

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

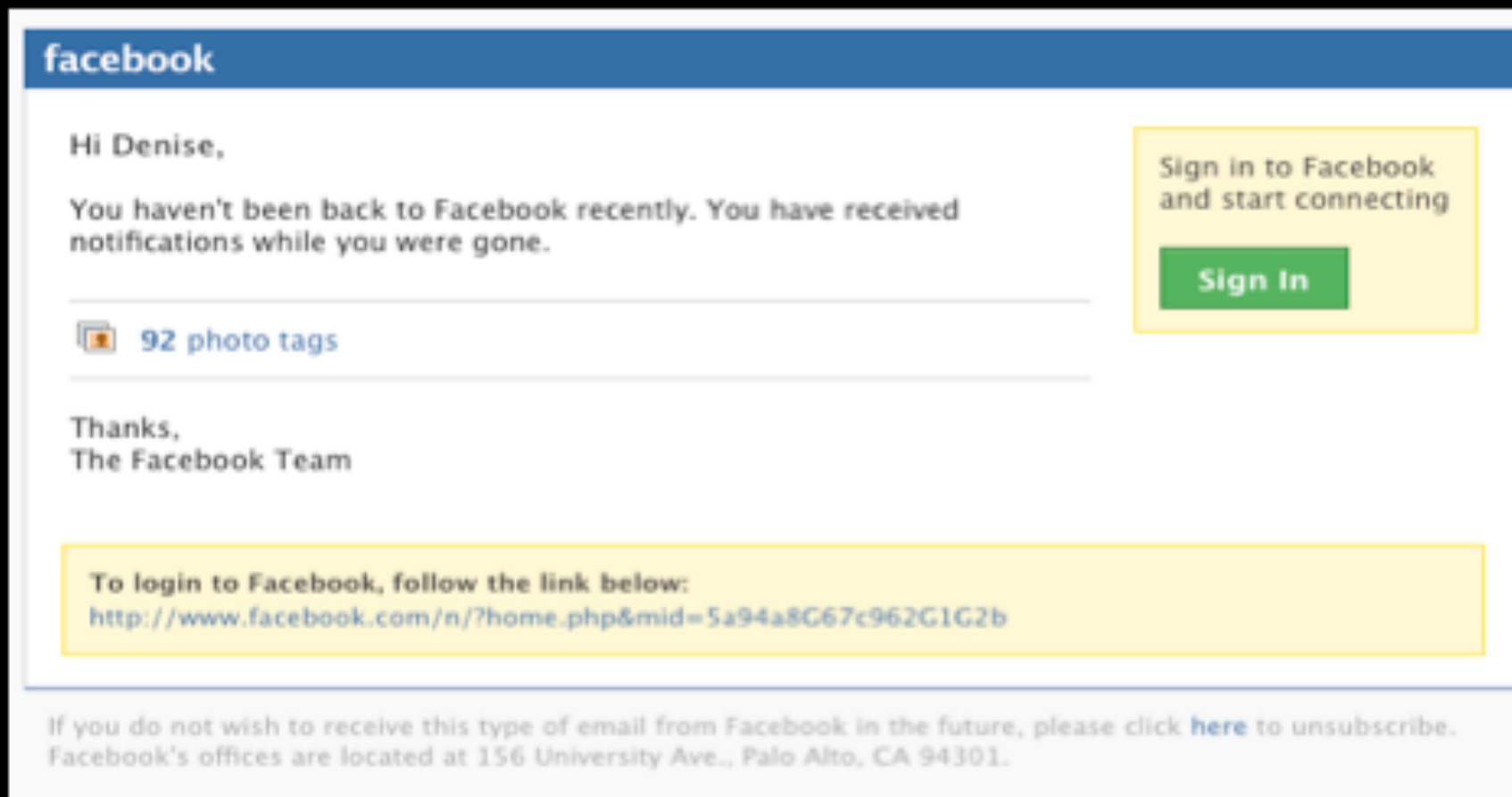
Big Data @ FB: Scope

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

