Evolution of Big Data Architectures@ Facebook

Architecture Summit, Shenzhen, August 2012

Ashish Thusoo

About Me

- Currently Co-founder/CEO of Qubole
- Ran the Data Infrastructure Team at Facebook till 2011
- Co-founded Apache Hive @ Facebook

Outline

- Big Data @ Facebook - Scope & Scale
- Evolution of Big Data Architectures @ FB
- Qubole

Big Data @ FB(2011): Scale

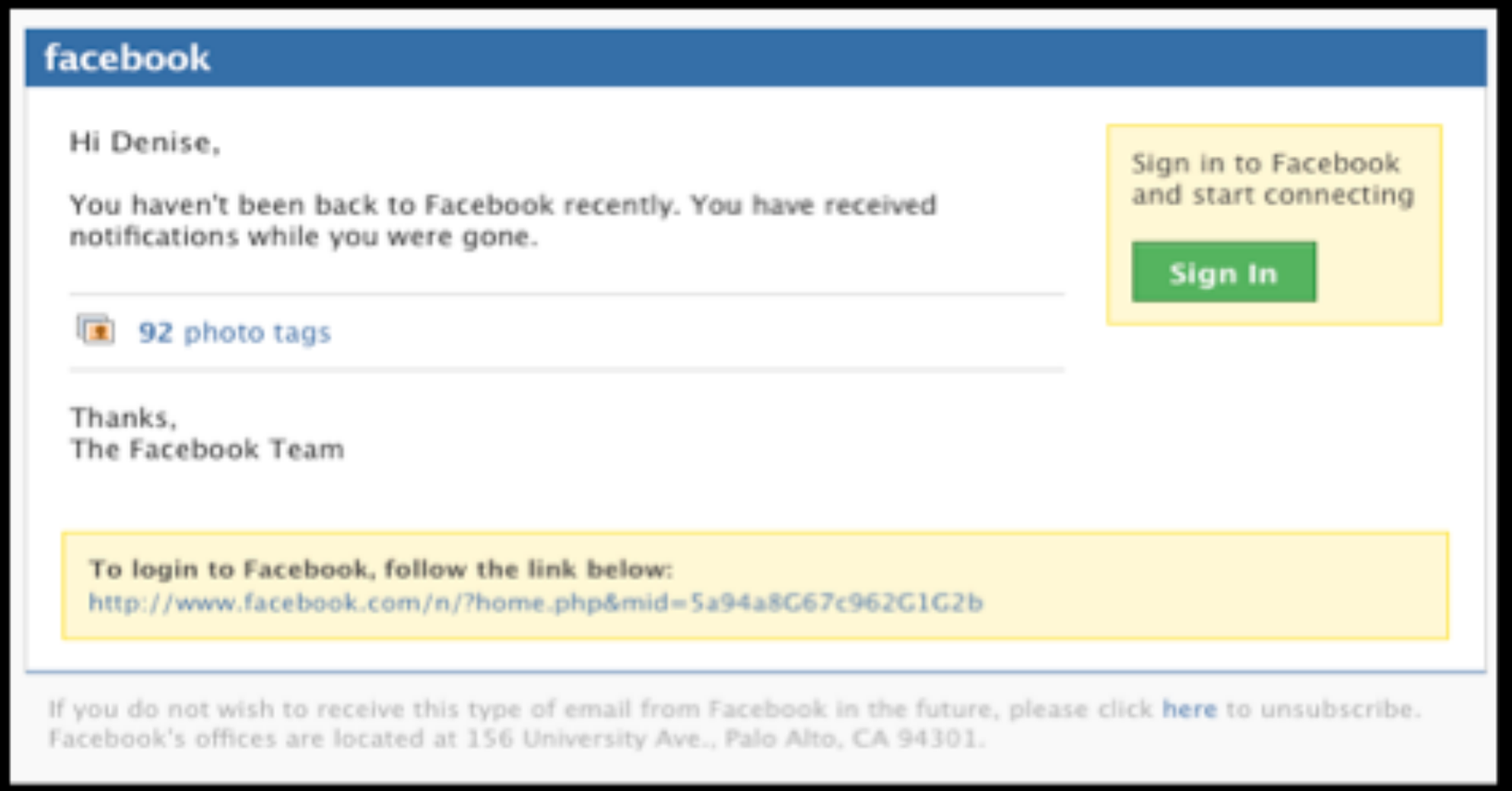
- 25 PB of compressed data ~ 150 PB of uncompressed data
- 400 TB/day (uncompressed) of new data
- 1 new job every second

Big Data @ FB: Scope

- Simple reporting
- Model generation
- Adhoc analysis + data science
- Index generation
- Many many others...

A/B Testing Email #1

A/B Testing Email #2



facebook

Hi Denise,

You haven't been back to Facebook recently. You have received notifications while you were gone.

