Demystifying Data Warehouse as a Service (DWaaS)

Kent Graziano, Senior Technical Evangelist @KentGraziano
Snowflake Computing

My Bio

- Senior Technical Evangelist, Snowflake Computing
- Oracle ACE Director (DW/BI)
- OakTable
- 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

About Snowflake

Founded in 2012 by industry veterans

Vision : A world with no limits on data

First data warehouse, taking full advantage of cloud computing

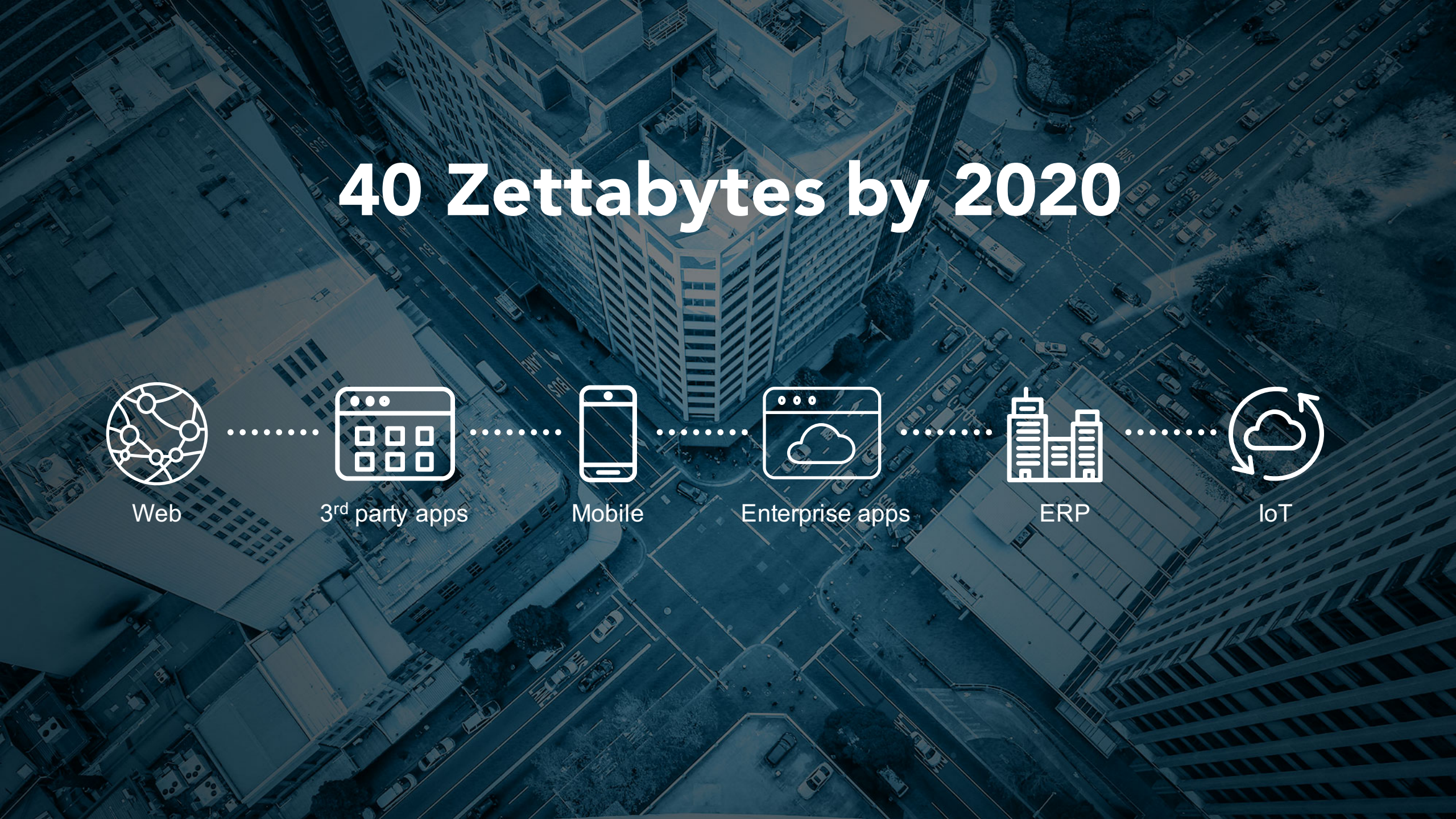
Founders and leadership team with significant experience

Agenda

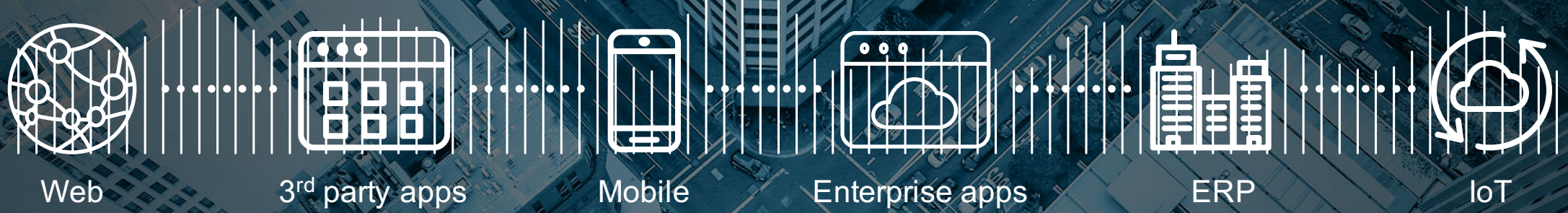
- Data Challenges
- What is a DWaaS?
- What can a DWaaS do for me?
- Features of a DWaaS
- Top 10 Features of Snowflake

