



# eBay M2M Email Analysis on Hadoop Platform

*Forest Su, Daniel Zhang*  
October 2011



# The Birth of eBay . . .

Started with a Broken Laser Pointer . . .



. . . sold for \$14.83 USD

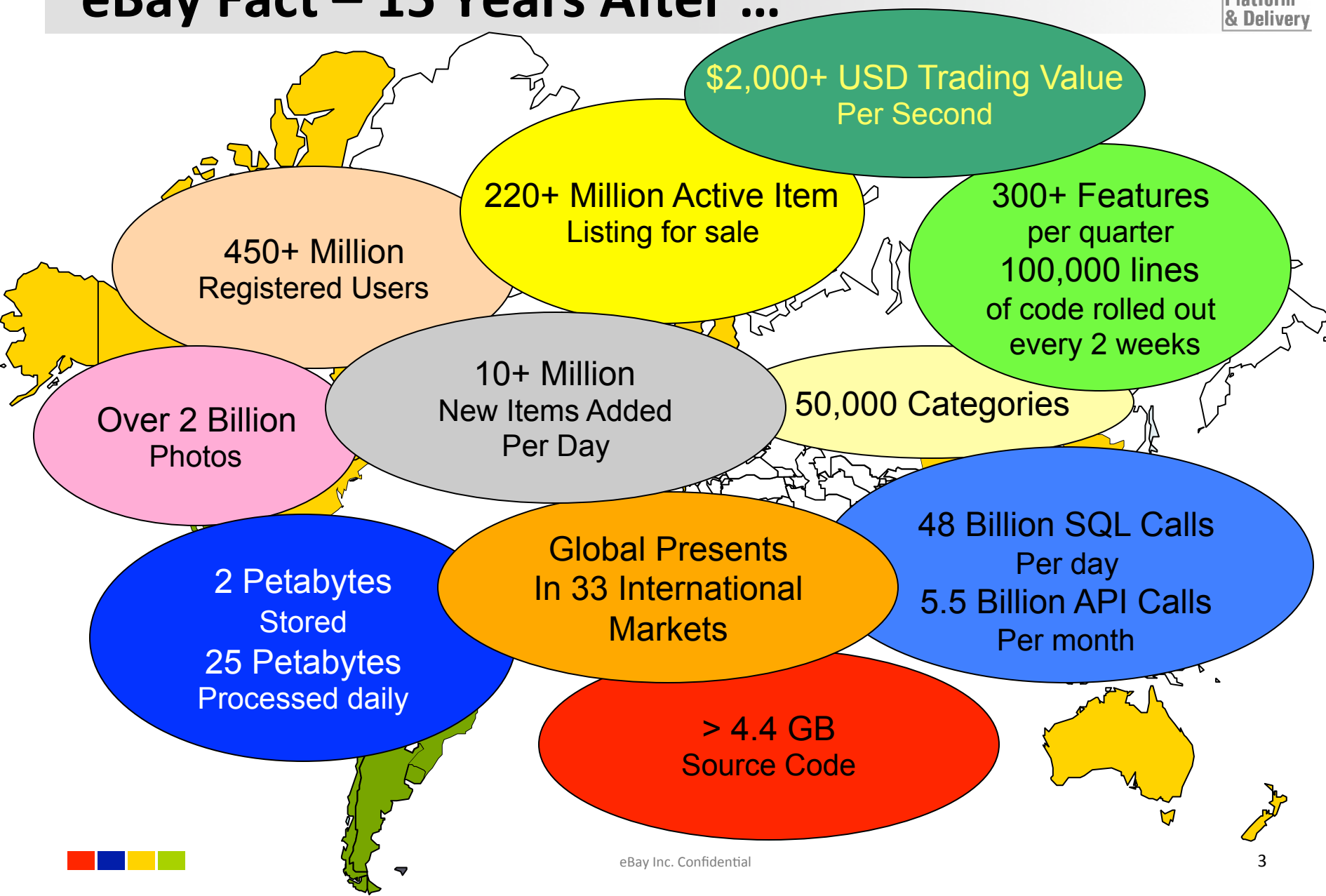
AuctionWeb was born on the Labor Day weekend in September 1995

eBay Founder



Pierre Omidyar

# eBay Fact – 15 Years After ...



# Velocity of trading



On an average day on eBay...

