Application and Platform Building of Big Data

Xinqiang Ma
Assoc. Prof, GAS
E-mail: xinqma@163.com
Study Groups: Dr. Li Danning, Dr.Li Dan, Mr. Chen yuqing

贵州科学院
Guizhou Academy of Sciences

In this age,Big Data

Is coming……

贵州科学院
Guizhou Academy of Sciences

- In pioneer days they used oxen for heavy pulling, and when one ox couldn’t budge a log, they didn’t try to grow a larger ox.

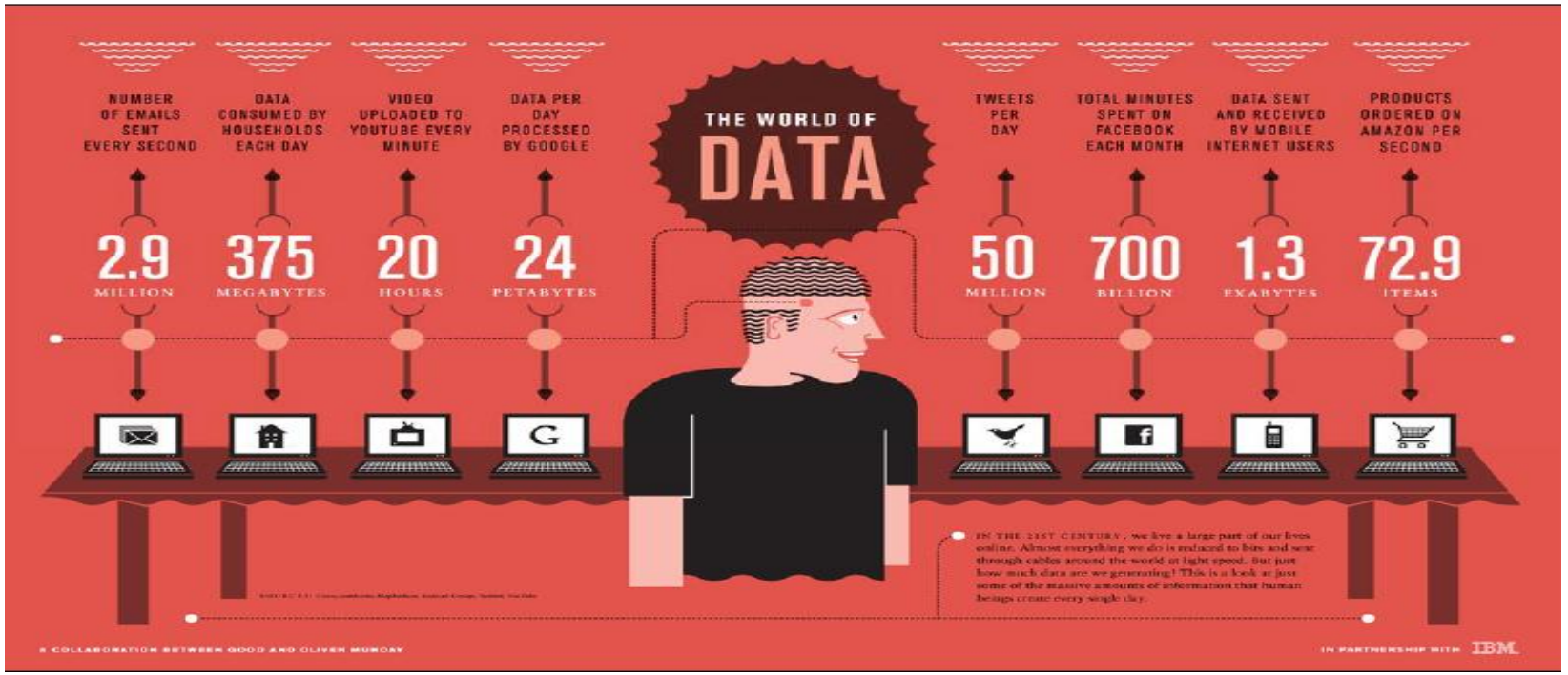
- We shouldn’t be trying for bigger computers, but for more systems of computers.