92 photo tags

Thanks,
The Facebook Team

Sign in to Facebook and start connecting

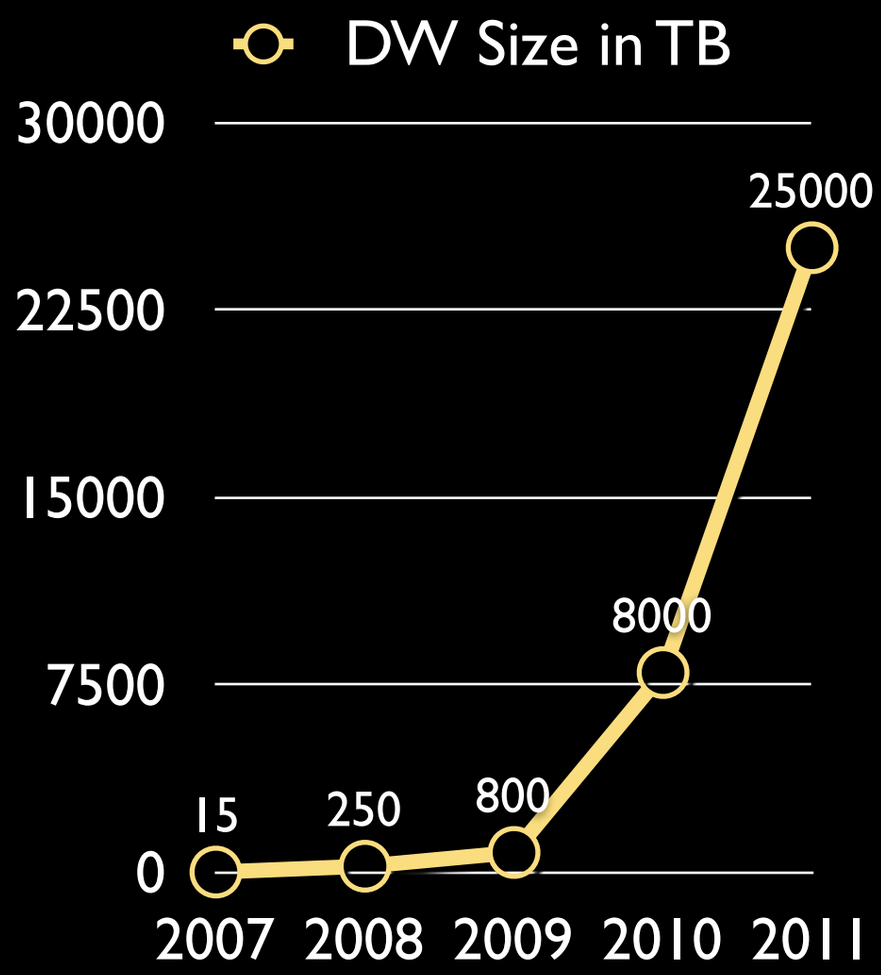
Sign In

To login to Facebook, follow the link below:
http://www.facebook.com/n/?home.php&mid=5a94a8G67c962G1G2b

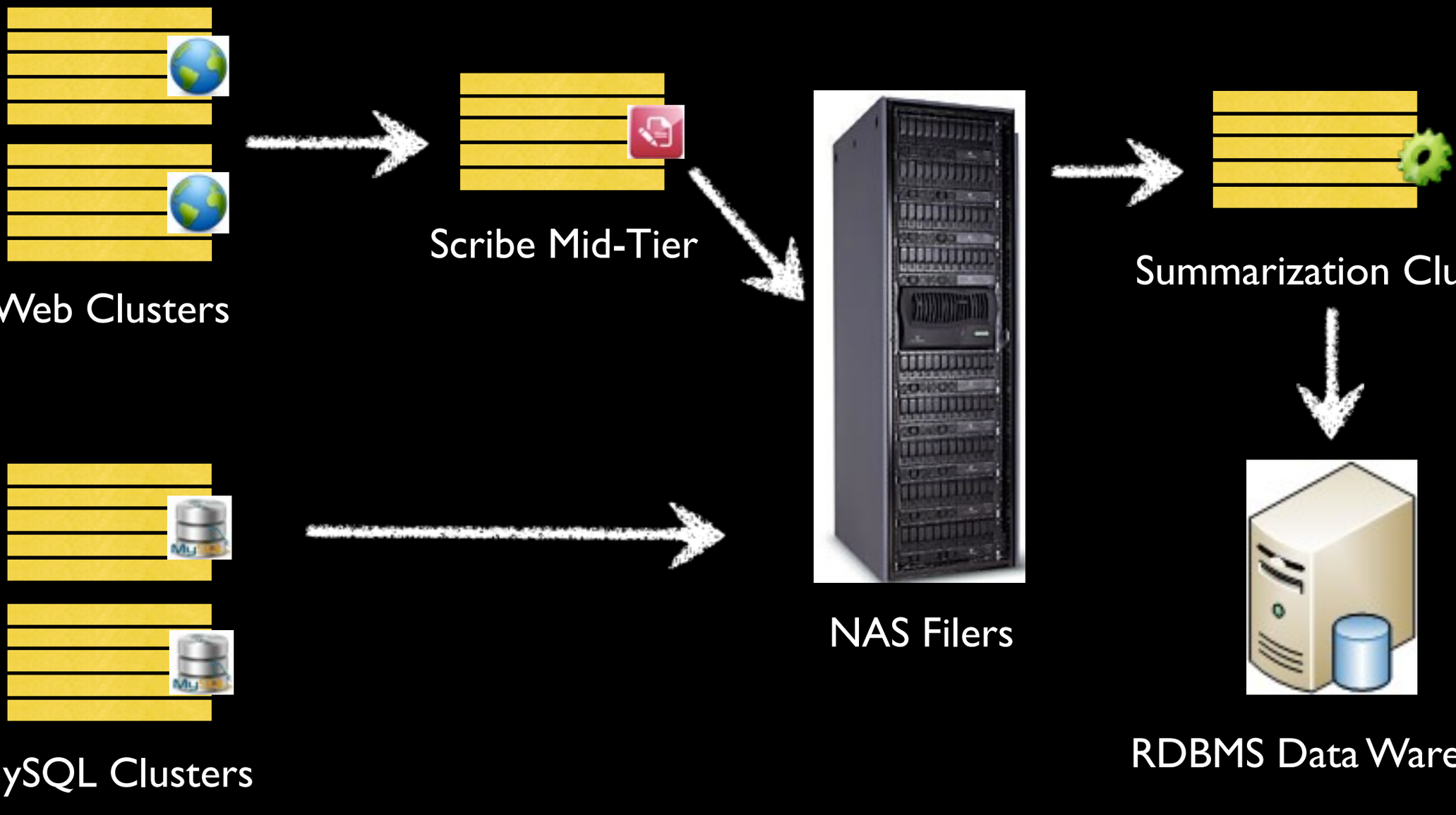
If you do not wish to receive this type of email from Facebook in the future, please click here to unsubscribe.
Facebook's offices are located at 156 University Ave., Palo Alto, CA 94301.

