facebook

facebook

HBase Today

Jonathan Gray
ApacheCon NA 2010
November 3, 2010

Agenda

1	Who is doing what with HBase
2	HBase 0.90
3	HBase 0.92 and beyond

HBase at Facebook

HBase at Facebook
Why HBase?

- **Tight integration with Hadoop**
  - Reap benefits of work done on HDFS
  - Exploit data locality with HDFS + HBase + Hive/MR

- **Proven architecture at very high levels of scale**
  - Some architectural issues with other NoSQL DBs

- **Bridges gaps between online/offline/archive**

Cluster 1

Scribe/Hadoop cluster

Cluster 2

Scribe/Hadoop cluster

5-15 minutes

5-15 minutes

Silver Cluster

Platinum Hadoop Cluster

Daily

Federated MySQL
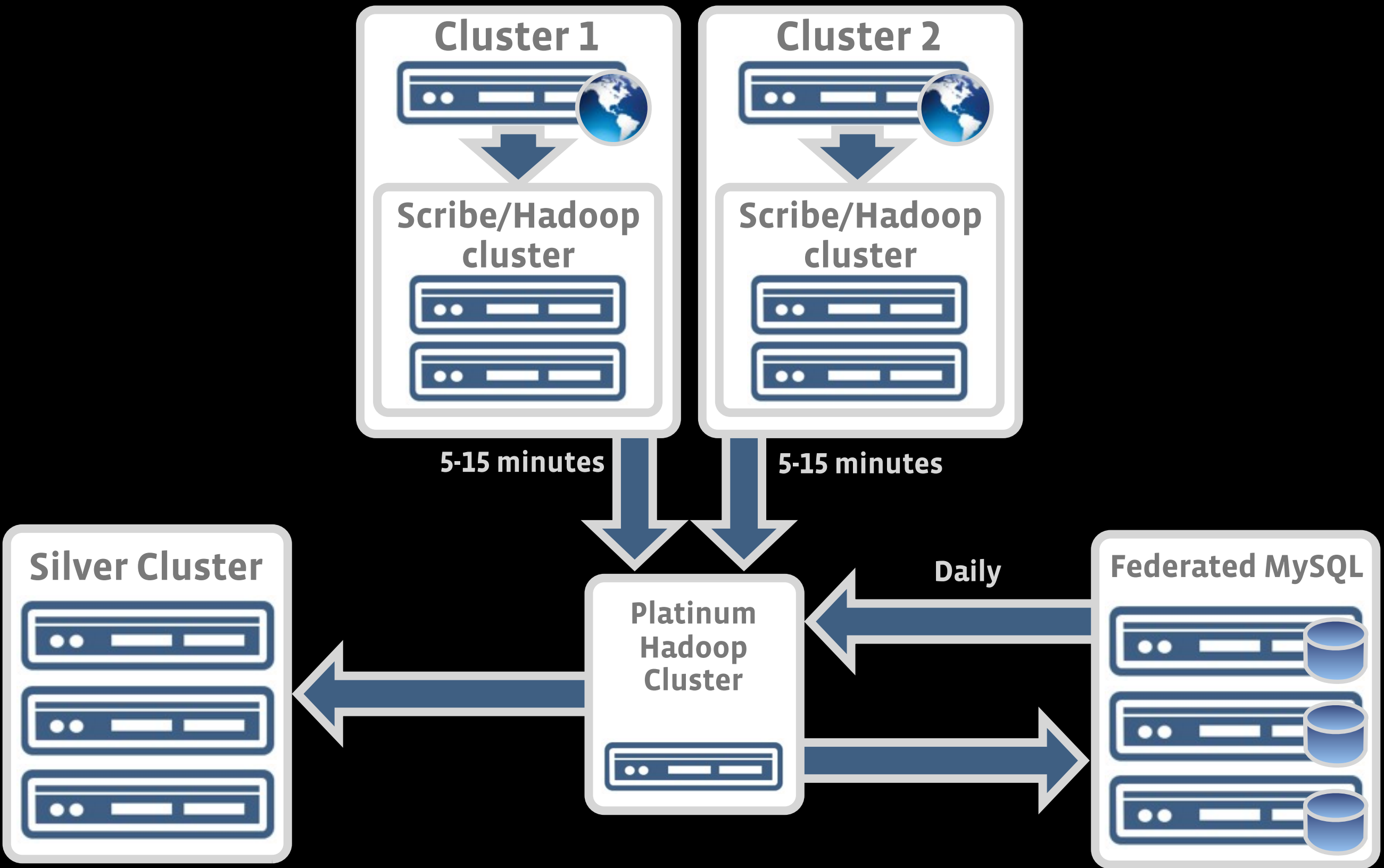
HBase at Facebook
Hive Data Warehouse

