1;

Snowflake Data Sharing Customer Use Cases

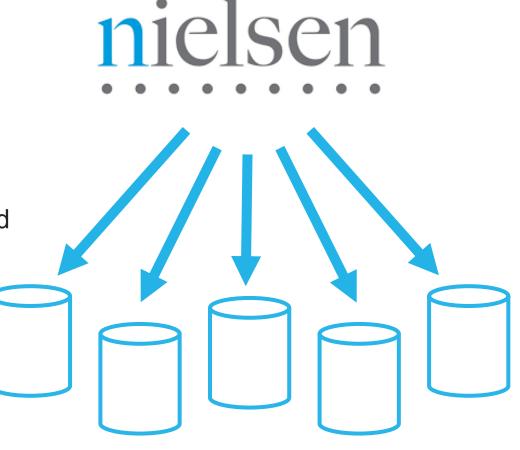
Use Cases for Snowflake Data Sharing

- Snagajob connects people with job openings
 - Uses StrideSpark for email reach out to potential candidates
 - Shares a view that lists likely matches for open jobs with StrideSpark
- StrideSpark runs queries on live data
 - Stridespark reaches out to likely candidates without delay
- Both Snagajob and StrideSpark get cost savings
 - No need to move data using complex data pipelines
 - No cost to Snagajob for using datasharing
 - StrideSpark pays for compute, but not for storage



Use Cases for Snowflake Data Sharing

- Nielsen is a global information, data, and measurement company
 - Nielsen knows "What People Watch, Listen To, and Buy"
 - Nielsen Marketing Cloud includes eXelate DMP which provides unified consumer profiles
 - Nielsen sells selective slices of their DMP data available to advertisers for particular marketing campaigns
- Nielsen plans to use data sharing for making detailed datasets available to subscribers
 - Lower friction, lower cost solution
 - Scalable operations



Agile Data Lifecycle

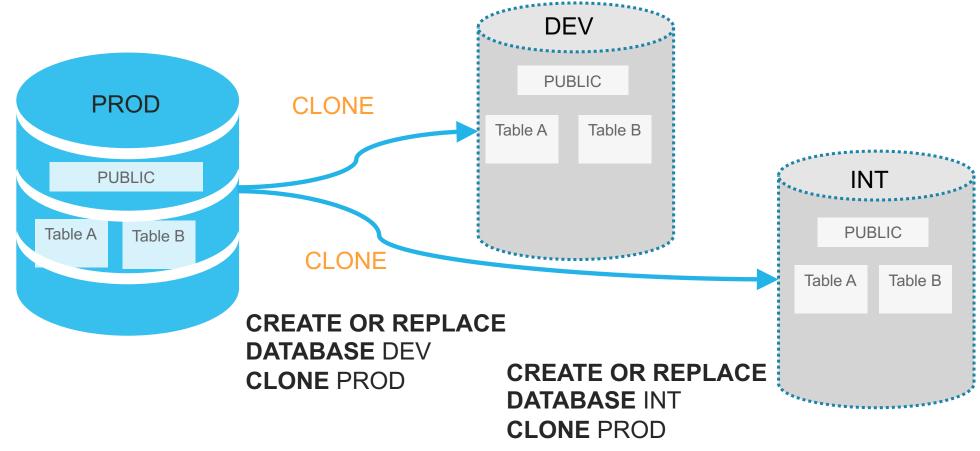


Agile Data Lifecycle

- Separation of Workloads
 - Individual virtual warehouse for each dev/test/prod functional area
- CLONE for dev/test on demand!
 - Full logical copy of the data, but uses no extra storage
 - Test/dev operations against clone have no effect on original data
 - Security
 - RBAC limits dev/test access to clone and not production data
 - Secure Views permit role- or user-based obfuscation / masking / projection
- Business Impact better quality code
 - Dev and test teams are working on data at scale, see true app performance
 - Full range of values means fewer surprises when app encounters live data