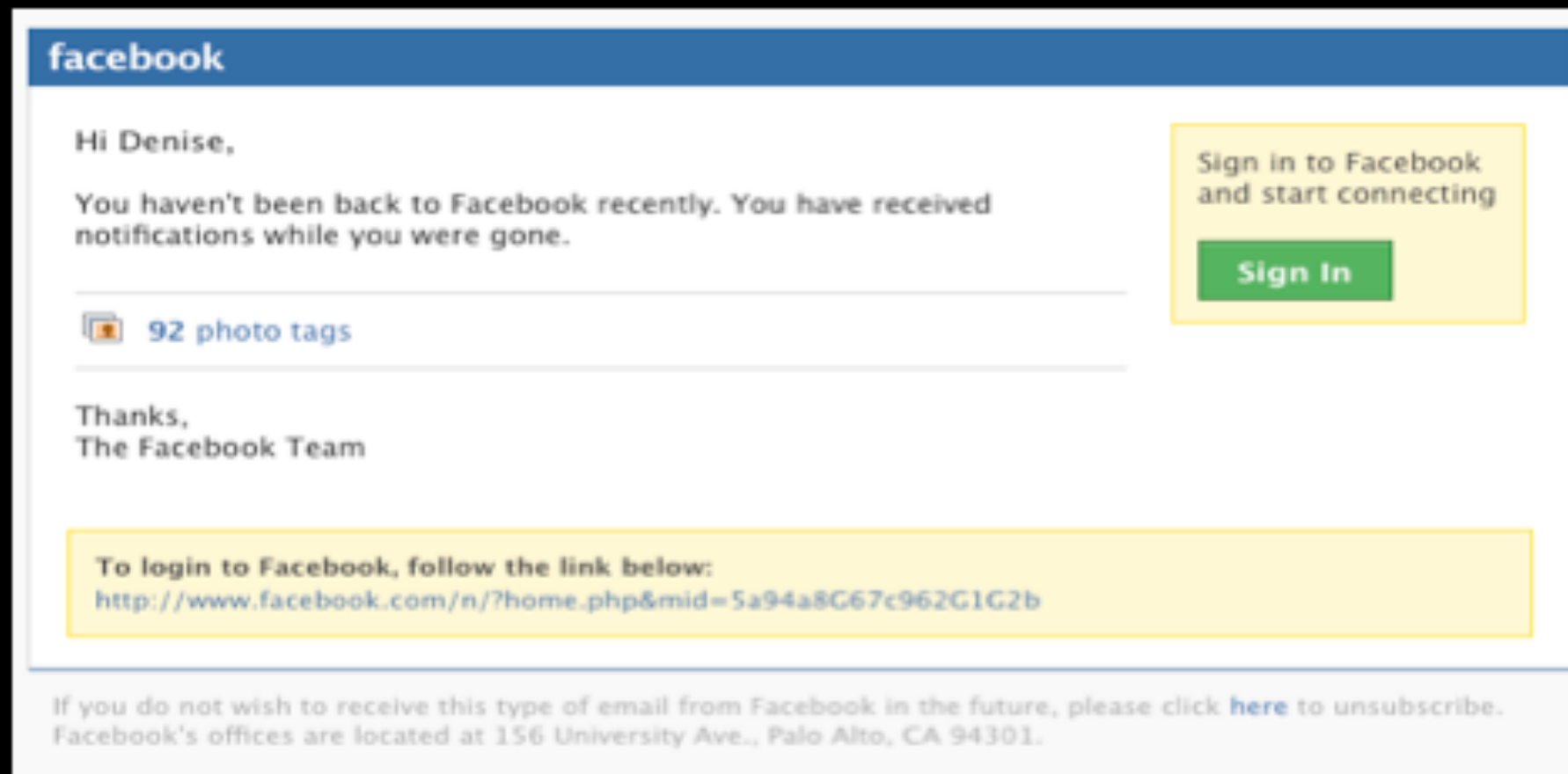
A/B Testing Email #1



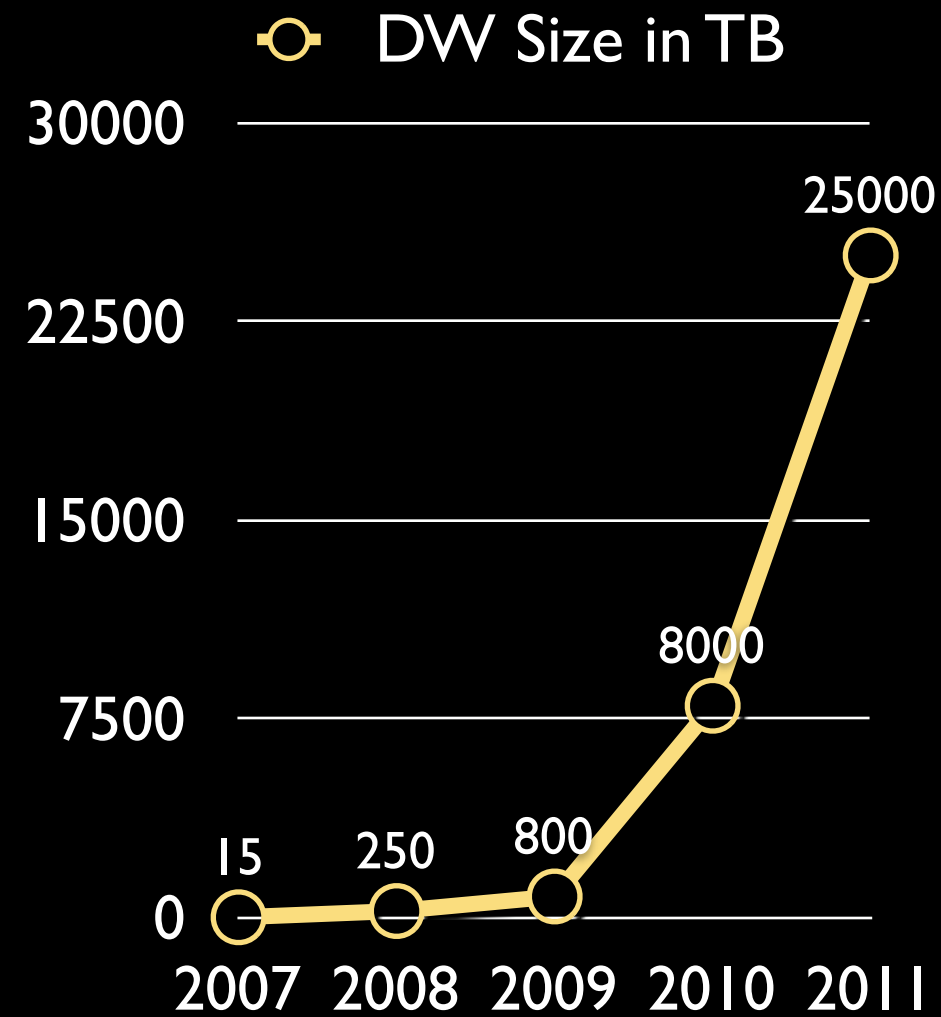
A/B Testing Email #2



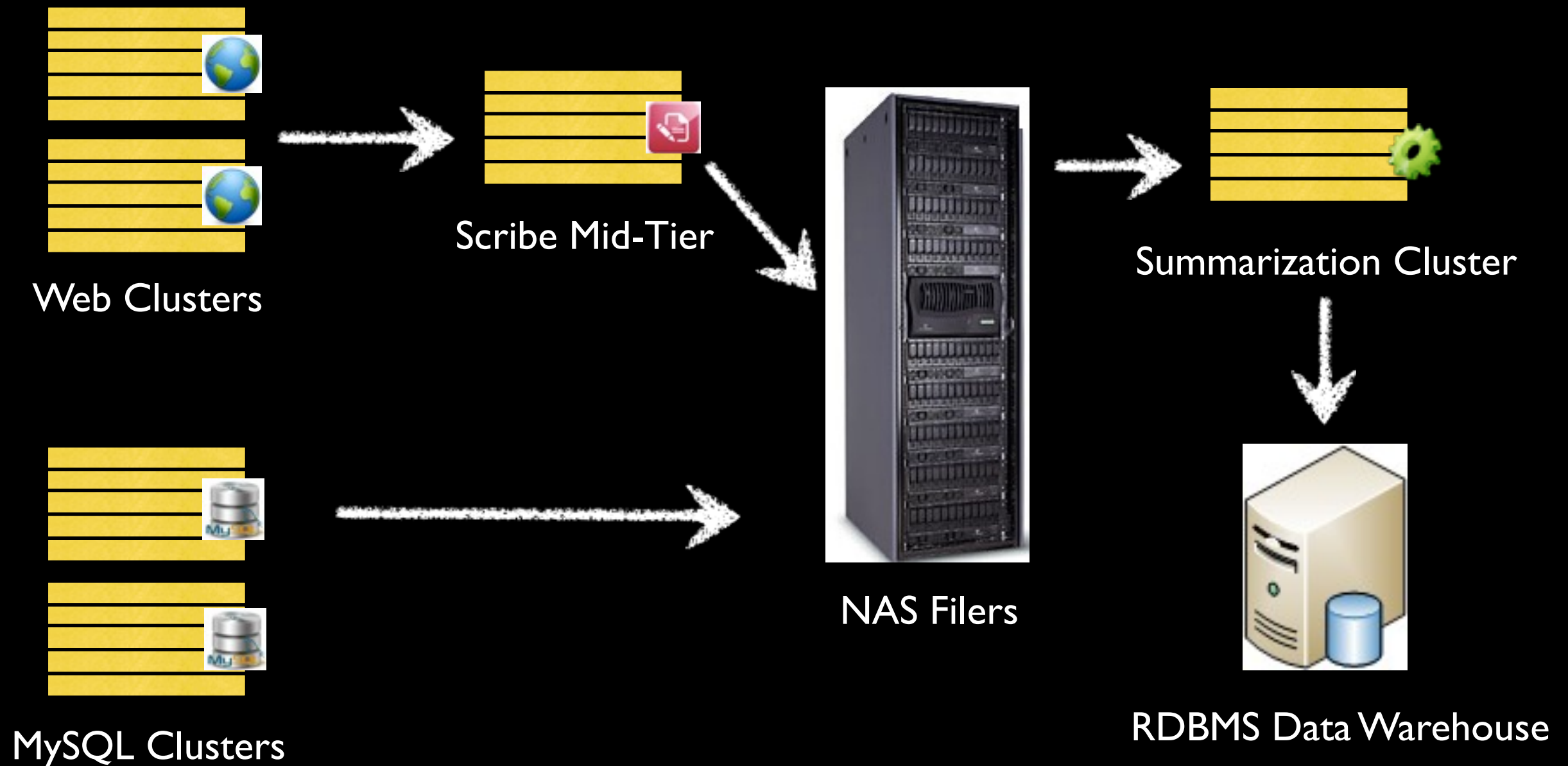
A/B Testing Email #2 is 3x Better



