Smater Planet
=
BigData & Analytics + IoT

Term Smarter Planet and how useful it is

BigData
  - Remind a little bit about Data WareHouse
  - What is different between DWH & BigData
  - What can BigData do for the future

Internet of Things
  - How did the world transform
  - What is the Internet of Things
  - IoT instances (Smart House)

How can BigData and IoT work together to make Planet smarter

Smarter Planet instances and its usability & helpfulness

Sharing the Smarter Planet ideas

- Electrical energy loses 40%-70% around the world due to inefficient grid systems
- Congested roadways in the U.S. cost $78 billion annually in wasted hours and gas
- Inefficient supply chain lead to lose $40 billion annually
- Climate changes
- Energy crisis
- Security issues

→ Operating things in the right way to get higher efficiency with lower cost

What are we facing today?

Smarter Planet provides a platform to apply what we know about infrastructure technologies and business insights to our clients' problems and opportunities.
The world is becoming instrumented and interconnected. IBM is positioned to infuse intelligence into these physical and technological infrastructures, systems and processes, which can improve the way businesses, governments, economies and societies run.

What is Smarter Planet?
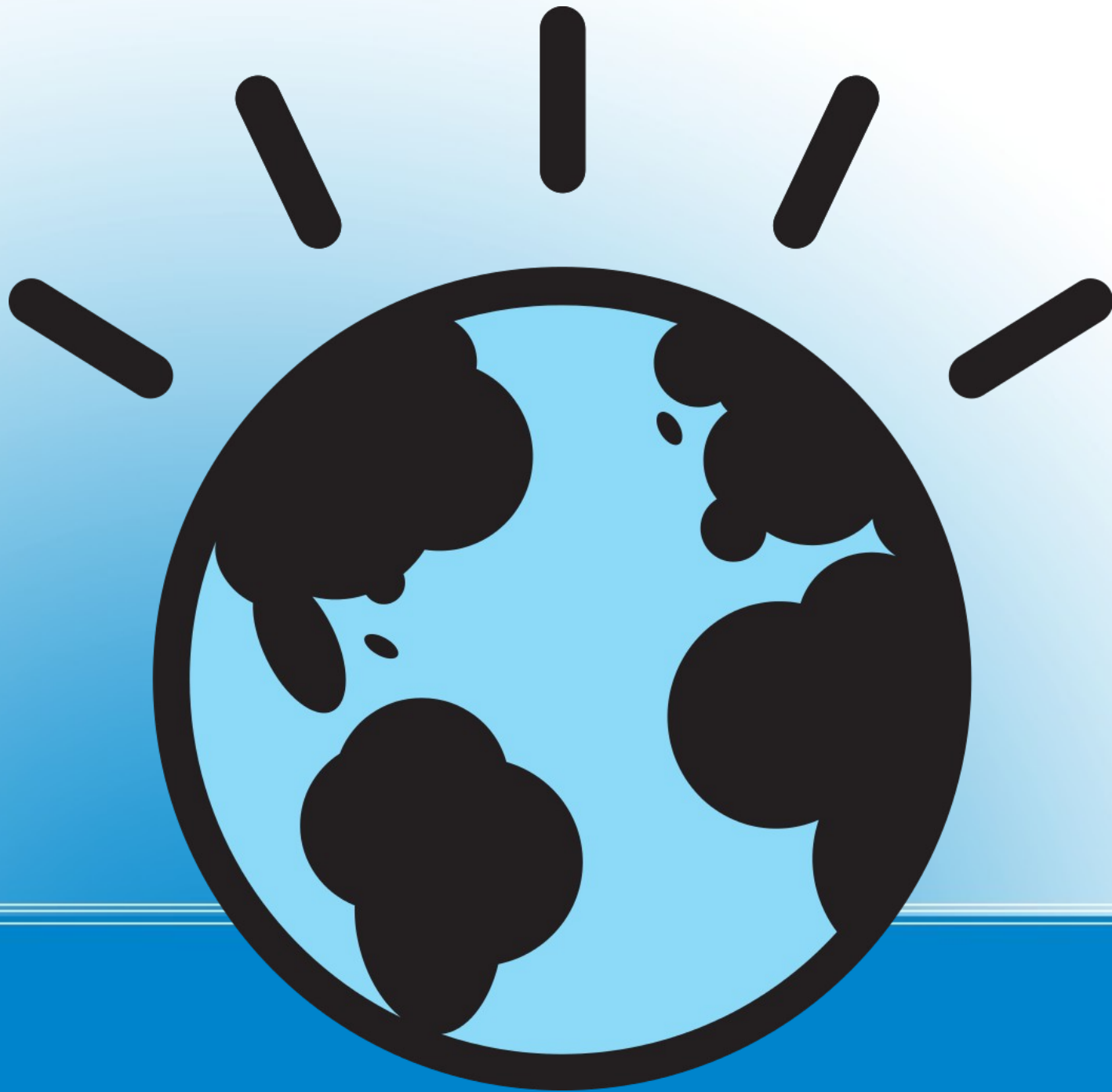
Energy   Traffic   Food   Infrastructure   Retail

