
การปฏิรูปข้อมูลขนาดใหญ่ (Big Data) เพื่อการพัฒนาประเทศ

Topics …

- Why Big Data ?
- Big Data Ecosystem
- Proposal of Gov. Big Data Architecture
- Movements of Government Big Data
- Toward the Big Data as a Service

Why Big Data ?

1 Minutes in Your Life!

DOMO

DATA NEVER SLEEPS

How Much Data Is Generated **Every Minute?**

Big data is not just some abstract concept used to inspire and mystify the IT crowd; it is the result of an avalanche of digital activity pulsating through cables and airwaves across the world. This data is being created every minute of the day through the most innocuous of online activity that many of us barely even notice. But with every website browsed, status shared, or photo uploaded, we leave digital trails that continually grow the hulking mass of big data. Below, we explore how much data is generated in one minute on the Internet.

EVERY MINUTE of the DAY

- **THE MOBILE WEB** RECEIVES **217** NEW USERS.
- **YOUTUBE** USERS UPLOAD **48** HOURS OF NEW VIDEO.
- **EMAIL USERS** SEND **204,166,667** MESSAGES.
- **GOOGLE** RECEIVES **OVER 2,000,000** SEARCH QUERIES.
- **FACEBOOK USERS** SHARE **684,478** PIECES OF CONTENT.
- **WORDPRESS** USERS PUBLISH **347** NEW BLOG POSTS.
- **571** NEW WEBSITES ARE CREATED.
- CONSUMERS **SPEND $272,070** ON WEB SHOPPING.
- **POURSQUARE** USERS **PERFORM 2,083** CHECK-INS.
- **TWITTER** USERS SEND OVER **100,000** TWEETS.
- **FLICKR** USERS ADD **3,125** NEW PHOTOS.
- **INSTAGRAM** USERS SHARE **3,600** NEW PHOTOS.
- **TUMBLR** BLOG OWNERS PUBLISH **27,778** NEW POSTS.
- **BRANDS & ORGANIZATIONS** ON FACEBOOK **RECEIVE 34,722** "LIKES."
- **APPLE** RECEIVES ABOUT **47,000** APP DOWNLOADS.

WITH NO SIGNS OF SLOWING, THE DATA KEEPS GROWING

These are just some of the more common ways that Internet users add to the big data pool. In truth, depending on the niche of business you're in, there are virtually countless other sources of relevant data to pay attention to. Consider the following:

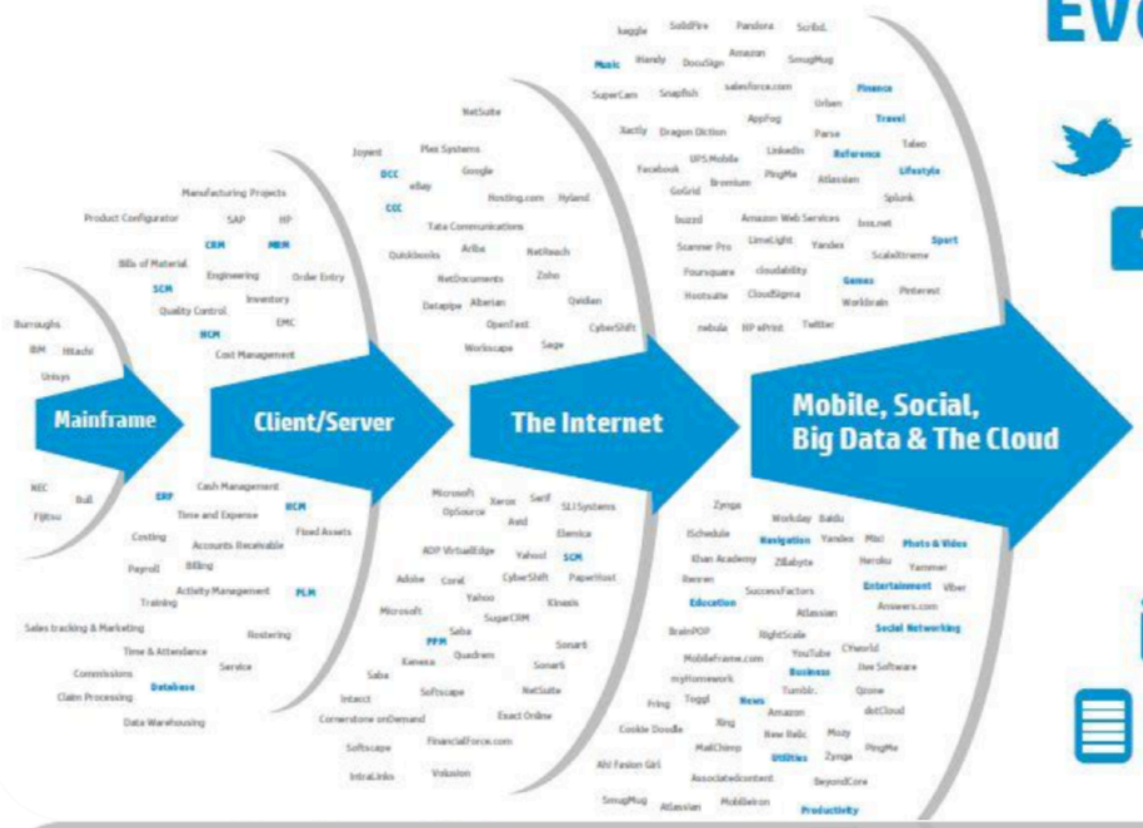
The global Internet population grew 6.59 percent from 2010 to 2011 and now represents **2.1 BILLION PEOPLE.**