Data challenges today

40 Zettabytes by 2020



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.***

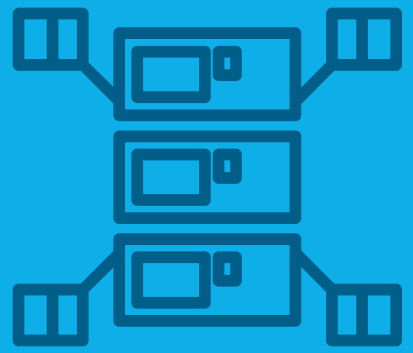
Source: Forrester

Silos & Islands

Data distributed across multiple systems,
difficult to bring together

Silos & Islands

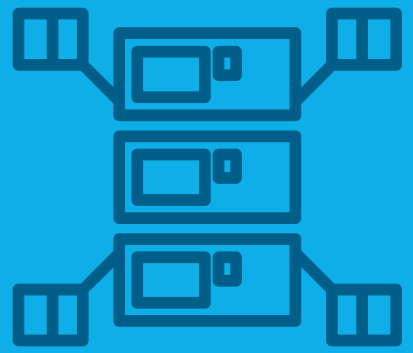
Silos & Islands



Complexity

Multiple systems, complex pipelines,
specialized skills and resources

Silos & Islands



Complexity

Silos & Islands

Complexity

Cost

Upfront capital costs, multiple copies of data, high cost to store data

Silos & Islands

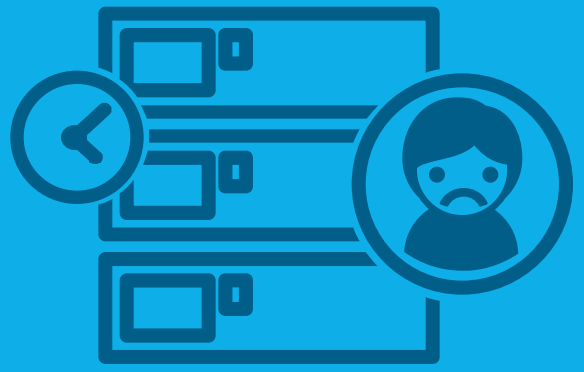
Complexity

Cost

Silos & Islands

Complexity

Cost



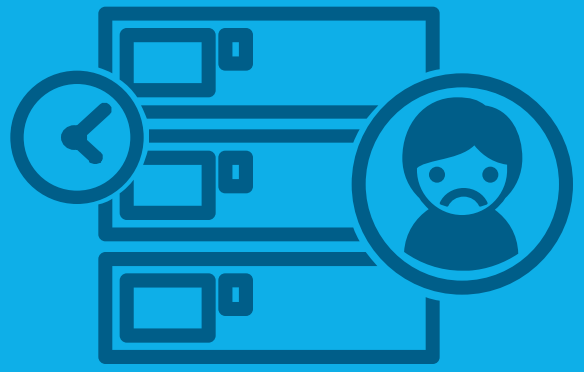
Delays

Data users forced to wait for access to data and analytics

Silos & Islands

Complexity

Cost



Delays

Silos & Islands Complexity Cost Delays

The evolution of data platforms

Relational database

Oracle, DB2, SQL Server

Data warehouse appliance

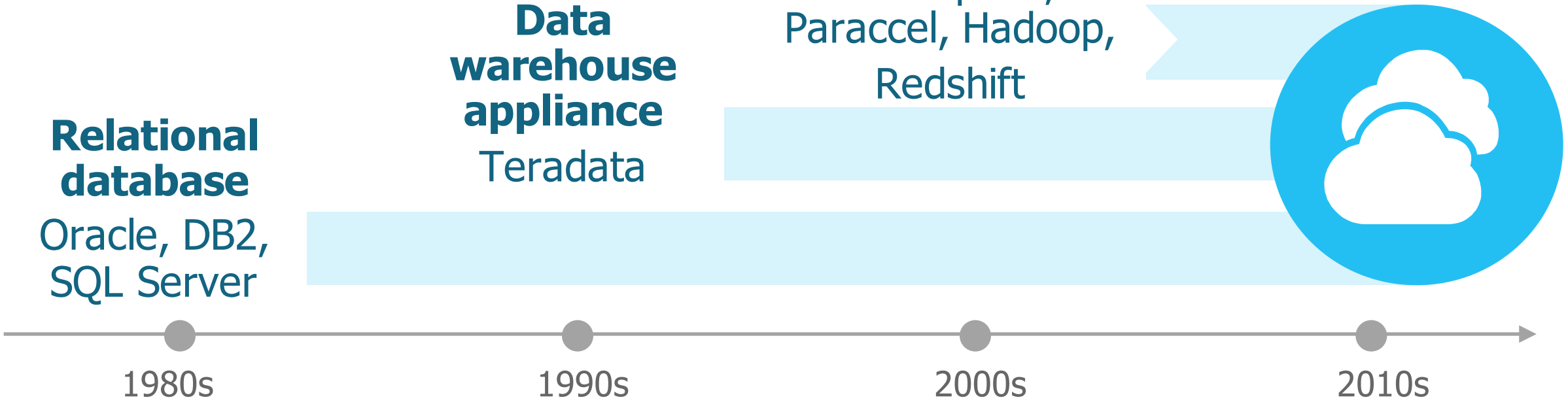
Teradata

Data warehouse & platform software

Vertica, Greenplum, Paraccel, Hadoop, Redshift

Cloud DWaaS

Snowflake



snowflake

What is a 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 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

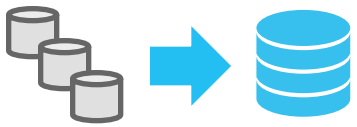
What to Expect from a DWaaS

- 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
- It should be easy to configure and manage
- It should provide a lower TCO
  - Cloud scale pricing

Common Scenarios



Datamart & data silo consolidation
Consolidate legacy datamarts to eliminate silos and support new projects

Integrated data analytics
Directly load structured + semi-structured data for reporting & analytics

Exploratory & ad hoc analytics
Direct access to data for SQL analysts & data scientists to explore data, identify correlations, build & test models

What's possible

Up to 200x faster reports that enable analysts to make decisions in minutes rather than days

Load and update data in near real time by replacing legacy data warehouse + Hadoop clusters

Developing new applications that provide secure (HIPPA) access to analytics to 11,000+ pharmacies

snowflake

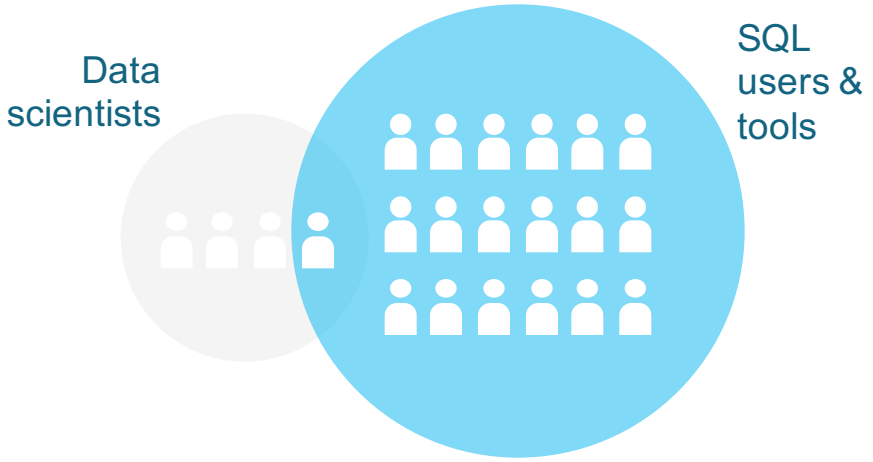
Introducing Snowflake

Snowflake:
Data Warehouse Built for the Cloud

Data Warehousing…

- SQL relational database
- Optimized storage & processing
- Standard connectivity – BI, ETL, …

