

# Seeing Beyond the Cloud:

New Applications,  
Opportunities, and Challenges  
in a Cloud-Centric World

- or -

Why Archives Are Exciting!

Nathaniel S. Borenstein  
Chief Scientist, Mimecast

# Cloud Computing: The ~~Next~~ Last Big Thing

*“The half-life of knowledge in technology is 6 months.”*

-- Ian Rudy, Everyday CIO

My focus/bias:

- email, email, email
- organizational effectiveness
- business

*This presentation: ½ theory, ½ examples*

# So, What's Next?

- Organizational Transformation
- Innovation driven by IT operations
- Enterprise-scale archives enabling
  - deeper business intelligence,
  - deeper self-insight,
  - and finer control.

# Experts are narrow oracles

- A prediction's credibility depends on the relevance of the predictor's expertise

**From:** daughter

**Subject:** granddaughters



My job: seeing what's next  
for a booming cloud company

- Doesn't mean I'm right!
- But improves the odds

# Consider what the cloud enables

- Technology rarely leaps, it evolves
  - Next big thing begets the next next big thing (usually)
  - Newly mainstream technologies promote innovation
- Cloud is or will soon be “newly mainstream”
  - The next generation of millionaire entrepreneurs will credit the cloud as background infrastructure
  - (and not just because it made things cheaper to build)

*The big question:*

## What's newly possible?

# What's Newly Possible?

A prerequisite question:

## What's most different?

At least Two Big Things:

- **Data Centralization**
- **Expertise Amortization**

# What is Data Centralization?



*“more centralized than yesterday”*

# Truly Centralized Data

- For the first time in history, it is possible to
  - store all kinds of data,
  - from all parts of an organization,
  - together, with unified efficient access to all of it.
- **Even IBM has never had that!**
- We face new vistas in business intelligence
- New applications can use this new resource.

*An app store for cloud applications?*





## A Tale of Expertise Amortization



Mimecast USA:  
The Waltham Watch Factory

**“All of This Has Happened Before and Will Happen Again”**

# Amortizing Expertise

- The average business is short on IT Expertise.
  - Cloud providers amortize this.
  - Already outperform in-house for commoditized services
- Next: Amortized business intelligence tools
  - Most businesses don't know what their data reveals
- Business data

➔ intelligence

➔ insight

➔ profit.

# Who Will Amortize Intelligence?

- Even though I work for a cloud provider:
  - I'm wary of too much power for them
  - I'm skeptical of their doing too many things (e.g. BI)
- Archive-driven apps can earn profits
  - for cloud providers
  - for third parties.
- *Open app stores, are better for customers*
  - *No monopolies*
  - *Specialized Expertise*
  - *But necessarily complex security*

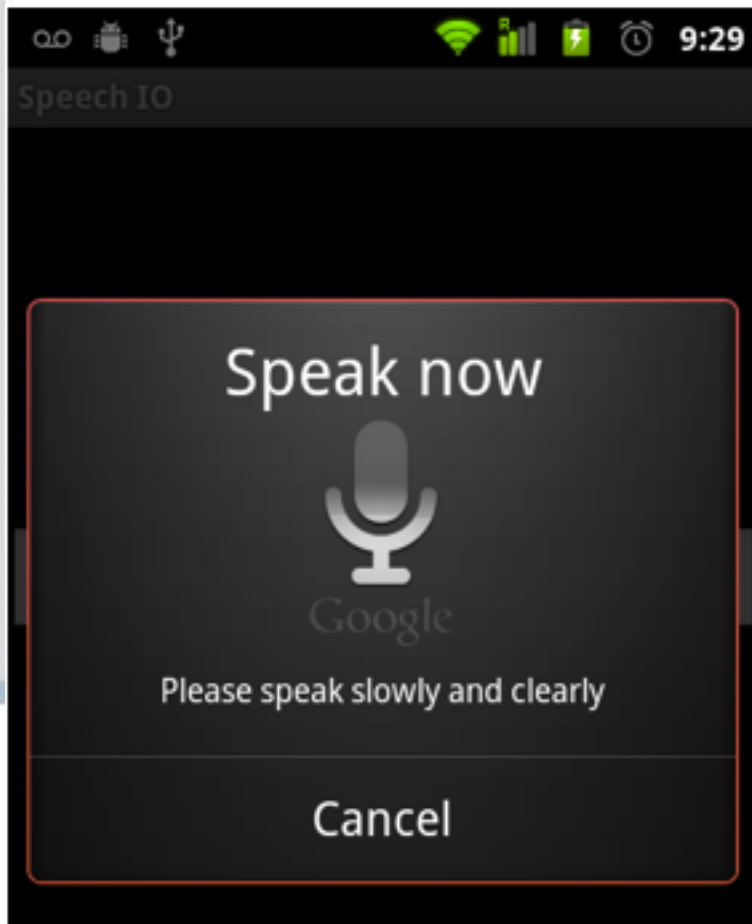
# Professional IT is Coming Even To Small Businesses