Intelligence   Government   Banking   Telcom   Oil

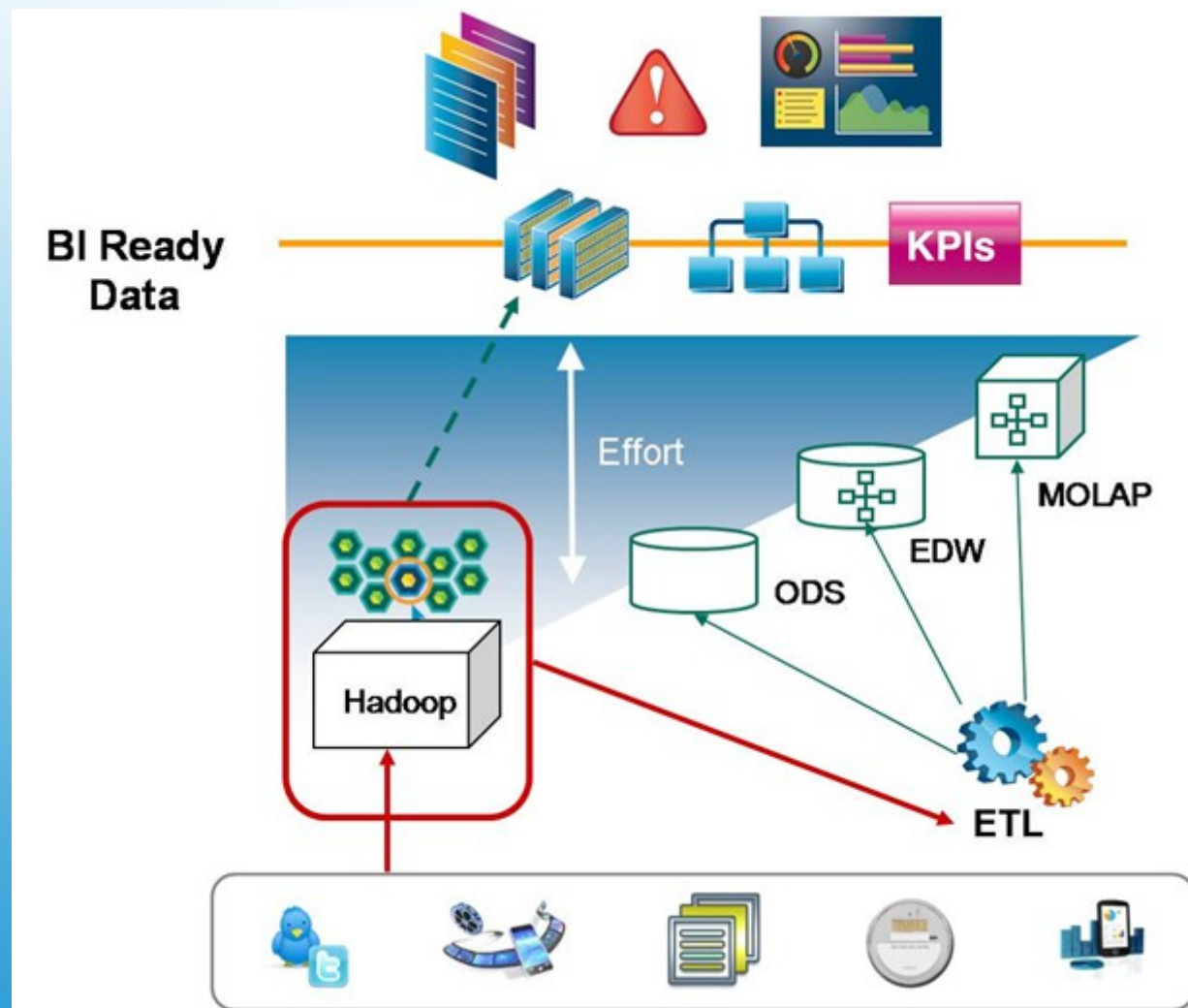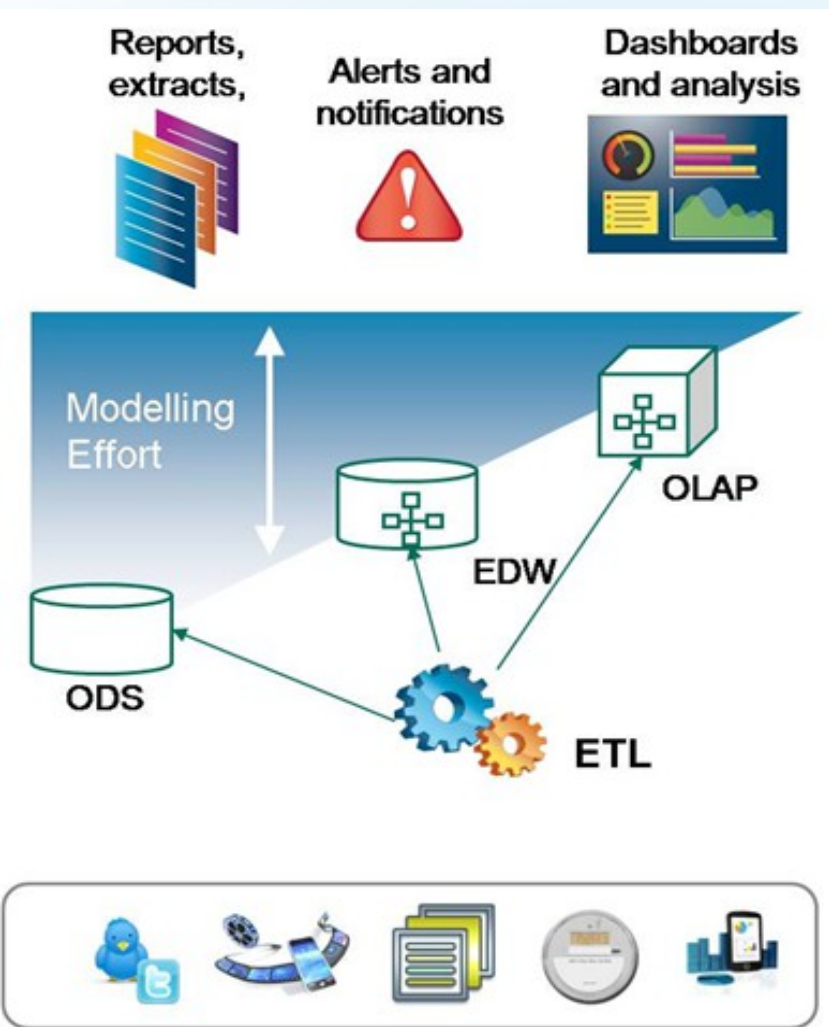Healthcare   Cities   Water   Public Safety   Buildings

Work   Rail   Products   Education   Cloud Computing

What is BigData & Analytics?