…for Everyone



- Existing SQL skills and tools
- "Load and go" ease of use
- Cloud-based elasticity to fit any scale

snowflake

The Snowflake difference

Performance

Multi petabyte-scale, up to 200x faster performance
and 1/10th the cost

Concurrency

Multiple groups access
data simultaneously
with no performance degradation

Simplicity

Fully managed with a
pay-as-you-go model.
Works on any data

snowflake

Data Warehouse built for the cloud

Complete SQL Database

Zero Management

All of your Data

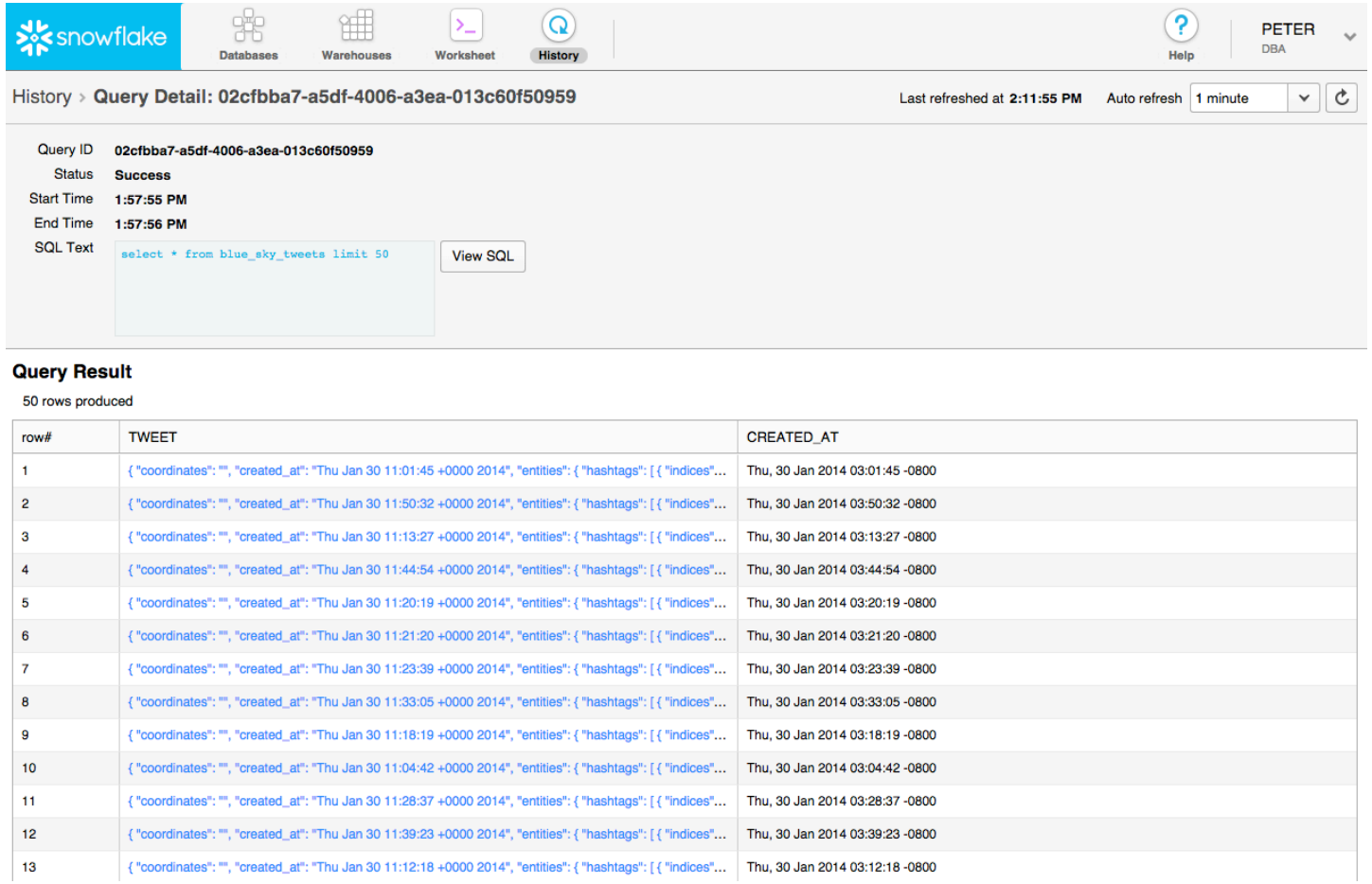
All of your Users

Pay only for what you use

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

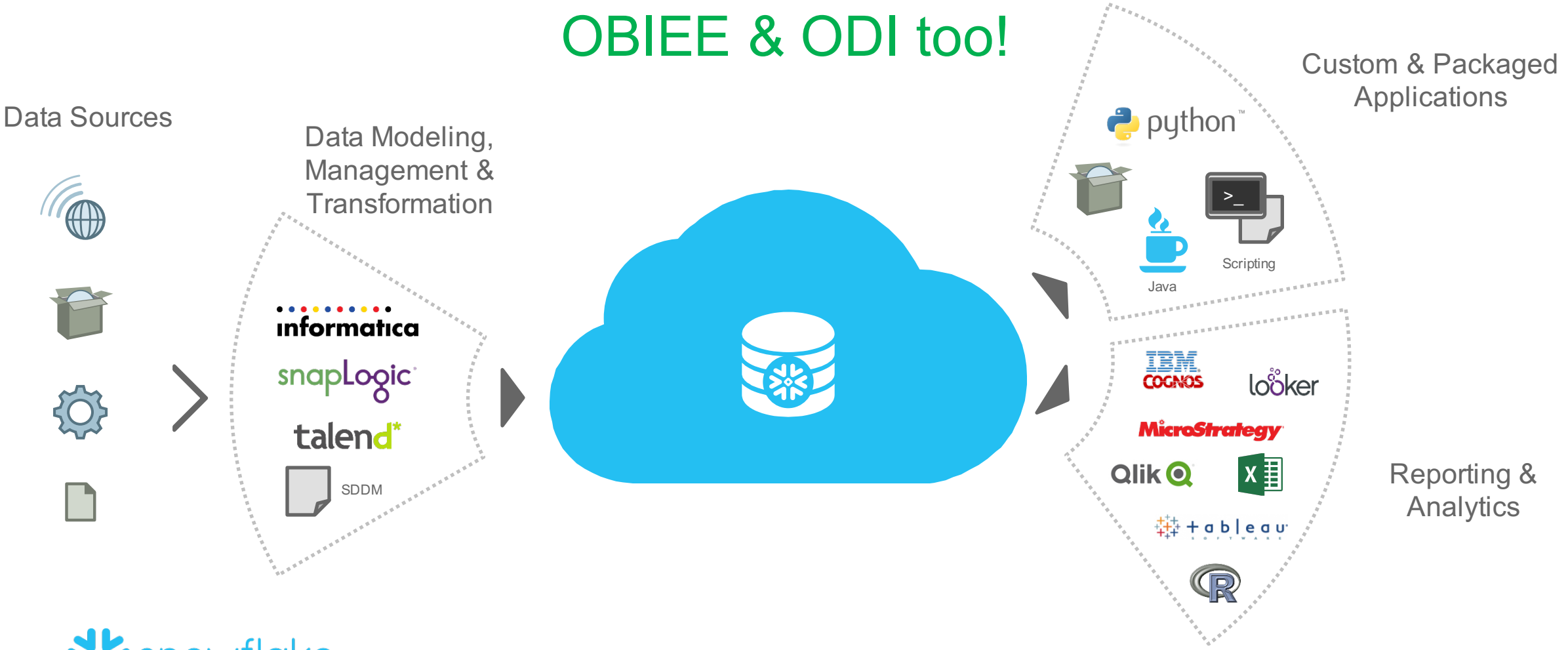
#10 – Persistent Result Sets

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



#9 Connect with JDBC & ODBC