- Cost of providing services goes down.
- Quality of services goes up.
- Cost of innovation is broadly amortized.
- A Mass Market for Business IT
- Longer Tail for business apps?

*A Taste of the Future:*

*Individuals with Smartphones*

# Business needs what I have!



- My phone knows *everything*, it seems
- Great for individuals
- Challenging for enterprises
- Organizations need
  - Complex data protection
  - Enterprise-scale data mining applications
  - Cross-device integration

# Why is Business Data Harder?

1. There's so much more of it
2. It comes in so many forms
3. It comes in so many contexts
4. It requires complex permissions

(for security and privacy)

*It's hard to set the context  
for knowing **everything!***

*Why is business data harder? #1:*

# There's so much more of it

- Legal and corporate documents
- Customer files
- Technical plans
- Reputation Information
- Artifacts Per Employee  
(APE) Ratio is Surely Rising





*Why is business data harder? #2:*

# It comes in so many forms

- Readable documents (text/Word/HTML/etc)
- Pictures (faxes, scans, security cameras...)
- Sounds (phone or meeting recordings, voice mail, security monitoring...)
- Video (training, videoconference recordings, webinars, security cameras)





*Why is business data harder? #3:*

# It comes in so many contexts

- Informal (Email, phone or in-person conversation or meeting recording)
- Formal (phone, meeting, presentation)
- Informal Document (meeting planning, shopping list, informal plan, racquetball reservation)
- Formal document (Legal document, medical or financial record, Project plan, HR Documents)
- Work Product/Drafts: (Legal doc, position paper, marketing doc, Web site, software, entertainment,
- Sensitive business-critical document



*Why is business data harder? #4:*

# Security and privacy require complex permissions

- Just me, or named individuals
- Team relationships
- Management relationships (up/down, all the way)
- Customer relationships (my accounts)
- HR relationships (employee sensitive material)
- Related job classifications or seniority levels
- Security clearance

# It's Harder for Business, but it's still possible!

- People are simple to model: content-consuming product-purchasing machines with friends
- Business is more complicated
- But IT has coped with more complex problems.
- A lot of specialized expertise must build up.
- But all your business data combined is probably much less than a micro-Google.

*If only you could use it intelligently!*

# Central Archiving is the Key

- Cloud computing is just the delivery model.
  - Albeit the first to make it practical!
- Need two kinds of apps, app stores:
  - Unified store & API simplify app-building
  - Some apps can run externally, use an API
  - Data-intensive apps must run “inside”
  - Need both to get full benefit

# External Apps Using an API

- Add new data and data types to an archive
- Customer-driven interaction and reporting
- Visualization of aggregated statistics
- Apps using search/analysis primitives provided by the archive
- Deeper analysis of smaller data (sub)sets

*Applications limited by quantity and bandwidth*

# Data Intensive Archive Apps

- Data-intensive apps must be “inside the cloud”
- Vendors need provisions for running customer and third party code without disruption.
- Permits deep analysis of enormous data sets
- Communications patterns and trends







Time for some examples...

# I. Instant Implicit Search (a)

Date: January 28, 2011 9:06:52 PM GMT  
From: John Doe <[jdoe@company1.com](mailto:jdoe@company1.com)>  
To: Mary Roe <[more@treefrog.com](mailto:more@treefrog.com)>  
Subject: Proposed alliance

Mary – I've been thinking

## Potentially relevant links:

-  7-May-09 Mary Roe-> Me  
*"You think too much"*
-  23-Sep-08 Me -> Mary Roe  
*"Our companies are natural allies"*
-  12-Dec-10 Mary Roe -> Me  
*"We should try again to get our companies to work together"*
-  8-Feb-09 Me -> Bill Boe  
*"I proposed to Judy last night!"*
-  Today: New law proposed to penalize drivers whose cars kill does
-  Yesterday: New challenge proposed to Roe v. Wade

# Instant Implicit Search (b)

Date: January 28, 2011 9:06:52 PM GMT  
 From: John Doe <[jdoe@company1.com](mailto:jdoe@company1.com)>  
 To: Mary Roe <[more@treefrog.com](mailto:more@treefrog.com)>  
 Subject: Proposed alliance

Mary – I've been thinking that it might make a great deal of sense for our two companies to jointly form an alliance with [DeathStar.com](http://DeathStar.com).