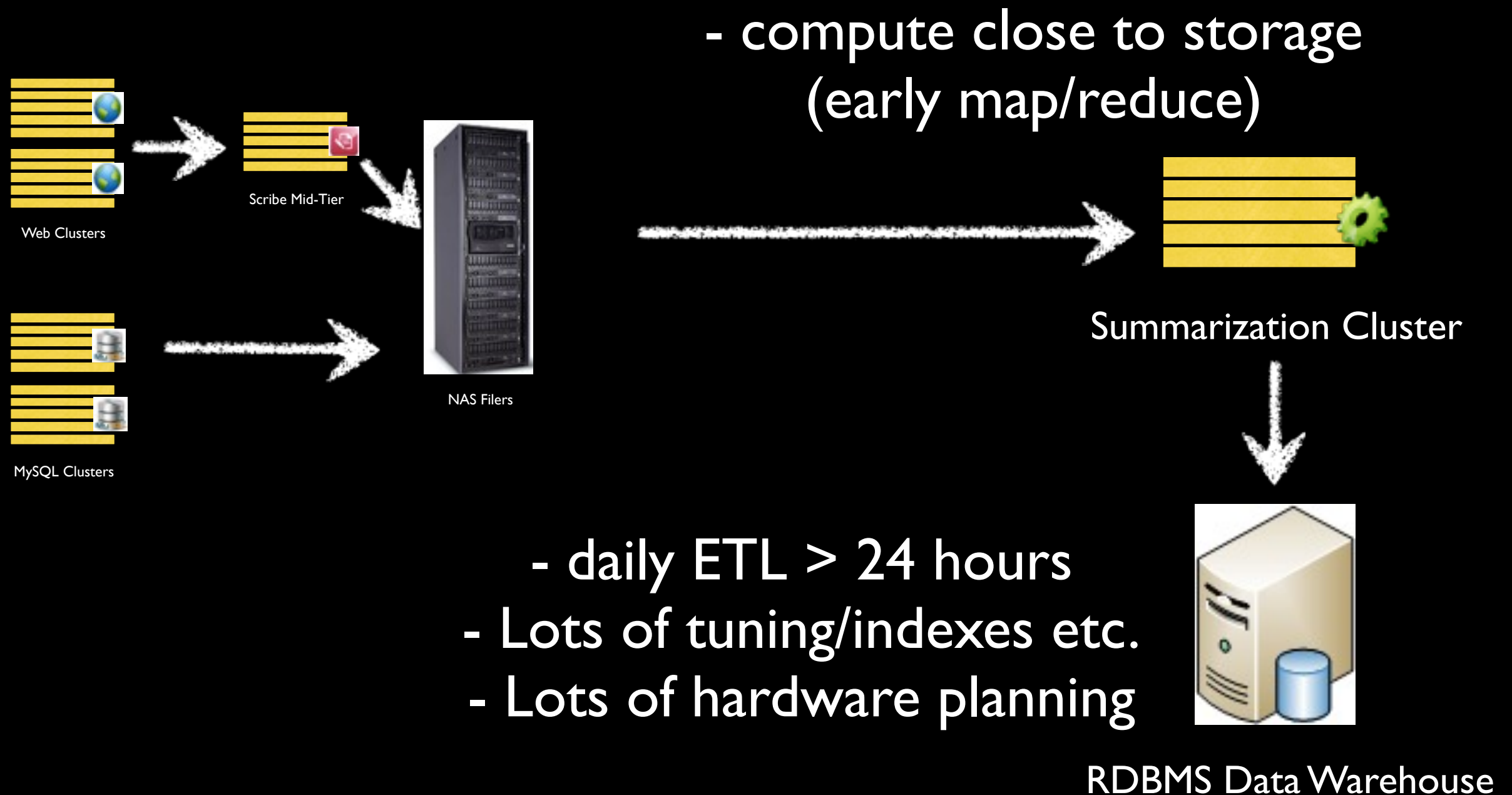
Evolution: 2007-2011



2007: Traditional EDW



2007: Pain Points

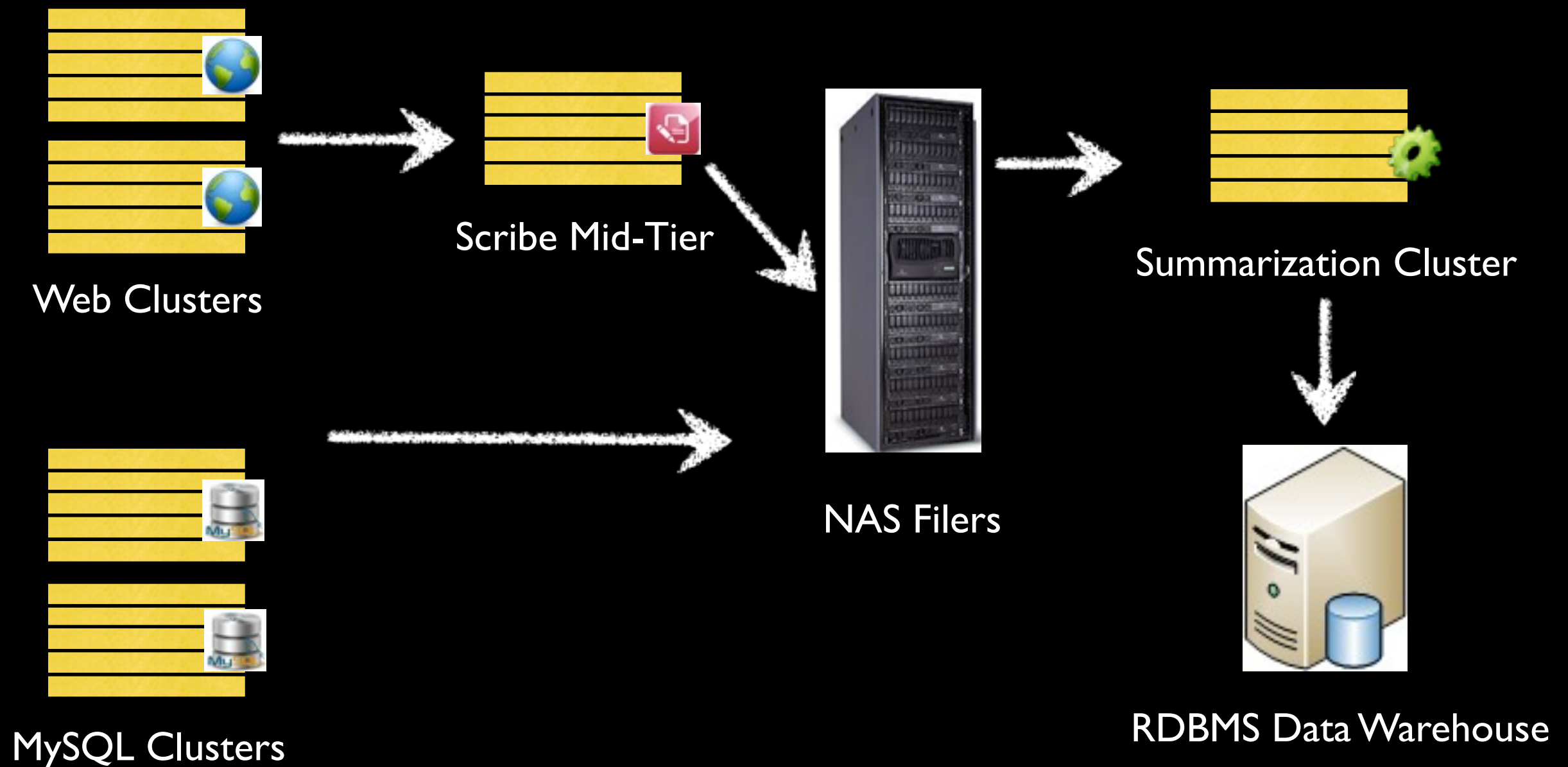


2007: Limitations

