

# YAHOO!

Storm-on-YARN: Convergence of Low-Latency  
and Big-Data

Andrew Feng ([afeng@yahoo-inc.com](mailto:afeng@yahoo-inc.com), September 2013)

# Agenda

- **Business motivation**
- Storm & YARN @ Yahoo
- Storm-on-YARN
  - › Demo
  - › Open source

YAHOO!

# Personalized Web

## YAHOO!

Search

Hi, Andrew Mail

[Show Ad](#)

- Mail
- My Yahoo!
- Finance
- Flickr
- Games
- Messenger
- Movies
- Music
- omg!
- Sports
- Weather
- Autos
- News
- Shine

More Y! Sites >

---

- [Facebook](#)
- [Gmail](#)
- [WSJ](#)
- [NPR](#)
- [More Favorites >](#)



### British prime minister loses Syria war vote

David Cameron says it is clear to him that the British people do not want to see military action. [His response to Parliament »](#)

1 - 5 of 95

Good deed — with a catch Gaga's see-through getup Best work decision? Man charging telemarketers Stunning vote on Syria war

---

All Stories News Local Entertainment Sports More ▾

**Yosemite Rim Fire: why there's more optimism about taming it**  
Although the Yosemite Rim Fire is still raging and expanding — now at 193,000 acres — there are reasons for optimism, several officials and fire experts say. But given that Yosemite is a [Christian Science Monitor](#)

**Midwest hot, dry spell brings back drought worries**  
DES MOINES, Iowa (AP) — A growing season that began unusually wet and cold in the Midwest is finishing hot and dry, renewing worries of drought and its impact on crops.  
Associated Press

**Hortonworks to seek IPO within two years, CEO says**  
(Reuters) - Enterprise software start-up Hortonworks plans to seek a public listing in 15 to 24

### Trending Now

- 1 Sandra Bullock
- 2 Military strike Syria
- 3 Leah Remini
- 4 Tylenol warnings
- 5 Scud missiles

### Watch the show »

- 6 Colorado marijuana ...
- 7 Harley-Davidson 11...
- 8 Holy Grail Jay Z
- 9 Indian rupee
- 10 Pressy smartphones

---

### Featured Videos

[Y! Screen](#)

90-year-old steam train runs on unusual fuel Student sets new McNuggets world record

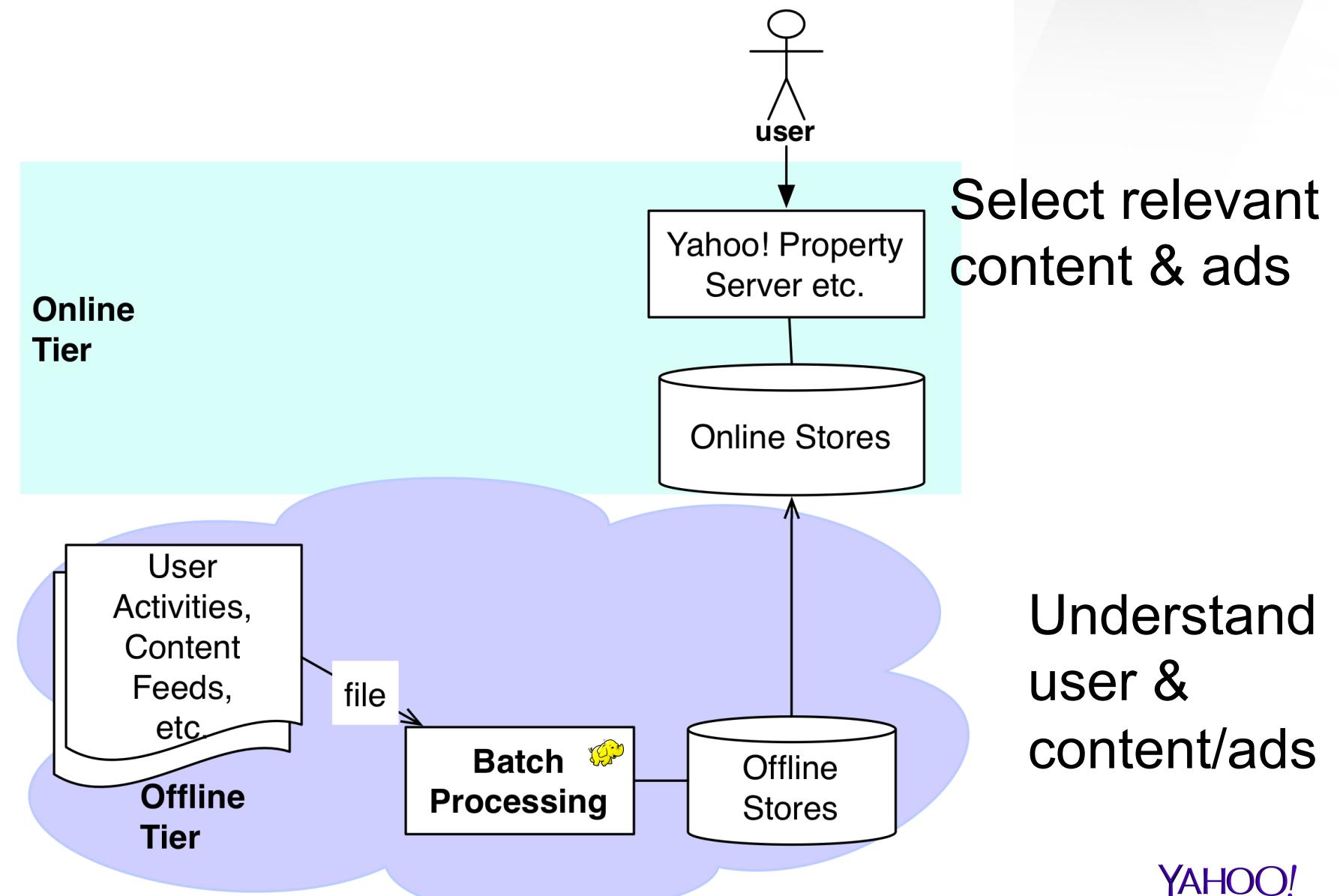
Revolutionary device keeps baby alive Dramatic waterspout forms off Croatian coast