A/B Testing Email #2 is 3x Better

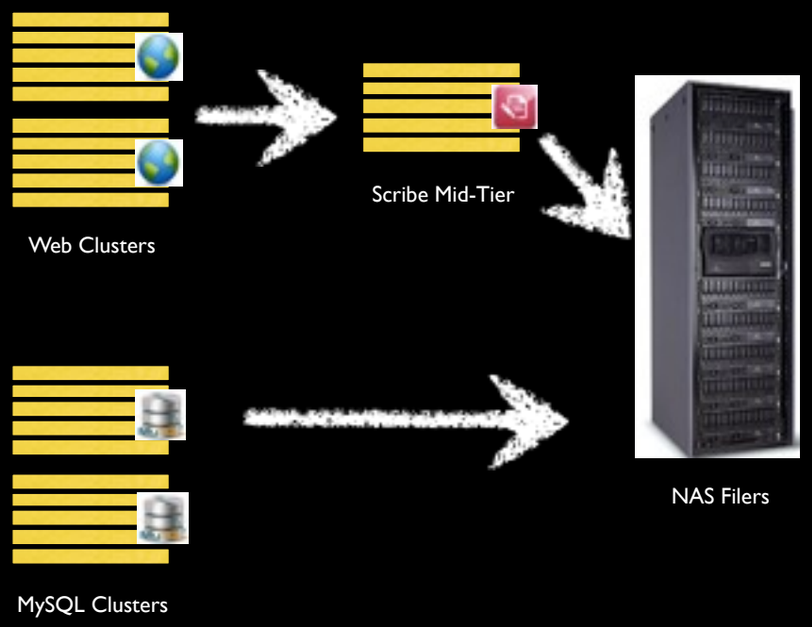
Evolution: 2007-2011



2007: Traditional EDW



2007: Pain Points



- compute close to storage
(early map/reduce)

Summarization Cluster

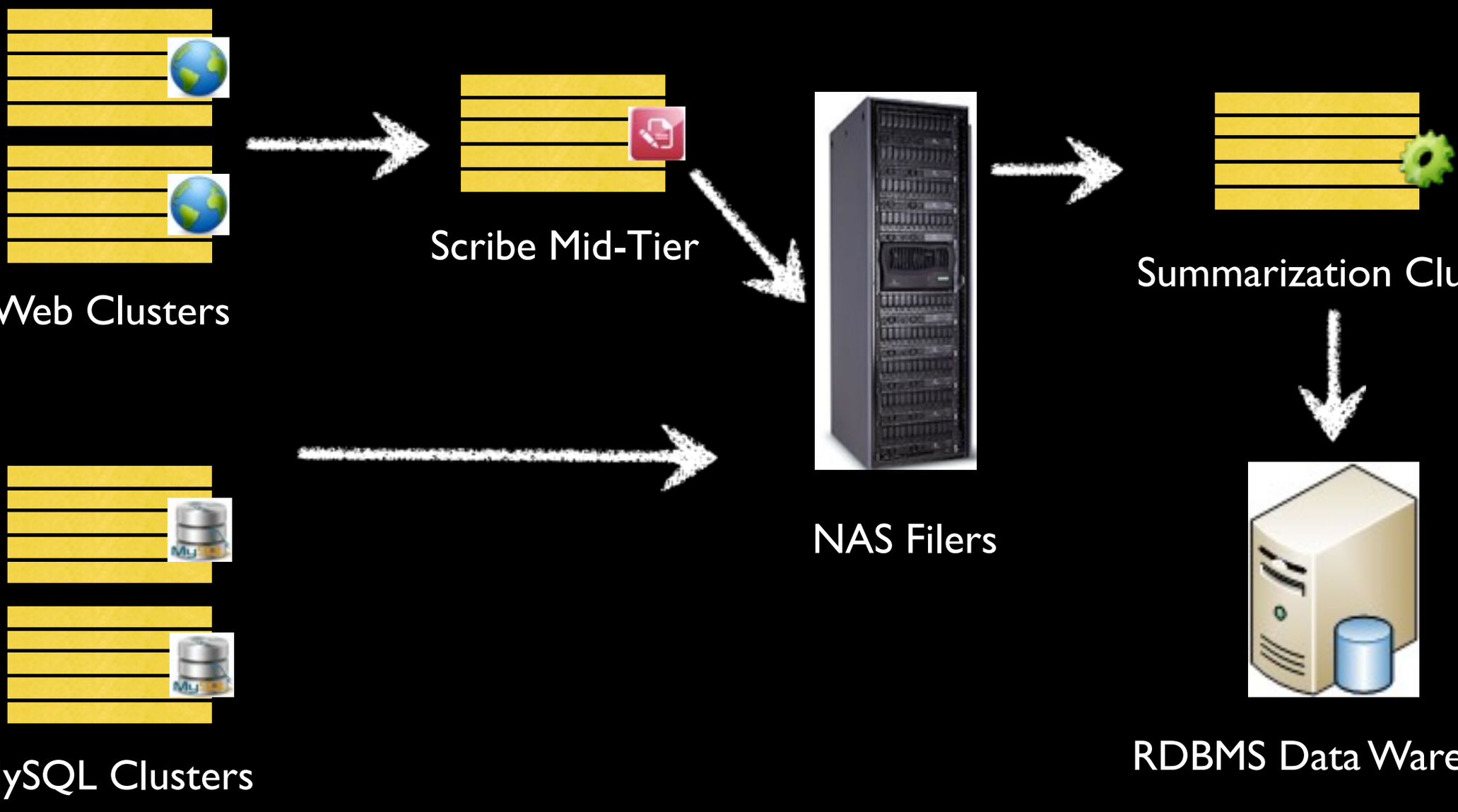
- daily ETL > 24 hours
- Lots of tuning/indexes etc.
- Lots of hardware planning