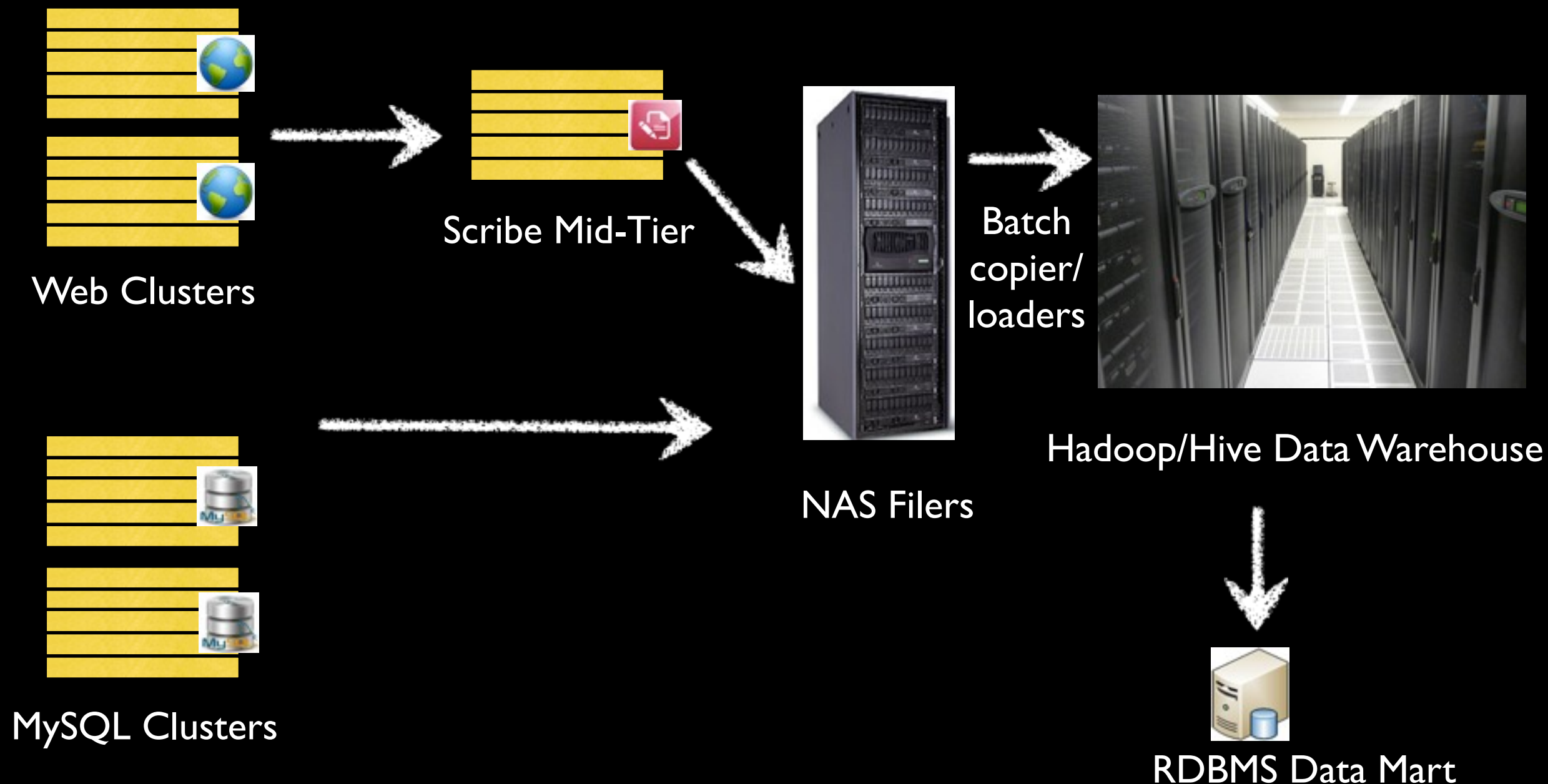
- Most use cases were in business metrics - data science, model building etc. not possible
- Only summary data was stored online - details archived away