A Diamond Ring is sold every **two minutes**

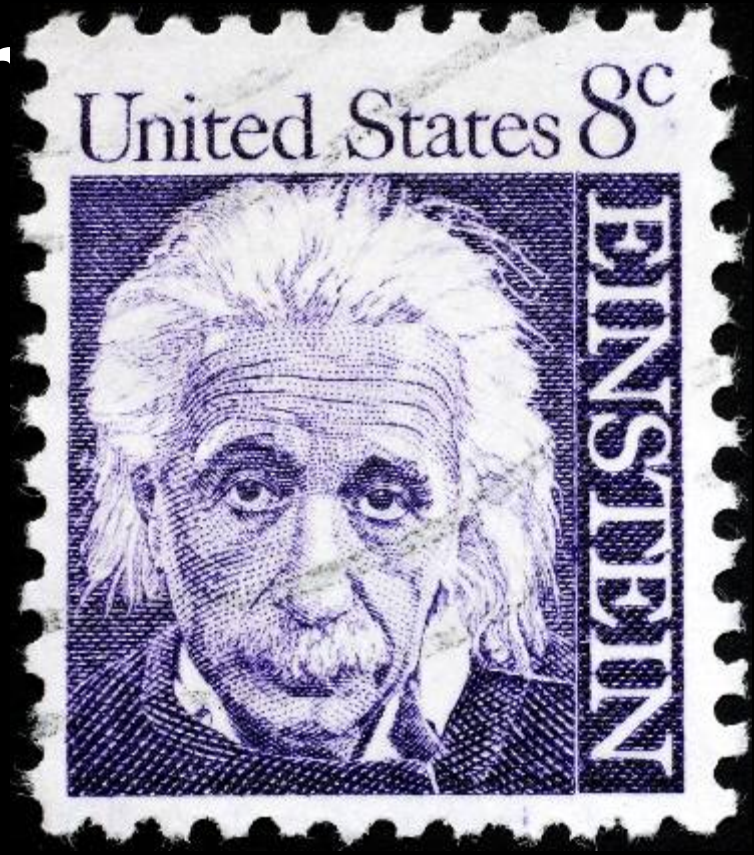
# Velocity of trading

On an average day on eBay...

Over **3600** MP3 players are sold



# Velocity of trading



On an average day on eBay...