OBIEE & ODI too!



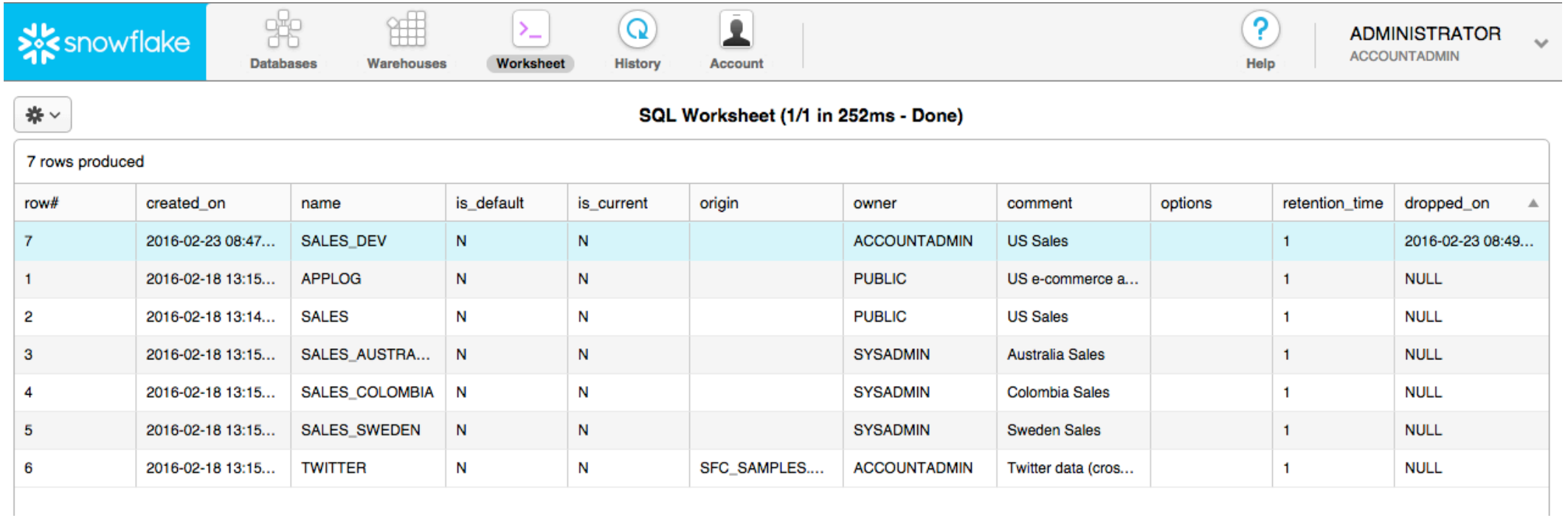
#8 - UNDROP

UNDROP TABLE <table name>

UNDROP SCHEMA <schema name>

UNDROP DATABASE <db name>

Part of Time Travel feature: AWESOME!



row#	created_on	name	is_default	is_current	origin	owner	comment	options	retention_time	dropped_on
7	2016-02-23 08:47...	SALES_DEV	N	N		ACCOUNTADMIN	US Sales		1	2016-02-23 08:49...
1	2016-02-18 13:15...	APPLOG	N	N		PUBLIC	US e-commerce a...		1	NULL
2	2016-02-18 13:14...	SALES	N	N		PUBLIC	US Sales		1	NULL
3	2016-02-18 13:15...	SALES_AUSTRA...	N	N		SYSADMIN	Australia Sales		1	NULL
4	2016-02-18 13:15...	SALES_COLOMBIA	N	N		SYSADMIN	Colombia Sales		1	NULL
5	2016-02-18 13:15...	SALES_SWEDEN	N	N		SYSADMIN	Sweden Sales		1	NULL
6	2016-02-18 13:15...	TWITTER	N	N	SFC_SAMPLES....	ACCOUNTADMIN	Twitter data (cros...		1	NULL