BigData – is not just a big data

"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. (Garner)

*** Analytics methods:
• Query and reporting
• Data mining
• Data visualization
• Predictive modeling
• Streaming data analytics
• …

What is BigData & Analytics?

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

2005
2020

6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day
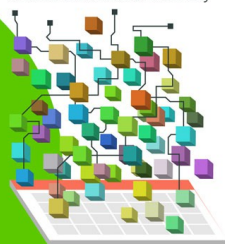
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Velocity
ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions

27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

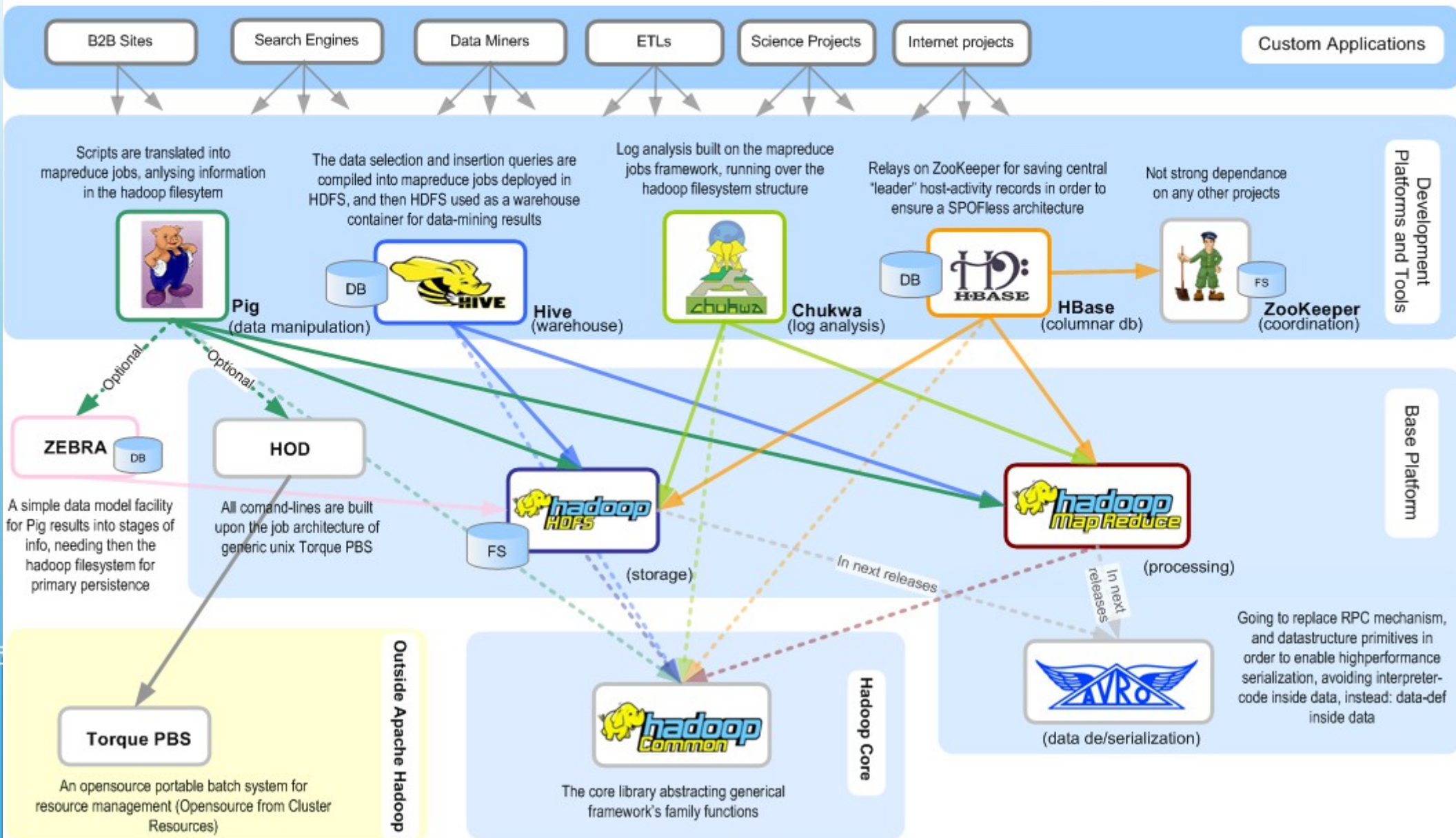Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