2008: Move to Hadoop



2008: Move to Hadoop

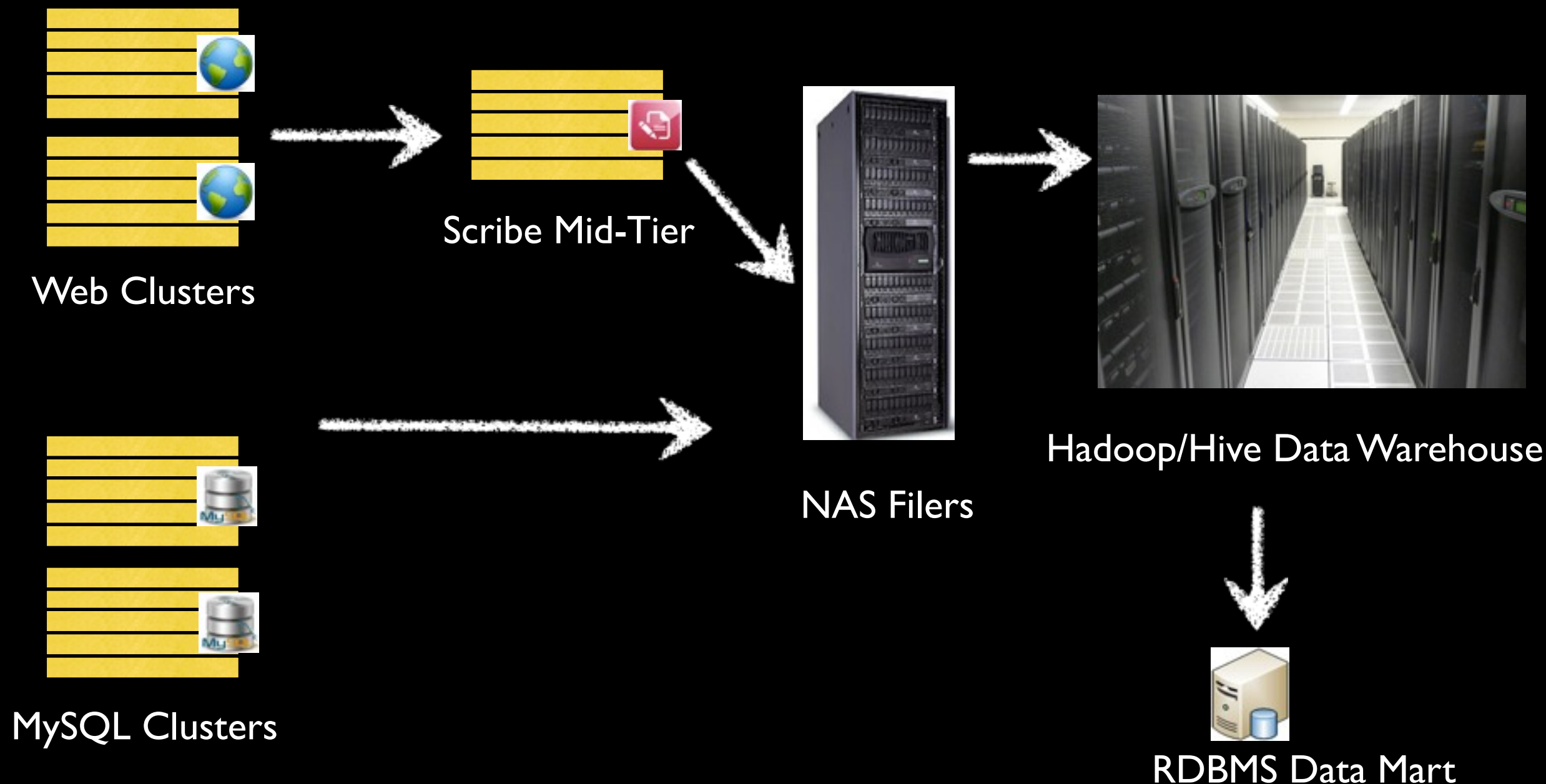


2008: Immediate Pros

- 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
```

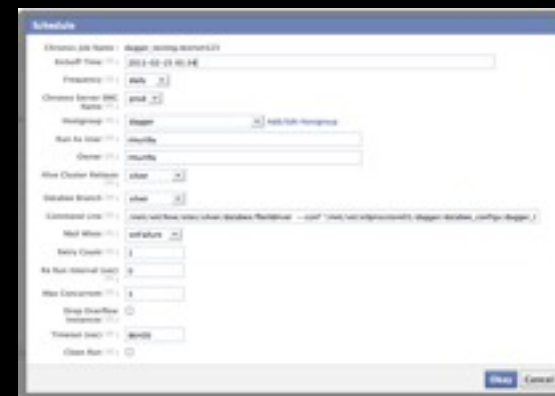
2009: Democratizing Data (Tools)

- 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

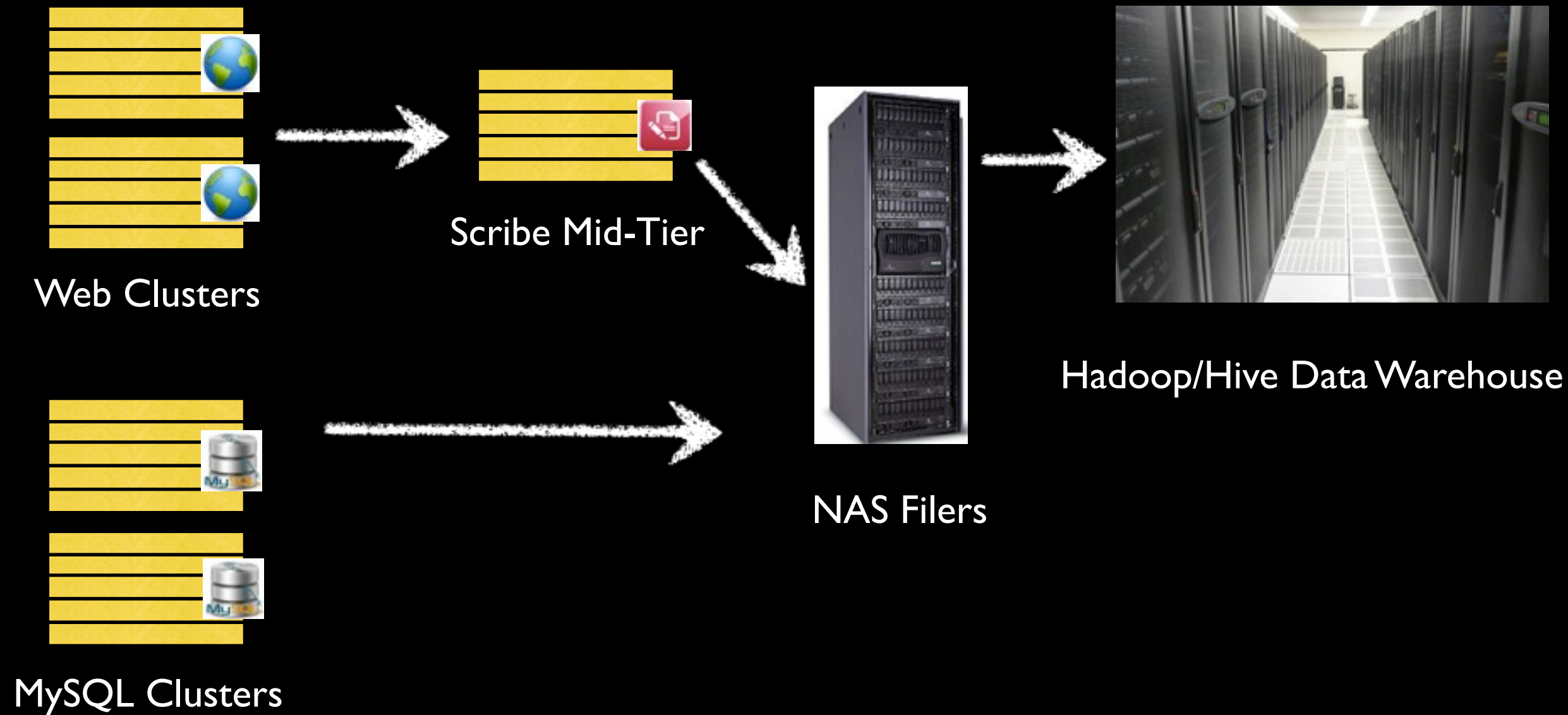
- Isolation to protect against Bad Jobs
- Fair sharing of the cluster - what is a high priority job and how to enforce it



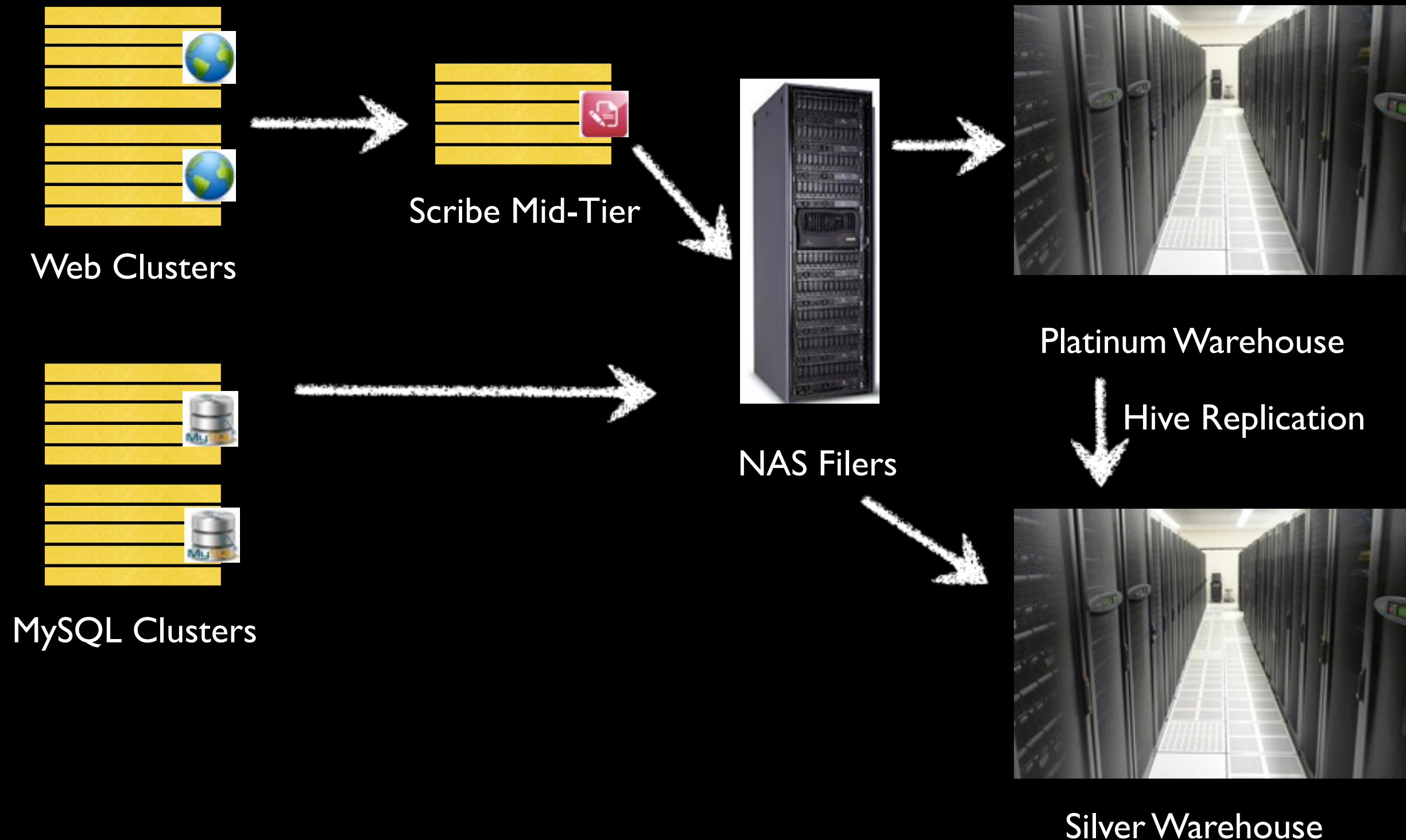
2010: Controlling Chaos

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

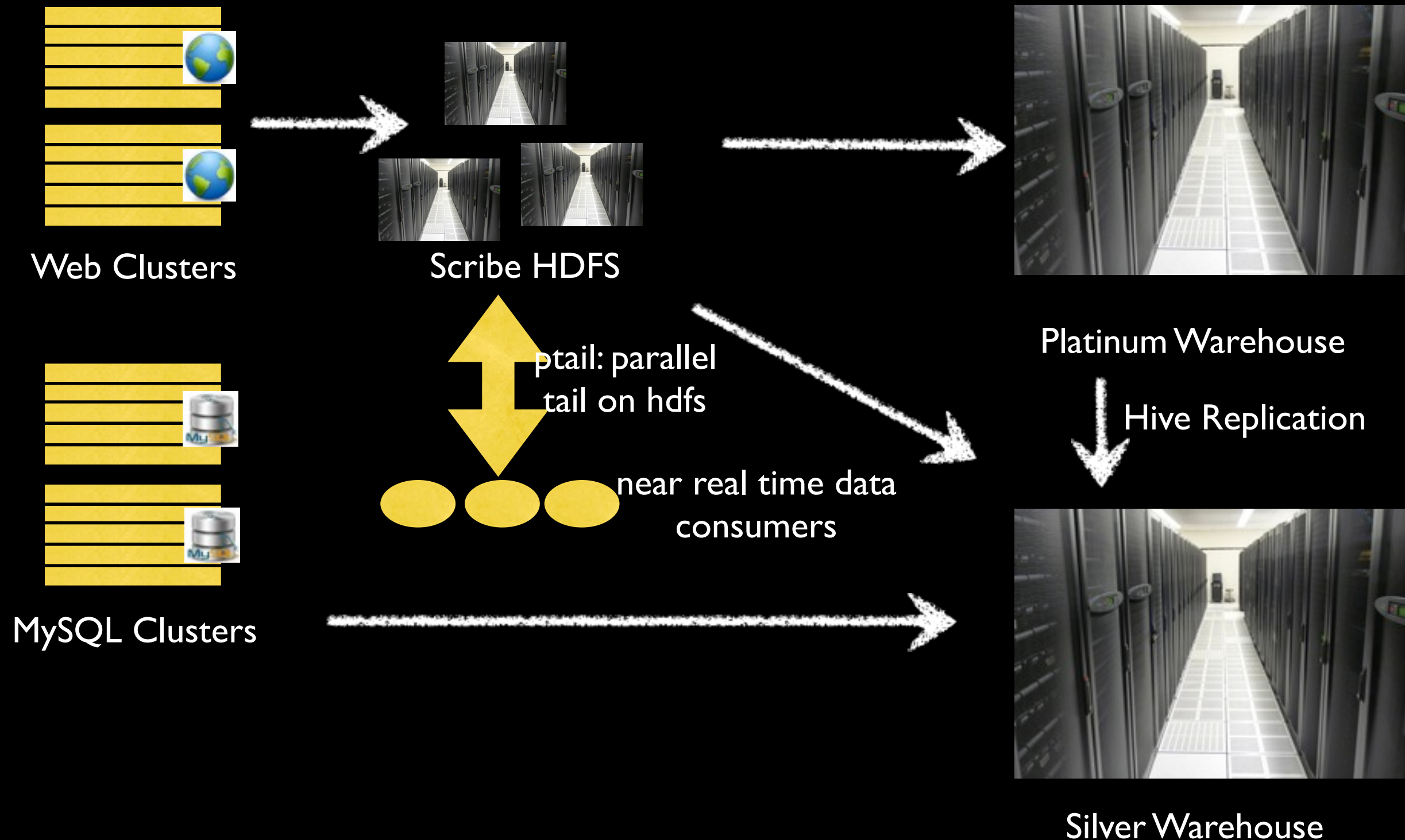
2010: Isolation



2010: Isolation

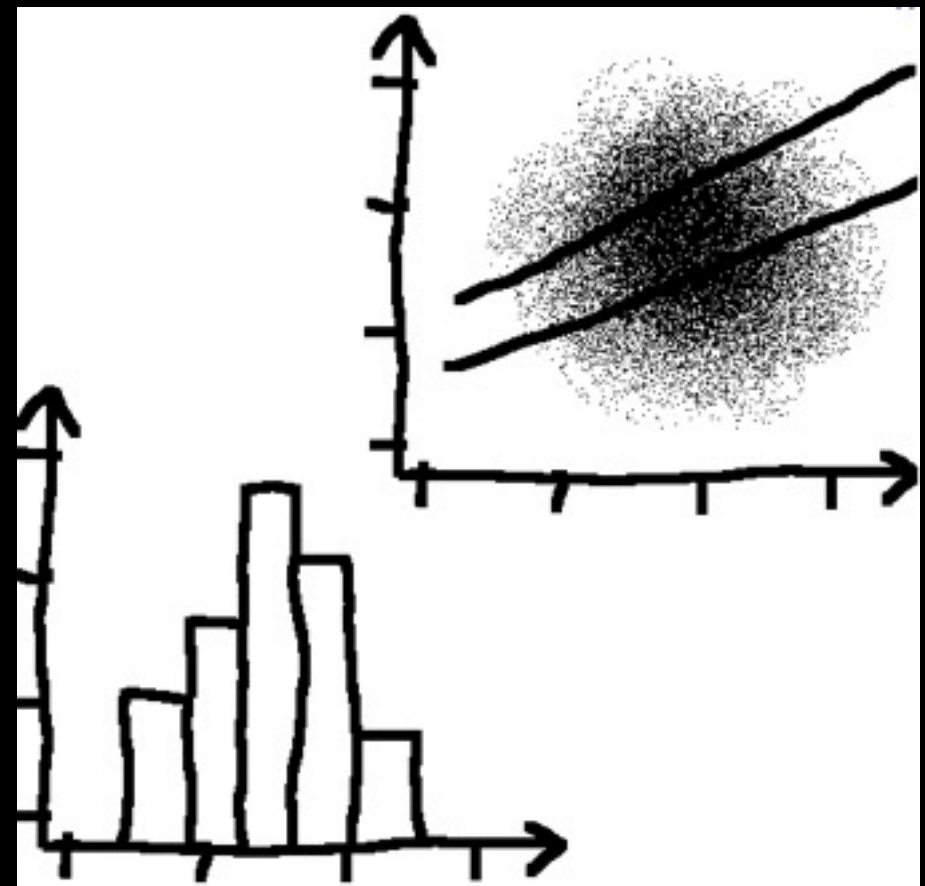


2010: Ops Efficiency



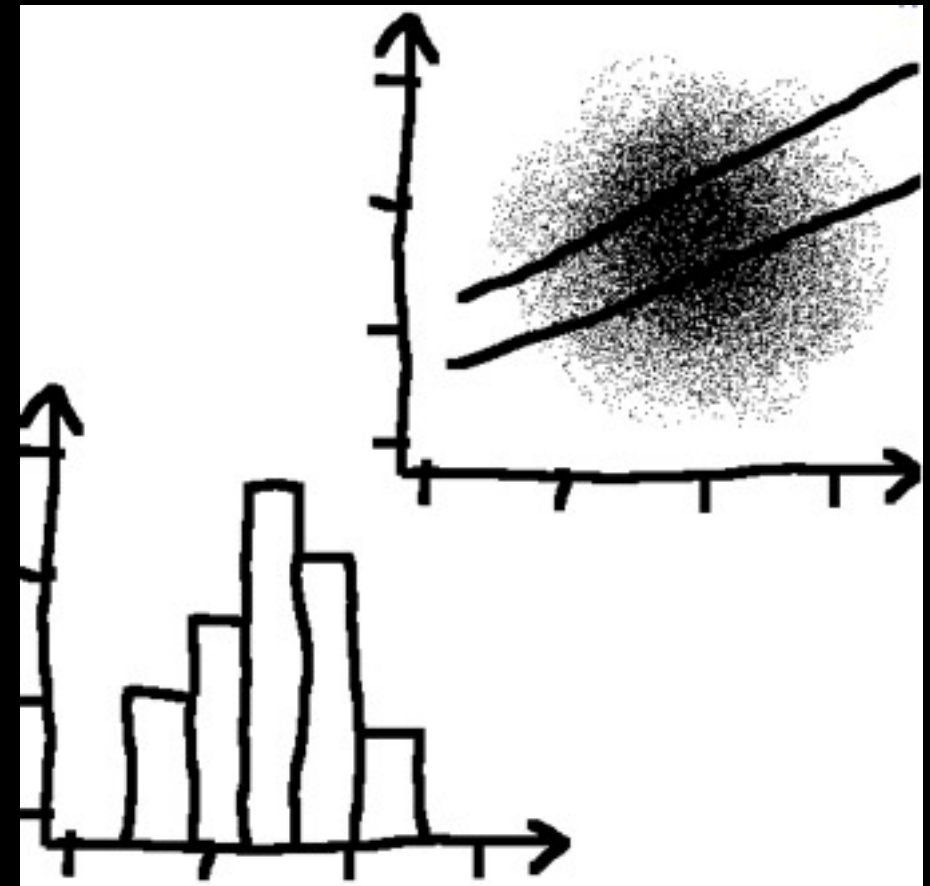
2010: Resource Utilization (Disk)

- 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



A photograph of an airport arrivals board. The board has a yellow header with the word 'Arrivals' and a small logo. Below the header, there are four columns of flight information. Each row represents a flight, with columns for flight number, origin, expected arrival time, and status. The status column shows 'On Time' for most flights. At the bottom of the board, there is a digital clock displaying '12:19'.

