Hadoop @ Yahoo!
an admin's perspective
Yahoo! and Hadoop

Yahoo! is:

- The largest contributor to Hadoop
- The largest tester of Hadoop
- The largest user of Hadoop
- A great place to do Hadoop development, do Internet scale science and change the world!
Hadoop @ Yahoo!

- > 82 PB data
- > 25,000 servers. (in clusters of size ~4000)
- New nodes have 8 cores with 24G RAM and 6 2T disks
SQL DB and Hadoop

Plan Executor
Parser
Operator Evaluator
Optimizer

Operator Execution Engine

Data Files
Index Files
System Catalog

Hadoop
Hadoop administrators

I manage everything from Hadoop down to Operating System

Static cluster configuration, no per job configuration: set and forget.

Ops

Hadoop Cluster

config
I run builds and put together install images.
I develop the code for Hadoop. I put in config knobs and have “rules of thumb” to set them.
SQL DB and Hadoop

Plan Executor
Parser
Operator Evaluator
Optimizer

Query Evaluation Engine

Operator Execution Engine

Data Files
Index Files
System Catalog

Hadoop
Pig
Hive
HBase
Zebra
Hadoop and self managed DB

- There is plenty to manage in Hadoop and there will be more in the future
Hadoop and self managed DB

- There is plenty to manage in Hadoop and there will be more in the future.
- Applying self managing DB techniques to Hadoop will not put anyone out of a job.