These users are real, and they are out there leaving data trails everywhere they go. The team at Domo can help you make sense of this seemingly insurmountable heap of data, with solutions that help executives and managers bring all of their critical information together in one intuitive interface, and then use that insight to transform the way they run their business. To learn more, visit www.domo.com.

DOMO

SOURCES: HTTP://NEWS.INVESTORS.COM/, ROYAL.PINGDOM.COM, BLOG.GROVO.COM, BLOG.HUBSPOT.COM, SIMPLYZESTY.COM, PCWORLD.COM, BIZTECHMAGAZINE.COM, DIGBY.COM

A new style of IT emerging



Every 60 seconds

- **98,000+** tweets
- **695,000** status updates
- **11million** instant messages
- **698,445** Google searches
- **168 million+** emails sent
- **1,820TB** of data created
- **217** new mobile web users

	Manufacturing and Natural Resources	Media/ Communications	Services	Government	Education	Retail	Banking	Insurance	Healthcare	Transportation	Utilities
Transactions	73%	62%	67%	67%	54%	93%	83%	81%	75%	79%	80%
Log data	44%	57%	58%	59%	54%	40%	66%	61%	33%	71%	60%
Machine or sensor data	53%	38%	35%	33%	31%	27%	27%	48%	42%	50%	40%
Emails /documents	27%	43%	43%	41%	46%	27%	34%	39%	17%	29%	20%
Social media data	32%	52%	39%	26%	54%	73%	27%	13%	-	50%	-
Free-form text	17%	24%	28%	30%	31%	20%	34%	35%	67%	21%	40%
Geospatial data	27%	14%	19%	19%	38%	27%	27%	26%	8%	29%	40%
Images	19%	24%	17%	11%	38%	13%	5%	16%	25%	7%	-
Video	8%	29%	12%	7%	31%	13%	-	6%	8%	7%	-
Audio	10%	19%	8%	4%	8%	-	-	6%	-	-	-
Other	8%	14%	13%	15%	8%	7%	10%	16%	42%	14%	-
n =	*59*	*21**	*127*	*27**	*13**	*15**	*41*	*31*	*12**	*14**	*5**

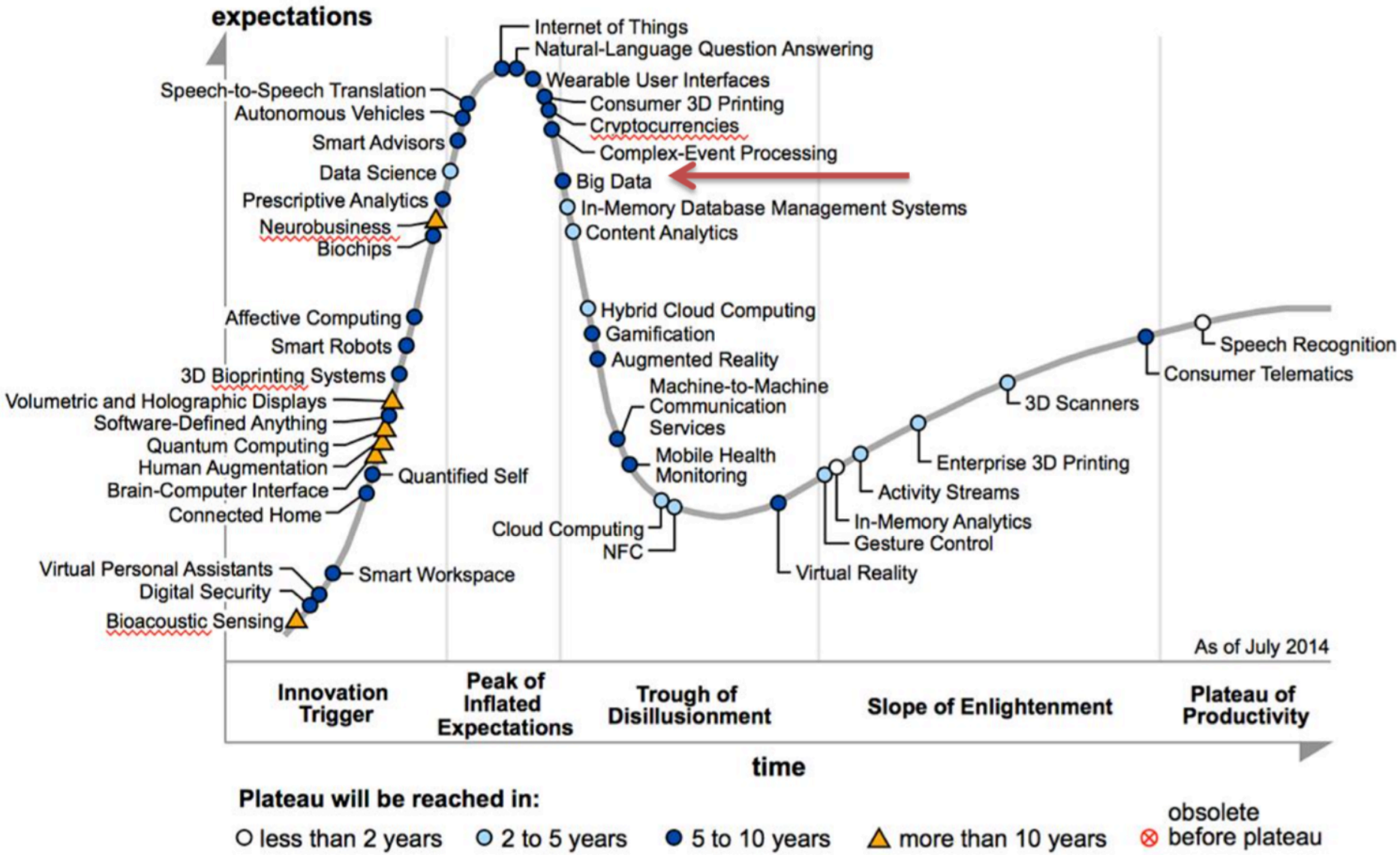
Note: Highlighted cells indicate the top three data types by industry.
Multiple responses allowed