Flight	Origin	Expected	Status
AA 111	Los Angeles	12:00	On Time
AA 112	Los Angeles	12:05	On Time
AA 113	Los Angeles	12:10	On Time
AA 114	Los Angeles	12:15	On Time
AA 115	Los Angeles	12:20	On Time
AA 116	Los Angeles	12:25	On Time
AA 117	Los Angeles	12:30	On Time
AA 118	Los Angeles	12:35	On Time
AA 119	Los Angeles	12:40	On Time
AA 120	Los Angeles	12:45	On Time
AA 121	Los Angeles	12:50	On Time
AA 122	Los Angeles	12:55	On Time
AA 123	Los Angeles	13:00	On Time
AA 124	Los Angeles	13:05	On Time
AA 125	Los Angeles	13:10	On Time
AA 126	Los Angeles	13:15	On Time
AA 127	Los Angeles	13:20	On Time
AA 128	Los Angeles	13:25	On Time
AA 129	Los Angeles	13:30	On Time
AA 130	Los Angeles	13:35	On Time
AA 131	Los Angeles	13:40	On Time
AA 132	Los Angeles	13:45	On Time
AA 133	Los Angeles	13:50	On Time
AA 134	Los Angeles	13:55	On Time
AA 135	Los Angeles	14:00	On Time
AA 136	Los Angeles	14:05	On Time
AA 137	Los Angeles	14:10	On Time
AA 138	Los Angeles	14:15	On Time
AA 139	Los Angeles	14:20	On Time