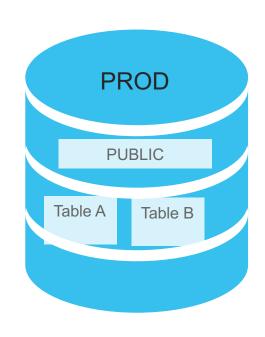
Scenario 1

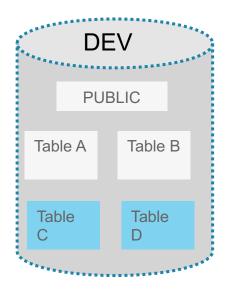
 Create development (DEV) and integration (INT) databases from production (PROD)

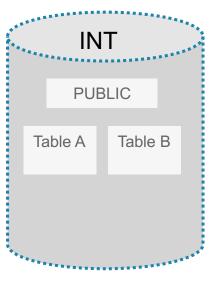


Scenario 2: new development

· Create two new tables, C and D, in the development (DEV) database

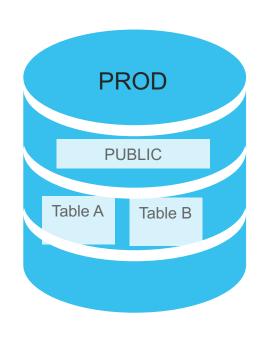


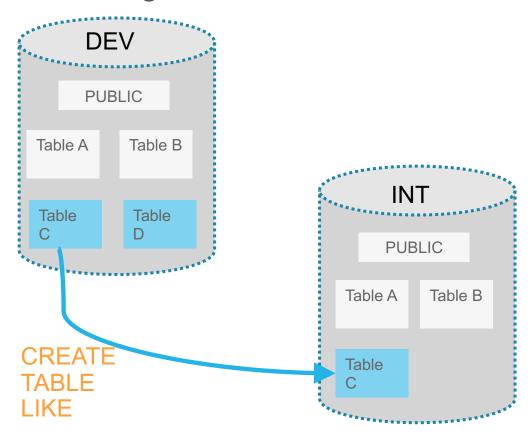




Scenario 2: new development

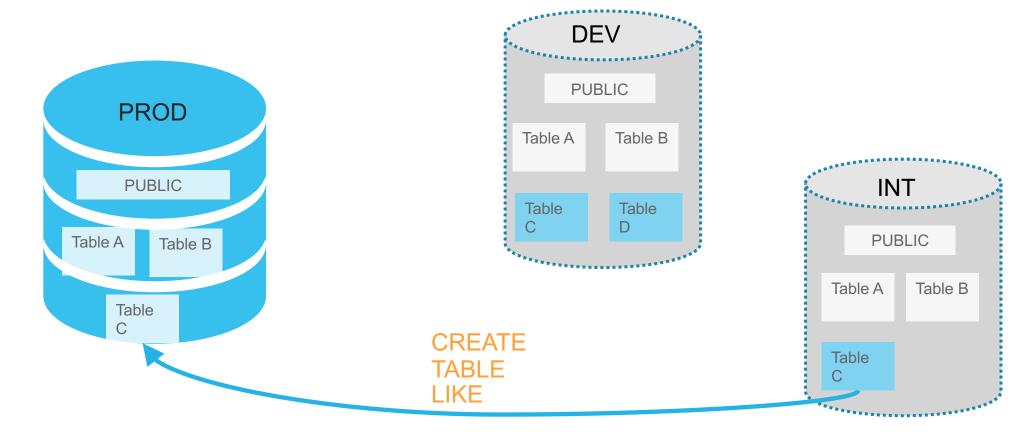
Mini-release: promote table C for integration testing





Scenario 2: new development

Deploy to production: promote table C to PROD database



Scenario 2: new development DEV **PUBLIC** Refresh DEV: get latest PROD data into DEV and INT Table B Table A DEV2 Table Table **PUBLIC PROD CLONE** Table A Table B CLONE **PUBLIC** INT Table Table CREATE OR REPLACE **PUBLIC** Table A Table B **DATABASE** DEV2 **CLONE PROD** Table A Table B CREATE OR REPLACE **Table TABLE** Dev2.TableD **Table CLONE** Dev.TableD