Source: Gartner (September 2013)

Gartner Survey Results on the Type of Data Analyzed According to Industry (Kart, Heudecker, and Buytendijk 2013)

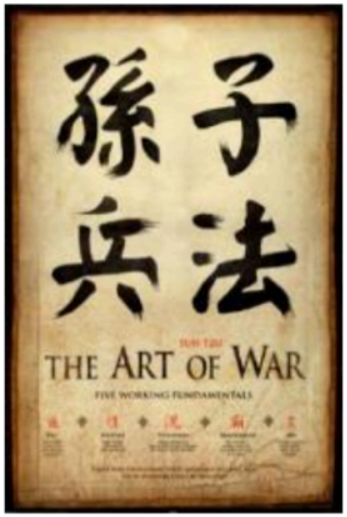
Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

"Gartner Inc."



Gartner's Hype Curve 2014

Why BigData?



Know thy self, know thy enemy. A thousand battles, a thousand victories.

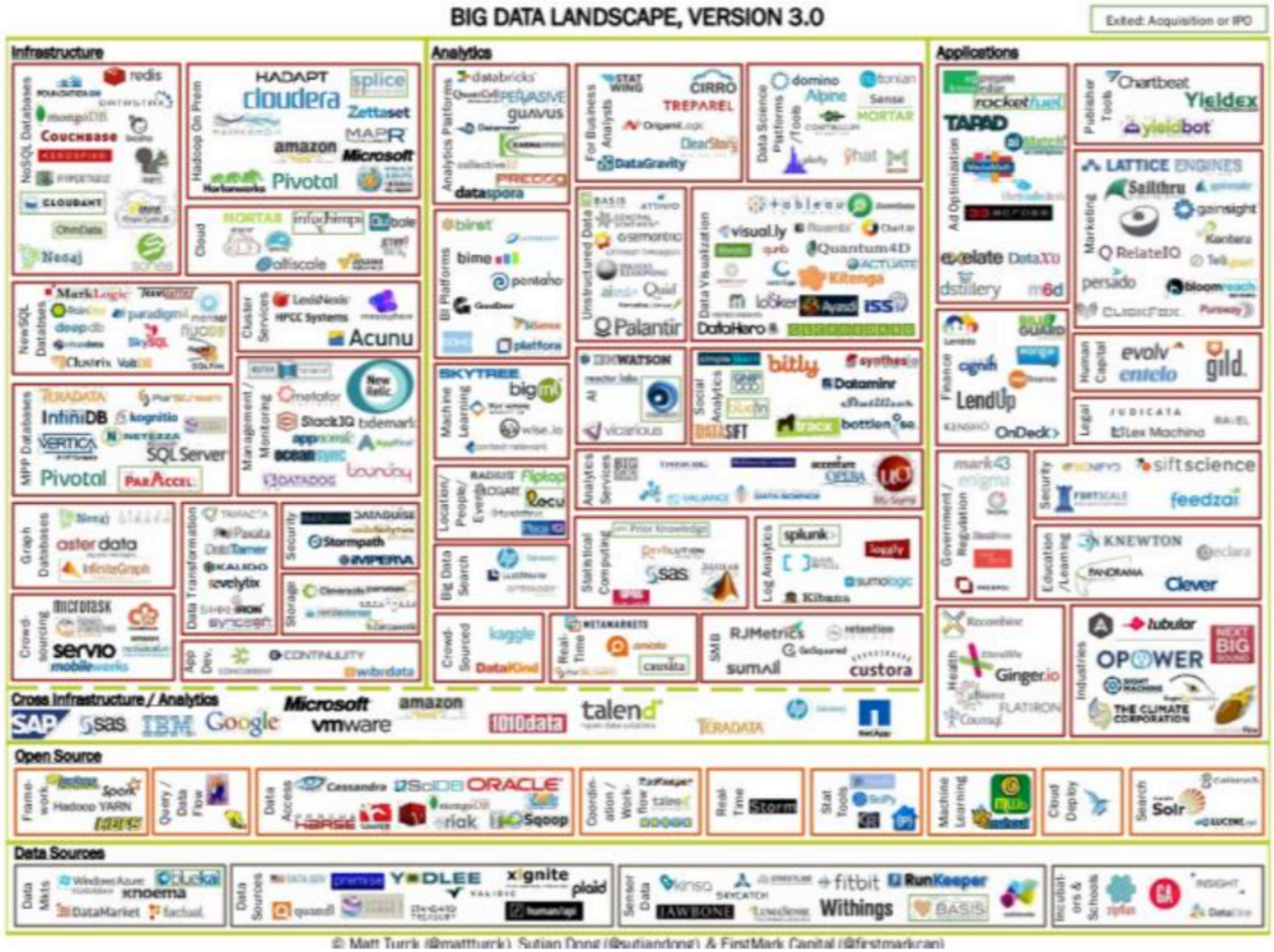
http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/big-data-cloud-technologies-brief.pdf)