Search videos

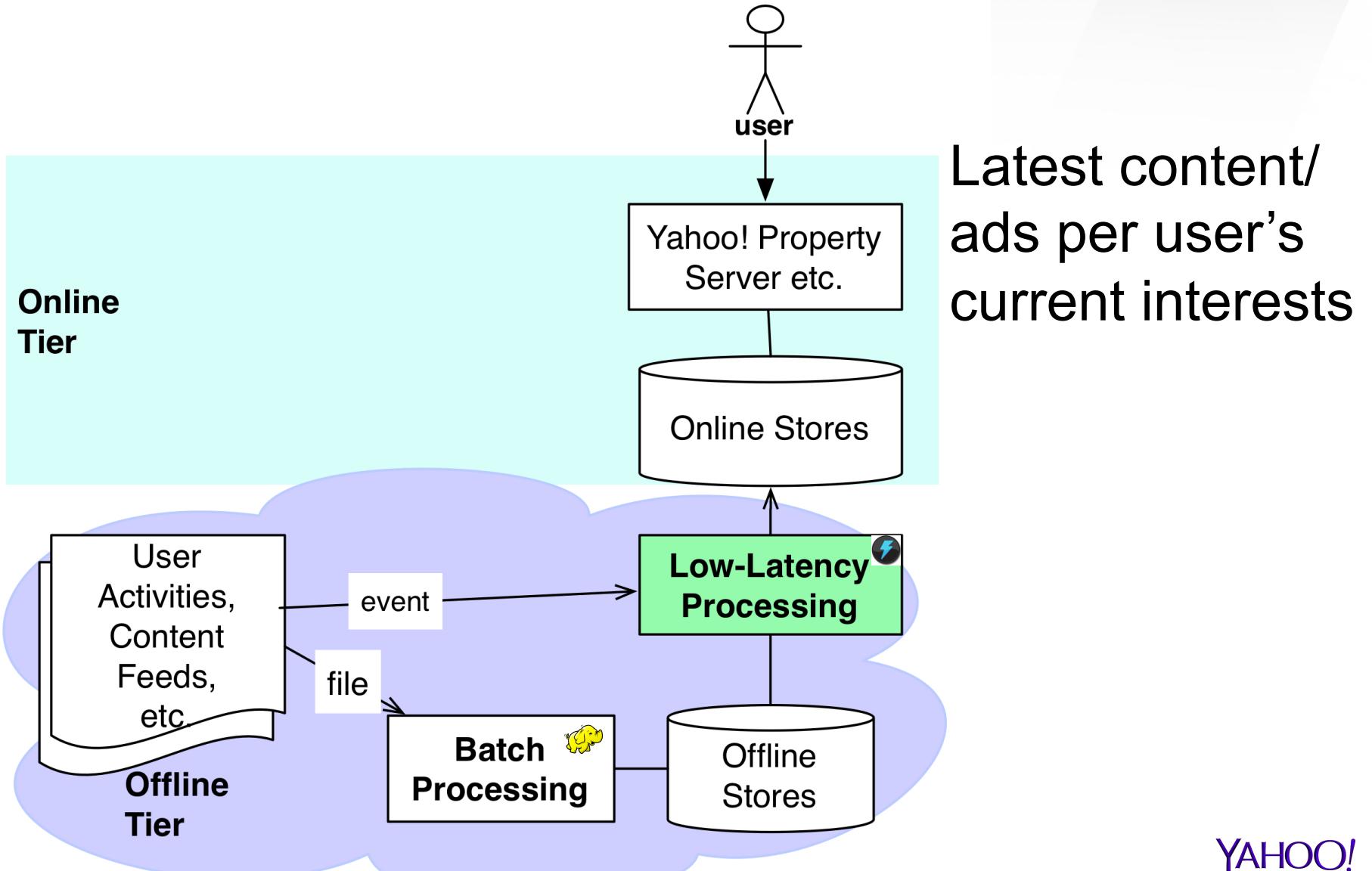
Sunnyvale 76°F Fair

**YAHOO!**

# Personalization w/ Hadoop

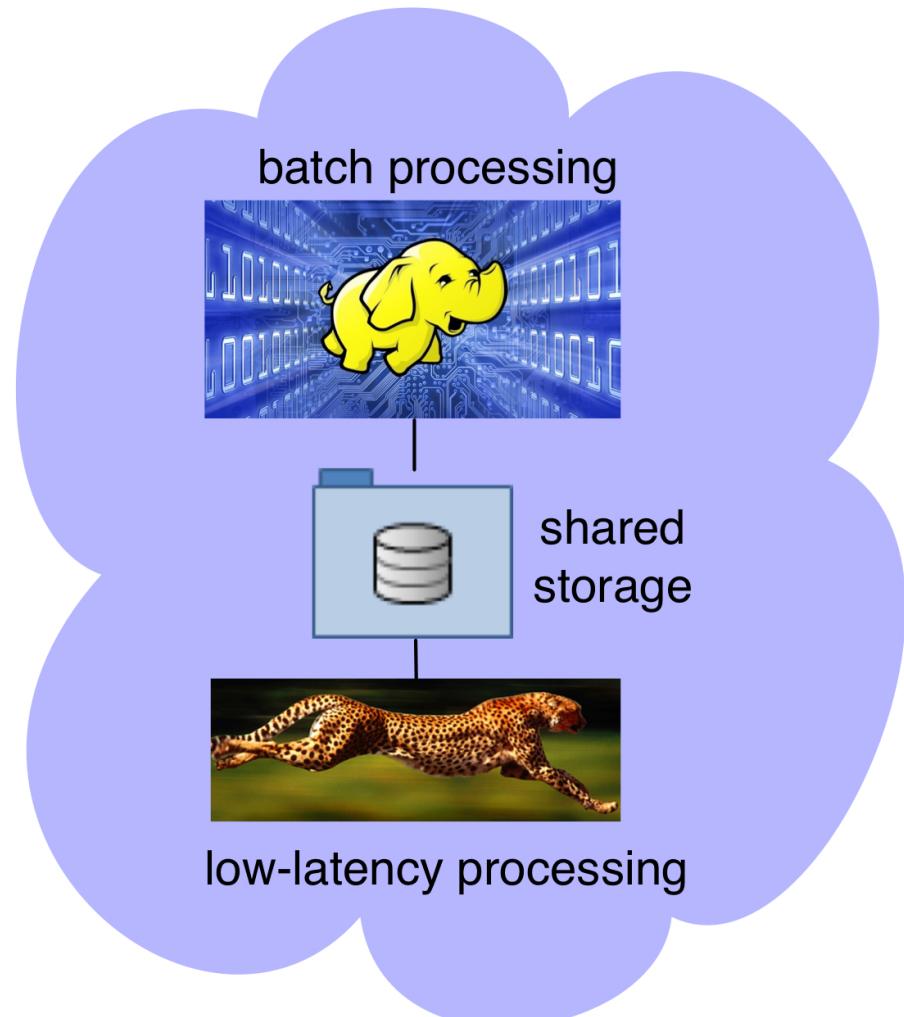


# Personalization w/ Low-Latency



# Big Data + Low Latency: Use Cases

- Personalization
- Ad targeting
- Ad budgeting
- Reporting
- Fraud detection
- Trending topics



YAHOO!

