

eBay's Hadoop Stack: Evolution and Revolution

Juhan Lee

Director of Hadoop Engineering / Platform

eBay Marketplaces




Welcome to today's online marketplace...

...the market that brings buyers and sellers together in an honest and open environment...

Welcome to eBay's AuctionWeb.

Welcome to our community. I'm glad you found us. AuctionWeb is dedicated to bringing together buyers and sellers in an honest and open marketplace. Here, thanks to our [auction format](#), merchandise will always fetch its market value. And there are plenty of great deals to be found!



[Take a look at the listings.](#) There are always several hundred auctions underway, so you're bound to find something interesting.

If you don't find what you like, take a look at our **Personal Shopper**. It can help you search all the listings. Or, it can keep an eye on new items as they are posted and let you know when something you want appears. If you want to let everyone know what you want, post something on our [wanted page](#).

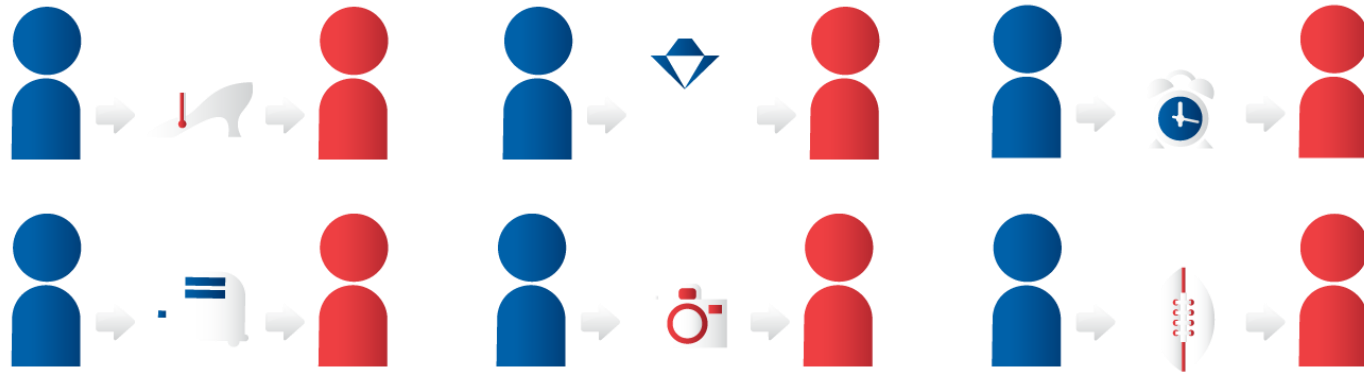
If you have something to **sell**, start your auction instantly.

Join our community. Become a registered user. Registered users receive [additional benefits](#) such as daily updates and the right to participate in our user feedback forum and the bulletin board.