The real value of big data is in the insights it produces when analyzed—discovered patterns, derived meaning, indicators for decisions, and ultimately the ability to respond to the world with greater intelligence.

- Improve product and service
- Increase customer satisfaction/behavior
- Improve operation efficiency
- Understand emerging market trends

Big Data Ecosystem

Big Data Ecosystem



Reference: http://dataconomy.com/understanding-big-data-ecosystem/

Big Data Eco system- Infrastructure

- **Hadoop-**
  - technologies designed for the storing, processing and analysing of data by breaking up and distributing data into parts and analysing those parts concurrently, rather than tackling one monolithic block of data all in one go.
- **NoSQL**
  - Stands for Not Only SQL
  - involved in processing large volumes of multi-structured data. Most NoSQL databases are most adept at handling discrete data stored among multi-structured data.
- **Massively Parallel Processing (MPP) Databases**
  - MPP databases work by segmenting data across multiple nodes, and processing these segments of data in parallel, and uses SQL.

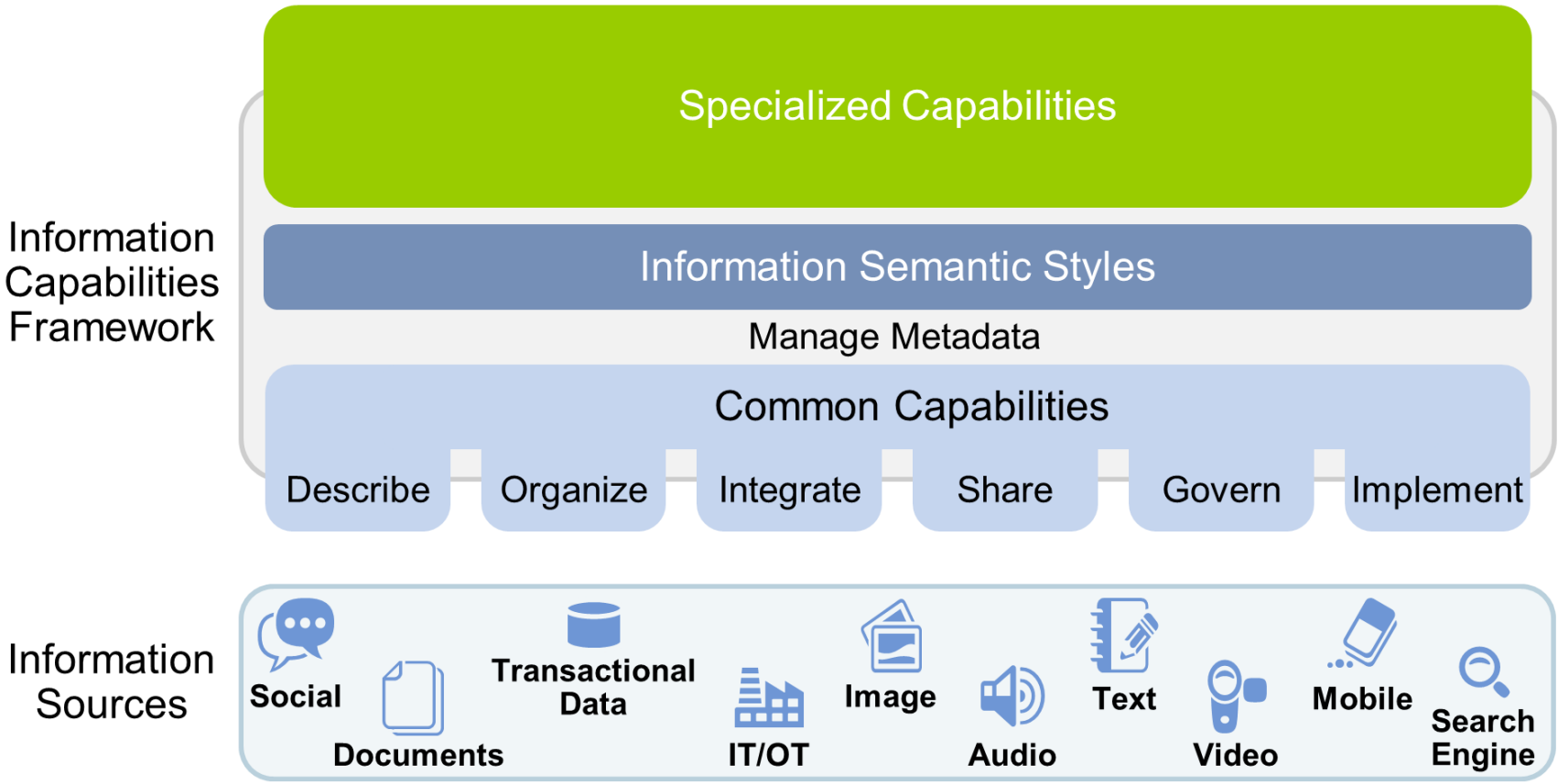
Reference: http://dataconomy.com/understanding-big-data-ecosystem/