- **Hive and MapReduce integration with HBase**
  - HBase tables as Hive tables

- **Incremental updates to warehouse from MySQL**
  - Availability in warehouse from 1 day to 15 mins

- **Random read access for sparse joins**

HBase at Facebook
Realtime Analytics

facebook

- **High frequency atomic counters**
  - 10K increments/node/sec with strong durability

- **Realtime tracking of demographics and topN**
  - Fast writes, pre-calculate base set of analytics

- **Archiving of historical data**
  - Cheap and scalable storage, never delete data

HBase at Facebook
Online serving database

- **Fast and scalable on "commodity" hardware**
  - Fast random writes, really fast streaming writes
  - Fully utilize dense nodes w/ 24TB of disks

- **Efficient random reads and range scans**
  - Optional caching layer depending on application

- **Bulk write to HDFS, read from PHP**

HBase in Production

HBase at Mozilla
Mozilla Socorro

- **Store and process Firefox crash reports**

- **Receive about 3.5 million reports (~450GB) / day**
  - Old system was only able to process 15%

- **Writes directly to HBase, reads directly from HBase**
  - MapReduce analytics on same cluster

HBase at StumbleUpon
Powering all of su.pr

- **Serves all user requests for su.pr without additional caching**

- **Same cluster used for tracking analytics for users**
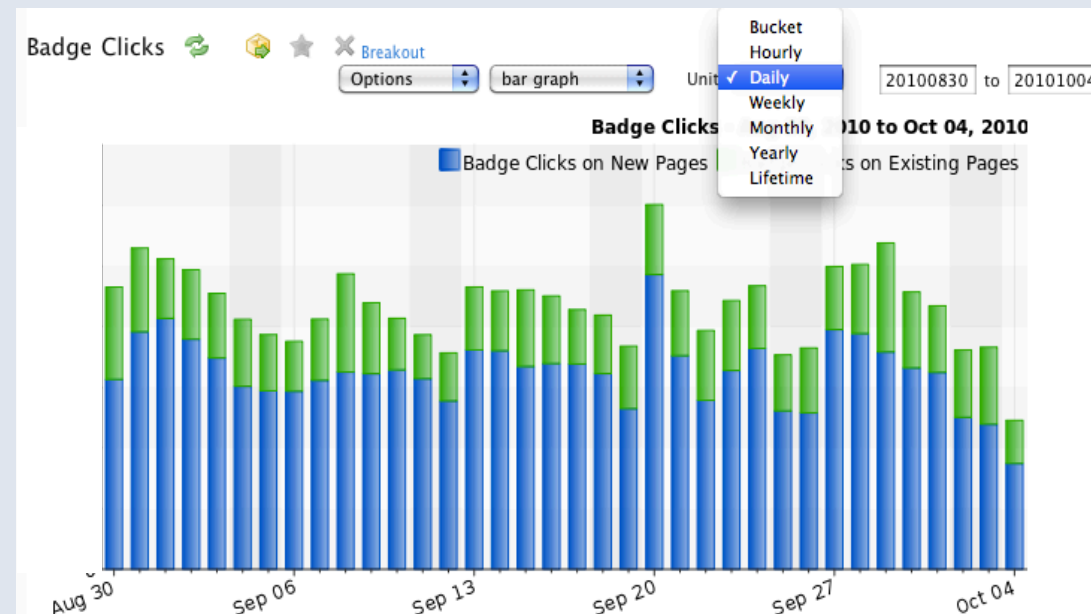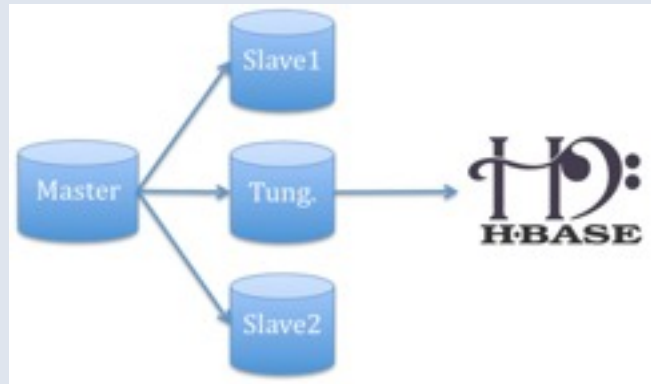
HBase at StumbleUpon