\$62
billion

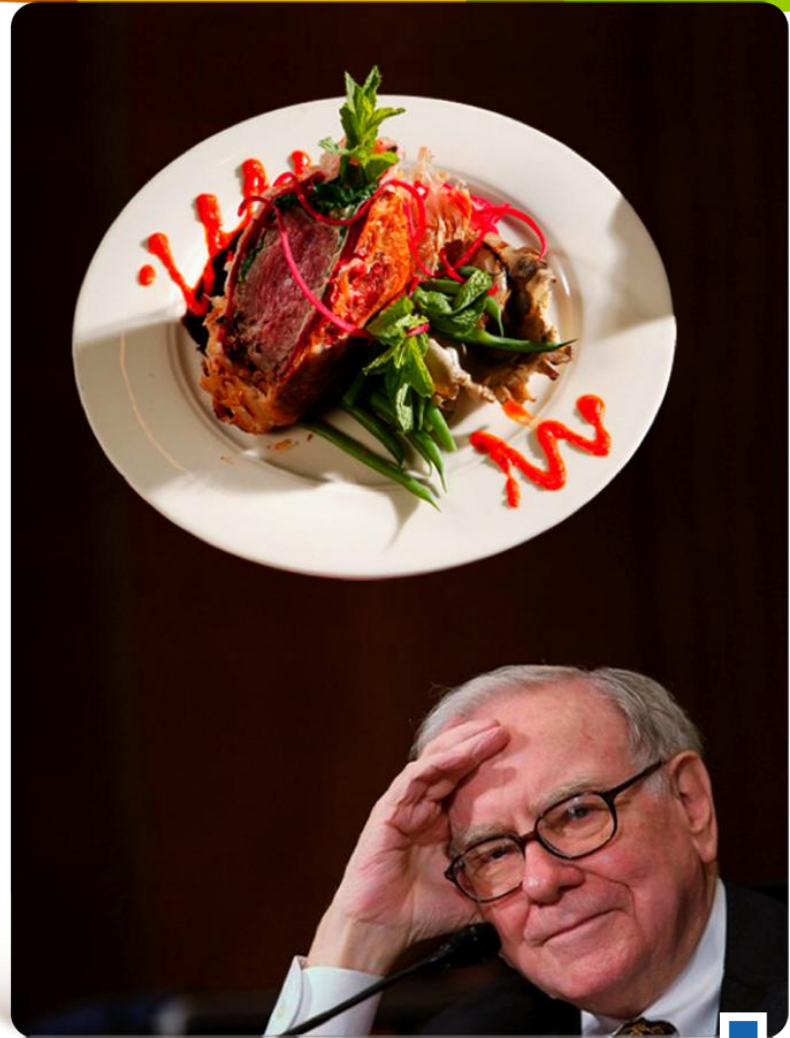
in merchandise sold in 2010



The world's largest online marketplace — where practically anyone can trade practically anything at any time.