RDBMS Data Warehouse

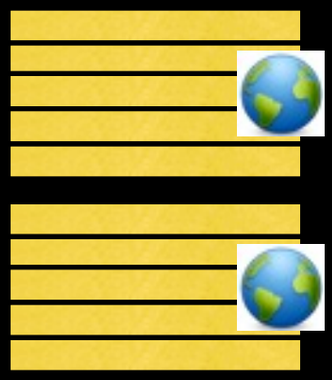
2007: Limitations

- Most use cases were in business metrics - data science, model building etc. not possible
- Only summary data was stored online - details archived away

2008: Move to Hadoop



2008: Move to Hadoop

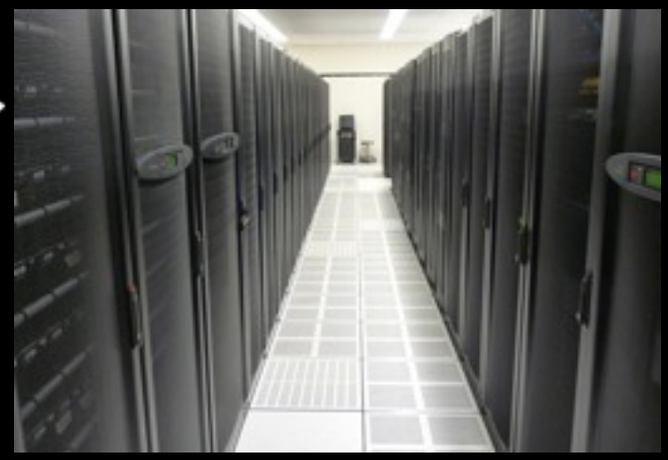


Web Clusters

Scribe Mid-Tier

NAS Filers

Batch copier/ loaders



Hadoop/Hive Data Warehouse

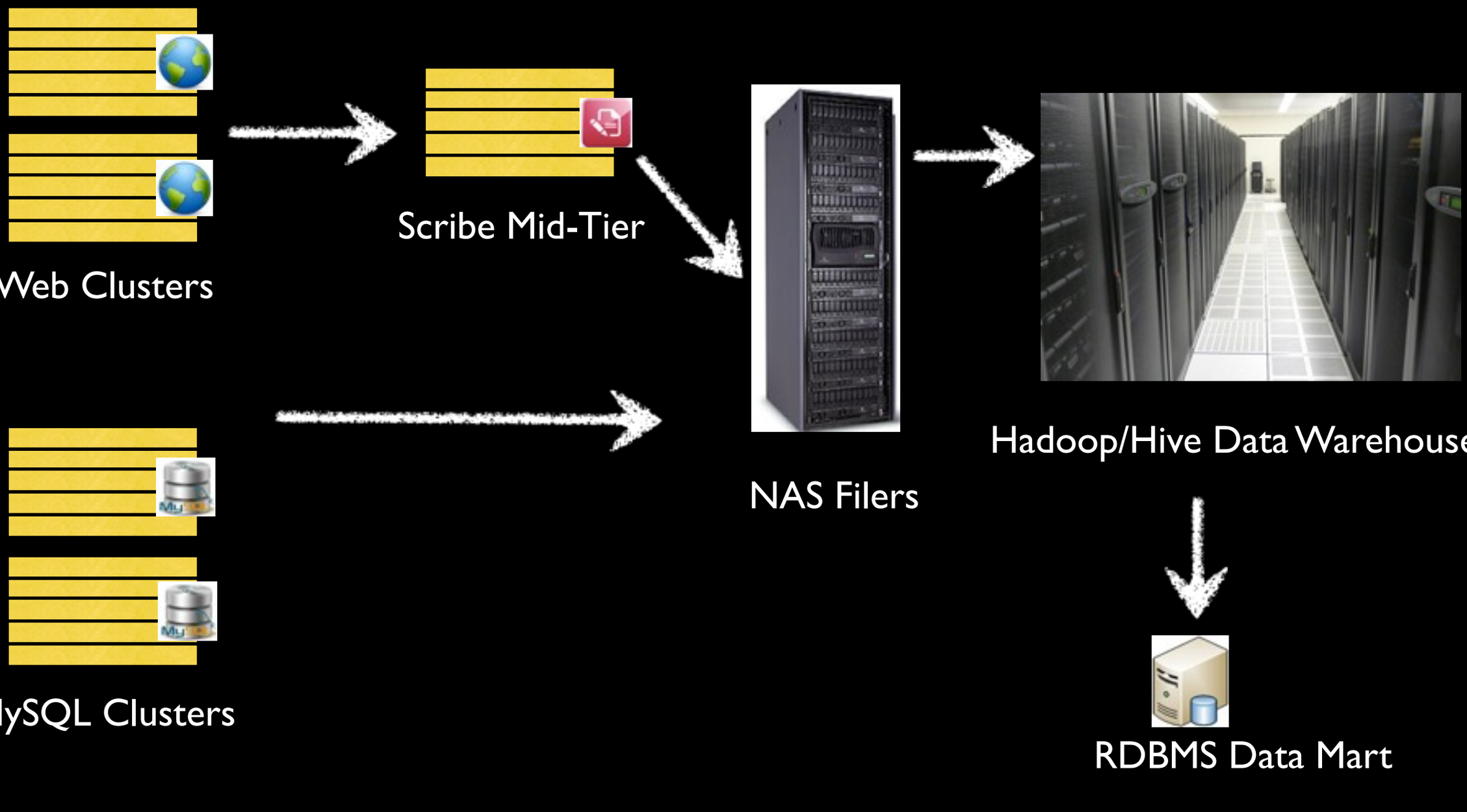
MySQL Clusters

RDBMS Data Mart

2008: Immediate Pros

- Data science at scale became possible
- For the first time all of the instrumented data could be held online
- Use cases expanded