Over **300** stamps are sold **every hour**

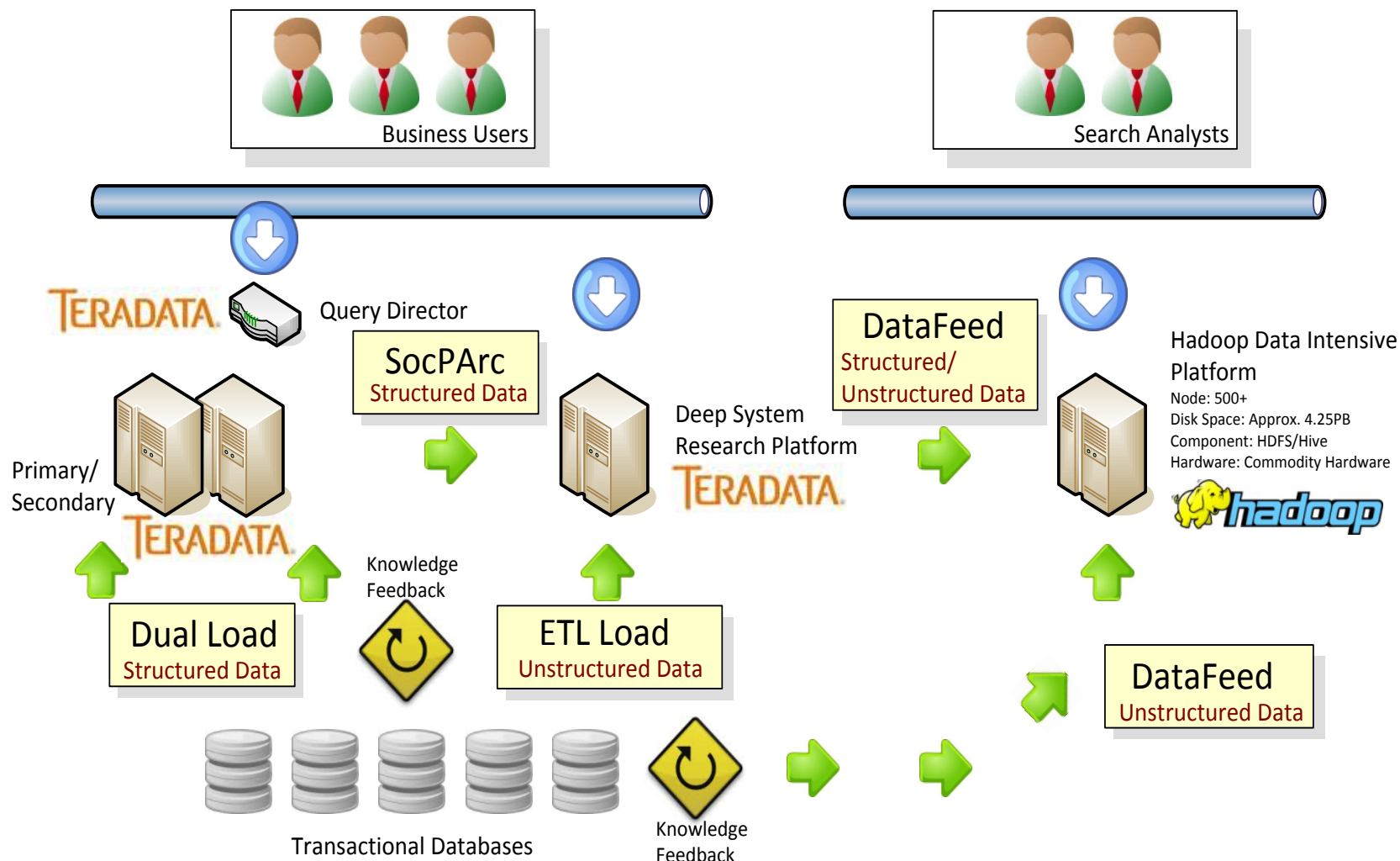
# Velocity of trading



On an average day on eBay...

An automobile is sold **every minute**

# Hadoop on the Landscape





# eBay Email Universe

- Members write to each other (“M2M”)
- Members write to eBay
- eBay writes to Members
  - Millions of times per day!
- Where are these messages?
  - In My Messages: Messages through eBay
  - In a user’s private system: Messages outside of eBay

# Single User Study

- What does a typical user send / receive?
- Follow a single user for 50 days during Oct/Nov 2010
- User Profile:
  - User subscribed to daily deal
  - User was member of:
    - eBay Bucks
    - eBay VIP
    - PayPal Advantage
  - User is mainly a buyer but also a casual seller

# Single User Study: Volume

Count	Grouping	Percentage	Percentage excl Daily Deal	Daily Average
53	eBay Daily Deal	25.1%		1.06
35	eBay Marketing	16.6%	22.2%	0.70
16	PayPal Marketing	7.6%	10.1%	0.32
87	Transaction Related	41.2%	55.1%	1.74
8	Watched item ending	3.8%	5.1%	0.16
9	Selling activity	4.3%	5.7%	0.18
3	Other	1.4%	1.9%	0.06
211	Total			4.22

# Single User Study: Visibility

Count	In MyMessages	Percentage	Percentage excl. Daily Deal
53	No (Daily Deal)	25.1%	
124	No	58.8%	78.5%
26	Inbox	12.3%	16.5%
8	Sent	3.8%	5.1%
211	Total		

Only 21.5% are visible from within eBay

# Email Data: Where is it?

Emails are stored in a mixed system:

- Metadata in DW
- Message body on NAS before 10/2010, on Grid (Low Cost Storage) since 10/2010

Email data not minable in current storage

- Queries on Email Subjects in Teradata suffer CPU skew
- Email body in individual files, does not scale at all

# eBay Sample Email



eBay sent this message to David Downie (sender\_user\_name).  
Your registered name is included to allow this message originated from eBay. [Learn more.](#)

## This member has a question for you.

Do not respond to the sender if this message requests that you complete the transaction outside of eBay. This type of offer is against eBay policy, may be fraudulent, and is not covered by buyer protection programs. [Learn More.](#)

Dear recipient\_user\_name,

what is the length of these jeans?  
thank you!!!