Big Data Eco system- Analytics

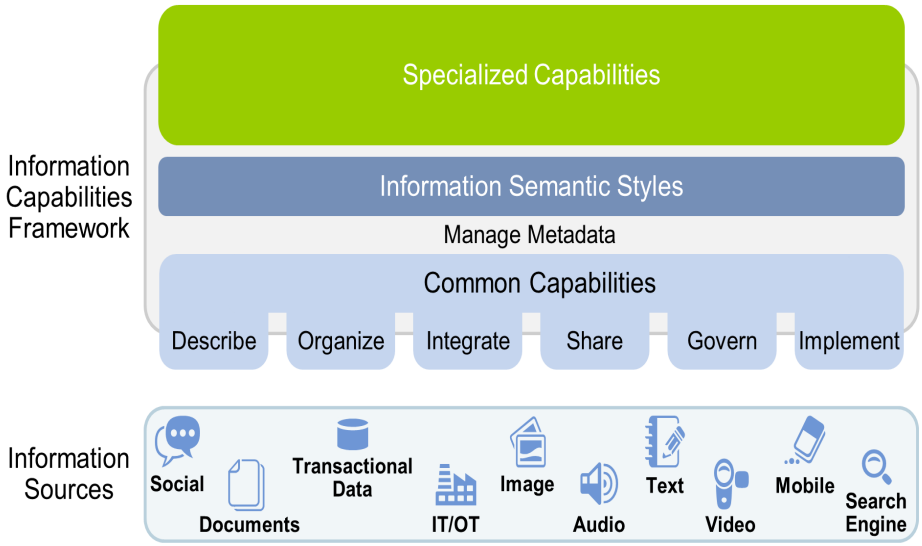
- **Analytics Platforms**
  - Integrate and analyse data to uncover new insights, and help companies make better-informed decisions.
- **Visualization Platforms**
  - visualizing data; taking the raw data and presenting it in complex, multi-dimensional visual formats to illuminate the information
- **Business Intelligence (BI) Platforms**
  - analyze data from multiple sources to deliver services such as business intelligence reports, dashboards and visualizations
- **Machine Learning**
  - machine learning is data the algorithm 'learns from', and the output depends on the use case. One of the most famous examples is IBM's super computer Watson, which has 'learned' to scan vast amounts of information to find specific answers, and can comb through 200 million pages of structured and unstructured data in minutes.

Reference: http://dataconomy.com/understanding-big-data-ecosystem/

Information Infrastructure Modernization Key Initiative Overview



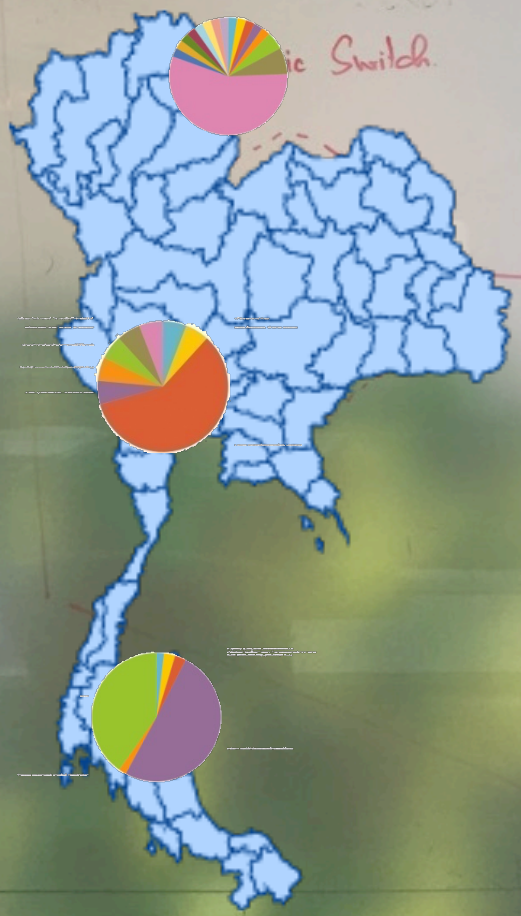
Information Infrastructure Modernization Key Initiative Overview