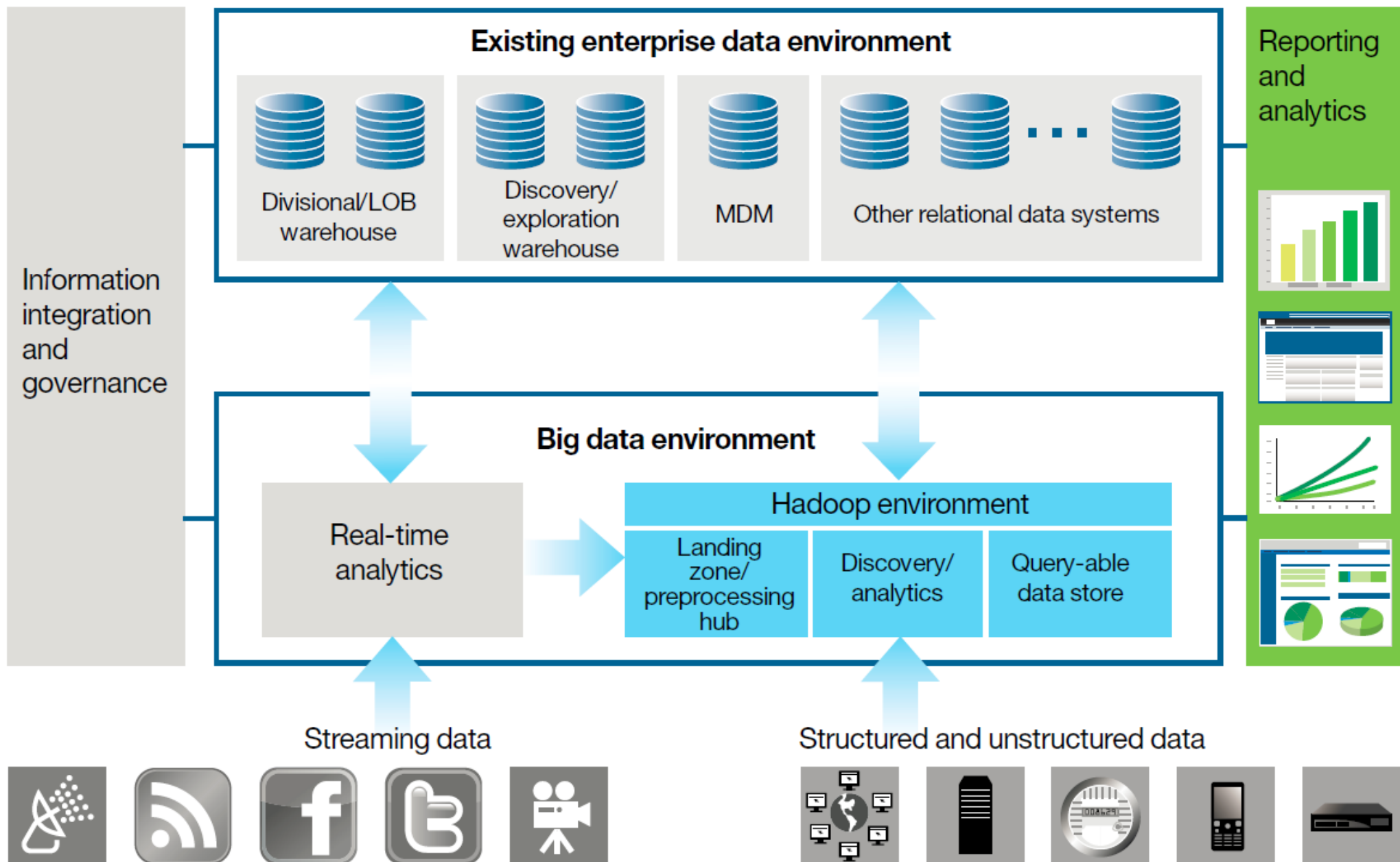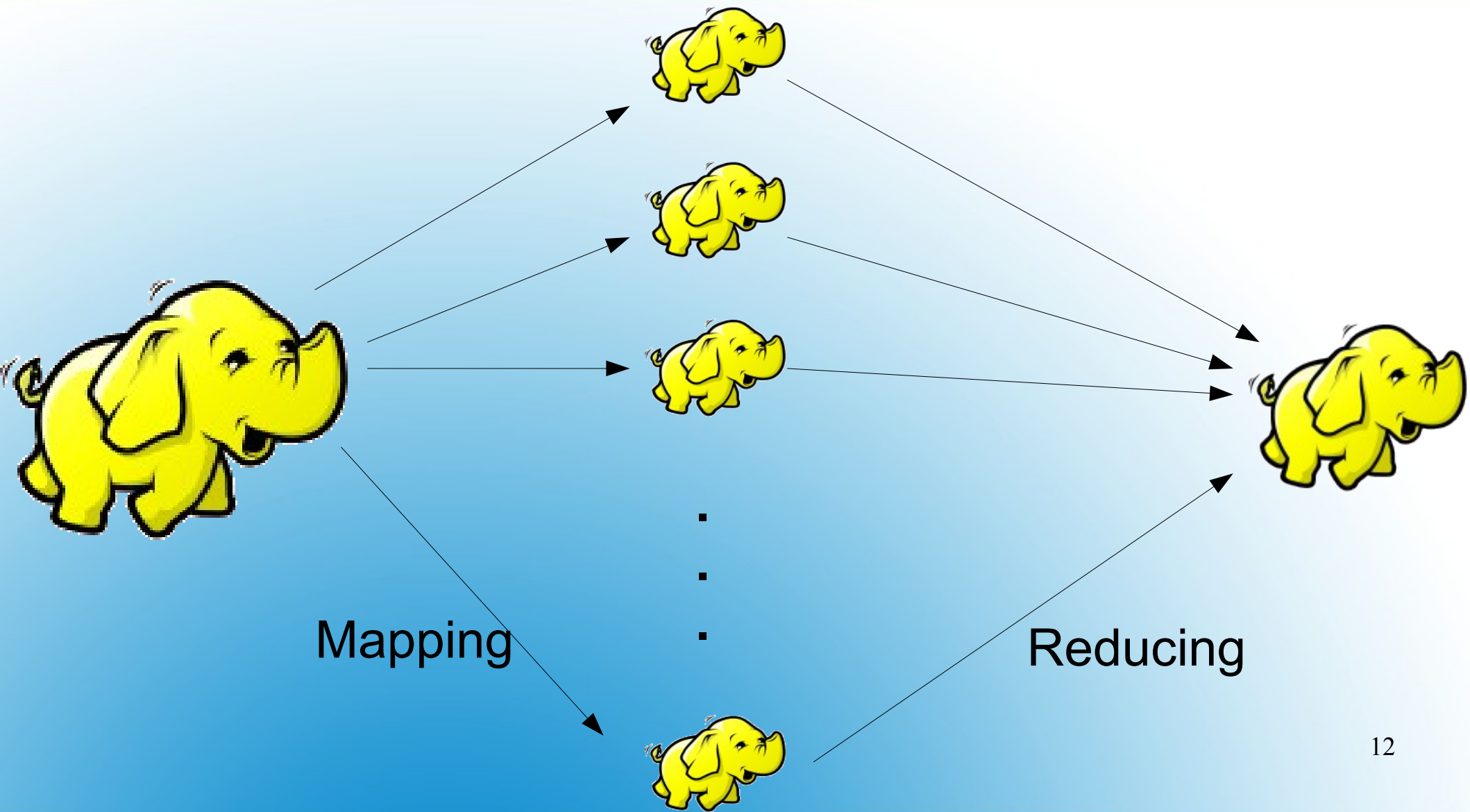**IBM**

