Big Data, Big Bets, Big Opportunities

Accelerate Discovery with Visual Analytics

Limit Surprise with Smart Data

Detect Patterns with Mass Correlation

FY11 FY12 FY13+

Ira A. (Gus) Hunt
Chief Technology Officer
18 October 2011
Our Mission

We are the nation's first line of defense. We accomplish what others cannot accomplish and go where others cannot go. We carry out our mission by:

Collecting information that reveals the plans, intentions and capabilities of our adversaries and provides the basis for decision and action.

Producing timely analysis that provides insight, warning and opportunity to the President and decisionmakers charged with protecting and advancing America's interests.

Conducting covert action at the direction of the President to preempt threats or achieve US policy objectives.
1. Keep our **Systems Up**
   - CIA doesn’t work when IT doesn’t work

2. Keep us on the **Cutting Edge**
   - Innovate and deliver solutions for Mission

3. **Mission First**
   - Focus on the business of CIA
<table>
<thead>
<tr>
<th>Big Bet</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Big Data</strong></td>
<td>Acquire, federate, and position for multiple constituencies to securely exploit. Grow the haystack, magnify the needles.</td>
</tr>
<tr>
<td><strong>2 Operational Excellence</strong></td>
<td>Innovate infrastructure operations and provisioning, create an authoritative source on our asset base, and run IT like a business.</td>
</tr>
<tr>
<td><strong>3 Serve CIA by supporting the IC</strong></td>
<td>Assume a leadership role in IC activities that benefit the CIA</td>
</tr>
<tr>
<td><strong>4 Talent Management</strong></td>
<td>Focus on continuous learning and diversity of thought, experience, background</td>
</tr>
</tbody>
</table>
## 5 Key Technology Enablers

<table>
<thead>
<tr>
<th></th>
<th>Advanced Mission Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>- World-class abilities to discover patterns, correlate information, understand plans and intentions, and find and identify operational targets in a sea of data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Enterprise Widgets and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>- A customizable, integrated and adaptive webtop that lets analysts, ops officers, and targeters to “have it their way”.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Security as a Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>- One environment, all data, protected and secure using common security services such as: ubiquitous encryption, enterprise authentication, audit, DRM, secure ID propagation, and Gold Version C&amp;A.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Enterprise Data Management—the Data Harbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>- An ultra-high performance data environment that enables CIA missions to acquire, federate, and position and securely exploit huge volumes data.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Cloud Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>- Ruthlessly standardized, rigorously automated, dynamic and elastic commodity computing environment. Massive capacity ahead of demand. Speed for mission need.</td>
</tr>
</tbody>
</table>
1. **Just Imagine if:**
   - Time and capacity **were not** constraints
   - **All data** was securely and appropriately available to those who need it
   - Data security was **automatic, seamless, and persistent**
   - Answers were computed **ahead** of the questions
   - Users had the tools they need to **do their job** their way
   - Sophisticated data engines **worked** for people not vice versa
   - IT was **Frictionless**

2. **We would Change the Conversation with Mission**
   - EIT becomes an **enabler**
   - The need for Mission to seek alternatives evaporates
   - Mission can focus on mission execution, not IT management
   - Overall Agency IT costs decline
KTE Focus

1. Provide the **Opportunity** to Change the Conversation
   - Speed of delivery, capacity ahead of demand, security ahead of need, answers ahead of questions

2. Deliver the **“Two-fer”**
   - Make Big Data real, while
   - Enabling Operational Excellence

3. Make IT **Frictionless** for Mission
   - Zero wait, zero effort
   - Like 110 out of the wall, plug in and go
1. **Capacity** ahead of demand

2. **Security** ahead of need

3. **Answers** ahead of questions

**Change the Conversation**

Make IT Frictionless for Mission
BIG

Our world by the numbers
7/25/2008
Google passes 1 Trillion Indexed URLs (1,000,000,000,000)
7/25/2008—Google passes 1 trillion URLs

$187
Cost per second of Ebay’s last outage
$16,156,800/Day
7/25/2008—Google passes 1 trillion URLs

789,400,000,000,000,000

$187/second—Cost of last Ebay outage
Current size of YouTube in bytes (789.4 PB)
7/25/2008—Google passes 1 trillion URLs

789.4 PB—Size of YouTube

2/4/2011

$187/second—Cost of last Ebay outage
IPv4 address space is exhausted, 4.3 billion addresses have been allocated
340 undecillion \( (10^{38}) \)

34,000,000,000,000,000,000,000,000,000,000,000,000,000

2/4/2011—IPv4 address space exhausted

7/25/2008—Google passes 1 trillion URLs

789.4 PB—Size of YouTube

$187/second—Cost of last Ebay outage
Size of the IPv6 address space
(340x10^{38})
2/4/2011—IPv4 address space exhausted

