Big Data
Challenges and Opportunities

Ira A. (Gus) Hunt
Chief Technology Officer

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.

It's a

Big Data

World

Google
> 100 PB
> 1T indexed URLs

FaceBook

> 800M users
> 100PB

YouTube

> 750PB
>200,000 4TB drives

World Population
> 6,987,139,094

6

Twitter
> 55B tweets/year
> 150M/day
>1700/sec

7

Global Text Messages
> 6.1T per year
> 193,000 per second
> 876 per person per year

8

US Cell Calls
> 2.2 T minutes/year
> 19 minutes / person / day
(uncompressed~1 YouTube/year)

9

3
Driving Forces

Social

Mobile

Cloud

14

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

16

Why We Care
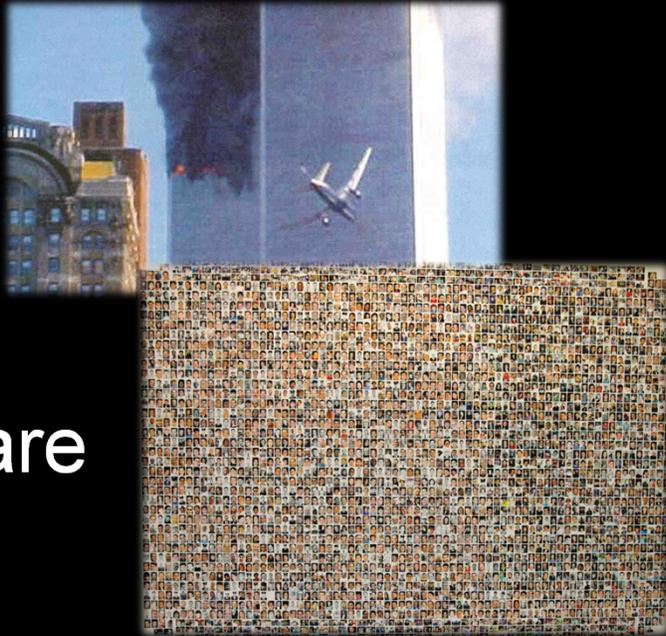
17
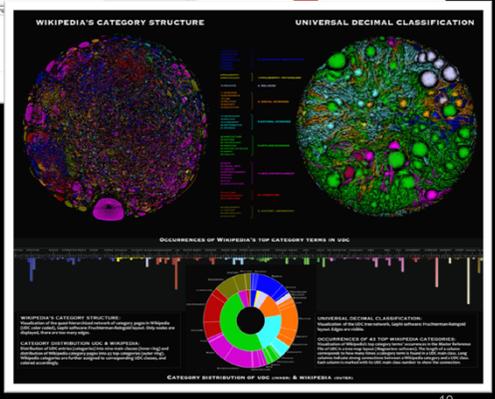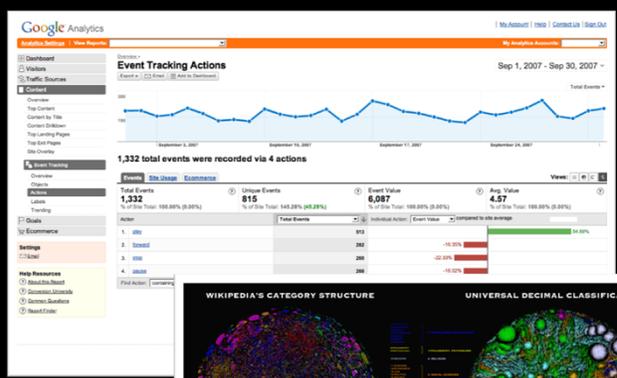
The Problem

20

Our Problem: Which 5K

1 Don't know the future **value** of a dot **today**

2 We cannot connect dots **we don't** have

3 The **old** collect, winnow, dissem **model fails** spectacularly in the **Big Data** world

The few cannot know the needs of the many

Secure the data, Connect the data, Empower the user

The

Challenge

Make

6,998,329,787

a small number

Why
is this important?