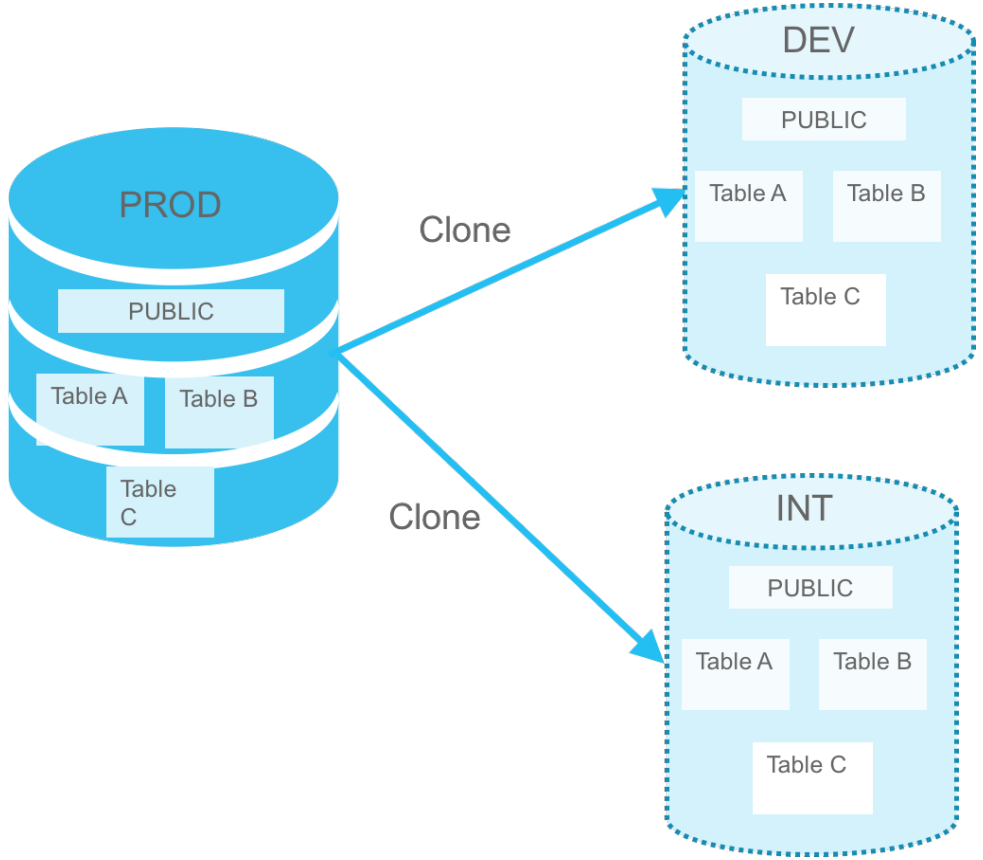
#7 Fast Clone (Zero-Copy)

- 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));
```



snowflake

#6 – JSON Support with SQL

Structured data
(e.g. CSV)

Apple	101.12	250	FIH-2316
Pear	56.22	202	IHO-6912
Orange	98.21	600	WHQ-6090

```
{       "firstName": "John",
        "lastName": "Smith",
        "height_cm": 167.64,
        "address": {
                "streetAddress": "21 2nd Street",
                "city": "New York",
                "state": "NY",
                "postalCode": "10021-3100"
        },
        "phoneNumbers": [
                { "type": "home", "number": "212 555-1234" },
                { "type": "office", "number": "646 555-4567" }
        ]
}
```

Semi-structured data
(e.g. JSON, Avro, XML)

All Your Data!

```
select v:lastName::string as last_name
from json_demo;
```

- Optimized storage
- Flexible schema - Native
- Relational processing

snowflake

#5 – Standard SQL w/Analytic Functions

```
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

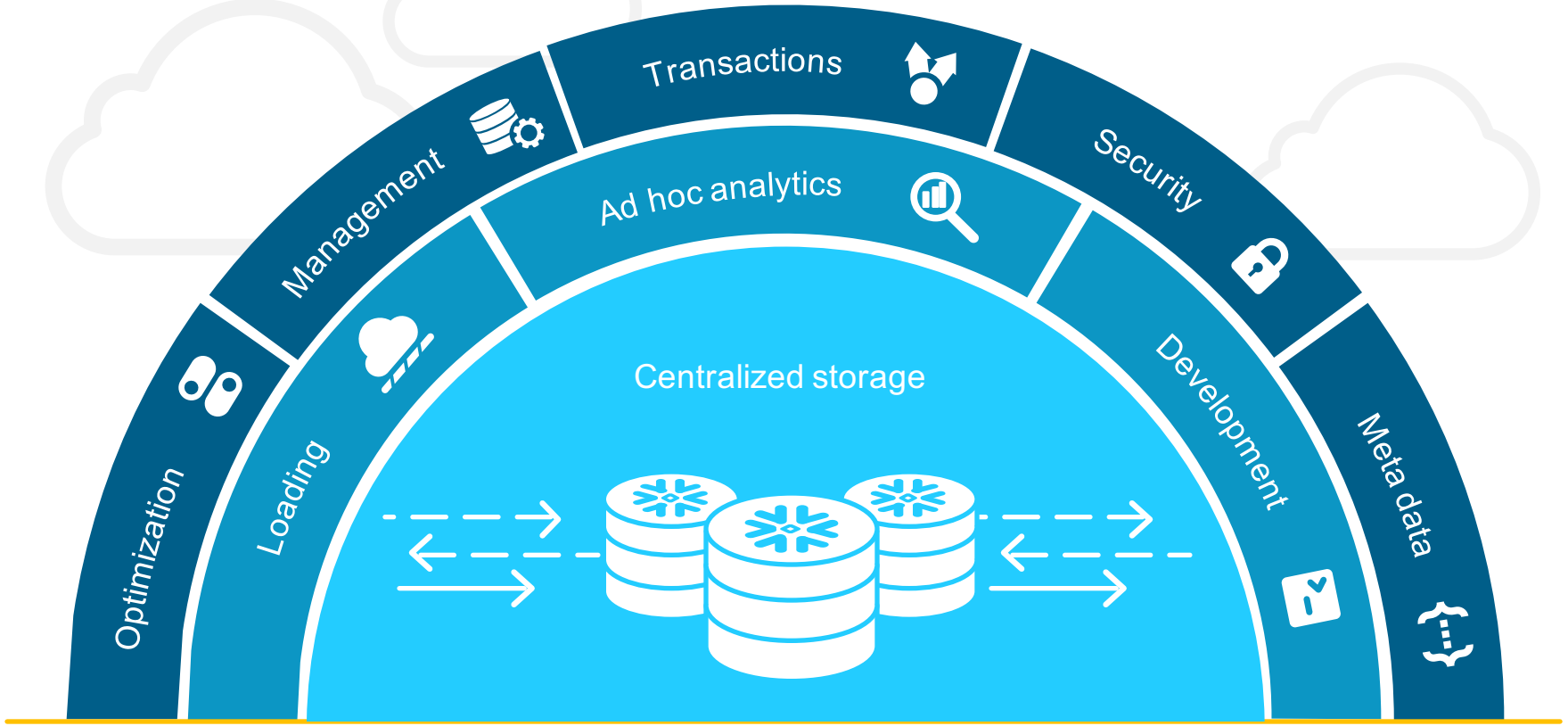
- Data definition language (DDLs)
- Query (SELECT)
- Updates, inserts and deletes (DML)
- Role based security
- Multi-statement transactions