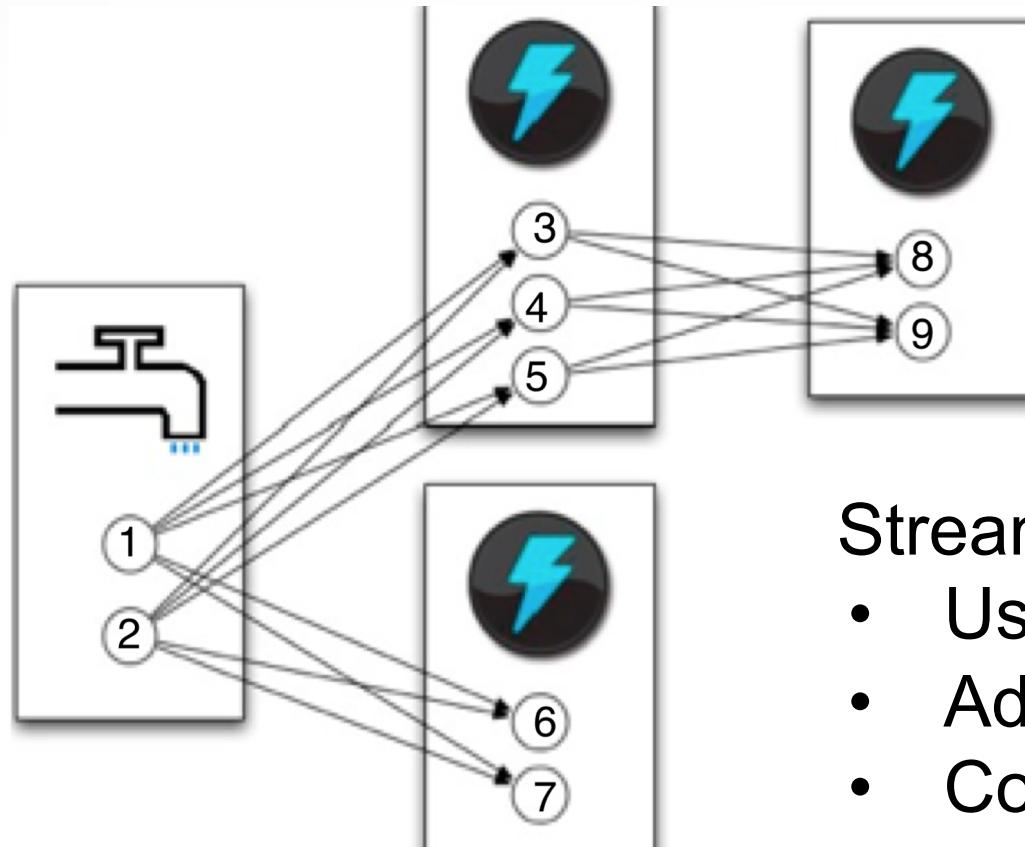
# Agenda

- Business motivation
- **Storm & YARN @ Yahoo**
- Storm-on-YARN & Demo

YAHOO!