Nano

Bio

Sensors

26

Sensors and The Internet of Things
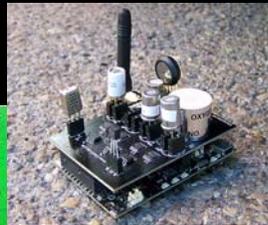
27

Sensors are BIG

1   Sensors are **unbounded**

2   Sensors are **promiscuous**

3   Sensors are **indiscriminate**

The Internet of Things is BIG

1   Everything is **Connected**

2   Everything **Communicates**

3   Everything is a **Sensor**

The inanimate is rapidly becoming sentient

Smarter Planet

Cars drive themselves

Machines know your needs

30

That's the

Really Big Data

challenge of our future

31

Technology is moving **faster** than government can keep up

32

How can we successfully navigate and **operate** in this new world??

33

Our Approach

1 **Know** the **Business**

2 **Set** an overarching **Strategy**

3 **Establish** a **Framework** for execution

4 **Fund** and Implement with **Intent**

---

4 Big Bets

1 Big **Data**
- Acquire, federate, and position for multiple constituencies to securely exploit. Grow the haystack, magnify the needles.

2 Operational **Excellence**
- Innovate infrastructure operations and provisioning, create an authoritative source on our asset base, and run IT like a business.

3 **Serve CIA** by supporting the **IC**
- Assume a leadership role in IC activities that matter to CIA
- Build capabilities assuming they will be shared

4 Talent **Management**
- Focus on continuous learning and diversity of thought, experience, background

5 Key Technology Enablers

1 Advanced **Mission Analytics**
  - World-class abilities to discover patterns, correlate information, understand plans and intentions, and find and identify operational targets in a sea of data

2 Enterprise **Widgets and Services**
  - A customizable, integrated and adaptive webtop that lets analysts, ops officers, and targeters to "have it their way".

3 **Security** as a **Service**
  - 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&A.

4 Enterprise Data Management--the **Data Harbor**
  - An ultra-high performance data environment that enables CIA missions to acquire, federate, and position and securely exploit huge volumes data.

5 **Cloud Computing**
  - Ruthlessly standardized, rigorously automated, dynamic and elastic commodity computing environment. Massive capacity ahead of demand. Speed for mission need.

Our Accelerated Technology Adoption Process

1 **Discover** the **Opportunities** (100)

2 **Evaluate** claims versus **Reality** (30)

3 **Pilot** with the **Mission** (10)

4 **Implement** (5)

Discover

Active External Engagement

VCs	Mission Link
Commercial Labs	Tech Connect
Government Labs	IC Partners
In-Q-Tel	Other Agencies
USG Contractors	Universities
Tech Expo	Road Trips
Showcase	Contracts

Evaluate

Unclassified and Classified Evaluation Facilities

iLab—unclassified, lots of data, variable hardware

Eval—high-side, on-desktop, real data, real users, defined hardware

NEAT—contracting mechanism to bring in capabilities from non-traditional vendors

Pilot

Real Problems, Real Users, Focused Outcomes

I2—the original IC "Cloud" proof of concept pilot

Mass Analytics Cloud (MAC)—high-side, big-data, real problems

Training—Cloudera, Hadoop, Developing for the Cloud

**Road Trips**—expose the pilot teams to best practices across sectors

Implement

Becoming part of our DNA
It's not just about Technology

- People and skills
- Architecture
- Governance
- Process
- Ruthless Standardization
- Complete change in Applications Development— think small, think horizontal
- Costing models
- Contracting models

Closing
Thoughts

42

Tectonic Technology Shifts

Traditional Processing	Mass Analytics/Big Data
Data on SAN	→ Data at processor
Move Data to Question	→ Move Question to Data
Backup	→ Replication management
Vertical scaling	→ Horizontal scaling
Capacity after demand	→ Capacity ahead of demand
DR	→ COOP
Size to peak load	→ Dynamic/elastic provisioning
Tape	→ SAN
SAN	→ Disk
Disk	→ SSD
RAM limited	→ Peta-scale RAM

It's all about SPEED!   Latency breeds contempt!!

A Few **Hard** Problems

- Pattern Discovery
- Correlation not Search—people, events, dates, locations, …
    - Boolean is broken
- "Curiosity" Layer
- Peta-scale in memory architectures
- Continuous, recursive, peta-scale recomputation
- Cloud encryption—key management
- Secure computing—assurance end-to-end
- Secure mobility

Challenges Ahead

- It's all about **speed**, latency breeds contempt
- Build a **continuous learning** organization
- Embrace **continuous change**
- Agility--become an "**Ahead of**" organization
- Software licensing—**metered use**, not ELAs