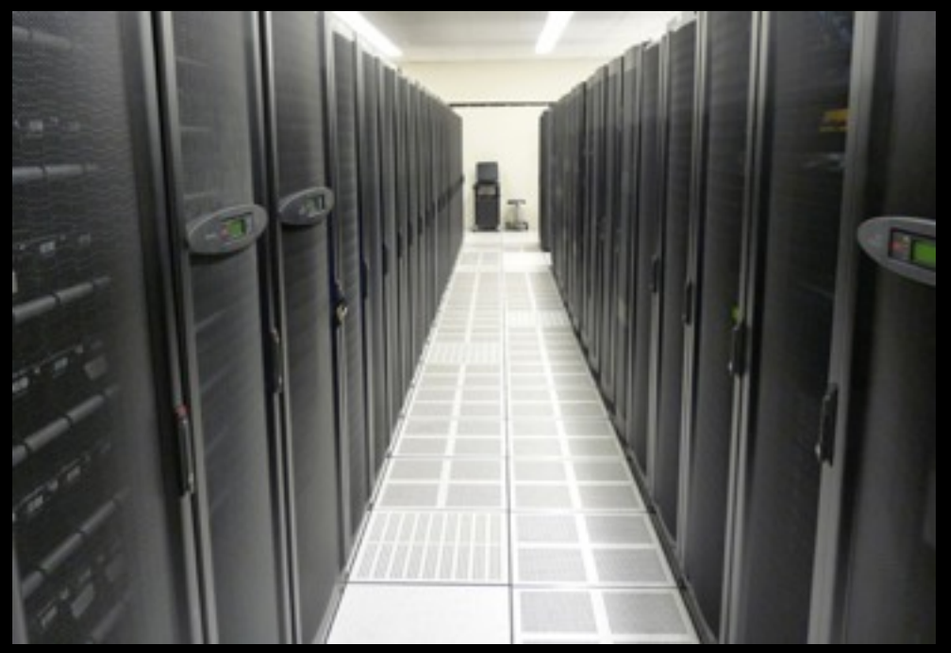
2009: Democratizing Data



2009: Democratizing Data

Databee & Chronos: Data Pipeline Framework

Nectar: instrumentation & schema aware data collection



HiPal: Adhoc Queries + Data Discovery

Hadoop/Hive Data Warehouse

Scrapes: Configuration Driven

2009: Democratizing Data(Nectar)

- Typical Nectar Pipeline
  - Simple schema evolution built in
  - json encoded short term data
  - decomposing json for long term storage

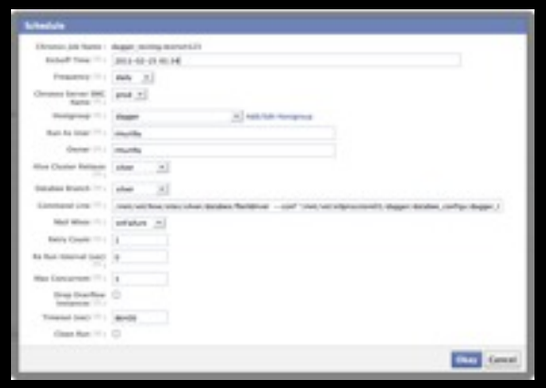
```
// This event has application name 'mobilelog' and app event type
// 'email_mms_upload'.
// NOTE: Make sure you use only one application name per new application.
// Also, app event type should not have any special characters or spaces,
// use underscores instead.  $sampling_rate is the scribe sampling rate and
// has a value between 0 and 100 - sampling is on userid

NectarAppSpecificEvent('mobilelog', 'email_mms_upload', $sampling_rate)

  ->addToOdsKeys(array('k1', 'k2'))        // if you want to add additional
                                           // ODS keys
  ->setODSSamplingRate(1)                  // default is 10000, meaning 1
                                           // in 10000 events is sent to ODS
  ->addToAppSection(array("key" => "val")) // can add different key value
                                           // pairs for different eventtypes
  ->log();                                 // need to explicit log app
                                           // specific events
```

2009: Democratizing Data (Tools)

- HiPal - data discovery and query authoring
- Charting and dashboard generation tools

2009: Democratizing Data (Tools)

- Databee: Workflow language
- Chronos: Scheduling tool



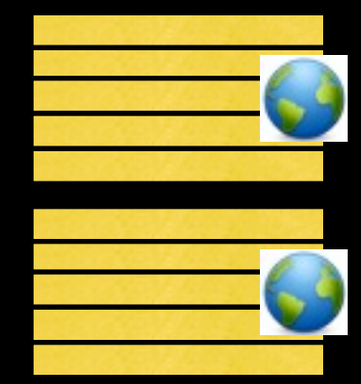
2009: Cons of Democratization

- Isolation to protect against Bad Jobs
- Fair sharing of the cluster - what is a high priority job and how to enforce it

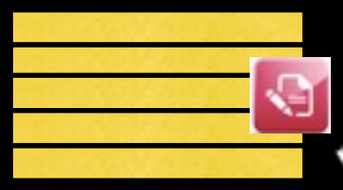
2010: Controlling Chaos

- Isolation
- Reducing operational overhead
- Better resource utilization
- Measurement, ownership, accountability