—Grace Hopper

Topics (Outline)

- 1. What is Big Data?
- 2. Application of Big Data
- 3. Big Data opportunities and challenges
- 4. How to Deal with Big Data?
- 5. What's Hadoop/MapReduce?
- 6. Big data players/Software Tools/Platforms
- 7. Examples

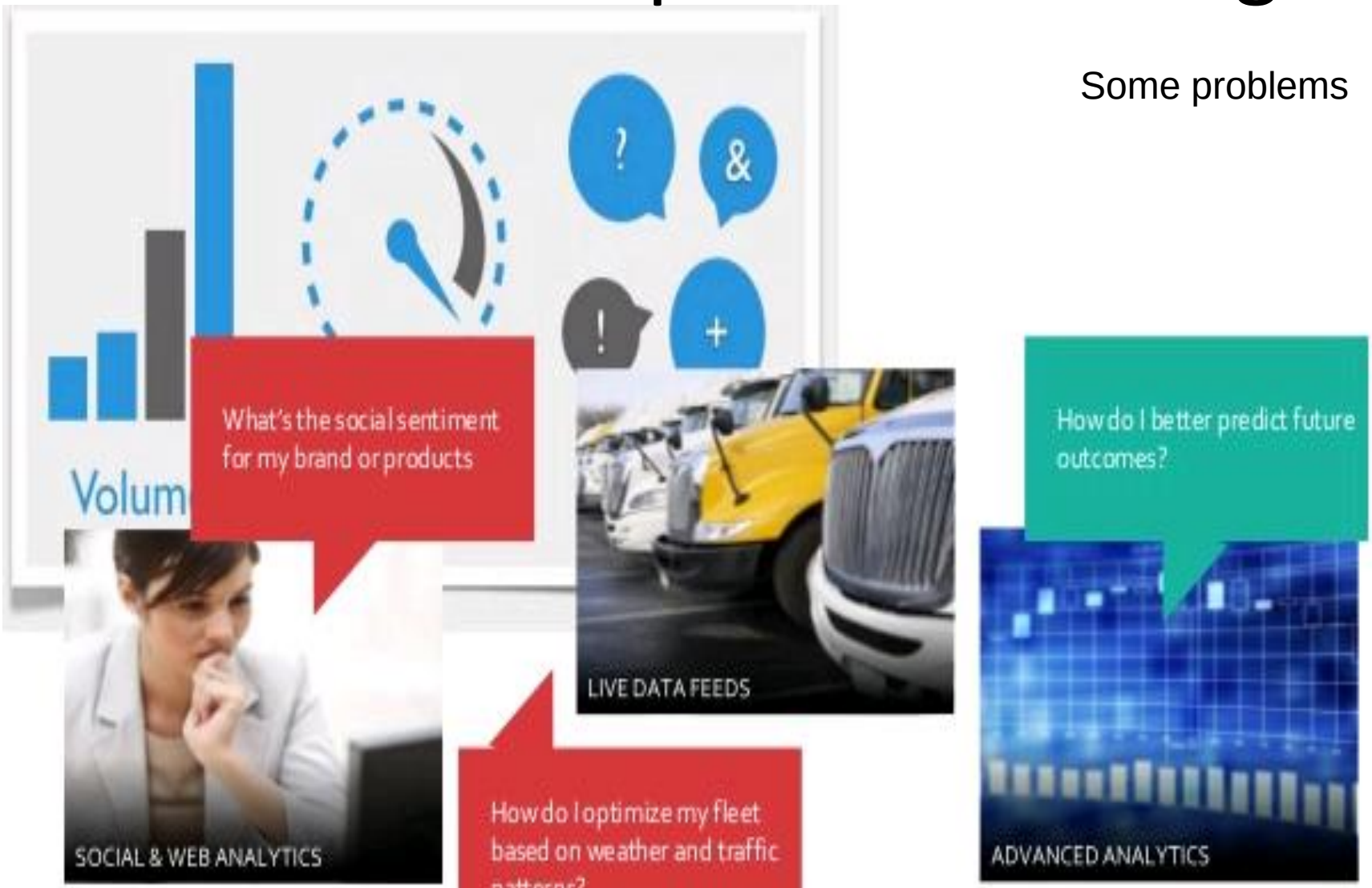
Data!