What is BigData & Analytics?

hadoop

A global view on the framework's subprojects dependencies

Basically every hadoop subproject relays on HDFS for input and output data and in MapReduce for processing it, in specific ways for different needs and focuses

Custom Applications

B2B Sites | Search Engines | Data Miners | ETLs | Science Projects | Internet projects

Development Platforms and Tools

Scripts are translated into mapreduce jobs, anlysing information in the hadoop filesytem

The data selection and insertion queries are compiled into mapreduce jobs deployed in HDFS, and then HDFS used as a warehouse container for data-mining results

Log analysis built on the mapreduce jobs framework, running over the hadoop filesystem structure

Relays on ZooKeeper for saving central "leader" host-activity records in order to ensure a SPOFless architecture

Not strong dependance on any other projects

Pig (data manipulation)

DB **Hive** (warehouse)

Chukwa (log analysis)

DB **HBase** (columnar db)

FS **ZooKeeper** (coordination)

Optional | Optional

Base Platform

ZEBRA DB

HOD

A simple data model facility for Pig results into stages of info, needing then the hadoop filesystem for primary persistence

All comand-lines are built upon the job architecture of generic unix Torque PBS

hadoop HDFS

FS

(storage)

In next releases

hadoop MapReduce

In next releases

(processing)

Going to replace RPC mechanism, and datastructure primitives in order to enable highperformance serialization, avoiding interpreter-code inside data, instead: data-def inside data

Outside Apache Hadoop

Torque PBS

An opensource portable batch system for resource management (Opensource from Cluster Resources)

hadoop Common

Hadoop Core

The core library abstracting generical framework's family functions

AVRO

(data de/serialization)

Existing enterprise data environment

Reporting and analytics

Divisional/LOB warehouse

Discovery/exploration warehouse

MDM

Other relational data systems

Information integration and governance

Big data environment

Real-time analytics

Hadoop environment

Landing zone/preprocessing hub

Discovery/analytics

Query-able data store

Streaming data

Structured and unstructured data

IBM InfoSphere BigInsights

Mapping

Reducing

How does Hadoop work?

Example: Google page rank search 'keyword'

Def Mapper('keyword')
    While site_list not null
    output(keyword, site_list.get_site());

Search('keyword', sitename)
    Output('keyword', sitename, site_rank);

Def reducer('keyword', sitename, site_rank)
    Output('keyword', <list> dessort_rank(site_list));

How does Hadoop work?
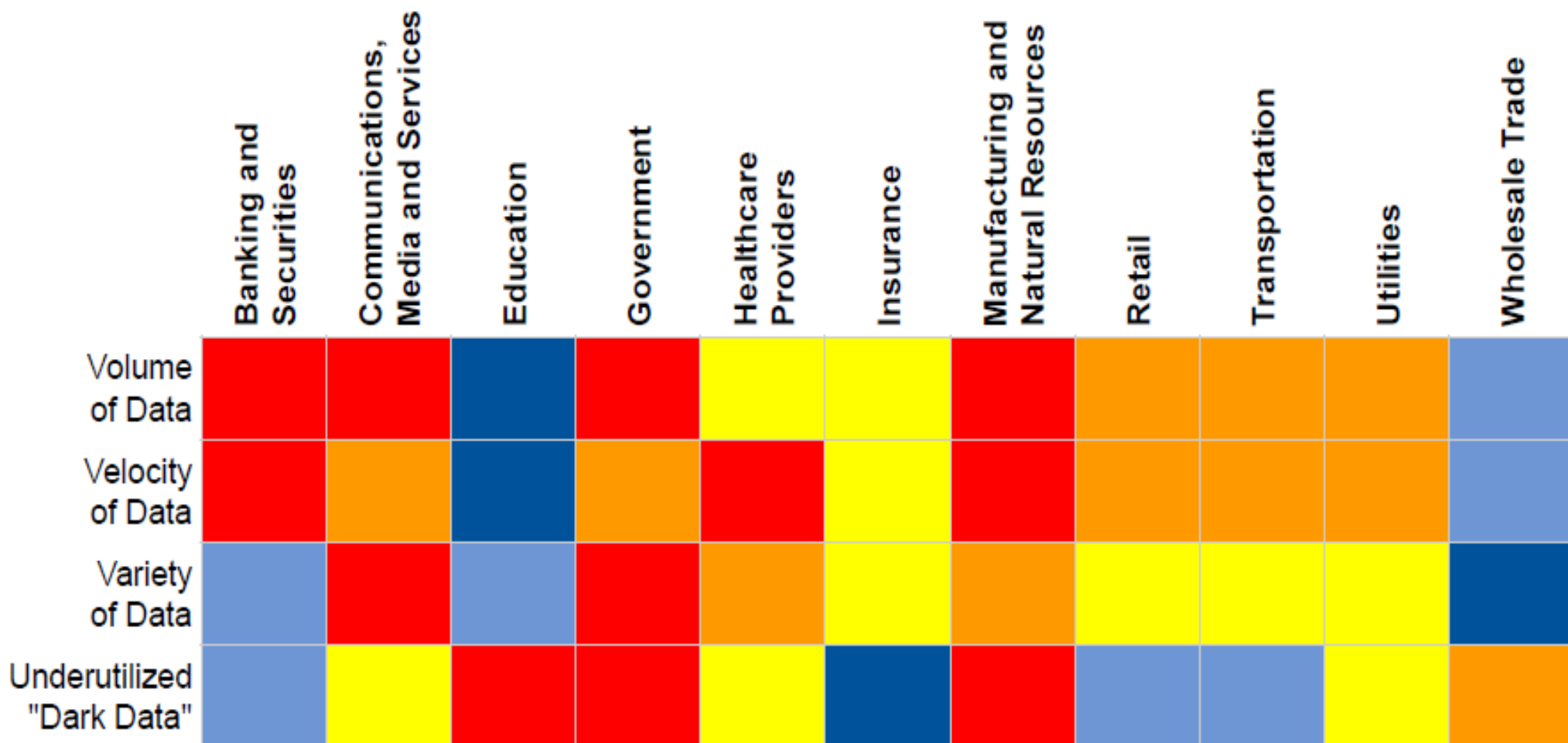
Potential big data opportunity on each dimension is:

