

Using Snowflake to Unify Data, People and Technology

David Kline, Senior Director of Data and BI

Videology



- > Leading converged, addressable advertising provider.
- > Turn data and inventory into guaranteed results for our clients.
 - > We process ~15 billion publisher requests (request for an ad) daily, which translates to ~50 million impressions (ad is shown) daily, tracking 30 – 50 billion behavioral attributes daily.
- > Strongly partnered with online brands like SnapChat, WhiteOps, Fox, AT&T, Comcast and many others.
- > Operate in 28 different countries.



Web



TV



Mobile



Problem Statement

Sizing

Due to design, we have never been able to right size the Previous Cloud Data Warehouse (PCDW) clusters.

Concurrency

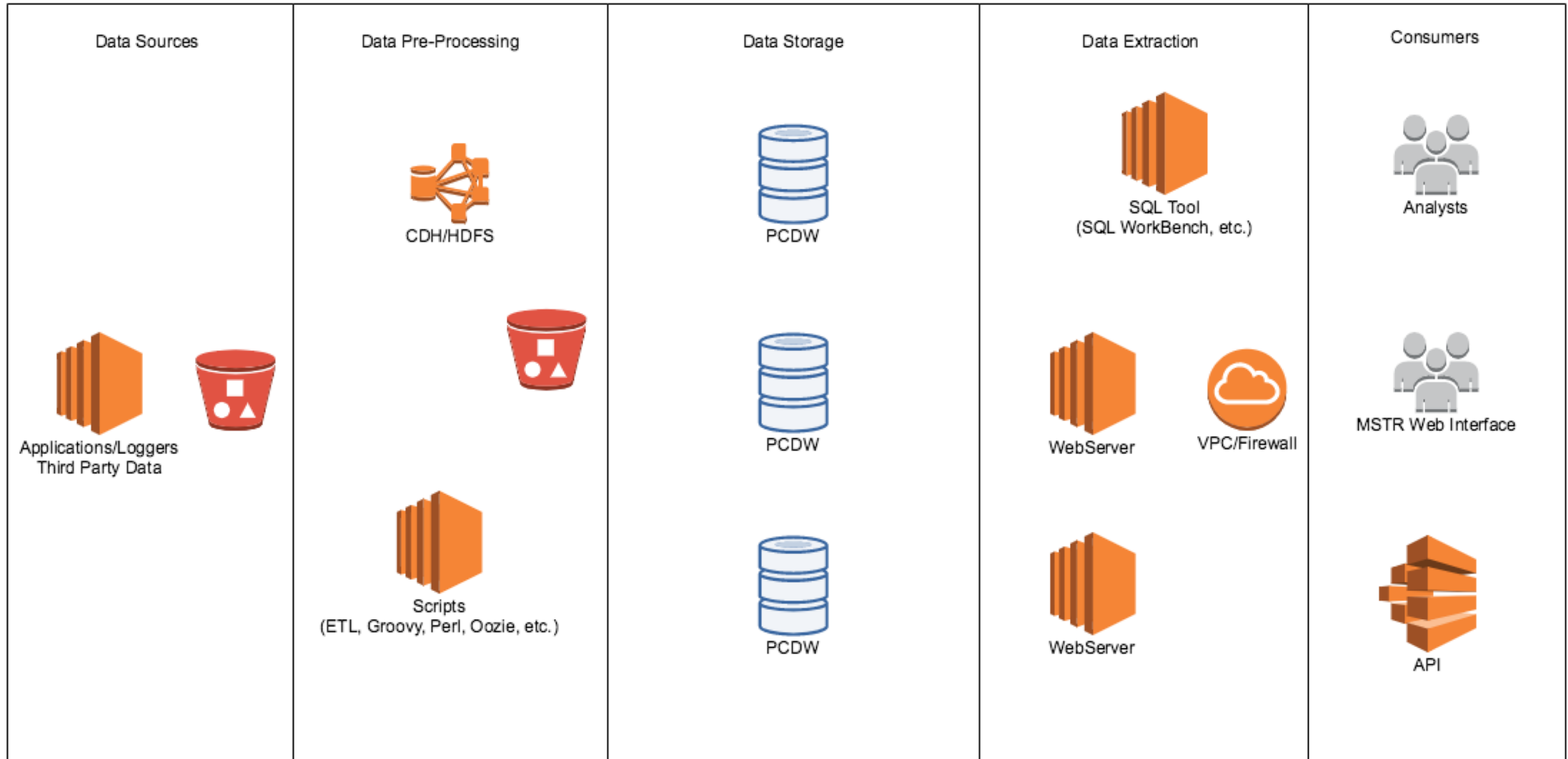
Due to concurrency limits, we have multiple PCDW clusters for different business operations leading to multiple people using multiple PCDW and duplicative storage.