AA 140	Los Angeles	14:25	On Time
AA 141	Los Angeles	14:30	On Time
AA 142	Los Angeles	14:35	On Time
AA 143	Los Angeles	14:40	On Time
AA 144	Los Angeles	14:45	On Time
AA 145	Los Angeles	14:50	On Time
AA 146	Los Angeles	14:55	On Time
AA 147	Los Angeles	15:00	On Time
AA 148	Los Angeles	15:05	On Time
AA 149	Los Angeles	15:10	On Time
AA 150	Los Angeles	15:15	On Time
AA 151	Los Angeles	15:20	On Time
AA 152	Los Angeles	15:25	On Time
AA 153	Los Angeles	15:30	On Time
AA 154	Los Angeles	15:35	On Time
AA 155	Los Angeles	15:40	On Time
AA 156	Los Angeles	15:45	On Time
AA 157	Los Angeles	15:50	On Time
AA 158	Los Angeles	15:55	On Time
AA 159	Los Angeles	16:00	On Time
AA 160	Los Angeles	16:05	On Time
AA 161	Los Angeles	16:10	On Time
AA 162	Los Angeles	16:15	On Time
AA 163	Los Angeles	16:20	On Time
AA 164	Los Angeles	16:25	On Time
AA 165	Los Angeles	16:30	On Time
AA 166	Los Angeles	16:35	On Time
AA 167	Los Angeles	16:40	On Time
AA 168	Los Angeles	16:45	On Time
AA 169	Los Angeles	16:50	On Time
AA 170	Los Angeles	16:55	On Time
AA 171	Los Angeles	17:00	On Time
AA 172	Los Angeles	17:05	On Time
AA 173	Los Angeles	17:10	On Time