- A modern information infrastructure helps organizations describe, organize, integrate, share and govern information assets independently of applications and use cases
- Effective information infrastructure provides these common capabilities, and facilitates their consistent use across the enterprise.

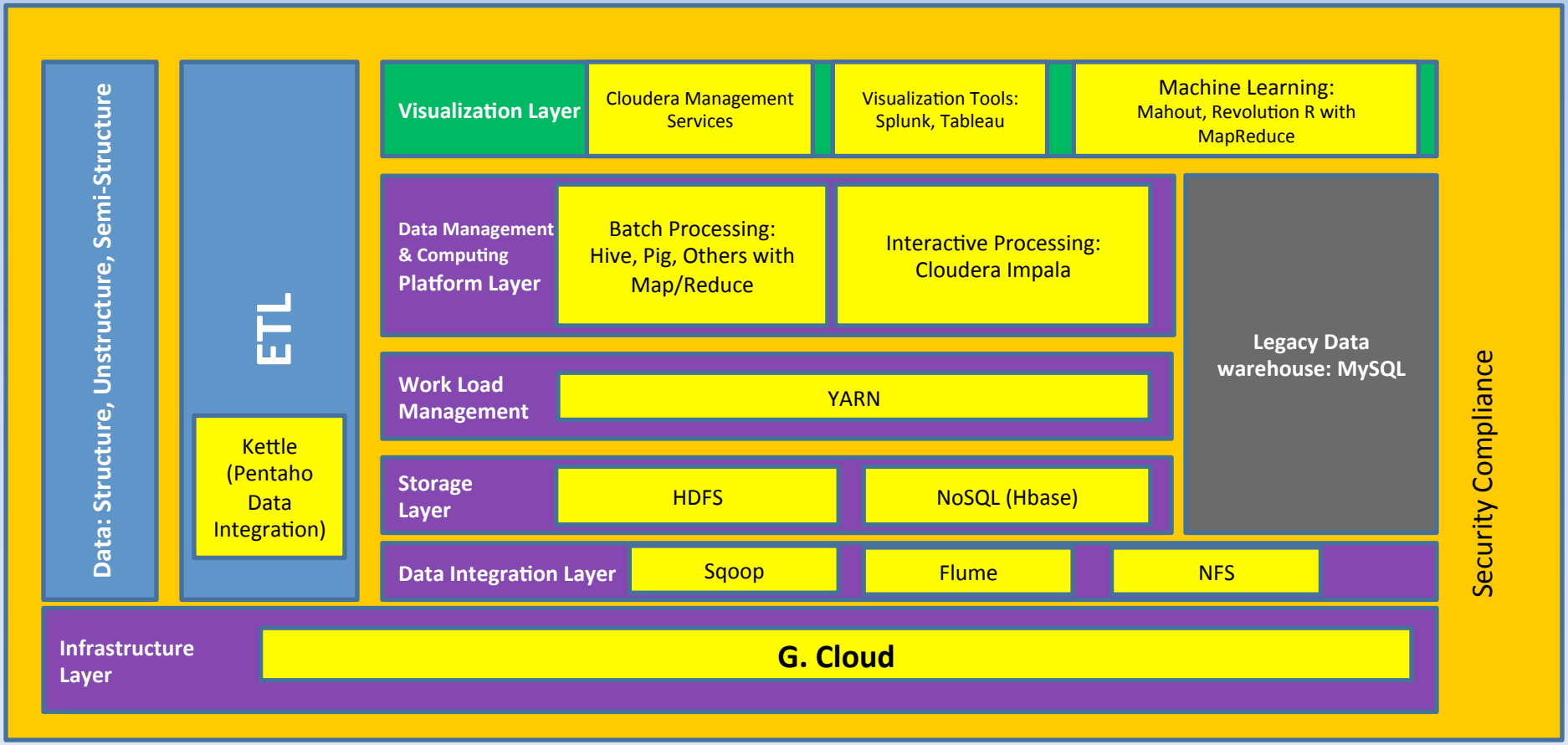
Proposal of Government Big Data Architecture

Proposal of Government Big Data Architecture



- Toward an analytical system using a distributed computing.
- Batch processing to historical data.
- Multi-Parallel Processing to Interactive query.
- Data insight: Patterns, Outliers.
- Support a legacy database system and an IoT platforms