# Storm: Distributed Stream Processing

<https://github.com/nathanmarz/storm>

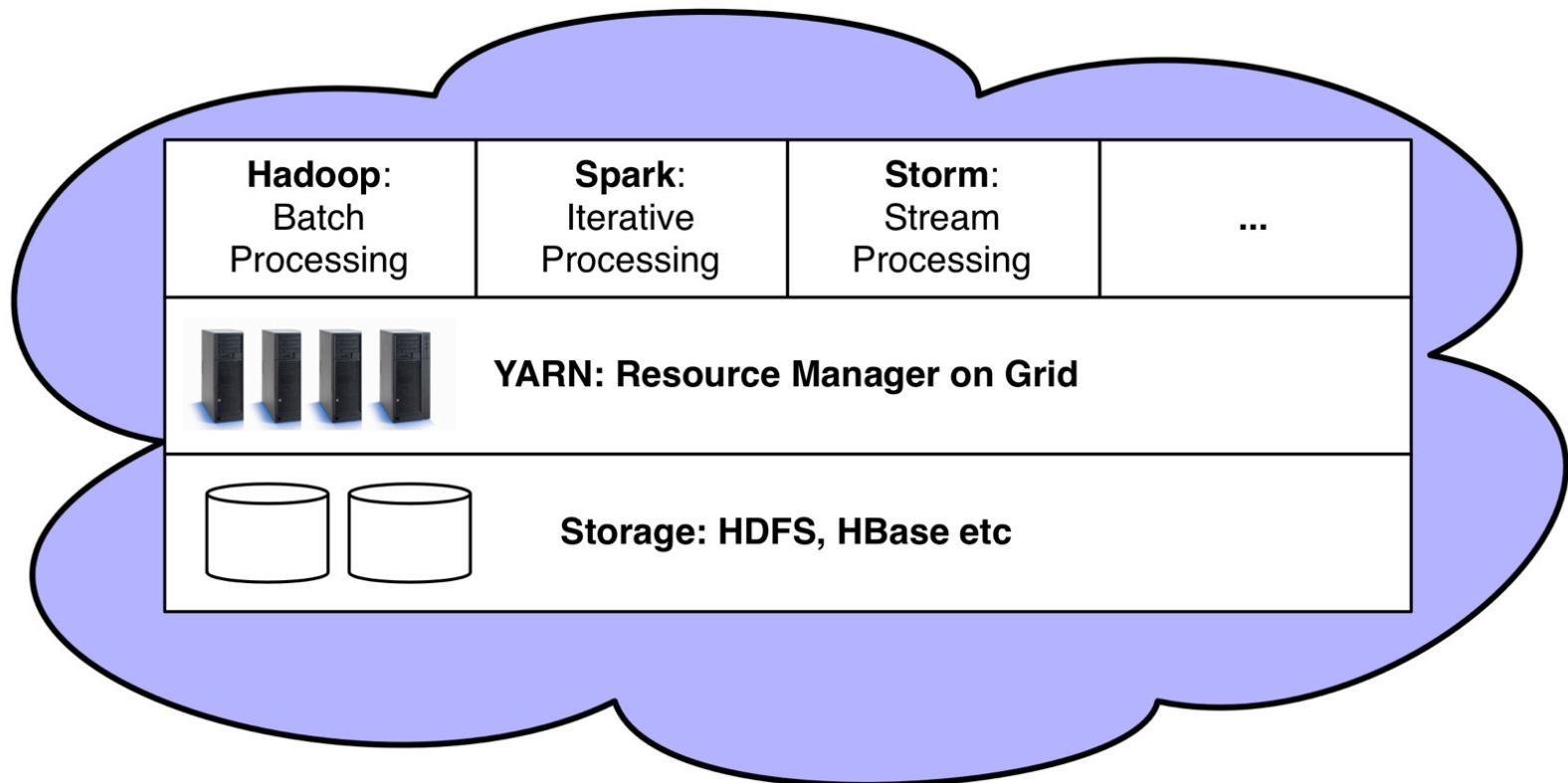


## Streams

- User activities
- Ad beacons
- Content feeds
- Social feeds
- ...

YAHOO!