tableau
looker
Qlik
Power BI
MicroStrategy
Excel
COGNOS

snowflake

#4 – Separation of Storage & Compute

Snowflake's multi-cluster, shared data architecture



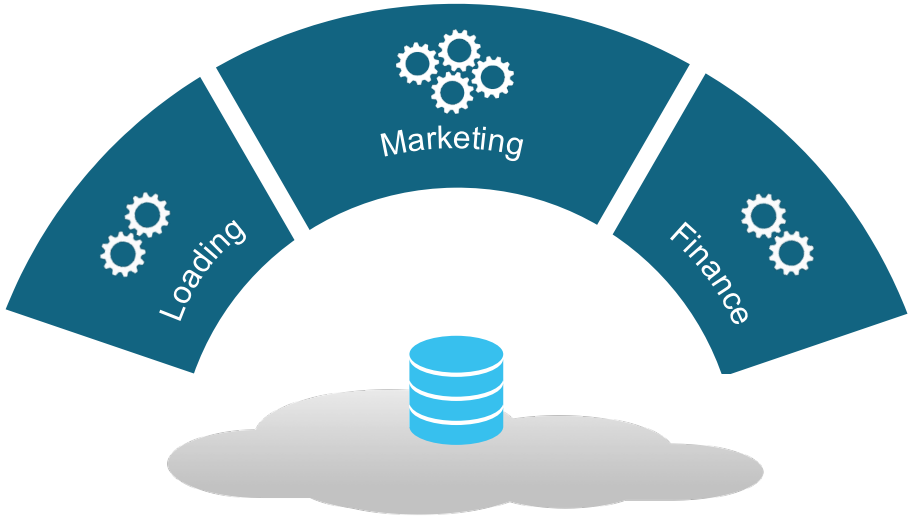
Service
Compute
Storage

Instant, automatic scalability & elasticity

snowflake

#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

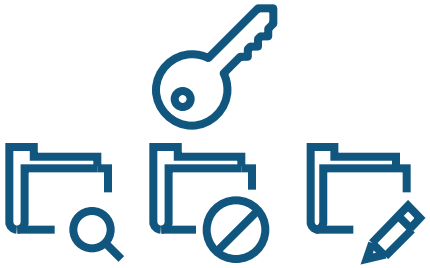
#2 – Secure by Design with Automatic Encryption of Data!

Authentication

Embedded multi-factor 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

AICPA Service Organization Control Reports
AICPA SOC aicpa.org/soc
Formerly SAS 70 Reports

SANS INSTITUTE

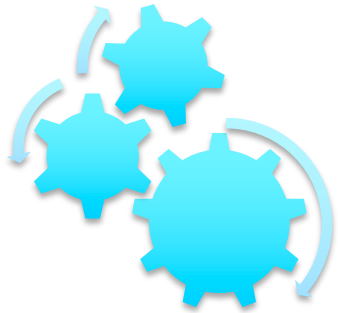
Certified against enterprise-class requirements

HIPPA Certified!

snowflake

#1 - Automatic Query Optimization

Zero Management

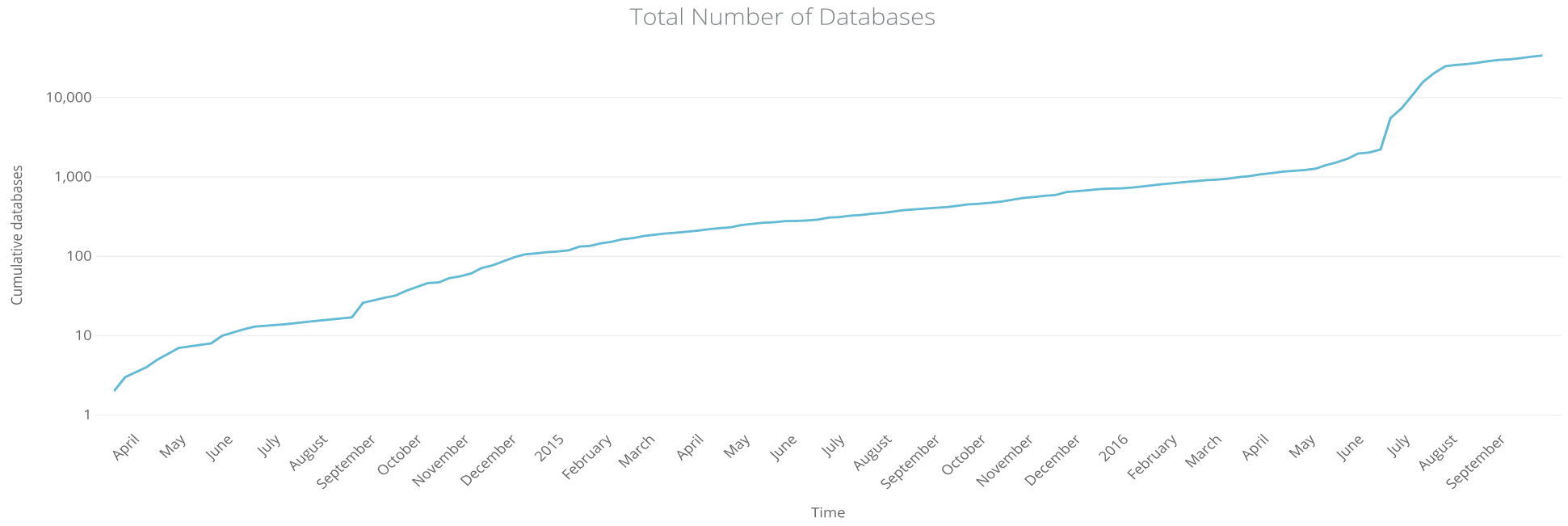


- Fully managed with no knobs or tuning required
- No indexes, distribution keys, partitioning, vacuuming,…
- Zero infrastructure costs
- Zero admin costs