Advertising platform

- **Raw data collected in HDFS**

- **Processed with Cascading and written into HBase**

- **Replicated to second cluster for machine learning**

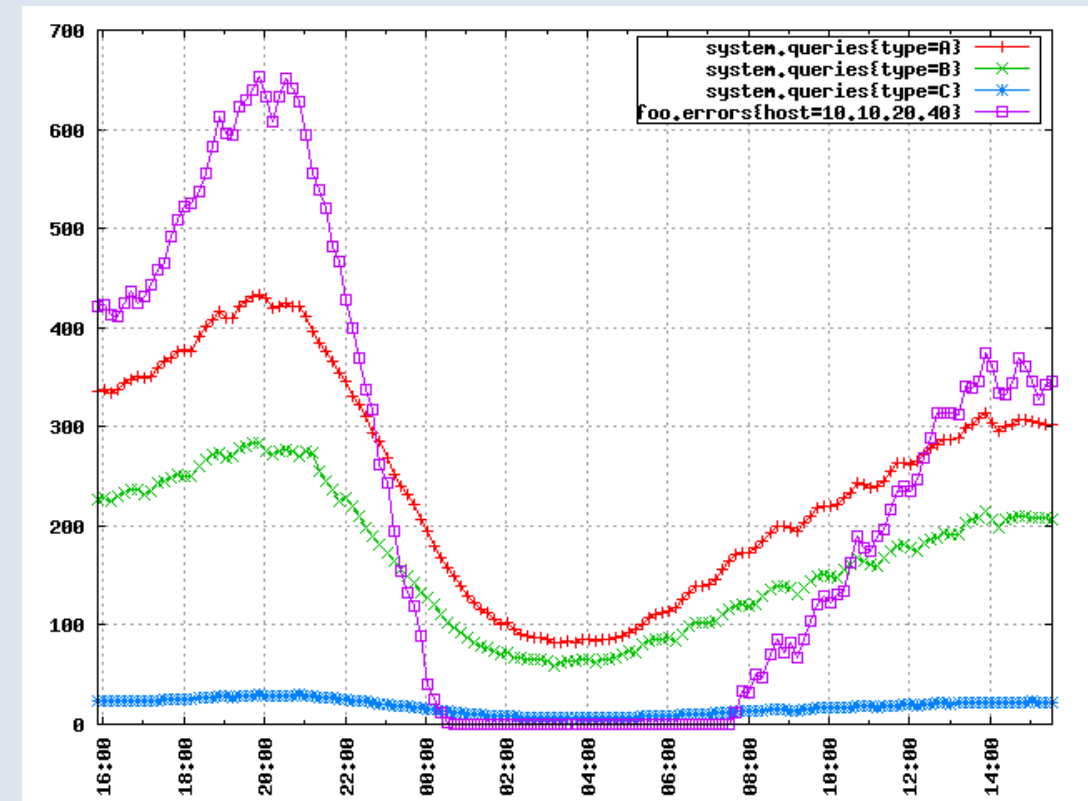Apache → HDFS → Cascading → HBase (Online) ⇄ HBase (Offline)