Other Cloning Use Cases

CLONE for Data Scientists

- · Quick and Safe sandbox for discovery and testing
- Combine with own virtual warehouse for complete isolation
- Business Impact better data science
 - More fine-grained data over longer time intervals
 - Deeper insights, better forecasting, more monetizable results

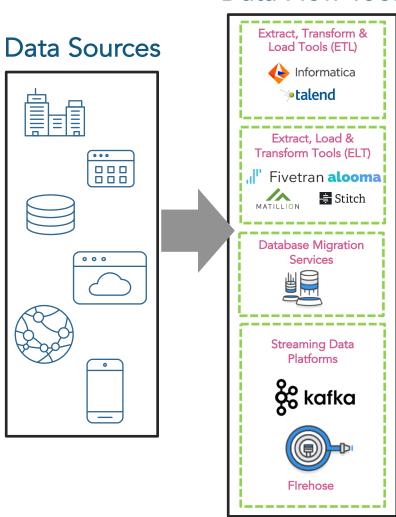
CLONE for Compliance

- Monthly, quarterly, annual clones financial reporting, auditing requirements
- Business Impact simpler compliance
 - Your "backups" are live and immediately available

Example Reference Architectures

Relational processing of multi-structured data

Data Flow Tools



Tables, CSV, JSON, XML, Avro, Parquet

Semi-/Structured Data

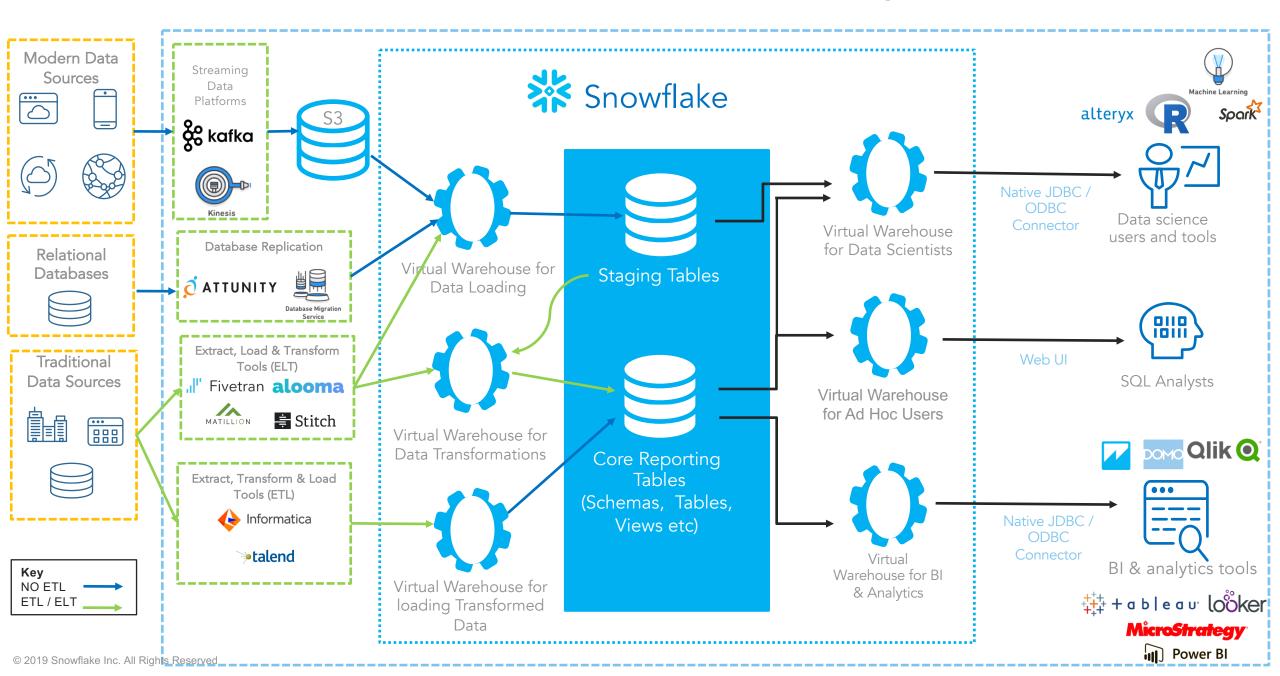
Relational Data
Processing

- Simple data architecture
- Structured & semi-structured data

Virtual

- Land directly into Snowflake
- Access and query with SQL
- Fully relational, ANSI, ACID
- High performance