Performance

In order to meet performance SLAs, fact tables were divided by high cardinality facets, little to no raw data storage.

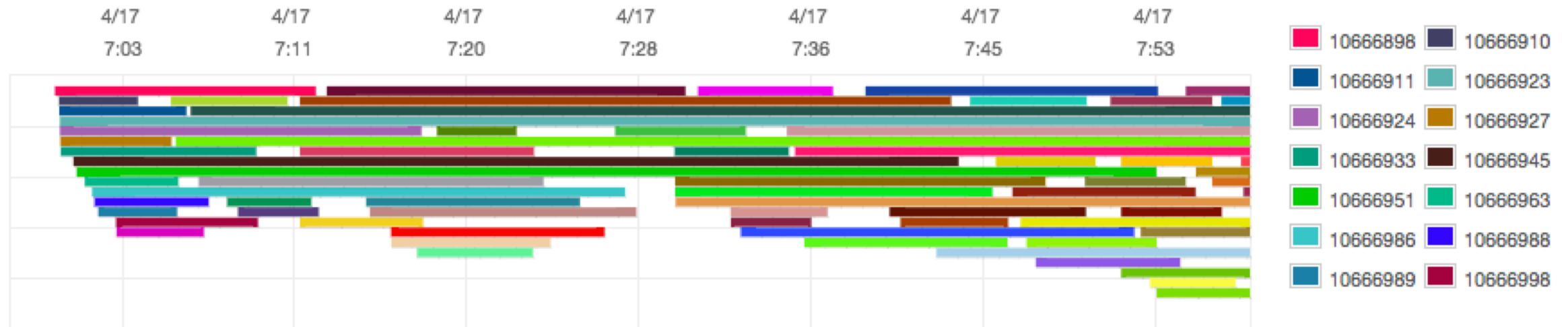


Previous Architecture



Challenges faced with Previous Cloud Data Warehouse (PCDW) platform

- Continuous challenge to SLAs to business
 - Stability issues with legacy PCDW platform
- Heavy ETL cycles caused domino effect
 - If ETL window was missed, MSTR queries would be impacted
 - Had to carefully design ETL and MSTR cube refresh schedule
- Unable to support ad-hoc, direct queries from analysts
 - Total reliance on cubes due to concurrency/performance challenges
- Scaling PCDW too disruptive
 - Read only mode for days whenever cluster resize was required to add capacity