Architecture of Big Data Hadoop on Cloud Computing



Powered by

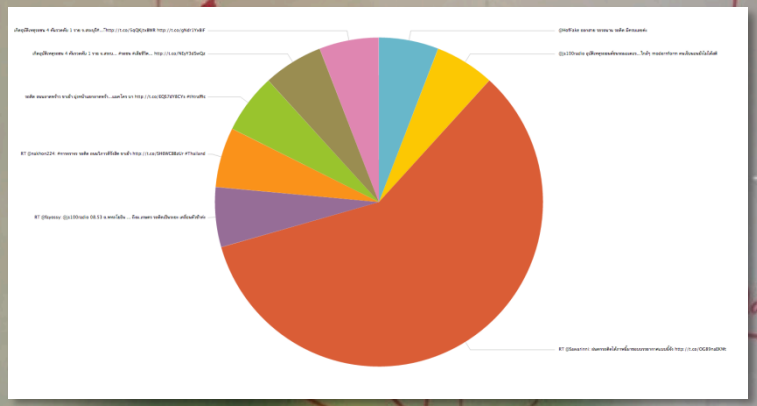
G-Cloud
Government Cloud Service

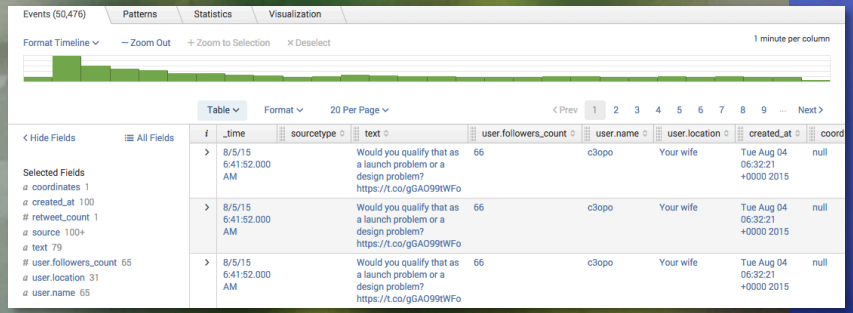
Movements of Government Big Data

Movements of Gov. Big Data

- Innovation:
  - ✓Sentiment Analysis of Government Services
  - ✓Video Analytics: Traffic Congestion
- Events

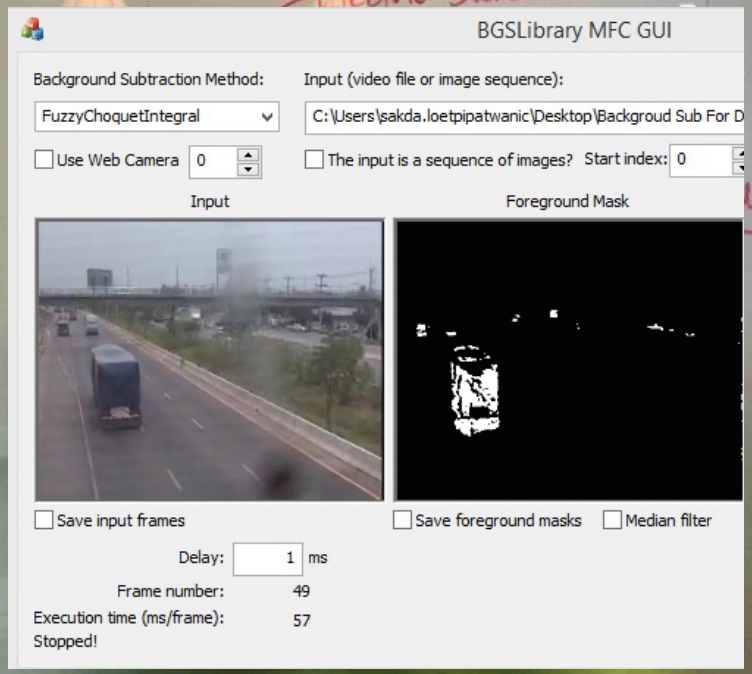
Sentiment Analysis of Government Services





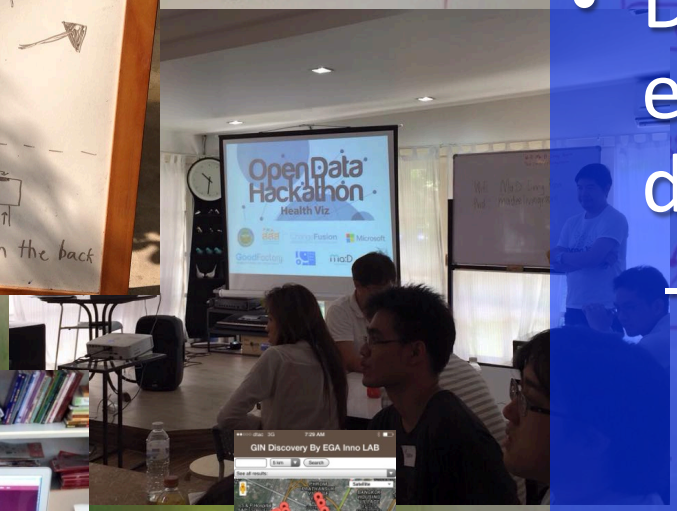
- Raw data: Twitter – Traffic congestion vs. Traffic control services in Bangkok.
- Objectives: rightly allocates and deploys manpower, tool, budgets and other resources.
- Technology: Flume with Twitter's APIs, Hadoop, Hive and Splunk with Hunk
  - Feed data from the Twitter in near real-time manner.
  - Explore and visualize raw big data without building fixed schemas or moving data to an in-memory store.
  - Drive deep analysis and find anomalies across terabytes or petabytes of raw data.