Data Warehouse Modernization using Snowflake



Evaluating Your Options

Evaluation Considerations

Not all cloud data warehouses are created equal

- Legacy data warehouse offered in the cloud
 - Might even be just a VM on a cloud platform
- Cloud data warehouse built on traditional architecture
 - Shared nothing
 - Shared disk
- Query services
- Cloud-native data warehouse as a service

IaaS, PaaS, or SaaS?

- Which are you looking for?
- Who handles upgrades?
 - Do you have staff to manage this?



Evaluation Considerations

Know your specific use case!

Support all your data

Do you have structured and semi-structured data?

Support all your users

- How many concurrent users and processes do you need?
- Do you have unpredictable usage spikes?
- Do you need to load and query at the same time?

Evaluation Considerations

Low or zero maintenance

- What are your DR and HA requirements?
- How much downtime can you tolerate?

What are your security requirements?

Do you have PII or PHI data?

Full relational DB with standard SQL

- Is the solution supportable with existing staff skills?
- Does it support your existing tools

What are your current and future business needs?

What is the cost model?

Do a POC – seeing is believing!



Snowflake in Action Today

Common customer scenarios



noSQL replacement

Replace use of noSQL system (e.g. Hadoop) for transformation and SQL analytics of multi-structured data

Data warehouse for SaaS offerings

Use Cloud DW as backend data warehouse supporting data-driven SaaS products

Data warehouse modernization

Consolidate legacy datamarts and support new projects

Delivering compelling results



Simpler data pipeline

Replace noSQL database with Snowflake for storing & transforming JSON event data





Faster analytics

Replace on-premises data warehouse with Snowflake for analytics workload

Data warehouse appliance: 20+ hours





Significantly lower cost

Improved performance while adding new workloads--at a fraction of the cost

Data warehouse appliance: \$5M + to expand



Snowflake:

added 2 new workloads for \$50K

Simplifying the Data pipeline







Staging

EDW Ana

Analysts & BI Tools



Game Event Data Internal Data

Third-party Data

Kinesis

Existing EDW

Analysts & BI Tools



Cleanse Normalize Trransform

15 minutes

Snowflake

11-24 hours



Scenario

Complex pipeline slowing down analytics



Pain Points

- Fragile data pipeline
- Delays in getting updated data
- High cost and complexity
- Limited data granularity



Send data from Kinesis to S3 to Snowflake with schemaless ingestion and easy querying



Snowflake Value

Solution

- >50x faster data updates
- 80% lower costs
- Nearly eliminated pipeline failures
- Able to retain full data granularity

DATA ANALYTICS AT EXTREME SCALE

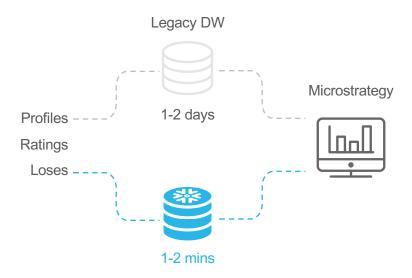
Scenario

- Financial institution with a huge focus on security
- Overburdened staff
- Business needs to run monthly reports that span 10 years of historical data
- No way to analyze semi-structured data

Pain Points

- Quoted \$500,000 to replace their existing hardware appliance
- 20+ hours to run reports
- Could not continue to scale
- Users unable to query while performing ETL