Data Warehousing as a Service in Action Today

Cloud-scale data warehouse

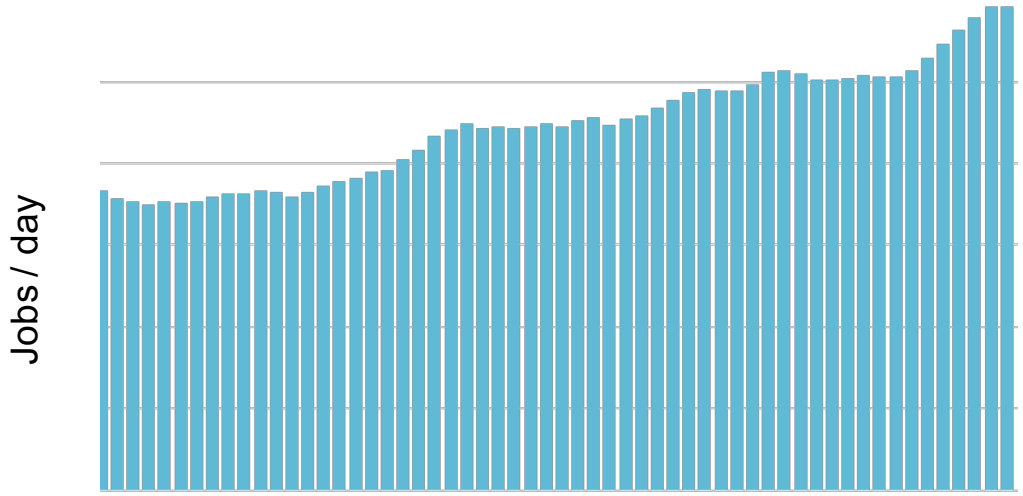
1,110,528,424	121,327,891	171,218,777
Total Number of Jobs	Total Number of Queries (SELECT)	Total Number of LOAD Queries (COPY)



Generated by looker on October 2, 2016 at 3:44pm PDT

snowflake

Steady growth in data processing



- Over 20 PB loaded to date!
  - Multiple customers with >1PB
- Multiple customers averaging >1M jobs / week
- >1PB / day processed
- Experiencing 4X data processing growth over last six months

snowflake

Customer results

accordantmedia

We can do **100 times** more queries per day, helping us give our clients richer analysis far more rapidly.

— Balaji Rao
VP Technology

JANA

Snowflake is faster, more flexible, and more scalable than the alternatives on the market. The fact that we don't need to do any configuration or tuning is great because we can **focus on analyzing data** instead of on managing and tuning a data warehouse.

— Craig Lancaster
CTO

iovation

With Snowflake, I'm able to spin up as many as I want on demand and to spin them down and not pay for those things that I'm not using.

— Kurk Spendlove
Director Engineering

SOASTA

Snowflake is awesomely fast, allows us to store data at a **low cost** and deploy exactly the compute capacity needed, and does all of that without requiring tuning or tweaking.

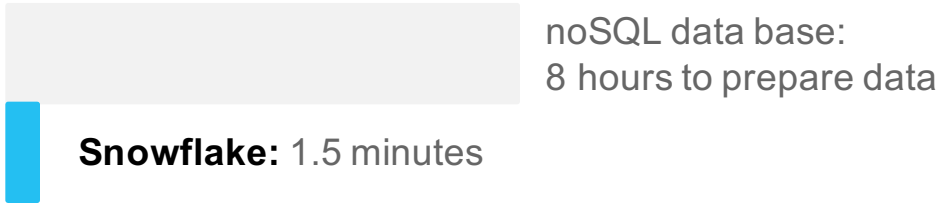
— Matt Solnit
CTO

snowflake

Delivering compelling results

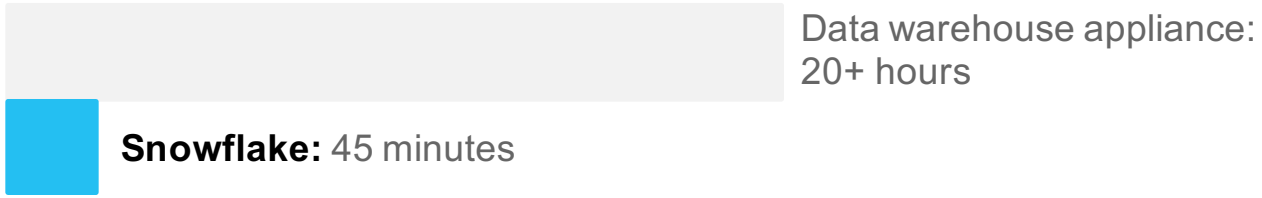
Simpler data pipeline

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



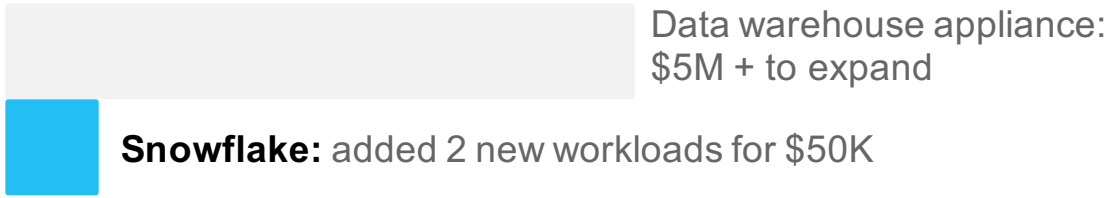
Faster analytics

Replace on-premises data warehouse with Snowflake for analytics workload



Significantly lower cost

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



snowflake

What does a good DWaaS enable?