- sender\_user\_name



### Mens Diesel Jeans - 32" Straight Leg Cut

Item Id: 190580993728

End time: Aug-07-11 01:50:12 PDT

Buyer: sender\_user\_name (3)

100.0% Positive Feedback

Member since Apr-27-09 in Australia

Location: Queensland, Australia

Listing Status: This message was sent while the listing was active.

## Marketplace Safety Tip

- Keep your money safe - never pay for items with cash or instant money transfer services, such as Western Union or MoneyGram. In the past some sellers have exploited these payment methods in order to defraud buyers, so eBay has banned them from the site.
- Received a Second Chance Offer email? [Double check that it's genuine](#): Second Chance Offer emails come directly from eBay and are shown in your My Messages.
- Trade safely. Beware of anyone who contacts you about [buying or selling outside of eBay](#). When you trade outside of eBay, you're not able to leave feedback or take advantage of protection programs and case resolution tools available on eBay. [Report](#) an inappropriate email.
- eBay helps guard your privacy and online safety when you use our messaging tools. If you and the member you're contacting haven't bought or sold from each other recently, our tools may make both of your email addresses anonymous.

Email reference id: [a03-dc3vydjpls#]\_[b0f4a5fe814d44e39d2434f0f4ede5db#]

Please don't remove this number. eBay customer support may ask you for this number, if you should need assistance.

[Learn More](#) to protect yourself from spoof (fake) emails.

Another eBay member sent this email to dividian1@hotmail.com through the eBay platform. eBay takes no liability for the sending of this email or its content.

Visit our [Privacy Policy](#) and [User Agreement](#) if you have any questions.

You can [report this message](#) as unsolicited (spam/spoof) email.

Copyright © 2011 eBay Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. eBay and the eBay logo are trademarks of eBay Inc. eBay Inc. is located at 2145 Hamilton Avenue, San Jose, CA 95125.



# Major Challenges

- Data is large
  - Data is in individual message files:
    - millions of files per day, > 1 billion files per year
- Data contains Personally Identifiable Information (PII)
  - Hadoop current release does not support effective
    - User Access Permissions -> Need Encryption

# Challenges Answered

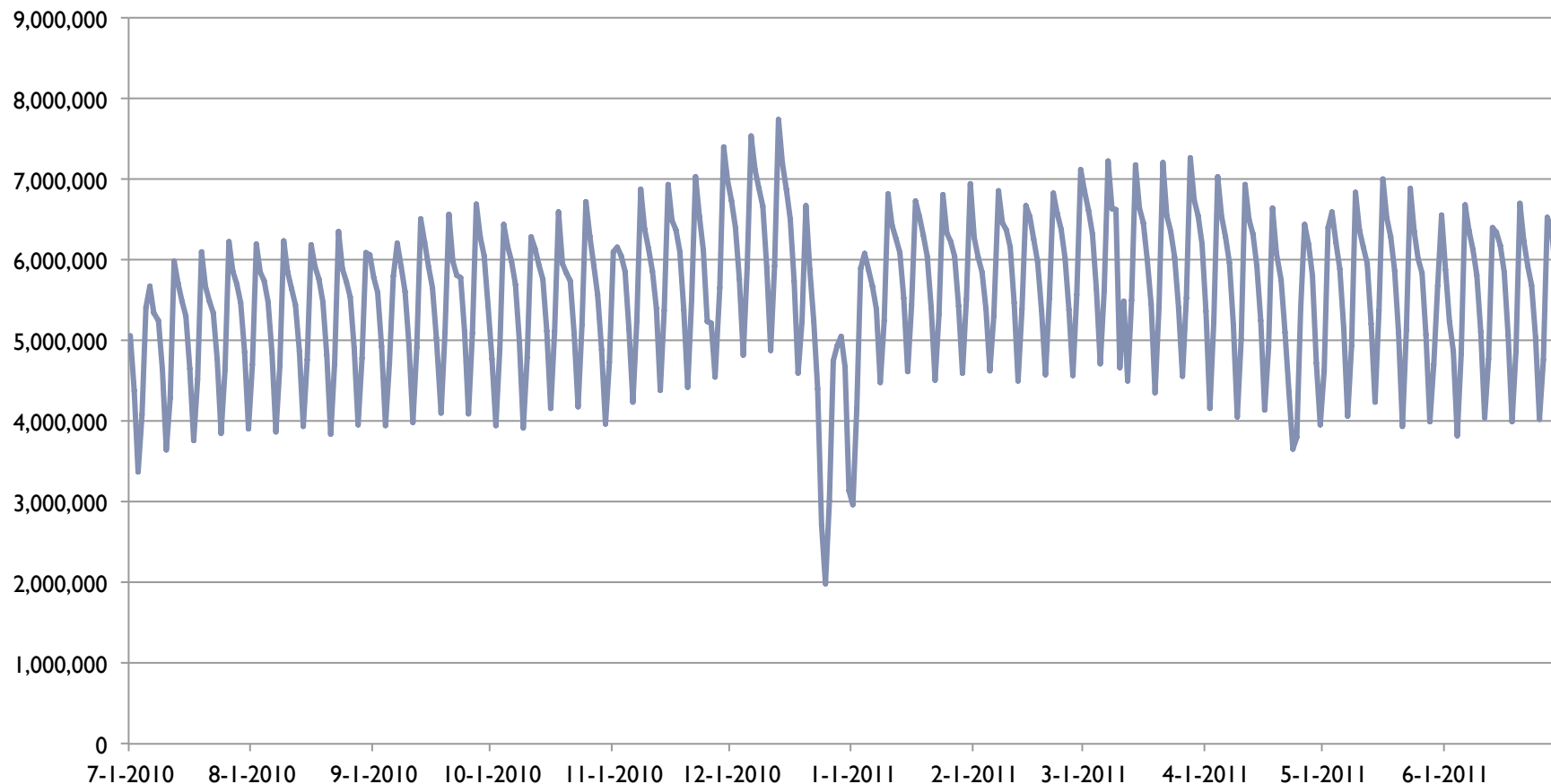
This copied over 1 billion files, about 15TB compressed