**\$2.63
million**

for a lunch with
Warren Buffett



Marketplaces



97 million

active buyers and sellers world wide

200+ million item

in more than 50,000 categories

2 billion page views

each day

250 million queries

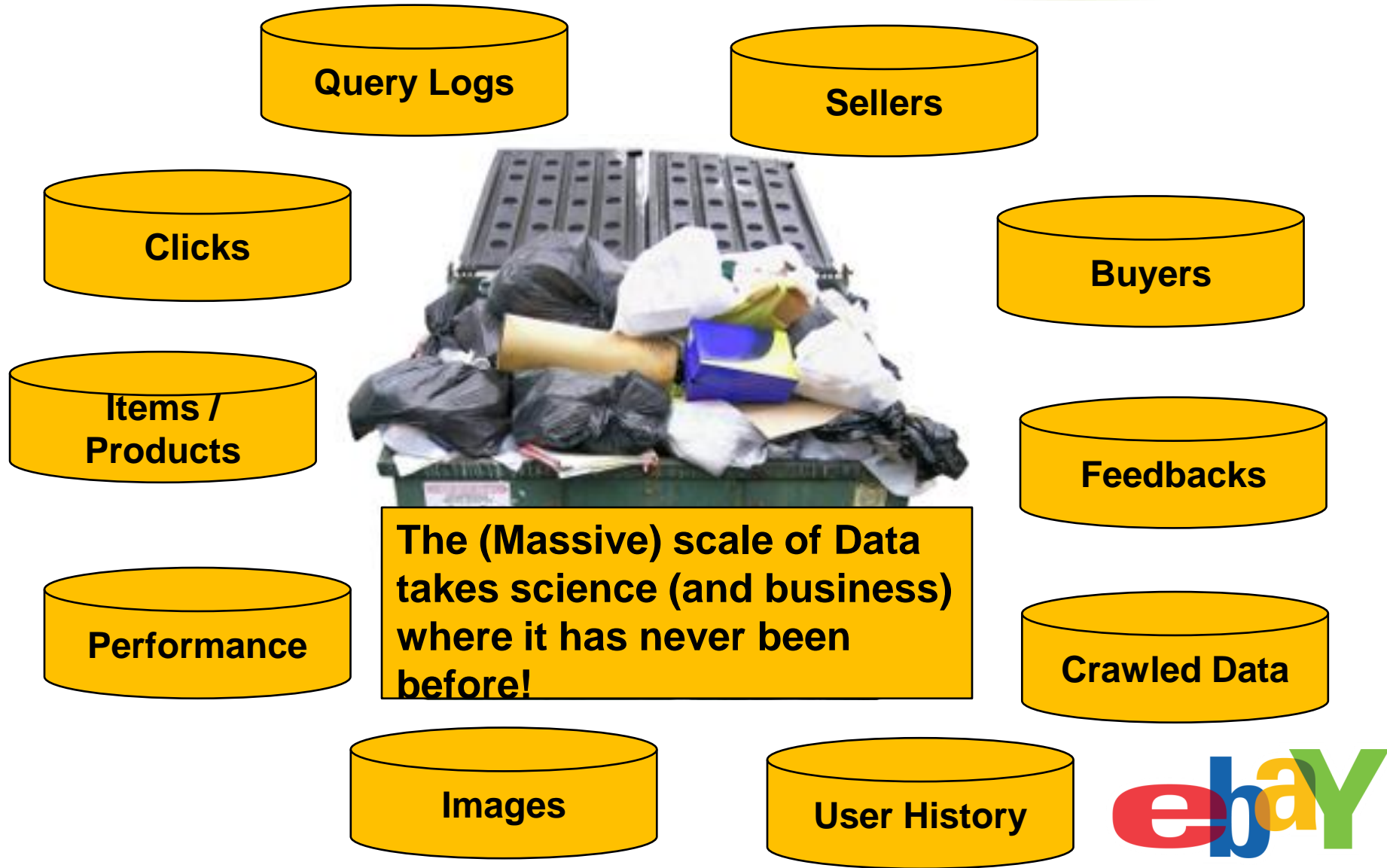
each day to our search engine



eBay Companies

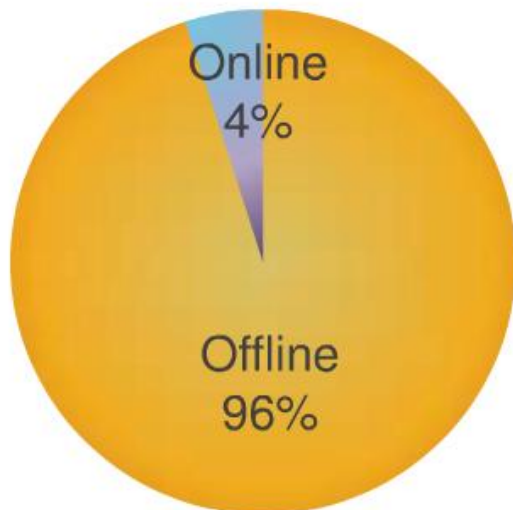


Data is Gold



Huge Opportunity: Taking the “e” out of commerce

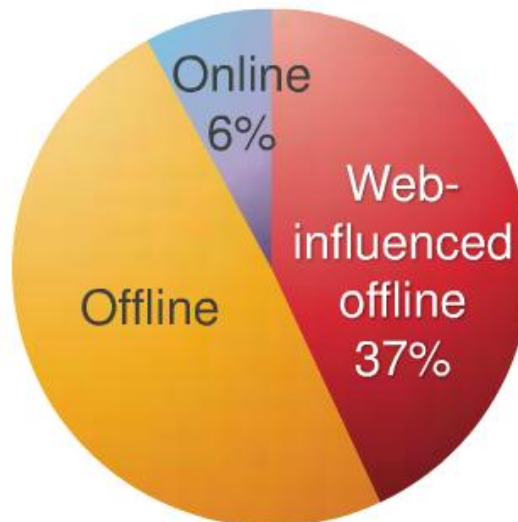
Yesterday



2008 = \$325B

Source: Forrester, Euromonitor and Economist Intelligence Unit

Today



Source: Forrester

Tomorrow



2013 = \$10T

Source: Economist Intelligence Unit





Hadoop at eBay

Hadoop @eBay



2009: 10 & 28 node clusters

2010: hundreds node cluster / multi PB

2011: thousands node cluster(s) / 10s PB

Usage / Growth Trends (2010 – 2011)