7/25/2008—Google passes 1 trillion URLs

789.4 PB—Size of YouTube

100 million gigabytes

340\times 10^{38}—Size of IPv6 address space

$187/\text{second}—\text{Cost of last Ebay outage}$
Size of Google’s index
144 million

7/25/2008—Google passes 1 trillion URLs

789.4 PB—Size of YouTube

2/4/2011—IPv4 address space exhausted

100 million gigabytes—Size of Google’s index

340x10^{38}—Size of IPv6 address space

$187/second—Cost of last Ebay outage
Number of Tweets per day
2/4/2011—IPv4 address space exhausted

7/25/2008—Google passes 1 trillion URLs

144 million—Number of Tweets per day

789.4 PB—Size of YouTube

1,700,000,000,000,000

100 million gigabytes—Size of Google’s index

340x10^{38}—Size of IPv6 address space

$187/second—Cost of last eBay outage
Number of objects in a small private startup database
7/25/2008—Google passes 1 trillion URLs

144 million—Number of Tweets per day

789.4 PB—Size of YouTube

90,000,000,000,000,000

1.7 trillion—Items in a startup’s DB

100 million gigabytes—Size of Google’s index

340x10^{38}—Size of IPv6 address space

$187/second—Cost of last Ebay outage
Facebook data holdings (90PB)
7/25/2008—Google passes 1 trillion URLs

144 million—Number of Tweets per day

789.4 PB—Size of YouTube

4,600,000,000

1.7 trillion—Items in a startup’s DB

100 million gigabytes—Size of Google’s index

340x10^{38}—Size of IPv6 address space

$187/second—Cost of last Ebay outage
Number of wireless devices in the world
1,000,000,000

2/4/2011—IPv4 address space exhausted
90PB—Facebook data holdings

7/25/2008—Google passes 1 trillion URLs

144 million—Number of Tweets per day

789.4 PB—Size of YouTube

1.7 trillion—Items in a startup’s DB

4.3 Billion—Mobile devices

100 million gigabytes—Size of Google’s index

340x10^{38}—Size of IPv6 address space

$187/second—Cost of last Ebay outage
PCs and Laptops
6,904,000,000 People
4,600,000,000 Phones
1,000,000,000 PCs
1. 1.3 Billion—China
2. 1.2 Billion—India
5. 311 Million—US
6. 237 Million—Indonesia
1. 1.3 Billion—China
2. 1.2 Billion—India
3. 800 Million—Facebook
5. 311 Million—US
6. 237 Million—Indonesia
1. 1.3 Billion—China
2. 1.2 Billion—India
3. 800 Million—Facebook
4. 500 Million—Microsoft Online
5. 311 Million—US
6. 237 Million—Indonesia
Connected World = Big Data

- 7/25/2008—Google passes 1 trillion URLs
- $187/second—Cost of last eBay outage
- 1 Billion—PCs and Laptops
- 789.4 PB—Size of YouTube
- 144 million—Number of Tweets per day
- 6.9 billion people
- 2/4/2011—IPv4 address space exhausted
- 340x10^{38}—Size of IPv6 address space
- 1.7 trillion—Items in a startup’s DB
- 4.3 Billion—Mobile devices
- 100 million gigabytes—Size of Google’s index
- 800 million Facebook users
- 1 Billion—PCs and Laptops
- $187/second—Cost of last eBay outage
- 4.3 Billion—Mobile devices
2
Driving Forces
Mobility  Cloud
Provides

Simple to use
Always on
Everywhere
User added functionality

Requires

Continuous Services
Unbounded Scaling

See Cloud
Continuous Services
Unbounded Scale
User added functionality

Big Data
Advanced Analytics

Radical Efficiency
Rigorous Standardization
Ruthless Automation
Failure Tolerant
Redundant by Design
Available by Default
Makes IT Frictionless
Frictionless IT

The new revolution in IT

Most significant paradigm shift to date

Bigger than mainframe to PC

Bigger than PC to LAN

Bigger than the Internet

Innovation engine of the future

Start Fast, Build Fast, Succeed Fast, Scale Fast

Or

Fail Cheap
Animoto and Amazon EC2

Amazon EC2 easily scaled to handle additional traffic

Peak of 5000 instances

Launch of Facebook modification.