# Hadoop Data Storage Use

Year	Month	Raw M2M GigaBytes	Clean M2M GigaBytes
2009	9-12	4,060	242
2010	1-12	14,959	944
2011	1-7	10,249	1,282
Total		29,268	2,467
2011	1	1,274	85
2011	2	1,271	81
2011	3	1,650	233
2011	4	1,484	219
2011	5	1,560	230
2011	6	1,506	217
2011	7	1,505	218

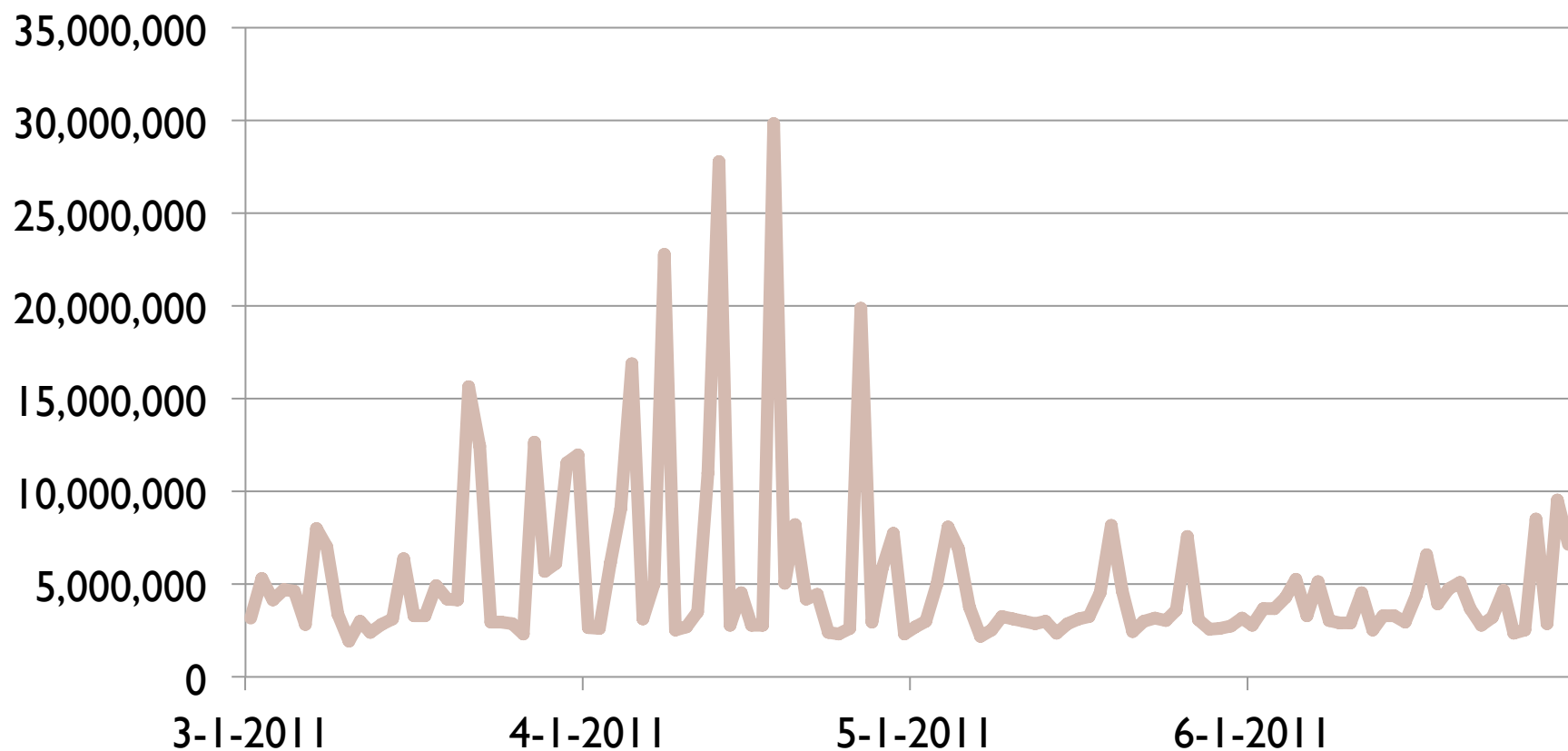
# eBay M2M Emails / Year

M2M 1 year



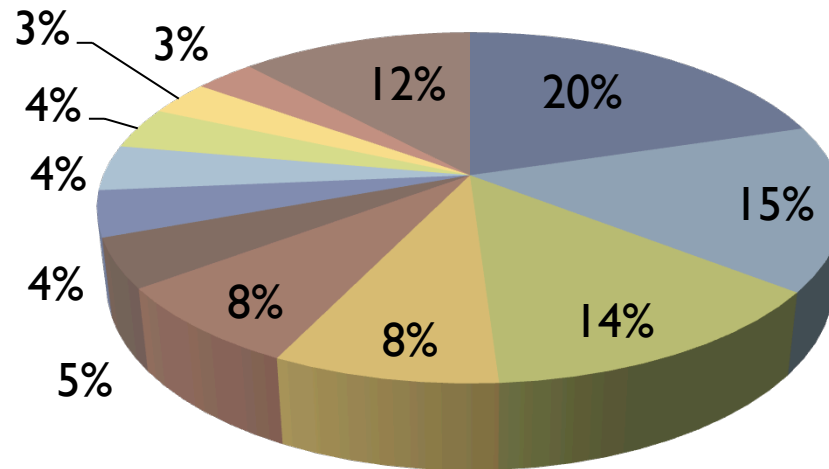
# eBay System Emails

## eBay System



# eBay System Emails

- Agreement Update
- Offers
- Seller's Return Policy
- Surveys
- Password / Email / Userid
- Welcome eBay / MyMessages
- Marketing
- Payments
- eBay Bucks
- UPI
- Cancelling Purchase
- Miscellaneous