2010: Isolation

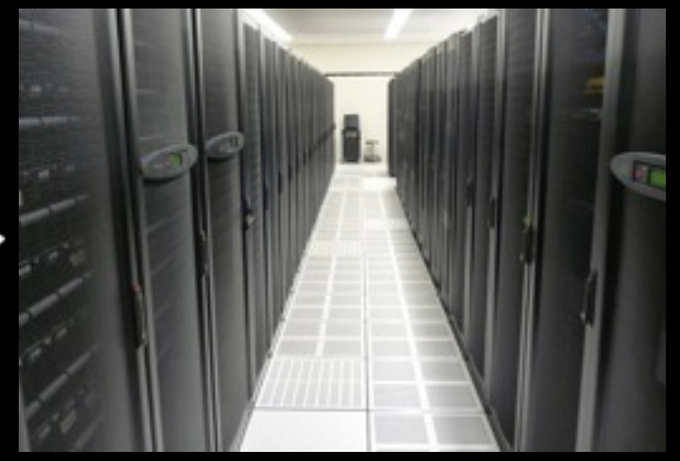


Web Clusters



Scribe Mid-Tier

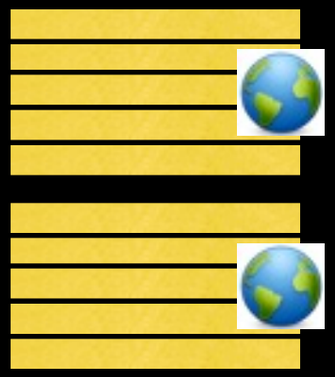
NAS Filers



Hadoop/Hive Data Warehouse

MySQL Clusters

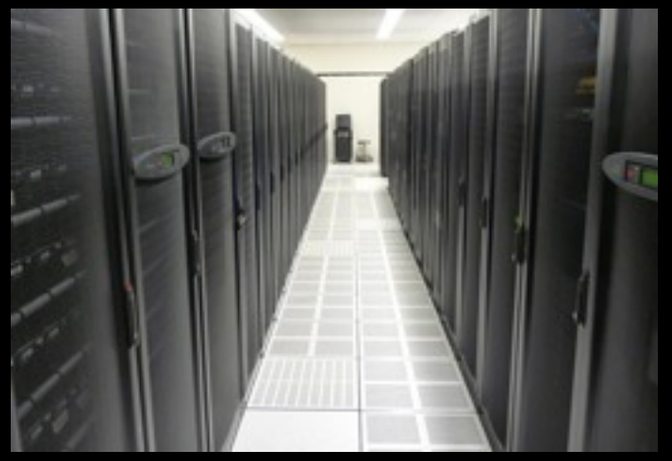
2010: Isolation



Web Clusters

Scribe Mid-Tier

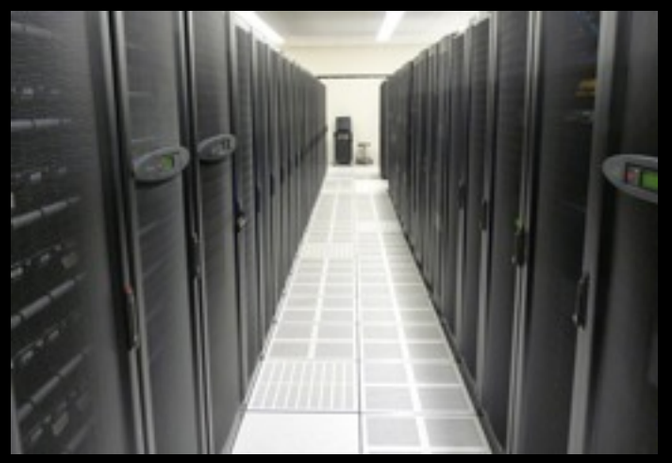
NAS Filers



Platinum Warehouse

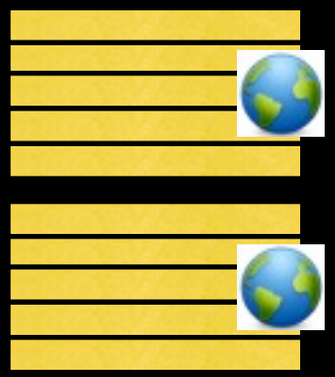
Hive Replication

MySQL Clusters

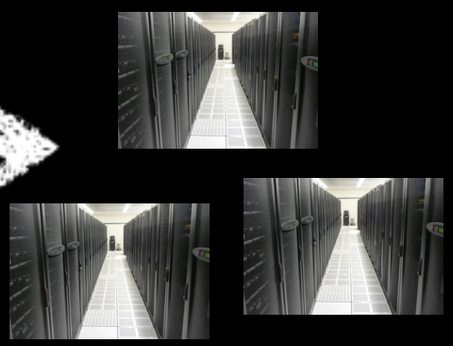