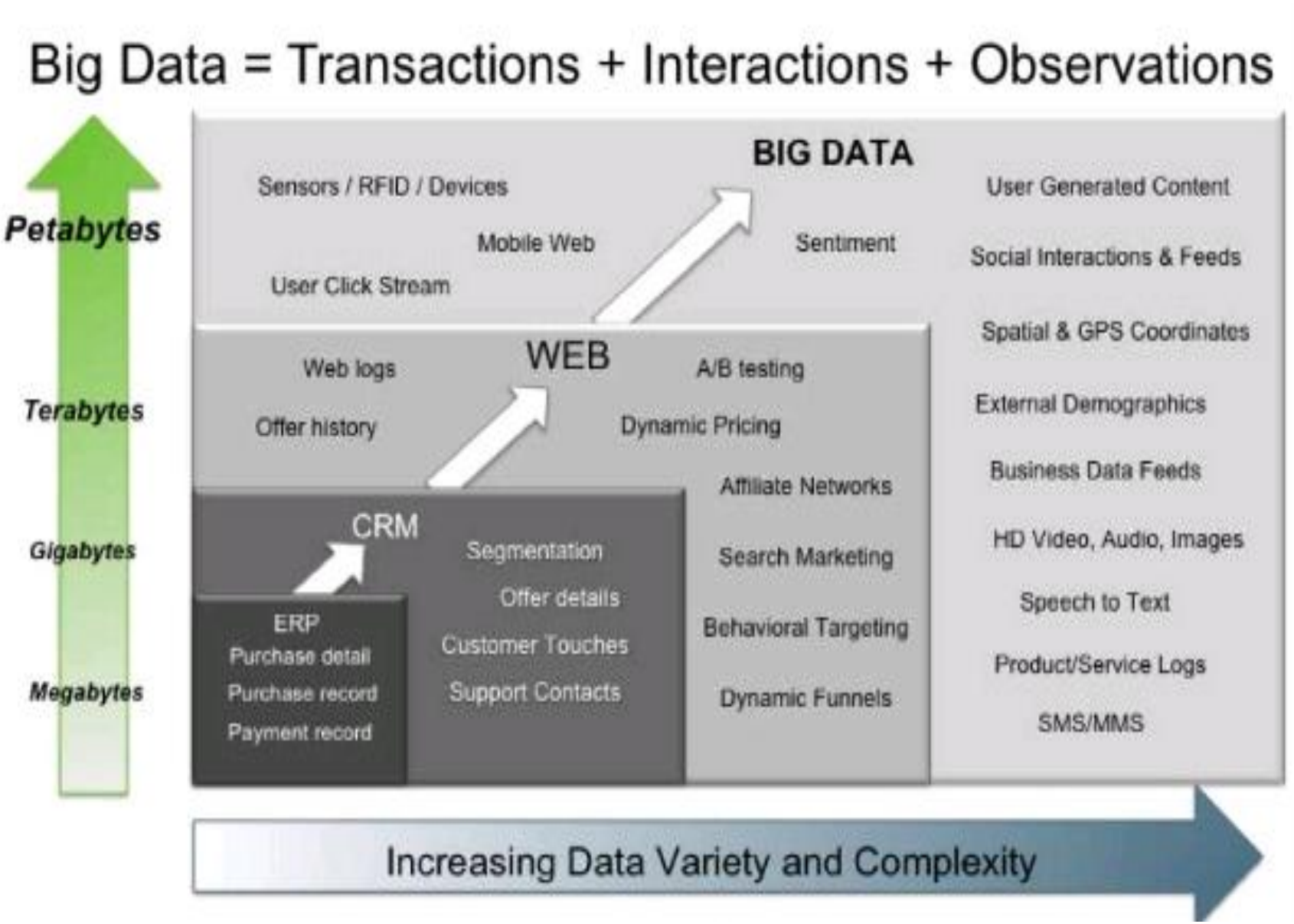


- We live in the data age. It’s not easy to measure the total volume of data stored electronically, but an IDC estimate put the size of the “digital universe” at 0.18 zettabytes in 2006 and is forecasting a tenfold growth by 2011 to 1.8 zettabytes.
- The trend is for every individual’s data footprint to grow, but perhaps more important, the amount of data generated by machines will be even greater than that generated by people. Machine logs, RFID readers, sensor networks, vehicle GPS traces, retail transactions—all of these contribute to the growing mountain of data.
- The good news is that Big Data is here. The bad news is that we are struggling to store and analyze it.