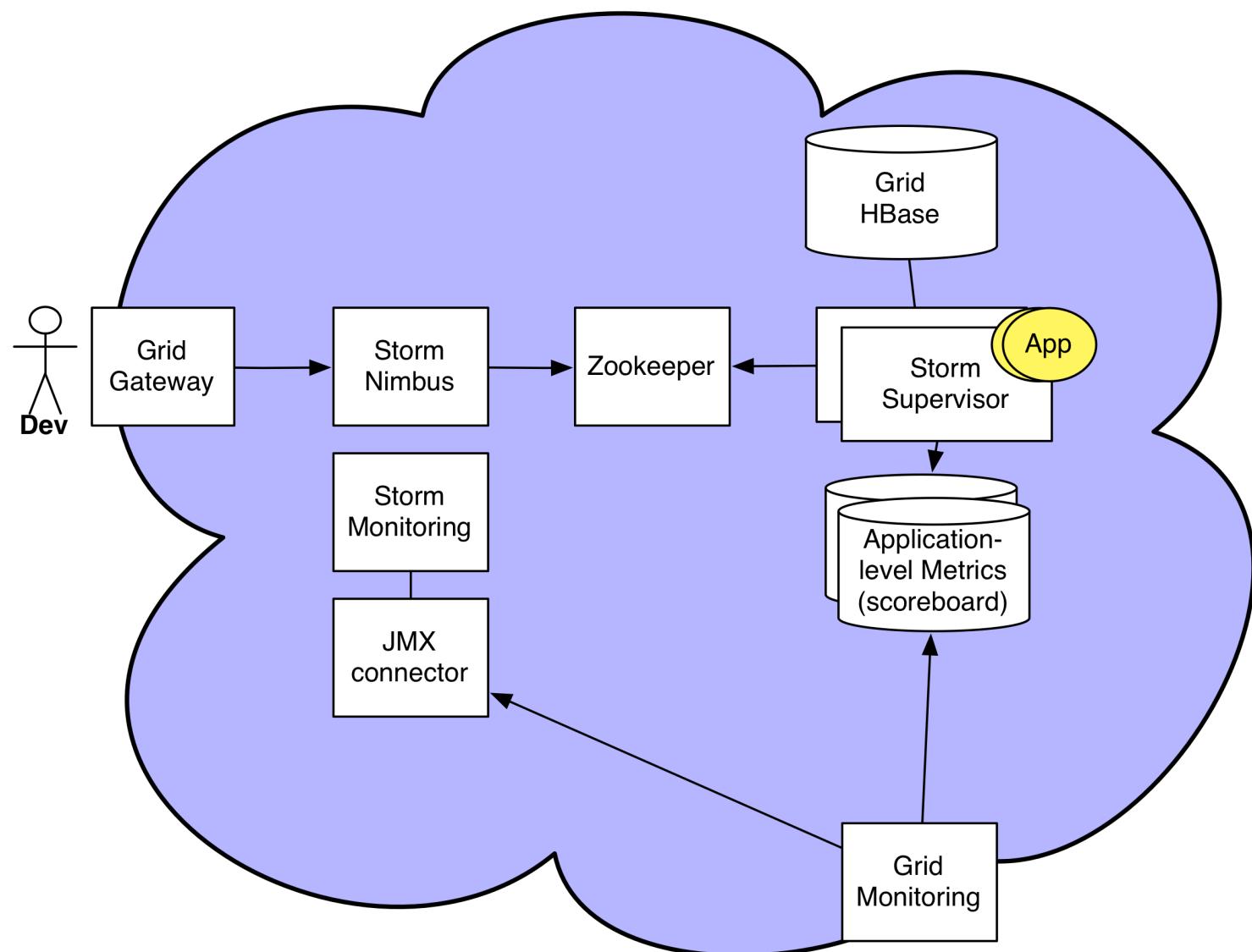
# YARN: MapReduce, Storm & Beyond



30k+ YARN nodes in Yahoo production clusters

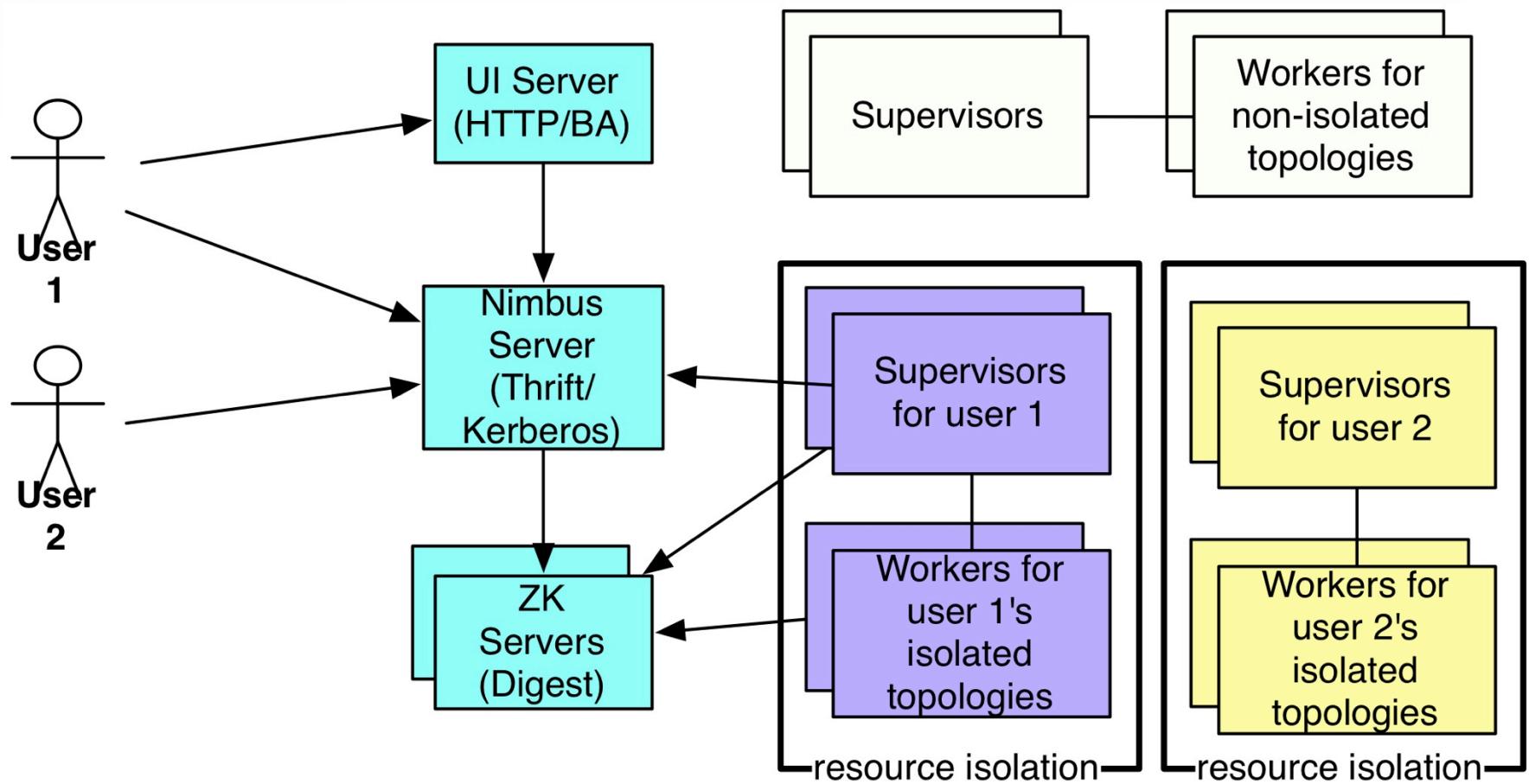
YAHOO!

# Storm Clusters on Hadoop Grid



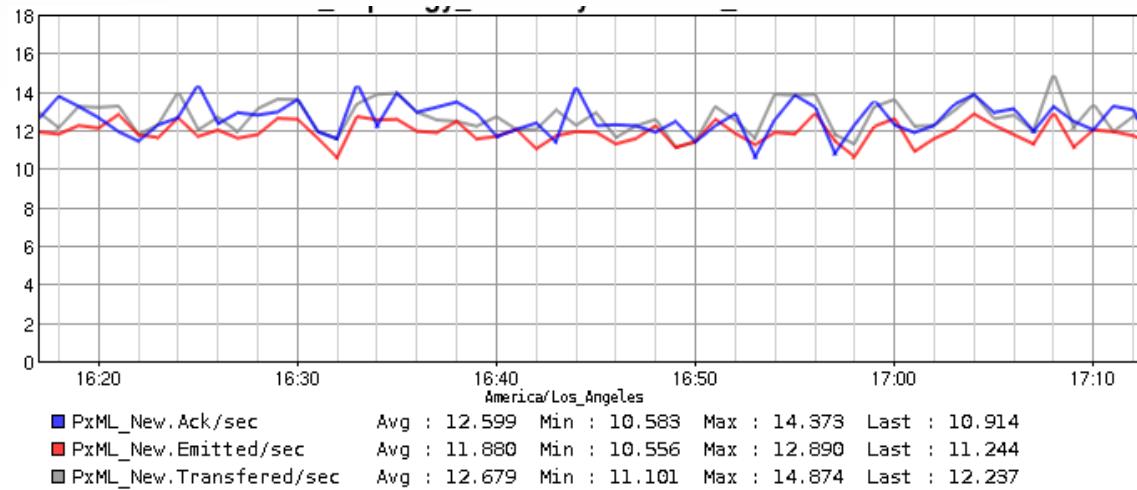
...JO!

# Storm Enhanced for Multi-tenancy: Security & Isolation

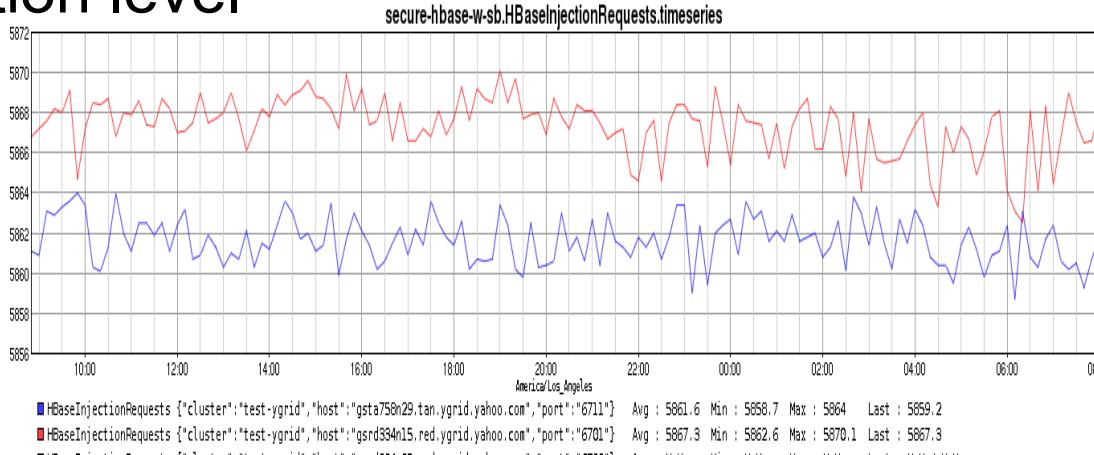


# Storm Dashboard on Grid

## Container level



## Application level



YAHOO!