- **Very hot** (compared with other industries)
- **Hot**
- **Moderate**
- **Low**
- **Very low** (compared with other industries)

Gartner
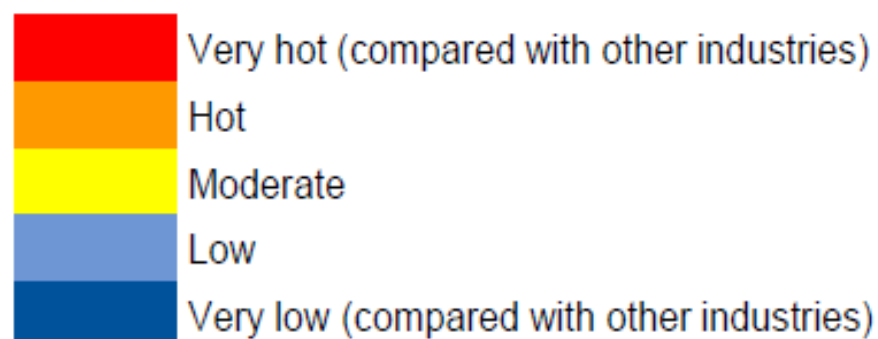
- 80% of the world's information is unstructured
- Unstructured information is growing at 15 times the rate of structured information

- *** What can BigData do for the future?
- Research in Science
- Education Service
- Consultant/Advisory Service
- Bussiness Insight
- Decision Making Assistant
- Prediction
- ...

What can BigData do for future?

Big Data Market Forecast by Sub-Type, 2011-2017 (in $US billions)
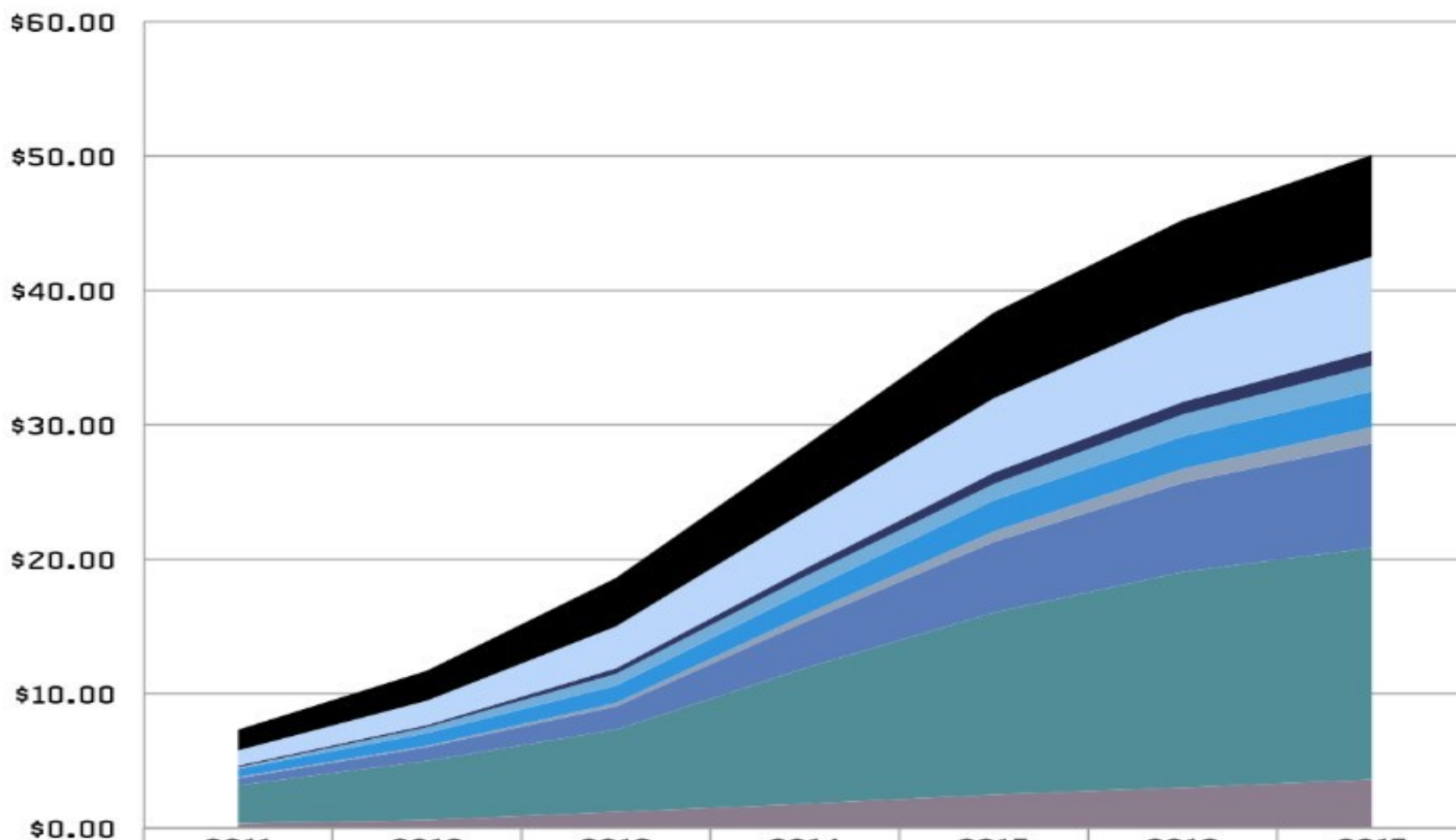
	2011	2012	2013	2014	2015	2016	2017
■ Compute	$1.53	$2.29	$3.65	$4.92	$6.40	$7.10	$7.60
■ Storage	$1.10	$1.75	$3.09	$4.20	$5.50	$6.40	$6.95
■ Networking	$0.15	$0.23	$0.42	$0.65	$0.85	$1.01	$1.15
■ Infrastructure Software	$0.14	$0.44	$0.83	$1.08	$1.25	$1.60	$1.90
■ SQL	$0.62	$0.88	$1.31	$1.75	$2.25	$2.45	$2.70
■ NoSQL	$0.07	$0.13	$0.29	$0.50	$0.80	$1.00	$1.20
■ Apps & Analytics	$0.52	$0.99	$1.69	$3.45	$5.29	$6.65	$7.75
■ Professional Services	$2.80	$4.42	$6.15	$10.10	$13.50	$16.00	$17.20
■ Cloud	$0.36	$0.62	$1.19	$1.82	$2.52	$3.05	$3.65