How to data capture and storage?

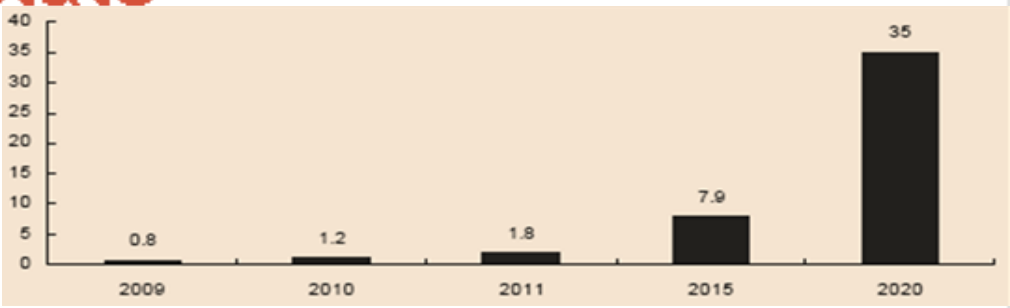
Some problems



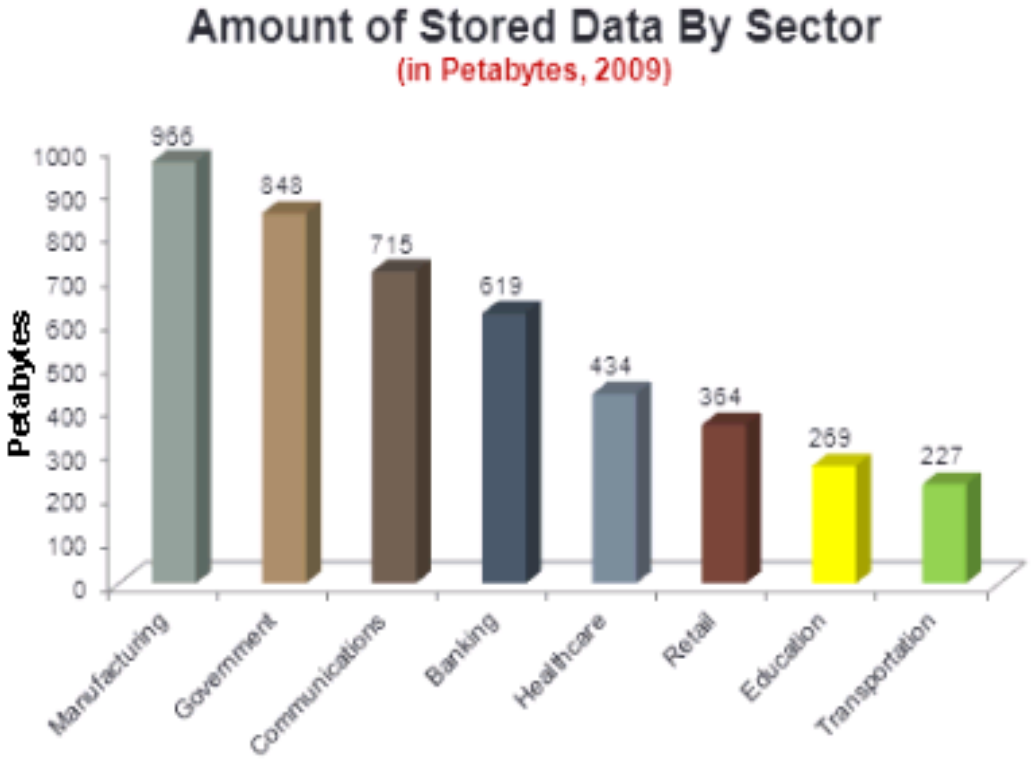


Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate.