Steady state of ~40 instances
Our Job

1. Leverage the Big Data world
2. Find the Information that Matters
3. Connect the Dots
4. Understand the Plans of our Adversaries
   - Prevent an attack, Save lives, Safeguard our national security
Why We Care
Why We Care
The Problem
**Our Problem: Which 5K**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Don’t know the future <strong>value</strong> of a dot <strong>today</strong></td>
</tr>
<tr>
<td>2</td>
<td>We cannot connect dots we <strong>don’t</strong> have</td>
</tr>
<tr>
<td>3</td>
<td>The <strong>old</strong> collect, winnow, dissem <strong>model</strong></td>
</tr>
<tr>
<td></td>
<td><strong>fails</strong> spectacularly in the <strong>Big Data</strong> world</td>
</tr>
</tbody>
</table>

The few cannot know the needs of the many

Secure the data, Connect the data, Empower the user
Let’s talk Cloud
Cloud Definitions

- **Cloud, n.** A visible mass of vapor, especially one suspended in the sky
- **Cloud, v.** To darken; obscure; threaten
- **Cloud, it.** The single most over-hyped term in the history of Information Technology
- **Cloud = 42**
Cloud Definitions

Public

Commercially available compute storage and hosting

Private

Operated, managed and controlled by a private enterprise

Hybrid

Private cloud that expands into the public cloud for occasional peak demand workloads
Cloud@CIA

Public

Private cloud that expands into the public cloud for occasional peak demand workloads

Managed, secured, audited, & operated under our control in our spaces!!!

Hybrid

Commercially available compute storage and hosting

X

X
Cloud, cia: Private, Large Scale
Cloud@CIA

Cloud, cia: Private, large scale,

Dynamic, Elastic

Computing, storage, and networking resources can be rapidly provisioned and released
Cloud, cia: Private, large scale, dynamic, elastic

Automated

Minimal management effort or service provider interaction--“Zero touch”
Cloud, cia:  Private, large scale, dynamic, elastic, automated,

Commodity

Built from commodity components, period. Follow Moore’s Law for both cost and performance
Cloud, cia: Private, large scale, dynamic, elastic, automated, commodity computing environment

An essential investment to enable our Missions
Why Cloud?

- **Strategic and technical direction**
  - Inexorable trend in computing
  - Commodity wins—PERIOD!

- **Efficiency and effectiveness**
  - Commodity pricing, commodity cost curve
  - Highly-automated management and provisioning
  - Rigorous standards at the OS and middle-ware levels
  - Green for “Free”

- **Speed and agility**
  - Capacity and scale ahead of demand
  - Dynamic and elastic provisioning
Unprecedented capabilities for mission

- Peta-scale capacity for peta-scale problems
- Big data, big analytics, big performance
- Connect the dots—find/stop the next underwear bomber or IED
- Accelerated targeting, Enhanced asset validation
- Correlation not search; verbs as well as nouns; patterns not just words
  - Boolean is broken
- Enhanced data analytics and visualization—volume is our friend
- Better security—encrypted enterprise
- COOP vs DR
Let’s talk Security
Security Matters

1. Share with **assurance**

2. They really are **out to get us**

3. It is a **life and death** matter
Security Maxims You Know
(by Heart)

1. Absolute security is absolutely impossible

2. People are the weak link

3. Current methods are insufficient and unsustainable
A Different Approach

1. Assume you’ll be had (ie: the trusted insider)

2. Minimize the Opportunity

3. Minimize the Loss

4. Elevate protections and Control your world
How many times do you **build** security?

Build **common** security services

Use them **everywhere**

Ruthlessly **protect** and **control** them
Security Services

1. **People**—Directory Services, Identity Management, PKI, Secure Identity Propagation

2. **Data**—Ingestion and Tagging, Signing, and DRM

3. **People & Data**—Access Control Service, Audit

4. **Enterprise Encryption**
Necessary Steps

1. Separate
   - data from applications
   - applications from security
   - security from rules
   - rules from data

2. Separate
   - roles and responsibilities from accesses
3. Implement transparent end-to-end encryption

4. Encrypt and sign all objects—data, people, devices

5. Enable shades of gray—it’s not a yes/no world

6. Distribute permission authority
<table>
<thead>
<tr>
<th>Step</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Build gold <strong>images</strong> then sign and <strong>vault</strong> them</td>
</tr>
<tr>
<td>2</td>
<td>Constantly <strong>reimage</strong> from the vault</td>
</tr>
<tr>
<td>3</td>
<td>Embrace elastic computing—make it a <strong>shell game</strong></td>
</tr>
<tr>
<td>4</td>
<td>Take advantage of <strong>replication management</strong></td>
</tr>
</tbody>
</table>
Missing Pieces

1. Transparent end-to-end encryption

2. Distributed key management in the Cloud

3. Secure identity propagation

4. Vaulted image management
Questions

The journey of a thousand miles begins with a single step.

Lao Tzu