Snowflake Value

120x faster – from 20 hours to 45 minutes

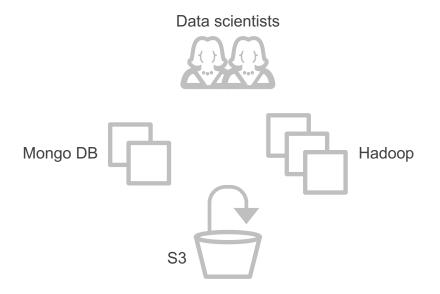
Ad-Hoc analytics available to all users in minutes

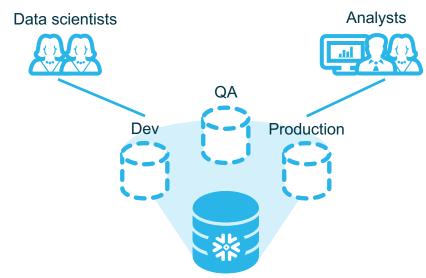


Deployed in a week during the busiest time of year

Data science and exploration









Scenario

Security organization forced to use delayed reporting



Pain Points

- Many data requests unaddressed
- 24 hour turnaround time on requests
- 24 hours to push new models live



Solution

Deploy Snowflake to accommodate analytics workloads



Snowflake Value

- 2 hours to push new models live
- Generated new research report
- Analysts can use data directly























































PROVEN BY OVER 2000 CUSTOMERS





































































Snowflake: A viable platform for the Enterprise



Moved Audience Manager platform from Netezza to Redshift and then to Snowflake



Moving enterprise reporting from Teradata to Snowflake



Delivering secure analytics to more than 11,000 pharmacies



Moving Production EDW from Redshift and Teradata onto Snowflake



Moved audience reporting to Snowflake, reference architecture for Buy Platform



Replacing Netezza with Snowflake to support PB scale analytics

What does a Cloud-native DWaaS Provide?

















Cost effective storage and analysis of GBs, TBs, or even PB's

Lightning fast query performance

Continuous data loading without impacting query performance

Unlimited user concurrency

Full SQL relational support of both structured and semi-structured data

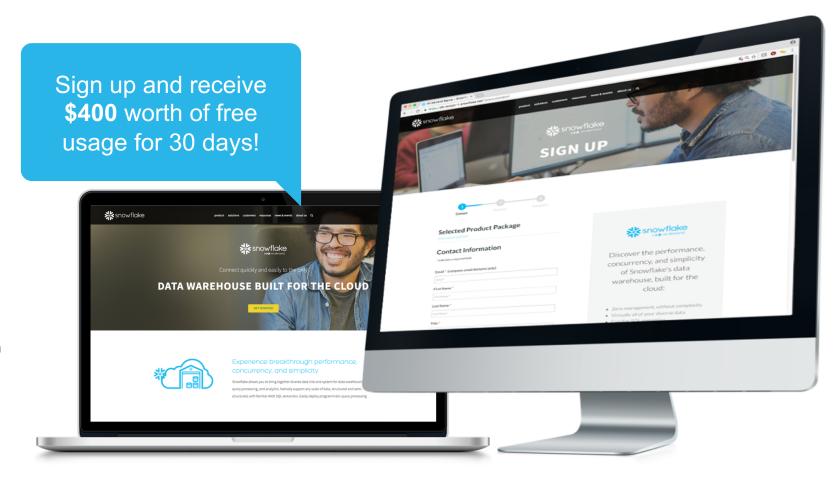
Support for the tools and languages you already use

Discover the performance, concurrency, and simplicity of Snowflake

As easy as 1-2-3!

- 01 Visit Snowflake.net
- O2 Click "Try for Free"
- O3 Sign up & register

Snowflake is the only data warehouse built for the cloud. You can automatically scale compute up, out, or down—independent of storage. Plus, you have the power of a complete SQL database, with zero management, that can grow with you to support all of your data and all of your users. With Snowflake On Demand™, pay only for what you use.







Web



3rd party apps



Mobile



Enterprise apps



ERP



IoT

Contact Info

Kent Graziano Snowflake Computing

Kent.graziano@snowflake.com On Twitter @KentGraziano

More info at http://snowflake.com

Visit my blog at http://kentgraziano.com





THANK YOU