Silver Warehouse

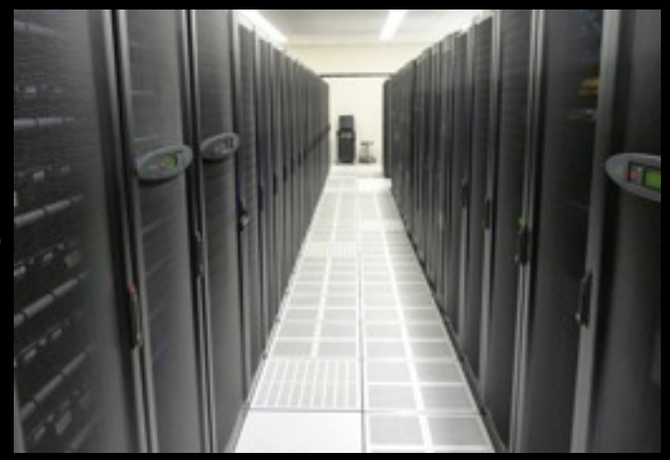
2010: Ops Efficiency



Web Clusters

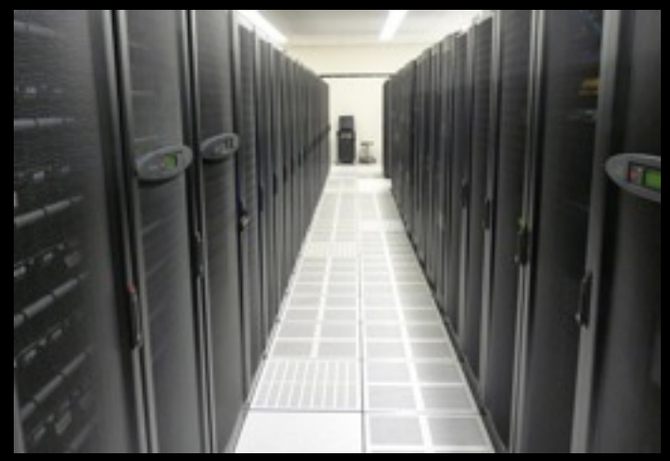


Scribe HDFS



Platinum Warehouse

Hive Replication



Silver Warehouse

MySQL Clusters

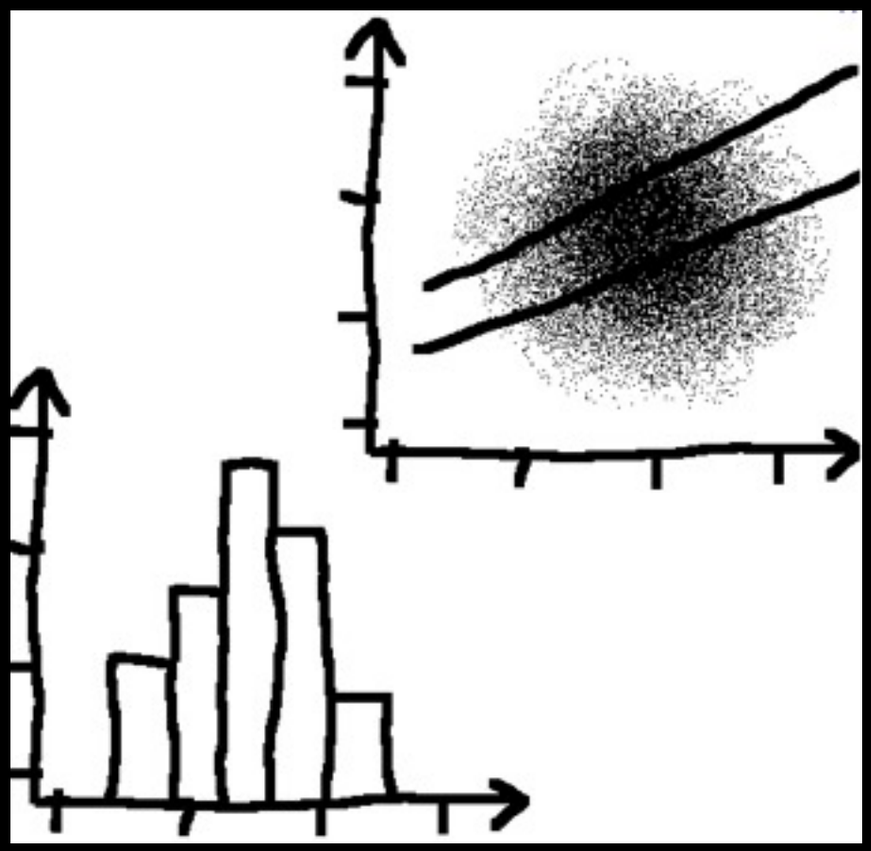
ptail: parallel tail on hdfs

near real time data consumers

2010: Resource Utilization (Disk)

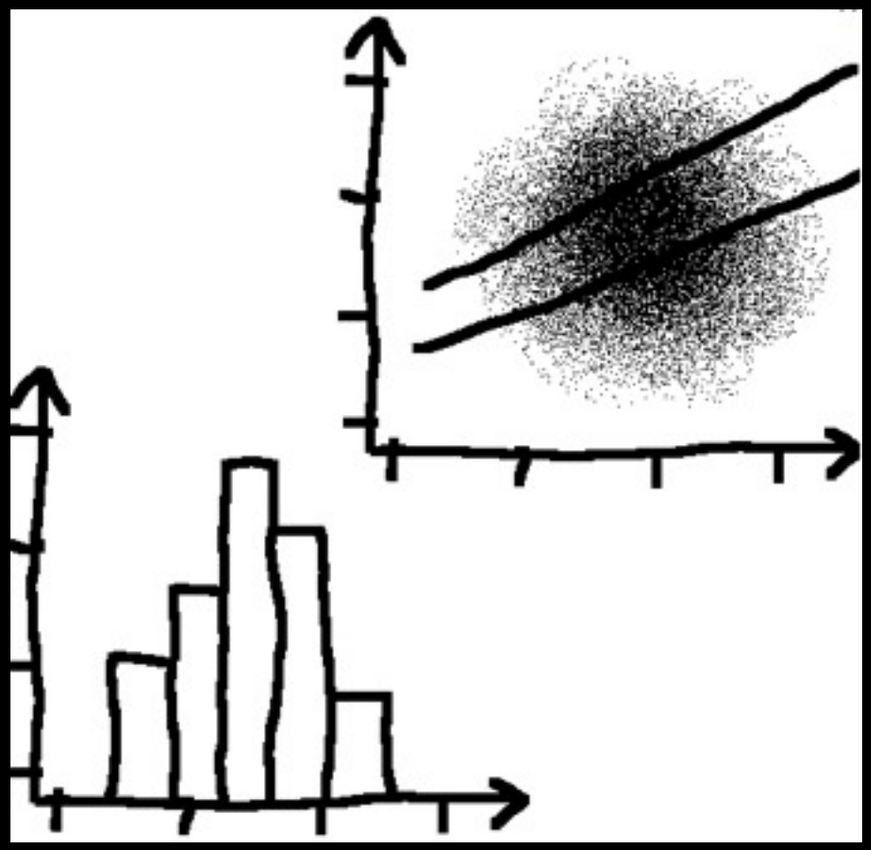
- HDFS-RAID: from 3 replicas to 2.2 replicas