Wikibon
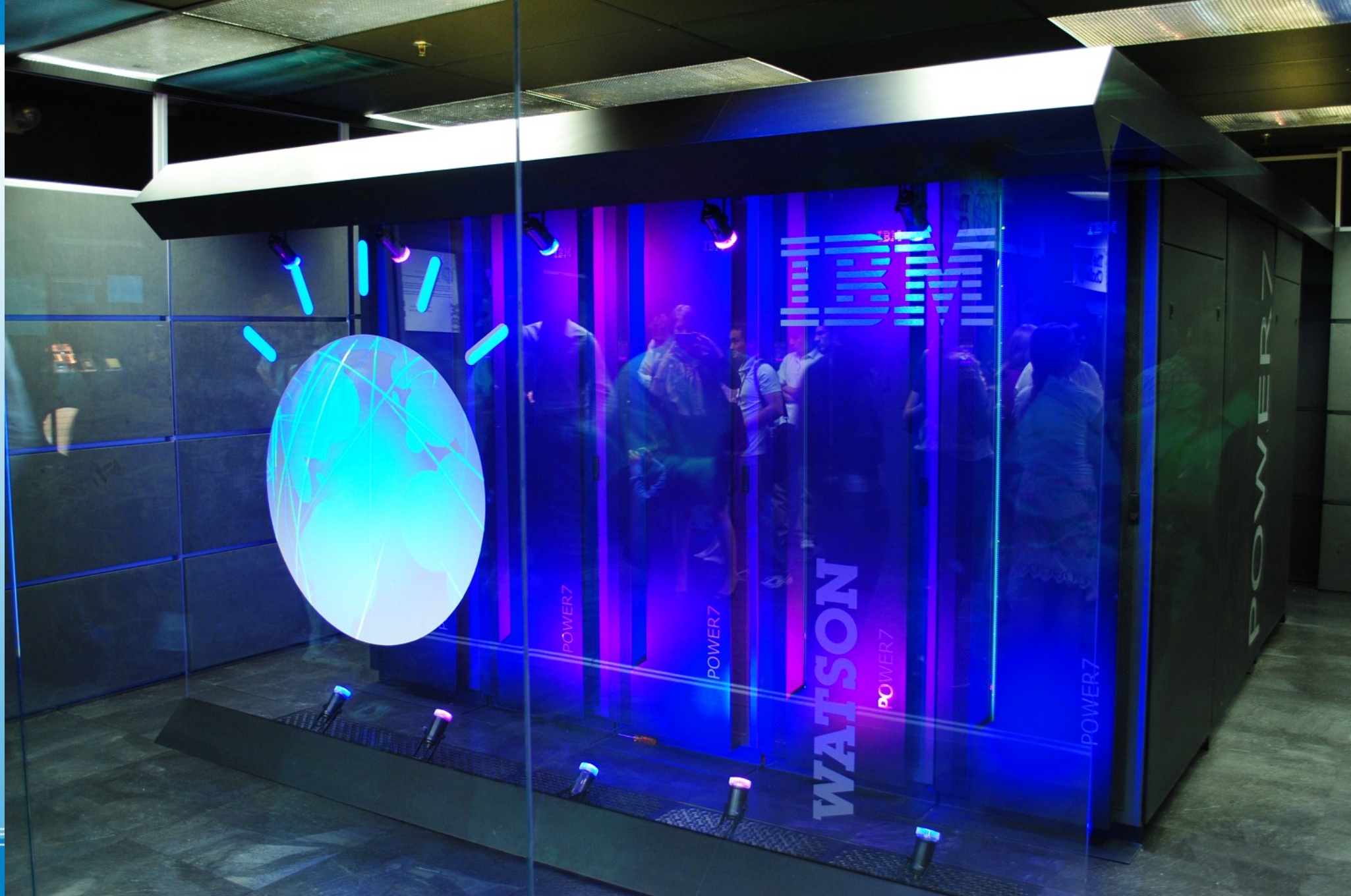
Cognitive System - IBM Watson

What can Watson do?
- Understand natural languages
- Dynamic learning
- Hypothesis generation

*** Watson services:
Watson Engagement Advisor
 → Advisory/Consultant servive
Watson Discovery Advisor
→ Research in Life Science (Healthcare, Climate, ...)