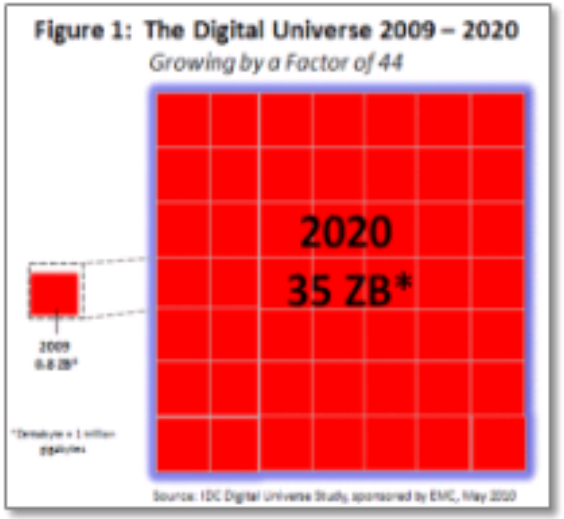
Some Big Data Stats



Rapid growth in data rates



Sources:
"Big Data: The Next Frontier for Innovation, Competition and Productivity."
US Bureau of Labor Statistics | McKinsey Global Institute Analysis



1 zettabyte?
= 1 million petabytes
= 1 trillion terabytes
= 1 quadrillion gigabytes

TB ⇨ PB ⇨ EB ⇨ ZB

What is Big Data?

- **Capturing and managing lots of information**
- **Working with many new types of data**

  Structure/Unstructured/ Semi structure
- **Exploiting these masses of information and new data types with new styles of applications**
- **Bigger than Terabytes**

  volume, variety, velocity, variability, value

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate.

Topics (Outline)

- 1. What is Big Data?
- 2. Application of Big Data
- 3. Big Data opportunities and challenges
- 4. How to Deal with Big Data?
- 5. What's Hadoop/MapReduce?
- 6. Big data players/Software Tools/Platforms
- 7. Examples

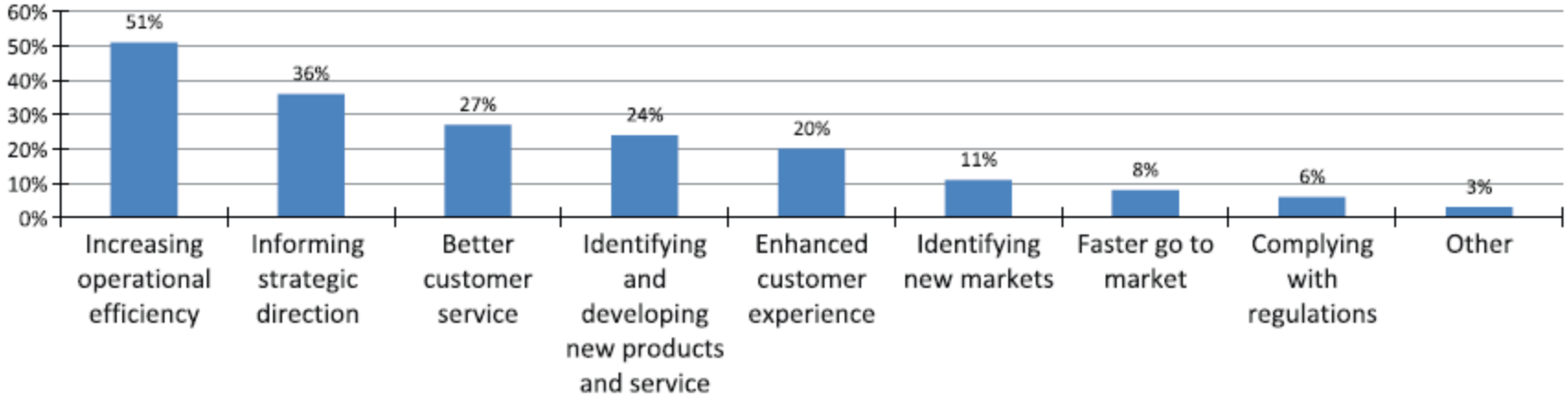
Application of Big Data

What This Means for You

Big Data can help a company do many things:

- Profile customers
- Determine pricing strategies
- Identify competitive advantages
- Better target advertising---Advertising recommendation
- Inform internal research and product development
- Strengthen customer service