## Potentially relevant links:

-  23-Sep-08 Me -> Mary Roe  
"Our companies are natural allies"
-  14-Jul-10 Steve Xu -> Me  
"[DeathStar.com](http://DeathStar.com) opportunities"
-  21-Oct-10 Steve Xu -> Bob Ray  
"[Competitive threat: Deathstar.com](http://DeathStar.com)"
-  2-Dec-10 Bob Ray -> Me  
"Figuring out [Deathstar.com](http://DeathStar.com)"
-  12-Jan-11 "[Deathstar.com](http://DeathStar.com) announces record profits"
-  15-Jan-11 "[Deathstar.com](http://DeathStar.com) declares largest-ever stock dividend"









# Instant Implicit Search (c)

Date: January 28, 2011 9:06:52 PM GMT  
 From: John Doe <[jdoe@company1.com](mailto:jdoe@company1.com)>  
 To: Mary Roe <[more@treefrog.com](mailto:more@treefrog.com)>  
 Subject: Proposed alliance

Mary – I've been thinking that it might make a great deal of sense for our two companies to jointly form an alliance with [DeathStar.com](http://DeathStar.com).

In particular, now that [Treefrog](http://Treefrog) has become an authorized reseller of Acme Widgets' product line, it seems sensible to me for us all to work together

## Potentially relevant links:

- 23-Sep-08 Me -> Mary Roe  
 "Our companies are natural allies"
- 14-Jul-10 Ali Bye -> Me  
 "New products from Acme"
- 21-Oct-10 Steve Xu -> Bob Ray  
 "Competitive threat: Deathstar.com"
- 12-Nov-10 Bob Ray -> Me  
 "Who do we know at Acme?"
- 23-Jan-11  "Deathstar.com sues Acme in \$23M patent brawl"
- 25-Jan-11  "Acme countersues Deathstar; CEO says 'This is war!'"

# Instant Implicit Search (d)

Date: January 28, 2011 9:06:52 PM GMT  
 From: John Doe <[jdoe@company1.com](mailto:jdoe@company1.com)>  
 To: Will B. Flexible <[willb@deathstar.com](mailto:willb@deathstar.com)>  
 Subject: Proposed alliance

Will— I just heard about the escalating war between Deathstar and Acme. I think we might be able to help you mop the floor with Acme and even some of their resellers such as those bozos at Treefrog.

## Potentially relevant links:

- 14-Jul-10 Ali Bye-> Me  
 "New products from Acme"
- 21-Oct-10 Steve Xu -> Bob Ray  
 "Competive threat: Deathstar.org"
- 12-Nov-10 Bob Ray -> Me  
 "Who do we know at Acme?"
- 23-Jan-11 "Deathstar.org sues Acme in \$23M patent brawl"  

- 25-Jan-11 "Acme countersues Deathstar; CEO says 'This is war!'"  

-  Dictionary: betrayal: disloyalty, treachery, bad faith, faithlessness, falseness, duplicity, deception, ...

## II. Overnight Implicit Search

Date: January 28, 2011 9:06:52 PM GMT  
From: Mimecast Implicit Search <[searchbot@company1.com](mailto:searchbot@company1.com)>  
To: John Doe <[jdoe@company1.com](mailto:jdoe@company1.com)>  
Subject: Links that might be useful

Based on the messages you sent and received yesterday, we have found the following links to information that we think you might find useful.

*14-Jul-10 Ali Bye-> Me "New products from Acme"*

*21-Oct-10 Steve Xu -> Bob Ray "Competitive threat: Deathstar.org"*

*12-Nov-10 Bob Ray -> Me "Who do we know at Acme?"*

*23-Jan-11 "Deathstar.org sues Acme in \$23M patent brawl"*

*25-Jan-11 "Acme countersues Deathstar; CEO says 'This is war!'"*

To further customize the search process, use this link: <http://company1.com/implicit/customize>

## II. Proactive Permission Management

From: Automated Search Robot  
To: Orlando Scott-Crowley  
Subject: Can we help Nathaniel?