# Storm Adoption @ Yahoo

- 70 topologies
  - › Content preprocessing
  - › Social signal processing
  - › Ad budget management
  - › Ad preprocessing
  - › User profiling
  - › ...
- 400k events per second
- 400 nodes
- 3 data centers

# Agenda

- Business motivation
- Storm & YARN @ Yahoo
- **Storm-on-YARN & Demo**

# What's Storm-on-YARN?

- Deploy Storm clusters onto machines managed by Hadoop YARN
  - › Standard Storm tools ... manage Storm topologies
- Enable Storm topologies to access datasets in Hadoop storage (ex. Hbase, HDFS)
  - › Avoid data movement b/w Hadoop cluster and Storm cluster

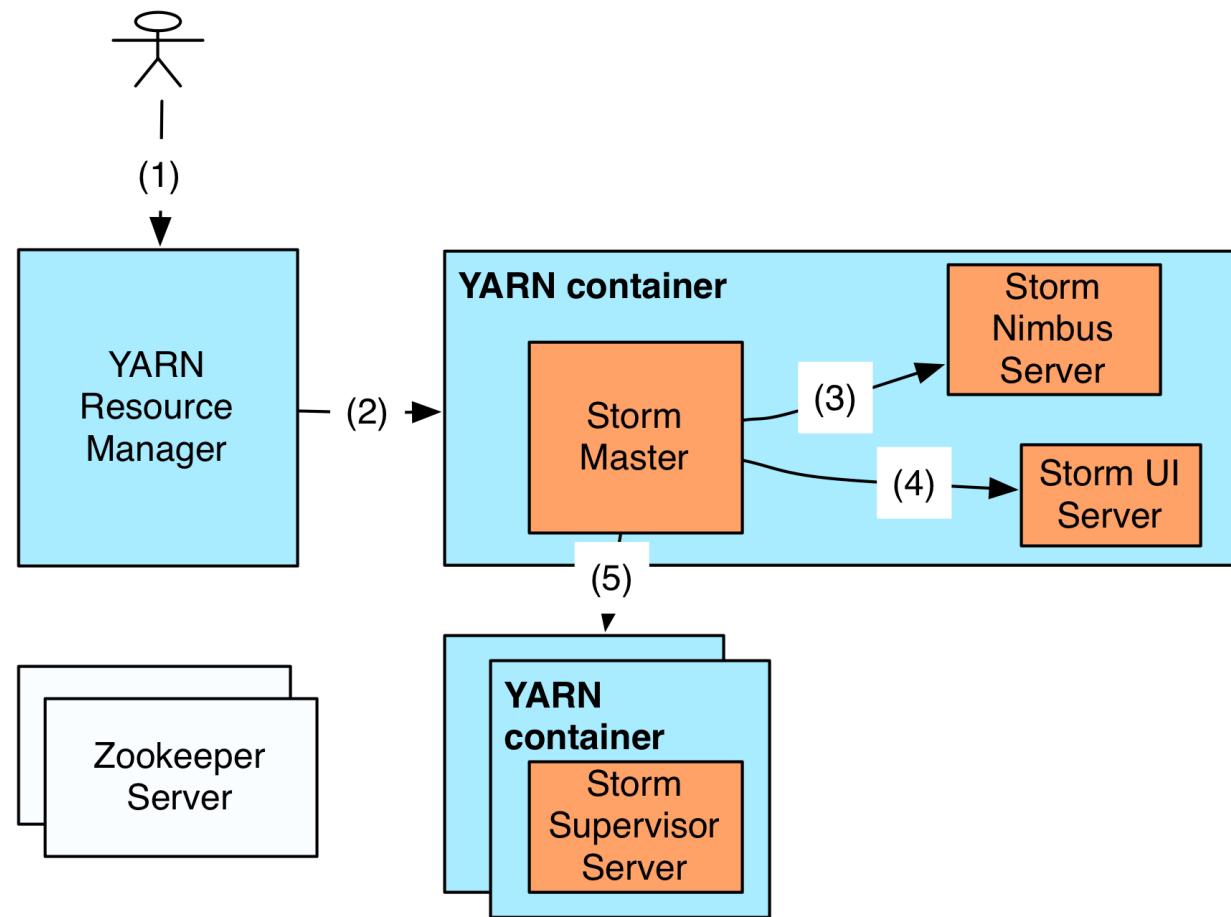
# Storm-YARN: 1-2-3

- 1. storm-yarn launch <master conf> <app name>**
  - › <app id>
- 2. storm-yarn getStormConfig <master conf> <app id>**
  - › <storm.yaml>
- 3. storm jar <topology jar> <topology name>**
- 4. storm kill <topology name>**
- 5. yarn application -kill <app id>**

YAHOO!