Big Data opportunities and challenges

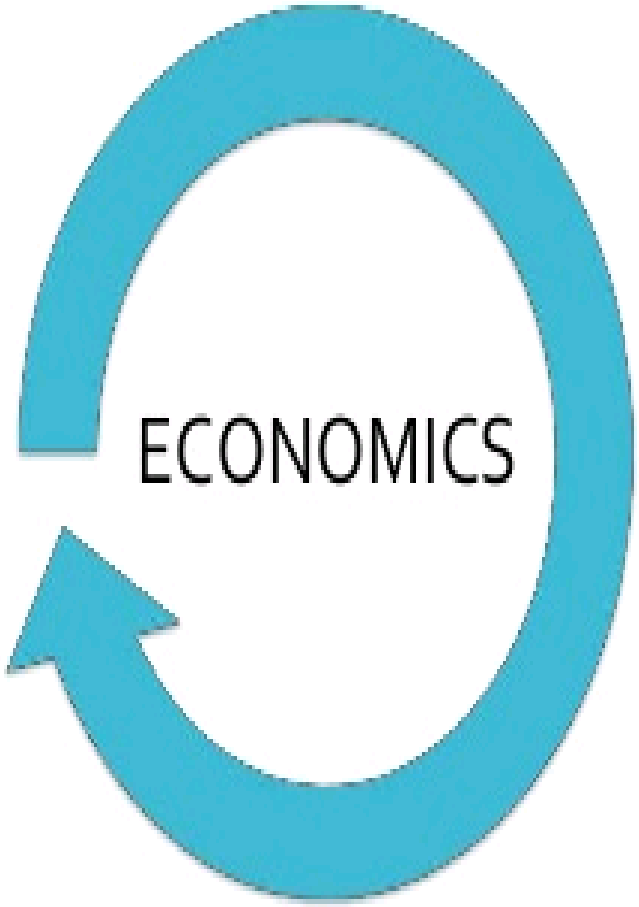


Big Data Opportunities: above 50% of 560 enterprises think Big Data will help them in increasing operational efficiency, etc.

- There are many advantages in business section that can be obtained through harnessing Big Data as illustrated in Fig, including increasing operational efficiency, informing strategic direction, developing better customer service, identifying and developing new products and services, identifying new customers and markets, etc.
- By liberal estimates , Big Data could produce $300 billion potential annual value to US health care, and €250 billion to European public administration. There will be $600 billion potential annual consumer surplus from using personal location data globally, and give a potential increase with 60%. Only in United States, Big Data produce 140,000 to 190,000 deep analytical talent positions and 1.5 million data-savvy managers. Undoubtedly, Big Data is usually juicy and lucrative if explored correctly.

Application of Big Data

BIG DATA
Economics



- Commodity hardware compatibility
- Reduction in storage cost
- Open source system
- The Web economy

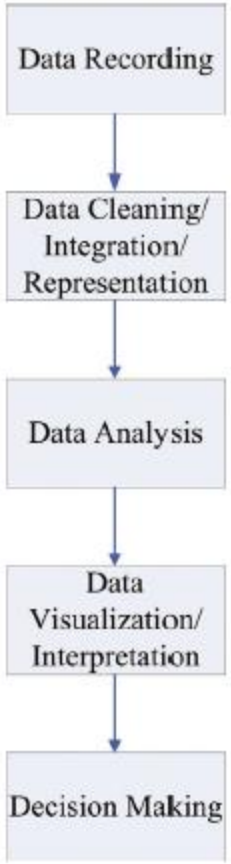
Big Data Solutions

- Cloudera: Cloudera Enterprise
- Microsoft:Windows Azure HDInsight Service
- Google:BigQuery
- Amazon: DynamoDB
- IBM: InfoSphere Streams/Netezza
- EMC: Greenplum
- TeraData: Aster MapReduce Platform
- Oracle: Hadoop/Mapreduce Big Data connectors
- Alibaba: AliCloud

libaba Gro
阿里巴巴集团

贵州科学院
Guizhou Academy of Sciences

Big Data opportunities and challenges



Knowledge discovery process.

- Opportunities are always followed by challenges. On the one hand, Big Data bring many attractive opportunities. On the other hand, we are also facing a lot of challenges when handle Big Data problems, difficulties lie in data capture, storage, searching, sharing, analysis, and visualization.