Video Analytics: Traffic Congestion



- Raw data:
  - DoH's traffic flows.
  - Gathers from vehicles detectors using microwave sensors.
- Objective:
  - Provides an alternative route to citizen.
- Technique: Image subtraction
  - Foreground = All – Background
- Next project
  - MapReduce computation with Java.

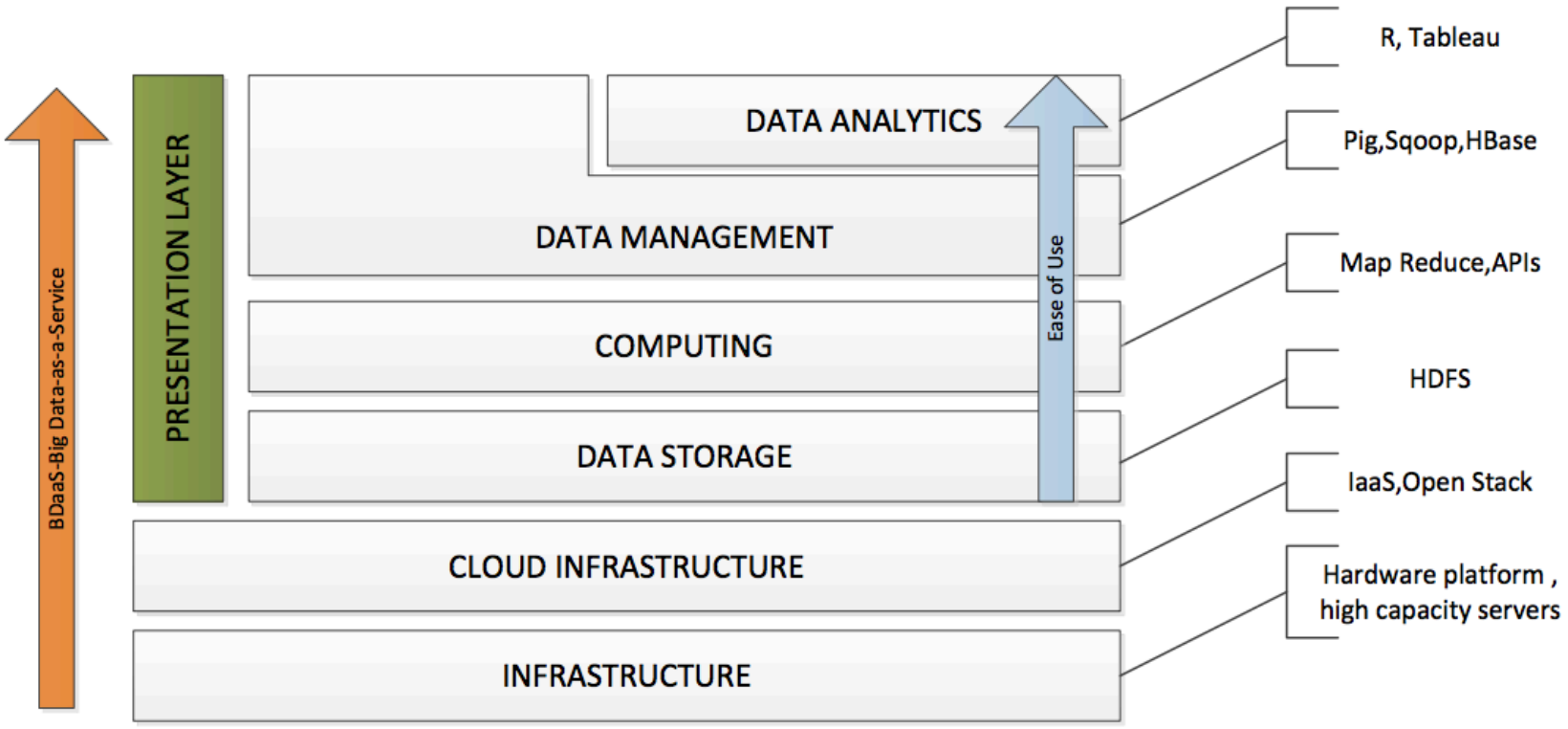
Events: Big Data & Open Data Hackathon



- Demand Side: EGA made engagement with developer communities.
  - Hosting a variety of Hackathons and other events to promote innovative online services.
- Supply Side:
  - Gov. Big Data Conference' 2015, today.

Toward the Big Data as a Service

Framework of Big Data as a Service



Adapted from http://web.mit.edu/smadnick/www/wp/2014-06.pdf