Enter Snowflake – Data warehouse built for the cloud

Performance



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

Concurrency



Multiple groups access data simultaneously with no performance degradation

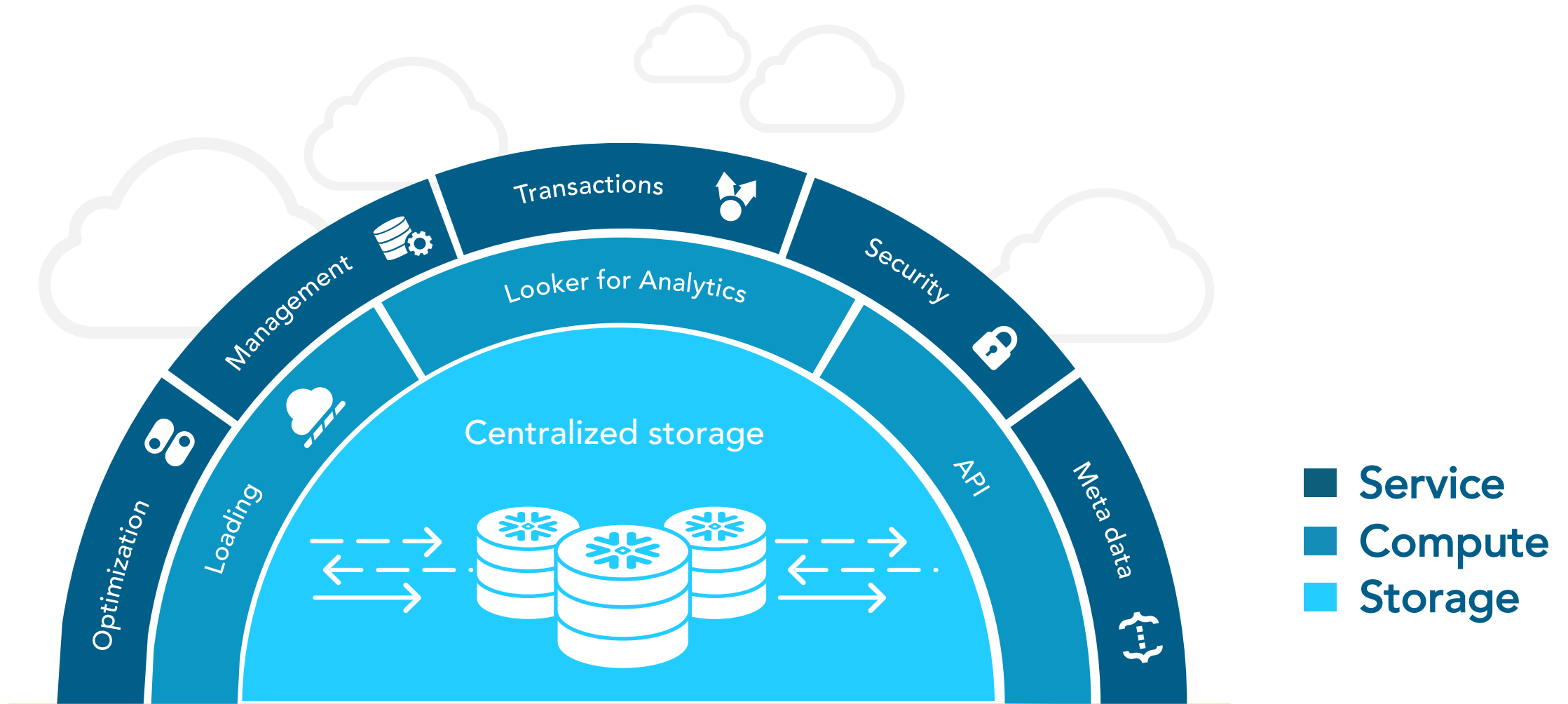
Simplicity



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



Snowflake's multi-cluster, shared data architecture



Instant, automatic scalability & elasticity

Why Snowflake works for Videology

Sizing

Storage and compute are completely separate allowing data storage in a single Snowflake instance with better compression (2 – 3x)

Concurrency

Multiple compute Warehouses for different purposes.

Resizing a warehouse is done in a matter of seconds

Single Instance for all users reducing overall data stored.