20X Active Users

30X Job Volumes

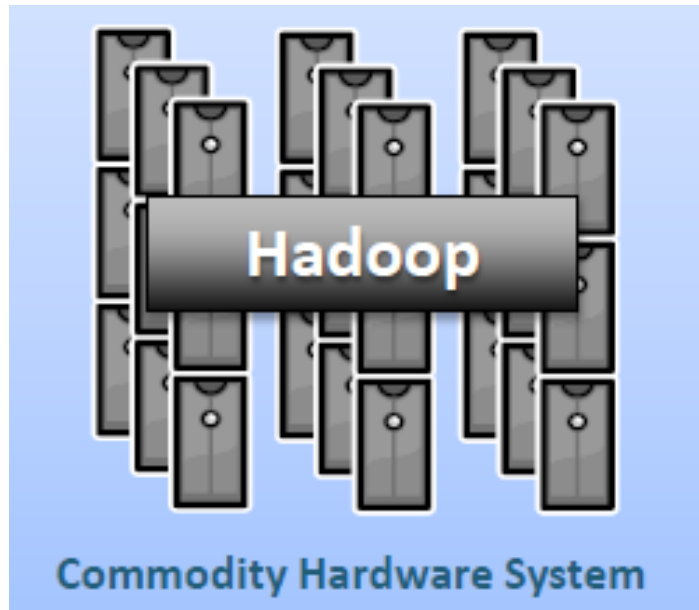
Pipeline for Business Critical Applications

Increasing Demands Across eBay Family

Proliferation of small, dedicated Hadoop clusters



Key Drivers



Massive amount of data!

Scalable, reliable, available data store and (flexible) data mining capability

Scale-out (linearly) for storage and processing

Realize 'big data' opportunity

Optimized for throughput

Commodity Hardware

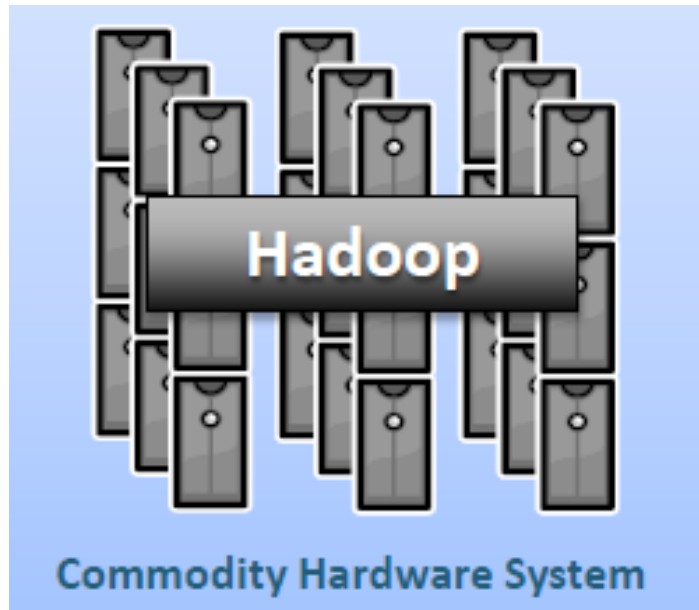
Tolerate Failures

Low cost of operations (OPEX & CAPEX)

Big business wins!



Technical Challenges



Massive amount of data!

Scalable(?), reliable(?), available(?)
data store and (flexible) data mining
capability

*Scale-out (linearly) for storage and
processing (?)*

Realize 'big data' opportunity

Optimized for throughput(?)

Commodity Hardware (?)