# Storm-YARN: Launch Cluster

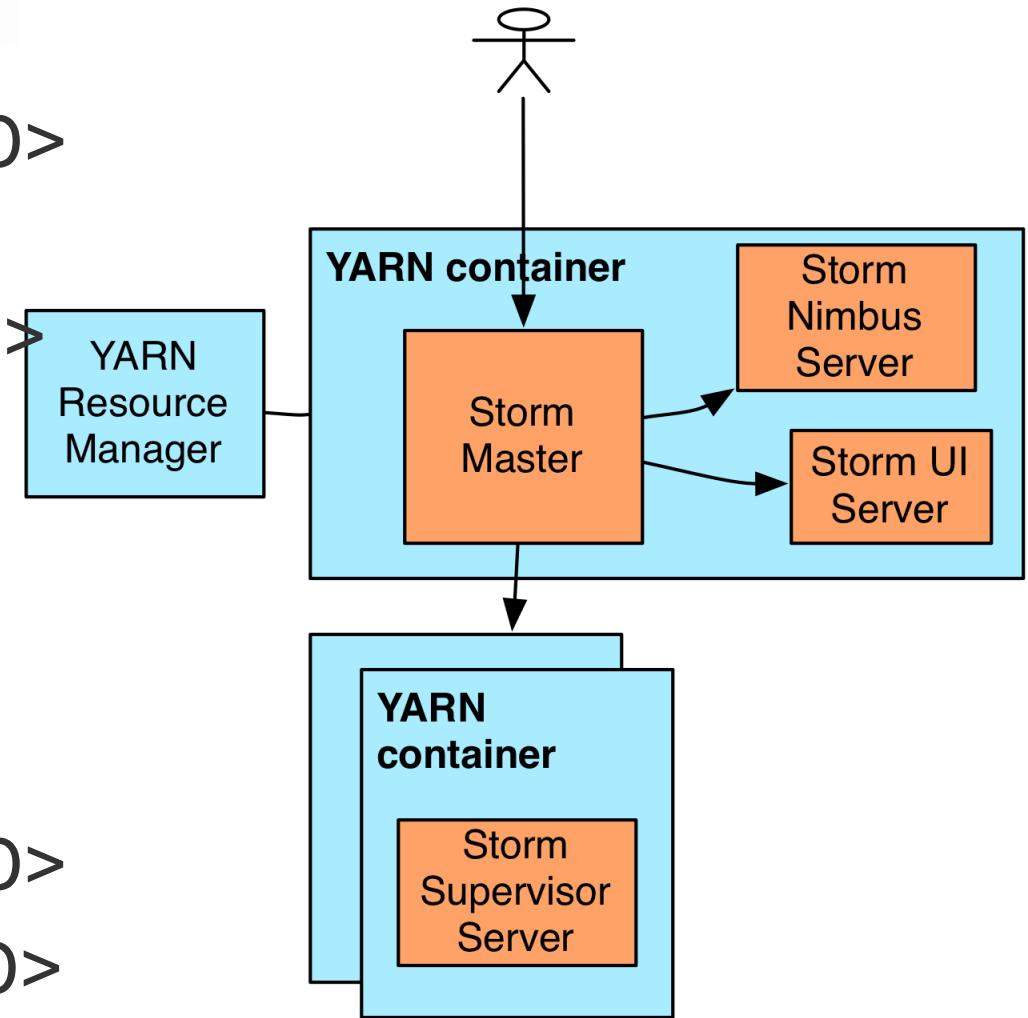
- storm-yarn launch <master conf>
  - › *Initial # of supervisors*
  - › *memory size of allocated container*
- Result: <app id> of the newly launched Storm cluster



100!

# Storm-YARN: Administrate Cluster

1. addSupervisors <appID> <count>
2. getStormConfig <appID>
3. startNimbus <appID>
4. stopNimbus <appID>
5. startUI <appID>
6. stopUI <appID>
7. startSupervisors <appID>
8. stopSupervisors <appID>



YAHOO!

# Storm-YARN: Demo



Logs for container\_1372875939789\_389289\_01\_000001

```

nimbus.log : Total file length is 15068 bytes.
stderr : Total file length is 8921 bytes.
stdout : Total file length is 9229 bytes.
ui.log : Total file length is 219 bytes.

SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/Releases/hadoop-0.23.8.2.1306220840-20130702-000/share/hadoop-0.23.8.2.1306220840/share/hadoop/common/lib/slf4j-log4j12-1.6.1.jar!/org/slf4j/impl/StaticLoggerBinder.
SLF4J: Found binding in [jar:file:/tmp/yarn-local/usercache/afeng/appcache/application_1372875939789_389289/filecache/93842046214450523/storm-0.9.0-wip21.zip/storm-0.9.0-wip21/lib/logback-classic-1
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
13/09/17 20:18:16 INFO yarn.MasterServer: Starting the AM!!!
13/09/17 20:18:16 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:16 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:17 INFO service.AbstractService: Service:com.yahoo.storm.yarn.client.AMRMClientImpl is initied.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tns.dir; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
13/09/17 20:18:17 INFO security.JniBasedUnixGroupsNetgroupMappingWithFallback: Falling back to shell based
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 INFO service.AbstractService: Service:com.yahoo.storm.yarn.client.AMRMClientImpl is started.
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: Storm master host:10.192.55.26
13/09/17 20:18:17 INFO yarn.MasterServer: launch nimbus...
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: starting nimbus...
13/09/17 20:18:17 INFO yarn.MasterServer: launch ui...
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: starting UI...
13/09/17 20:18:17 INFO yarn.MasterServer: launch 1 supervisors
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: adding 1 supervisors...
13/09/17 20:18:17 INFO yarn.StormAMRMClient: Added 1 supervisors, but not requesting containers now.
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: Running: /home/gs/java/jdk/bin/java -server -Detorm.home=/storm/storm-0.9.0-wip21/ -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.
13/09/17 20:18:17 INFO yarn.StormMasterServerHandler: Running: /home/gs/java/jdk/bin/java -server -Detorm.home=/storm/storm-0.9.0-wip21/ -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.
13/09/17 20:18:17 INFO yarn.MasterServer: Got a registration response minimumCapability {}, memory: 512, , maximumCapability {}, memory: 8192, ,
13/09/17 20:18:17 INFO yarn.StormAMRMClient: Max Capability memory: 8192
13/09/17 20:18:17 INFO yarn.MasterServer: Starting HB thread
13/09/17 20:18:17 INFO yarn.MasterServer: Starting launcher
13/09/17 20:18:17 INFO yarn.StormAMRMClient: Creating new ContainerRequest with memory: 8192 and 1
13/09/17 20:18:17 INFO yarn.StormAMRMClient: Requested container ask: Capability[memory: 8192]Priority[{Priority: 0}]ContainerCount[1]
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:17 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:17 INFO yarn.MasterServer: Starting Master Thrift Server
13/09/17 20:18:19 INFO yarn.MasterServer: HB: Received allocated containers (1)
13/09/17 20:18:19 INFO yarn.MasterServer: HB: Supervisors are to run, so queuing (1) containers...
13/09/17 20:18:19 INFO yarn.MasterServer: LAUNCHER: Taking container with id (container_1372875939789_389289_01_000002) from the queue.
13/09/17 20:18:19 INFO yarn.MasterServer: LAUNCHER: Supervisors are to run, so launching container id (container_1372875939789_389289_01_000002)
13/09/17 20:18:19 INFO yarn.StormAMRMClient: launchSupervisorOnContainer() for containerid=container_1372875939789_389289_01_000002
13/09/17 20:18:19 INFO yarn.StormAMRMClient: Connecting to ContainerManager via UGI with security mode
13/09/17 20:18:19 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:19 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:19 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:20 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tns.dir; Ignoring.
13/09/17 20:18:20 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.security.token.service.use_ip; Ignoring.
13/09/17 20:18:20 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:20 WARN conf.Configuration: yarn-site.xml:an attempt to override final parameter: hadoop.tmp.dir; Ignoring.
13/09/17 20:18:20 INFO yarn.StormAMRMClient: launchSupervisorOnContainer: startRequest prepared, calling startContainer. container_launch_context {}, container_id {}, app_attempt_id {}, application_id {}, id: 13/09/17 20:18:20 INFO yarn.StormAMRMClient: Supervisor log: http://gsrd203n04.red.ygrid.yahoo.com:8042/node/containerlogs/container_1372875939789_389289_01_000002/afeng/supervisor.log

```

# Storm-YARN Open Source: github.com/yahoo/storm-yarn

- Code released for early access

- › under the Apache 2.0 License
- › move to apache.org later

- Welcome contribution!