- RCFile: Row columnar format for compressing Hive tables



2010: Resource Utilization (CPU)

- Continuous copier/ loaders
- Incremental scrapes
- Hive optimizations to save CPU



2010: Monitoring(SLAs)

- Per job statistics rolled up to owner/group/team
- Expected time of arrival vs Actual time of arrival of data
- Simple data quality metrics

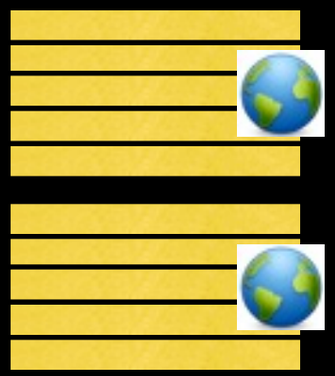
2011: New Requirements

- More real time requirements for aggregations
- Optimizing resource utilization

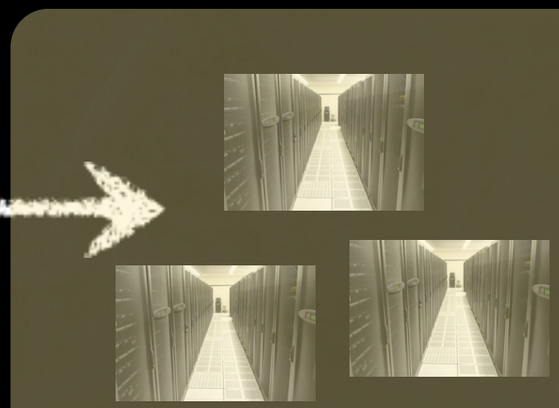
2011: Beyond Hadoop

- Puma for real time analytics
- Peregrine for simple and fast queries

2011: Puma



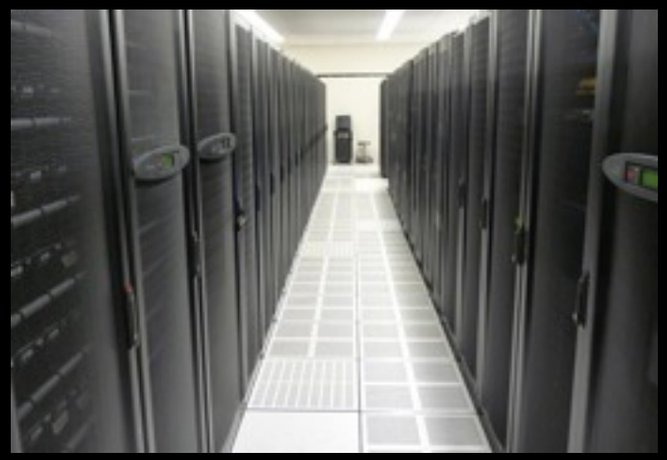
Web Clusters



Scribe HDFS

ptail: parallel tail on hdfs

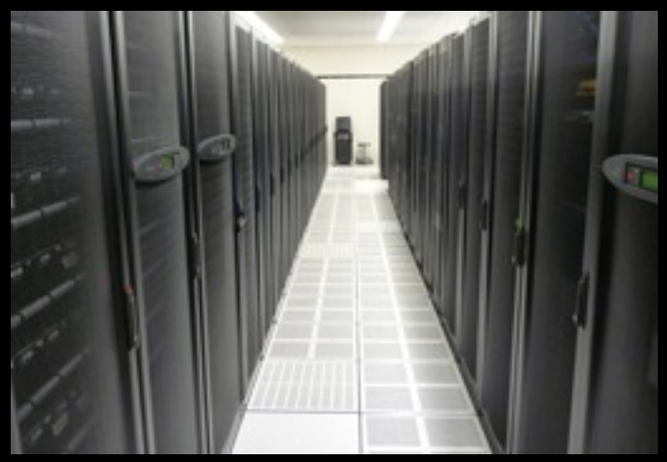
near real time data consumers



Platinum Warehouse

Hive Replication

MySQL Clusters

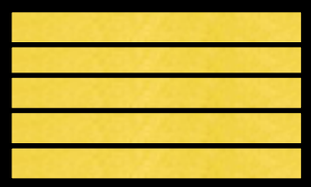


Silver Warehouse

2011: Puma

Scribe HDFS

ptail: parallel tail on hdfs



Puma Clusters

Hbase Cluster

Some takeaways

- Operating and optimizing Data Infrastructure is a hard problem
- Lots of components from log collection, storage, compute, query processing, tools and interfaces
- Lots of choices within each part of the stack

Qubole

- Mission:
  - Data Infrastructure in the Cloud made Easy, Fast and Reliable
  - We take care of operating and optimizing this infrastructure so that you can focus on your data, analysis, algorithms and building your data apps

Qubole - Information

- Early Trial(by invitation):
  - www.qubole.com
- Come talk to us to join a small and passionate team
  - jobs@qubole.com
- Follow us on twitter/facebook/linkedin