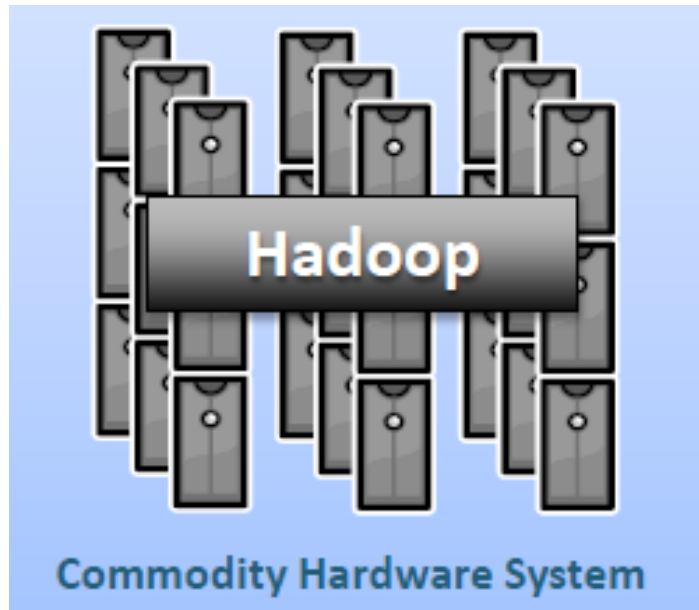
Tolerate Failures (?)

Low cost of operations (?)

**Problems amplified
at larger scale!**



Operational Challenges



Non-standard environment

Multiple distributions & OS & H/W

Manual processes

System reliability

Frequent outage

Bugs

Compliance requirements

Network

High cost of operations!



Evolving Hadoop Stack @eBay



Key feature work

- Redundant Name Node
- Support large stream / HBase
- Data expiration / Garbage collection

Make it scalable

- Better resource management / scheduling
- Virtual cluster management
- Metadata explosion



Evolving Hadoop Stack @eBay



Keep things simple

- 1 Large, shared cluster
- 1 unified eBay Hadoop distribution -> meets / exceeds business needs (private branch)
- Standard operating environment -> common software stack
- Commodity servers -> no special H/W systems

Keep it secure

- Add security features -> integrate with corp sec infrastructure
- Multi-tenancy



Evolving Hadoop Stack @eBay



Automate everything

- **Provisioning / deployment / configuration updates**
- **Keep the repair / recovery simple**

You (developers) own the business, you support the site

- **Motivation for you to automate failure handlings**

Metrics

- **Throughput / Utilization**
- **Monitoring / Chargeback**



Transforming Hadoop Platform @eBay



Move from primarily batch, offline platform to online platform

- **eBay's new search engine - Cassini**
- **Others**
 - **Social applications**
 - **Mobile applications**
 - **Crawler / Catalog system**



Cassini – eBay's new Search Engine



- **Modern Search Engine**
 - Significantly larger index than the current one with much more metadata & history
 - Faster data acquisition and feature implementation
 - Do heavy work at the index build time
 - Platform for the next wave of innovations
- **Most ambitious core engineering project at eBay**
 - Entirely new codebase w/ 100+ engineers over 18 months
- **Going live in 2012**



Cassini – eBay's new Search Engine

Built on Hadoop / HBase Platform

- **Hadoop:**
 - Distributed index generation
 - Refreshed every hour
 - Re-score or re-classify entire site inventory
- **HBase:**
 - Bulk and incremental (near-real time) updates to the index
 - Stores eBay's items
 - Fast item reads for index construction
 - Fast item reads / writes for item annotation



HBase Innovations @eBay



- **Performance enhancements**
 - Load balancing
 - Small job efficiency
 - M/R scheduling optimization
 - Efficient HDFS merge
- **Improve reliability and scalability**
 - Large table operations
 - Fix 'region-in-transition' problem
 - Region availability
 - Higher number of regions / servers
- **Features**
 - Multi-tenancy
 - Isolation / Security







Thanks!

<http://www.ebaycareers.com/Home.html>