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

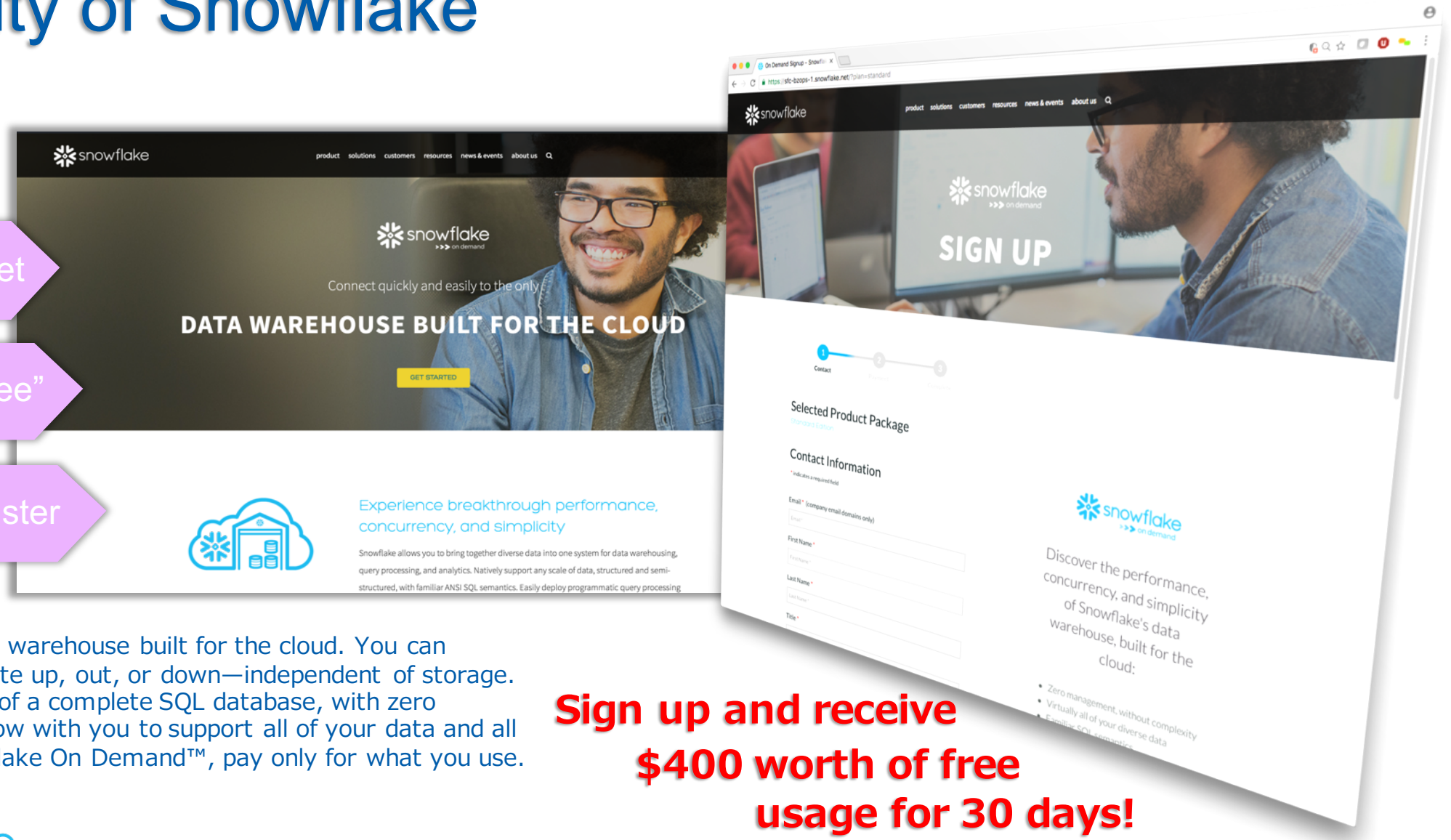
Support for the tools and languages you already use

Making Data Warehousing Great Again!

Discover the performance, concurrency, and simplicity of Snowflake

As easy as 1-2-3!

1. Visit Snowflake.net
2. Click "Try for Free"
3. Sign up & register



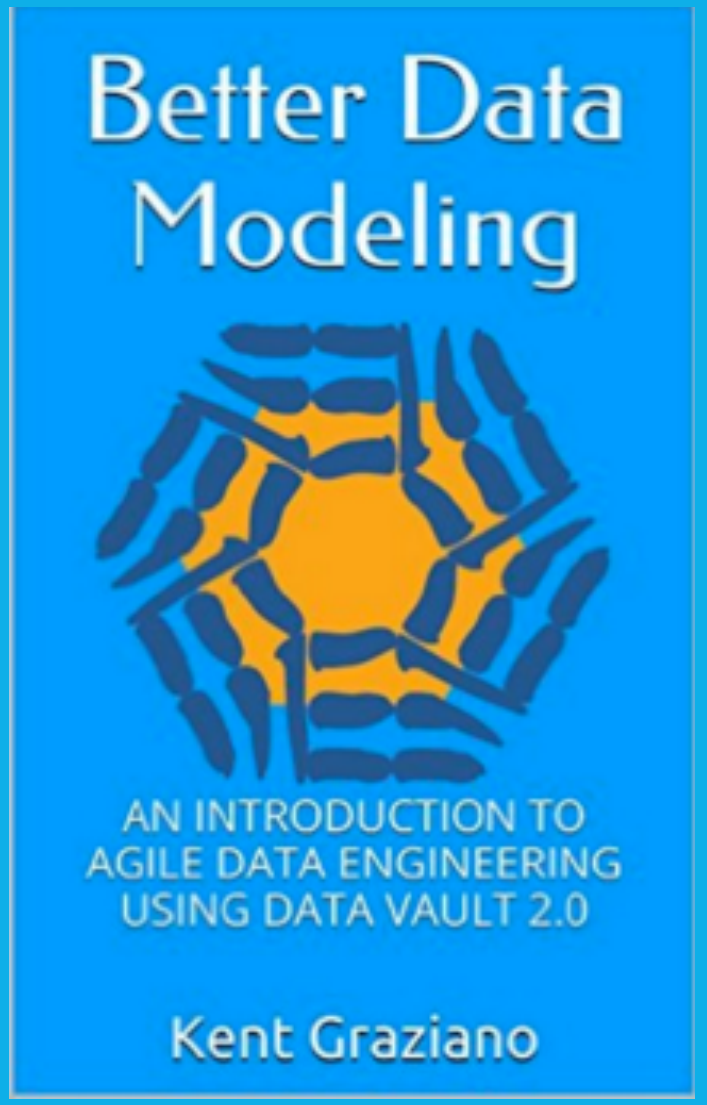
Snowflake is the only data warehouse built for the cloud. You can automatically scale compute up, out, or down—independent of storage. Plus, you have the power of a complete SQL database, with zero management, that can grow with you to support all of your data and all of your users. With Snowflake On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

Sign up and receive $400 worth of free usage for 30 days!

VISA MasterCard AMERICAN EXPRESS DISCOVER NETWORK

snowflake

SHAMELESS PLUG:



Available on Amazon.com

Introduction to Agile Data Engineering

http://www.amazon.com/Better-Data-Modeling-Introduction-Engineering-ebook/dp/B018BREV1C/

Contact Information

Kent Graziano
Snowflake Computing
Kent.graziano@snowflake.net
On Twitter @KentGraziano

More info at
http://snowflake.net

Visit my blog at
http://kentgraziano.com

The Data Warrior
Changing the world, one data model at a time. How can I help you?

Thank you