- › Submit proposals
- › Sign Apache style CLA
- › Submit git pull requests

The screenshot shows the GitHub interface for the repository `storm-yarn`. At the top, there are download options: `Clone in Mac`, `ZIP`, `HTTP`, `SSH`, and `Git Read-Only`. The URL `git@github.com:yahoo/storm-yarn.git` is also displayed. Below this is a navigation bar with dropdowns for `branch: master`, `Files`, `Commits`, and `Branches`. The main area shows a list of files and their commit history:

File	Last Commit	Description
<code>bin</code>	6 days ago	read-link simplified [anfeng]
<code>lib</code>	11 days ago	almost ready for integration test [anfeng]
<code>src</code>	a day ago	Merge pull request #12 from anfeng/master [revans2]
<code>CLA.pdf</code>	a day ago	Added a copy of the CLA for contributors to sign. [revans2]
<code>LICENSE.txt</code>	a month ago	Initial commit. Works with Hadoop 2.0.4-alpha and storm-0.9.0-wip17. [revans2]
<code>README.md</code>	2 days ago	Nimbus hostname should always be decided by StormMaster [anfeng]
<code>create-tarball.sh</code>	a month ago	Initial commit. Works with Hadoop 2.0.4-alpha and storm-0.9.0-wip17. [revans2]
<code>pom.xml</code>	8 days ago	Integrated test now available [anfeng]

YAHOO!

# Storm-YARN: *mvn test*

## 1. storm-yarn launch

- › ./conf/storm.yaml
- › --stormZip lib/storm.zip
- › --appname storm-on-yarn-test
- › --output target/appId.txt

## 2. storm-yarn getStormConfig

- › ./conf/storm.yaml
- › --appId application\_1372121842369\_0001
- › --output ./lib/storm/storm.yaml

## 3. storm jar

- › lib/storm-starter-0.0.1-SNAPSHOT.jar
- › storm.starter.WordCountTopology
- › word-count-topology

## 4. storm kill

- › word-count-topology

## 5. storm-yarn shutdown

- › ./conf/storm.yaml
- › --appId application\_1372121842369\_0001

Storm UI

<http://127.0.0.1:7070>

### Topology summary

Name	Id	Status	Uptime
word-count-topology	word-count-topology-1-1372210293	KILLED	38s

### Topology actions

[Activate](#) [Deactivate](#) [Rebalance](#) [Kill](#)

### Topology stats

Window	Emitted	Transferred	Complete latency (ms)
10m 0s	20200	10860	0.000
3h 0m 0s	20200	10860	0.000
1d 0h 0m 0s	20200	10860	0.000
All time	20200	10860	0.000

### Spouts (All time)

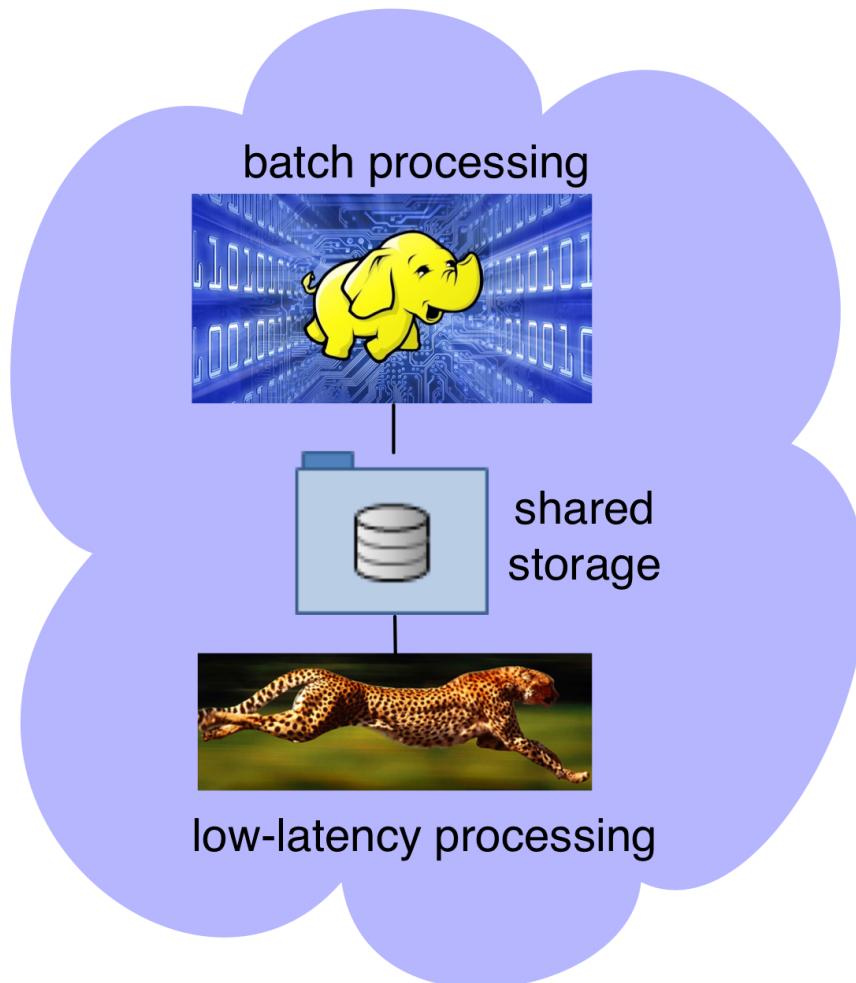
Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)
spout	5	5	1460	1460	0.000

### Bolts (All time)

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)
count	12	12	9340	0	0.005	0.113
split	8	8	9400	9400	0.001	0.056

YAHOO!

# Conclusion



- YARN empowers the emergence of big-data & low-latency processing
- Open source from Yahoo
  - › Storm-yarn @ [github/yahoo](https://github.com/yahoo)
  - › Spark-yarn @ [spark-project.org](https://spark-project.org)

YAHOO!

# We Are Hiring!

<http://careers.yahoo.com/>