Cognitive System - IBM Watson

Vision of a Smarter Planet where sensors, intelligence, and connectivity are embedded into every human artifact, every element of the natural world, and even every physical person. (IBM)

A network of physical objects that contain embedded tech to communicate, sense, and interact with internal states or external environment (Gartner)

19

What is Internet of Things (IoT)?

What is Internet of Things (IoT)?

What is Internet of Things (IoT)?

Today's Wireless Landscape
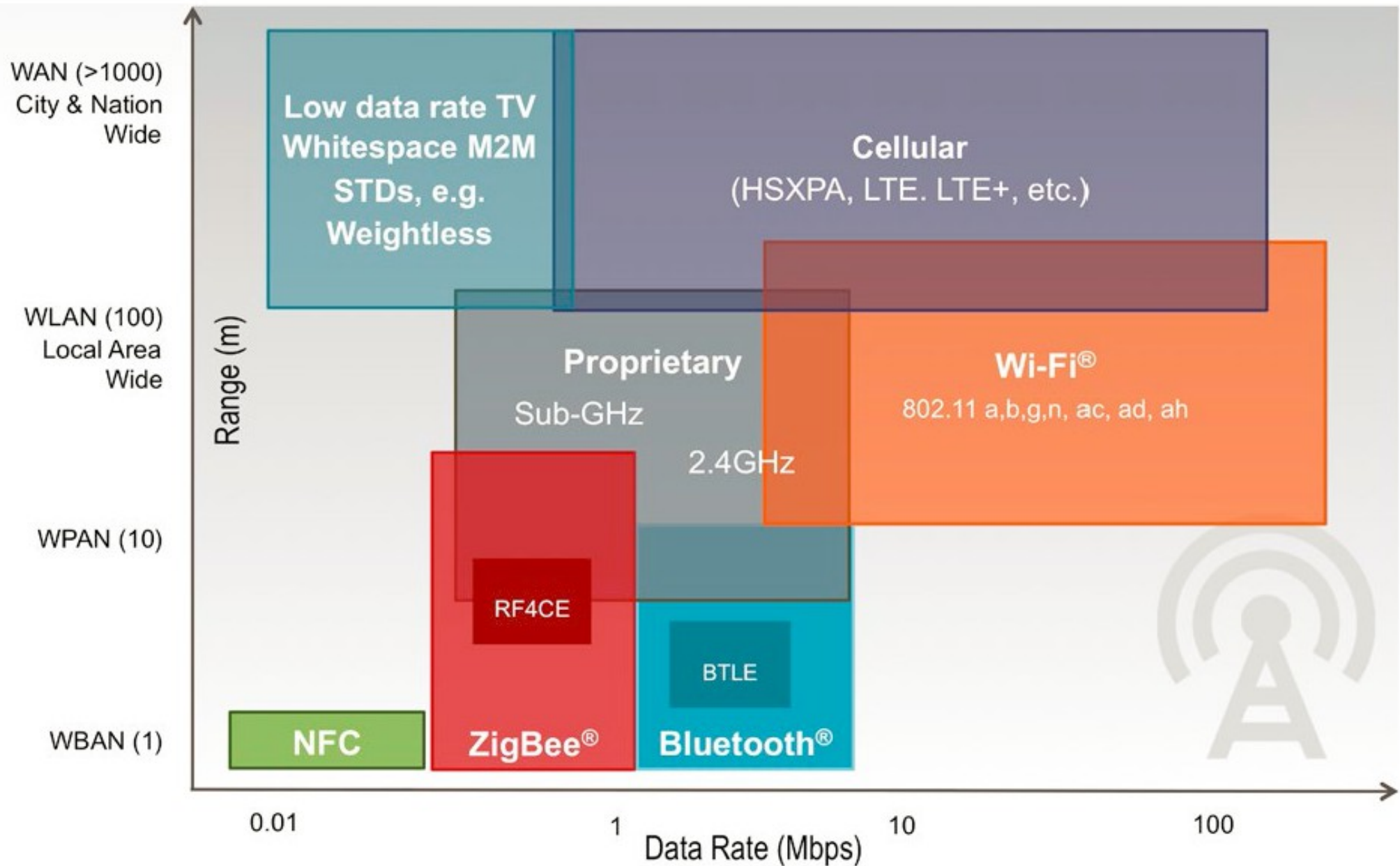
THE BIG
OPPORTUNITY

CONNECTED DEVICES GROWING AT 300%[3]

DATA GROWING AT 10X

DATA GROWTH BY 2016 – 90% UNSTRUCTURED[2]

200B
Devices
2020[1]

15B
Devices
2015

2B
Devices
2006

[1]IDC, Intel, United Nations
[2]IDC Digital Universe Study, December 2012
3- McKinsey Global Institute

How can they work together?

Smarter Social Networking
Smarter Cities (Dutch, Da Nang, …)
Smarter Advertisment/Marketing
Search Engine
Africa's grand challenges with Watson's help

Actualised Smarter Planet Project

Sharing your Smarter Planet ideas

- Forbes

- BigData University

- IBM BigData Hub

- IBM.com

- Intel.com

- Hadoop/Cloudera/Hortonwork

- IoT European Research Cluster (IERC)

Reference