The automated search robot has determined that it is likely that it would help Nathaniel Borenstein to do his job if we showed him an email message in your archive:

Active link

12-22-01 **Bobby Valentine** *How the Mets will win it all this year*

Do you want to permit Nathaniel to see this message?

Yes, just this message

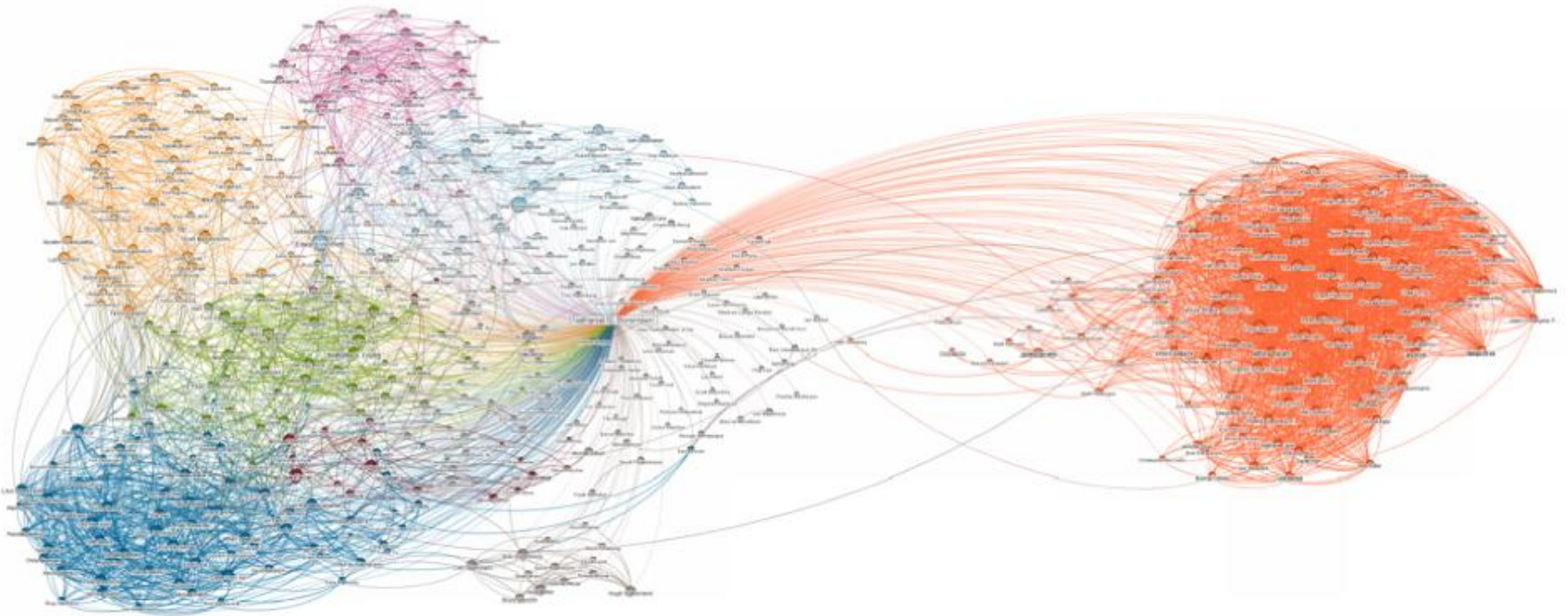
Yes, and show him anything helpful  
in the future without asking

No, do not show it to him

No, and never ask again about  
showing things to Nathaniel

## IV. Enterprise Social Network Analysis

- You've probably seen diagrams like this one from LinkedIn, showing one person's network:



A map of just one person's network is surprisingly useful!