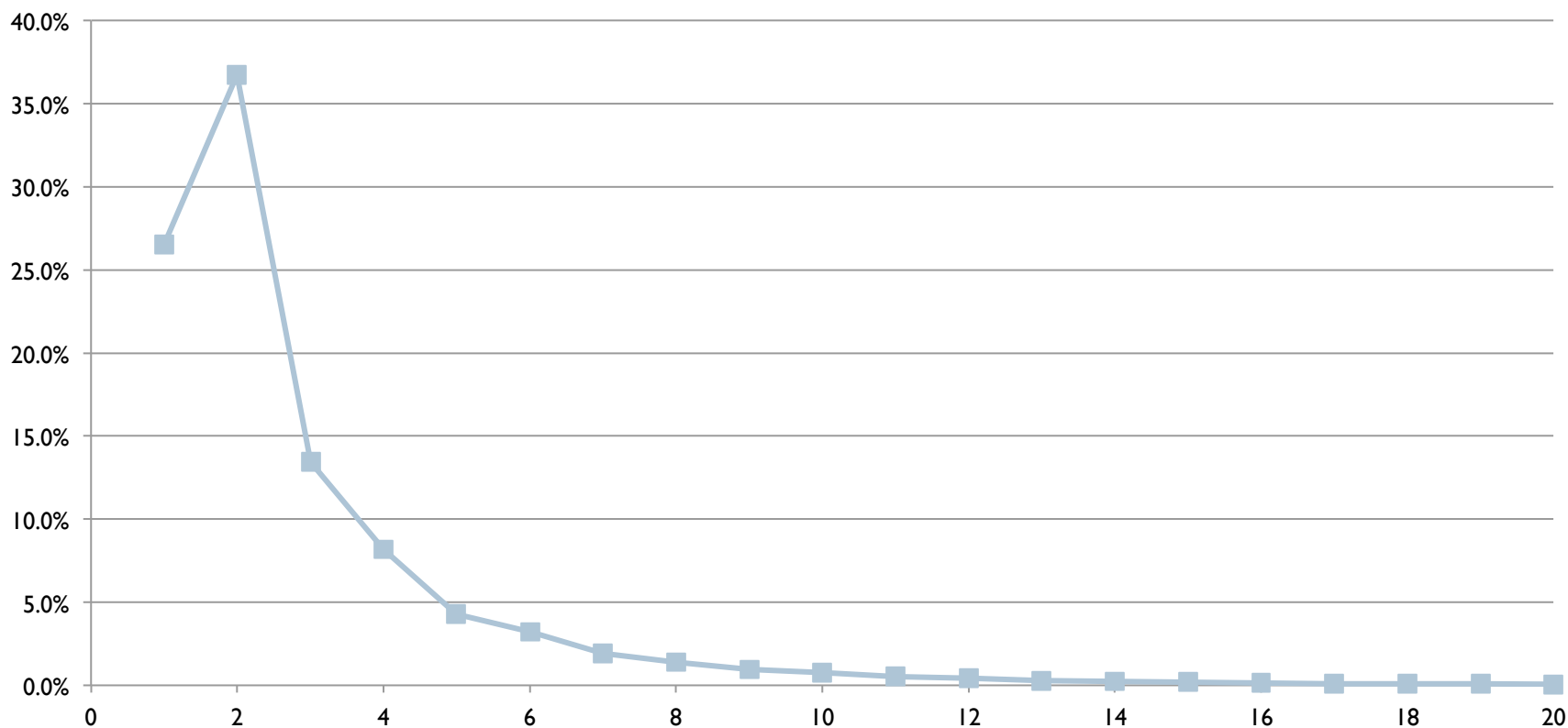




# eBay System Emails

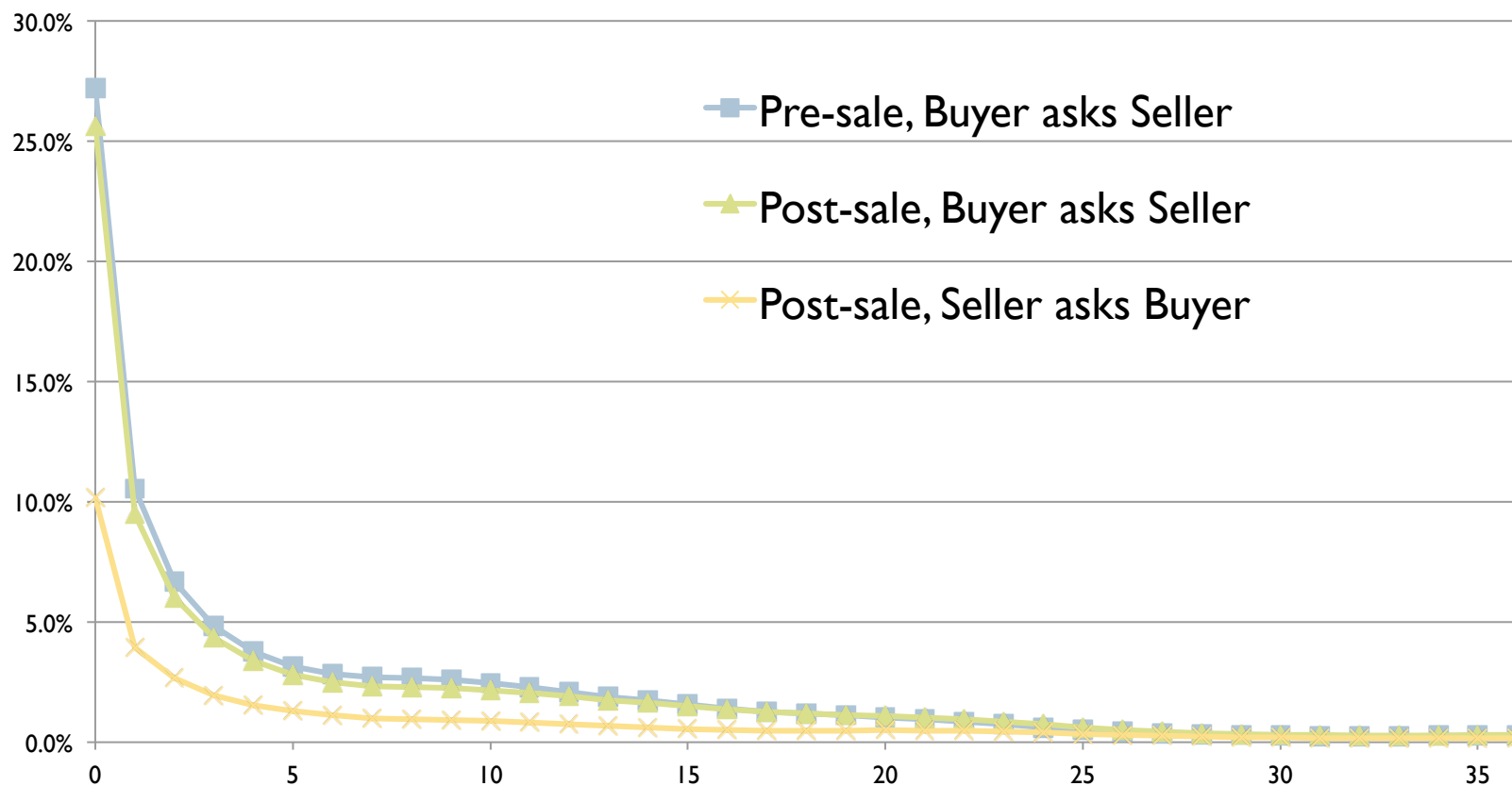
How many emails per conversation ?

**Messages per conversation**



# eBay Emails

In an M2M conversation, how long for the first reply?



# eBay M2M Emails: Profanity

Profanity detection: Identify messages containing language not becoming an eBay member

Sample messages found with simple filters:

- *f@#\$ off you mother f@#\*#\$ and do wot you want your not getin any money you c#%\$ you sad prick*
- *You f@#!#\$ c#%\$! Stop pissing around! Don't waste my time you son of a b#\$%!*
- *im saying that i want these f@#\*#\$ blake griffins. i will come and kill you if you do not AGREE!!!!!! so please ship these to me for 42 dollars and a couple south beach lebron replicas. if you do not comply, i will have to kill you. i know where you live.*

# eBay M2M Emails: Next Steps

The study of this data is yet as its infancy. Sample future use cases are:

- Identify buyers extorting refunds from sellers by threatening to leave negative feedback or to open a case (Ongoing: TnS / Buyer Abuse team)
- Identify sellers creating BBEs before negative feedback is left, or also in cases where no feedback is left.
- Identify sellers creating more than the average number of emails (ASQs), with the intent to determine on how to help these sellers to improve our buyers' experiences.



# Q&A





北京站 · 2012年4月18~20日  
[www.qconbeijing.com](http://www.qconbeijing.com) (11月启动)

QCon杭州站官网和资料  
[www.qconhangzhou.com](http://www.qconhangzhou.com)

全球企业开发大会

INTERNATIONAL  
SOFTWARE DEVELOPMENT  
CONFERENCE