HBase at StumbleUpon
Open Source HBase Extensions

- **MySQL-to-HBase Replication**

- **openTSDB - an Open Source time series database built on top of HBase (release tomorrow)**

HBase for Lily CMS
Scalable CMS with HBase and Solr

- **Data stored and archived in HBase**

- **HBase Indexing Library for secondary indexes**

- **HBase RowLog for Solr WAL and reliable queueing**

HBase for Hadoop Users

Increasing interest from big Hadoop users

- **HBase now officially support by Cloudera**
  - Training, consulting, and packaged with CDHv3

- **In production at Twitter, Meetup, and Adobe**

- **In development at Yahoo, eBay, and NavTeq**

HBase 0.90

HBase 0.90
Production Ready

- **HBase 0.90 RC next week**

  Testability, Stability, and Durability release

  - Strong durability
  - Strong consistency
  - Atomic operations
  - HDFS bulk writes
  - MR scans with data locality

HBase 0.90
Testability

- **60 new test classes, 13k new lines of unit tests**

- **Refactor of major components for testability**

- **HBaseTestingUtility**
  - Start ZK, HDFS, MR cluster inside unit tests
  - Start multiple Masters and RegionServers
  - Able to kill components, inject faults, etc…

HBase 0.90
Stability

- **Rewrite of master failover and regions-in-transition**

  - **Utilize ZooKeeper to hold important state**
    - "Gremlin" scripts for load and failure testing

  - **Successfully in production at many levels of scale**
    - Many with 10, 20, or 40 node clusters
    - Popular at 100 nodes and up to 700 so far

HBase 0.90
Durability

- **HDFS Appends and WAL improvements**

  - **No data loss under RS/DN failure**
    - Can completely lose any two nodes at once
    - Automatic data recovery

  - **Tunable durability for increased performance**
    - Configurable deferred log flushing

HBase 0.90
0.89 Development Releases

- **Released three early-cuts as 0.89 releases**

- **Get latest code into hands of users early**
  - No replacement for testing in the wild

- **Already being used in production**
  - Early issues already fixed before initial release rather than in 0.90.1

HBase 0.90

New Features

- **Master-Slave inter-DC Replication**

- **Batch operations for Put, Get, and Delete**

- **HFileOutputFormat improvements**

- **Significant Query Optimizing**
  Efficient reads in rows with millions of columns and versions

HBase 0.92 and beyond

HBase 0.91
Next development release

- **Early release of 0.92 branch in mid-November**

  - Coprocessors

  - Security

  - Configurable compaction policies

HBase 0.92
Feature and Performance Release

- **Expected release in early January 2011**

  - Coprocessors and Security
  - Flush/Compact/Split improvements
  - Multi-Master Replication
  - Improved load balancing algorithms
    - Focus on data locality

Beyond HBase 0.92
Moving towards HBase 1.0

- **Online dynamic configuration changes**

- **Continued improvement to load balancing**

- **Coprocessor-based Functionality**
  - Aggregate Functions
  - Secondary Indexing
  - Realtime MapReduce

Beyond HBase 0.92

Hadoop Improvements

- **AvatarNode**

- **RaidNode**

- **HighTide**

- **HDFS performance improvements**

HBase
More Information

- **HBase.org**

- **Mailing lists and Freenode IRC**

- **HBase User Group and HBase Hackathon**
  - meetup.com/hbaseusergroup
  - meetup.com/hackathon
  - and at your local Hadoop meetup

facebook