# Deeper Social Network Analysis

- Enterprise organizational network explorers are more complex, interactive
- It's easy to see patterns; are they useful?
- Usually a few people are key connectors
- And most key connectors...  
...get bad performance reviews, *are laid off first!*
- *Cloud archive-enabled social network apps can lead to better HR management!*



# V. Communication Pattern Analysis

- Warning: 5 employees may be in a conspiracy
- But beware of surprise parties, etc. (Dr. B. Aven, CMU)

The screenshot shows the Science News website interface. At the top left is the logo for 'the Atlantic'. The main header features the 'Science News' logo with the tagline 'MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC'. To the right of the logo are several category links: ATOM & COSMOS, BODY & BRAIN, EARTH, ENVIRONMENT, GENES & CELLS, HUMANS, LIFE, and MATTER & ENERGY. Below the header is a navigation menu with links for HOME, NEWS, FEATURES, BLOGS, COLUMNS, DEPARTMENTS, RSS FEEDS, and E-MAIL ALERTS. The main content area displays the article title 'Information flow can reveal dirty deeds' and a sub-headline 'Analysis of Enron e-mails reveals structure of corrupt networks'. On the left side, there is a profile for Alexis Madrigal with a photo and links to his posts and email. At the bottom of the screenshot, there are social media sharing icons for '+ SHARE', Facebook, LinkedIn, and Twitter, along with a 'Previous' link.

## Enron Emails Reveal What a Web of Deceit Really Looks Like

JUL 13 2011, 11:54 AM ET | 12

+1 12 Recommend 60



# Organizational Matchmaking

- Joe in Boston, Su in Beijing, and Padma in Bangalore are researching the same thing
- They haven't met, and aren't likely to.
- An archive can notice their common interests and introduce them.
- New teams beget new discoveries, products, and profits.

*Archive-based BI will become  
a key part of your team!*



## VI. Information Security Analysis

- How many sensitive artifacts do you have?
- Which are most widely available internally?
- Who has access to the most sensitive objects?
- How has your document security evolved?
- Which access rights make the least sense?

## VII. Regulatory Compliance Analysis

- Today, we scan emails for credit card #'s, etc.
- Tomorrow: Scan all for HIPAA violations, etc.
- Potential cost savings for audits/certifications
  
- Which employees need extra training?
- Where are risks lurking unnoticed?

# From Archives to Information Banks

- Archives are where information goes to die
- Information Banks are highly centralized repositories that repurpose your information and pay interest
- Information Banks have API's, diverse applications
  - Archives are safe deposit boxes

*It's real today.*

# From Information Banks to Organizational Memory

- Information Banking: When Archives Come Alive
- Organizational Memory: When They Get Smart
  - *Where in this organization can I find the answer?*
- OM is a decades old dream of efficiency
  - Examples: Answer Garden, Mr. Wizard

*It's finally almost here!*

# So, Everything's Just Great?

- Not so simple, of course
- More power raises the stakes
- Provider selection is more important than ever
  - Bad guys can make more mischief with more data
  - Security, Privacy, Availability, Reliability, .....
- Don't want employees to feel spied on
  - Aggregation increases the sensitivity of each datum
- What role for the government?

# What's that about provider selection?

- Provider selection is a decision that *must be taken with inadequate information!*
- Customer references are vital.
- Monitoring as outsourcing is vital.
- Emerging standards for evaluating cloud providers will help eventually, perhaps.

# It Matters How Employees Feel

- Advanced archival analysis will fail without employee support.
- Give them fine-grained Opt-Out
- Employees deserve private contexts.
- Kids' teachers, medical appointments,  
(generally being human)

*Yes, there's a risk, but...*

# Resist the Temptation to Spy

- You're trying to cope with the information you already have, not to collect more.
- You need your employees' help
- Spying is not an antidote for mistrust
- Tell users how to opt out of monitoring  
... and you'll monitor nearly everything
- You're not doing this to catch bad employees.

*But opt-out overuse  
is a useful red flag.*



# And then there are governments...

- Must everything be discoverable?
- Potential for conflicts of interest
- But you have to follow the law...  
... in many, many jurisdictions!

- *Unprecedented need for sophisticated compliance, and irreversible deletion tools*

*Best left to the cloud providers!*



# Implications for Universities and Research Labs

- Institutions inclined to sharing could do even more
- Organizational matchmaking on a global scale, or even a consortium scale
- Challenges: Coordination of archive providers, management of permissions



# The Bottom line

- In certain ways, consolidate archives will grow smarter than we are.
- Corporations can't afford to ignore this stuff
- Pioneers can get competitive advantage
  - Or arrows in their backs!
- You've got cloud infrastructure, now exploit it!
- Don't bet against efficiency and optimization!

Questions? Comments?

Nathaniel Borenstein  
<nsb@mimecast.com>