AA 174	Los Angeles	17:15	On Time
AA 175	Los Angeles	17:20	On Time
AA 176	Los Angeles	17:25	On Time
AA 177	Los Angeles	17:30	On Time
AA 178	Los Angeles	17:35	On Time
AA 179	Los Angeles	17:40	On Time
AA 180	Los Angeles	17:45	On Time
AA 181	Los Angeles	17:50	On Time
AA 182	Los Angeles	17:55	On Time
AA 183	Los Angeles	18:00	On Time
AA 184	Los Angeles	18:05	On Time
AA 185	Los Angeles	18:10	On Time
AA 186	Los Angeles	18:15	On Time
AA 187	Los Angeles	18:20	On Time
AA 188	Los Angeles	18:25	On Time
AA 189	Los Angeles	18:30	On Time
AA 190	Los Angeles	18:35	On Time
AA 191	Los Angeles	18:40	On Time
AA 192	Los Angeles	18:45	On Time
AA 193	Los Angeles	18:50	On Time
AA 194	Los Angeles	18:55	On Time
AA 195	Los Angeles	19:00	On Time
AA 196	Los Angeles	19:05	On Time
AA 197	Los Angeles	19:10	On Time
AA 198	Los Angeles	19:15	On Time
AA 199	Los Angeles	19:20	On Time
AA 200	Los Angeles	19:25	On Time

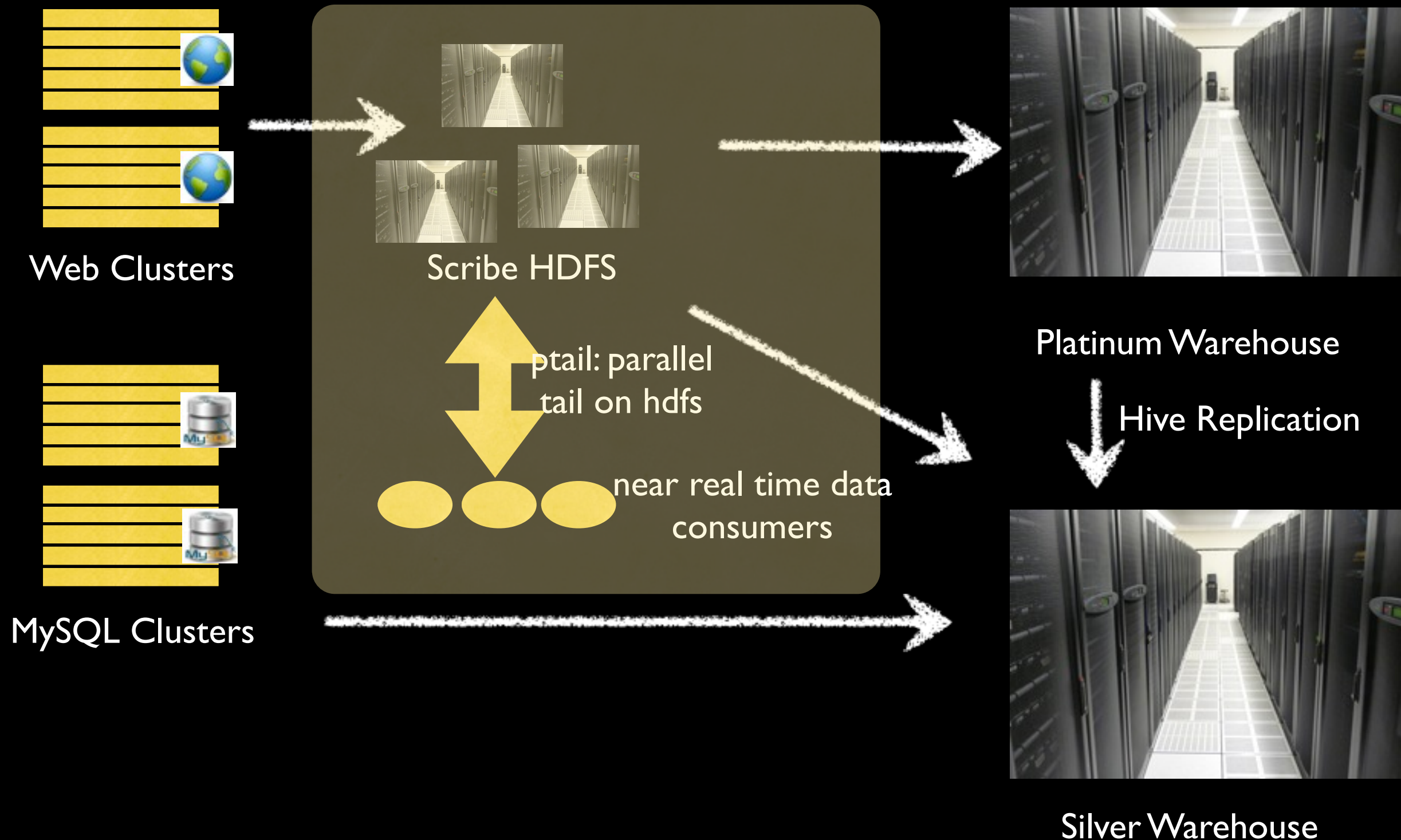
2011: New Requirements

- 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



2011: Puma



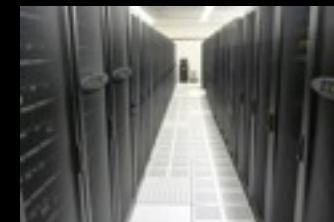
Scribe HDFS



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