Performance

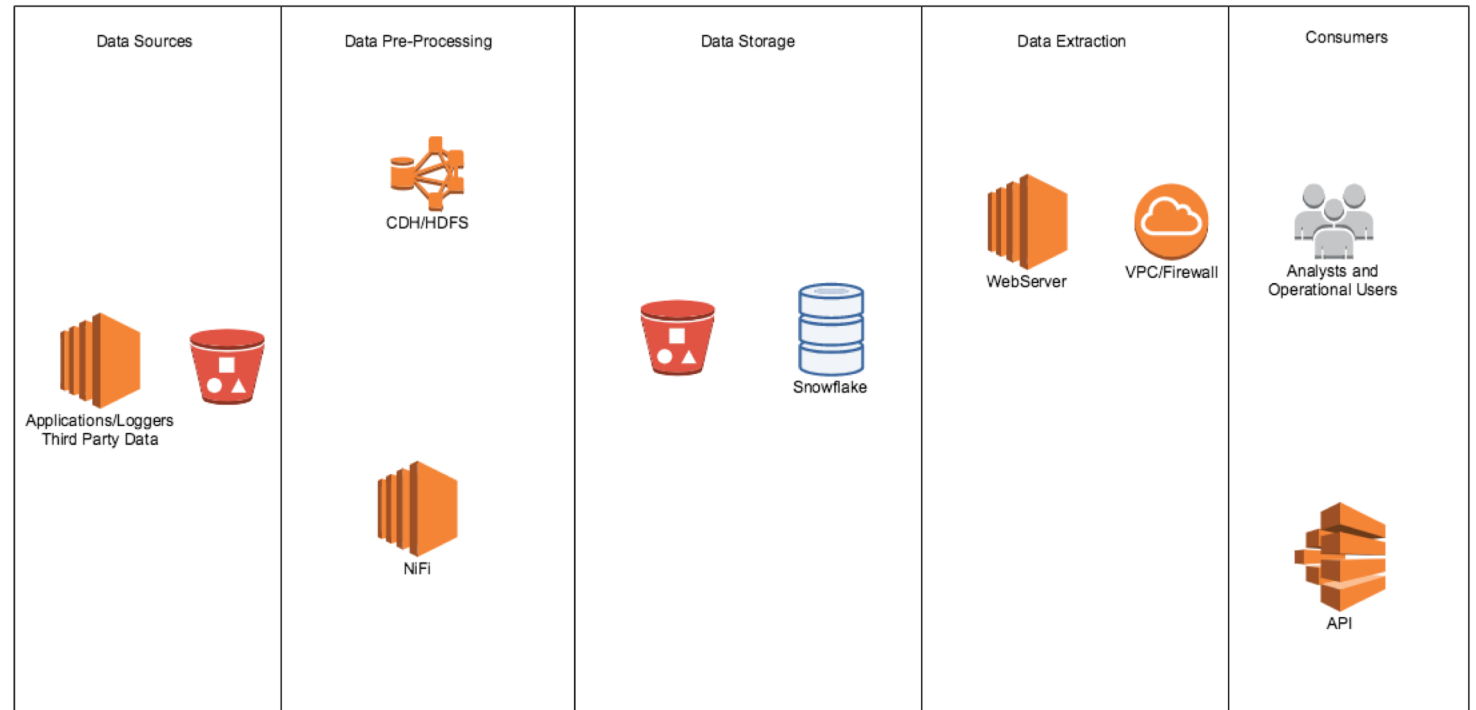
Data can be stored raw or close to it and still be performant allowing for better what if analysis

Uses cases for semi-structured data were uncovered.



Architectural Changes

- Have pre-processing steps drop to Snowflake S3 bucket.
- Replace 3 PCDW clusters with Snowflake (40K/mo savings)
- Eliminate MSTR and allow for true analytics with a single instance of Looker.
- 3 month transition from PCDW to Snowflake.



Use Cases – Time Travel

Time Travel - the careless developer's best friend.

- Restore a table to its state before a bad update.
- Identify changes to the data.

```
select h.*
```

```
(select * from mytable at(timestamp => dateadd('day',-1,current_date)::timestamp)) h
```

```
join mytable t on h.key_id = t.key_id
```

```
where h.field_1 <> t.field_1;
```

- Clone a lower environment at a certain release date.
 - Costs savings of nearly 100% per cloned environment.
 - No compromises necessary (looks like production with only differences stored).
 - Clone a set of tables or entire databases in seconds or minutes not days or weeks.
 - Only costs are changes to cloned data.



Use Cases – Semi Structured Data (JSON)

Semi-structured use cases

1. Tracking change history of a different database

```
{"budget_spent": {"new": 78.12, "old": 78.105}, "budget_spent_today":  
{ "new": 1.845, "old": 1.83}, "delivery_units_spent": {"new": 5183.58,  
"old": 5182.6 }, "delivery_units_spent_today": { "new": 120.54,  
"old": 119.56}}
```

2. Storing data of indeterminate length for fast querying

- Storing data as a json object has significant query performance improvements over arrays.
 - Select * from mytable where array_contains("searchterm",arrayfield); (slow)
 - select * from mytable where jsonfield:"searchterm" is not null; (fast)



Final Thoughts

Unifying Data

- *Single instance allowing for all data storage supporting all uses cases.*
- *Raw, semi-processed, and fully aggregated data in one location.*

Unifying People

- *All users pulling from the same source and the same data.*
- *Reduced maintenance freeing up valuable technology resources.*

Unifying Technology

- *Reduced number of supporting systems for both ingest and extract of data.*
- *Greater general adoption due to the single central storage philosophy.*

All of the above indirect benefits and cost savings with a 30 – 40% direct costs savings.