Challenges

- **Information growth**
- **Processing power**
- **Physical storage**

  disk capacity increase dramatically

  100 MB/S read from disk (bottle neck)

  data seeking time is slow than data transferring
- **Data issues**
- **Costs**

Topics (Outline)

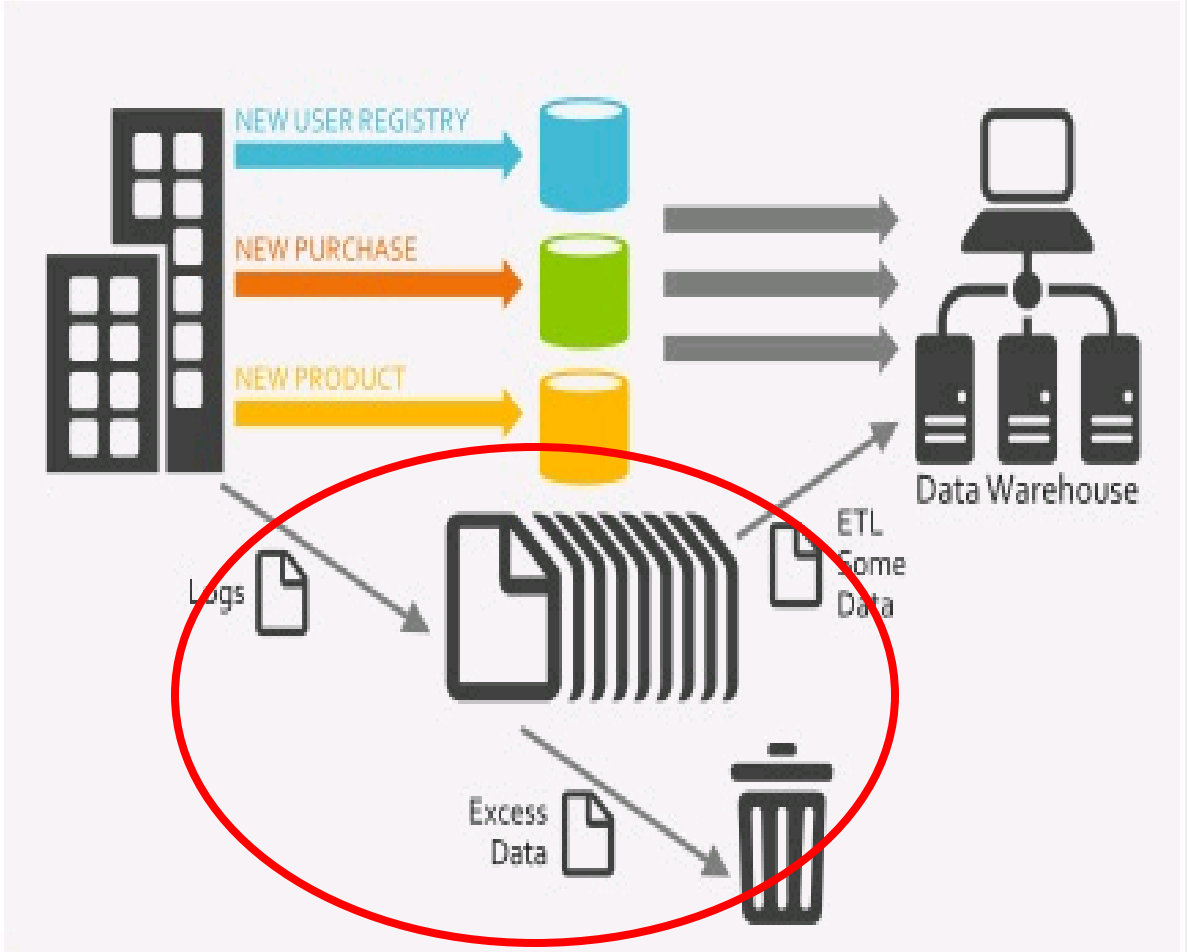
- **1. What is Big Data?**
- **2. Application of Big Data**
- **3. Big Data opportunities and challenges**
- **4. How to Deal with Big Data?**
- **5. What's Hadoop/MapReduce?**
- **6. Big data players/Software Tools/Platforms**
- **7. Examples**

Main steps in adopting an analytical system

- 1. What Will We Analyze?
- 2. Do We Buy or Build?
- 3. Are We Ready to Invest?
- 4. Do We Understand the Impact?

OPERATIONAL DATA

With current implementation



Serial Parallel Extract-Transform-Load

OPERATIONAL DATA

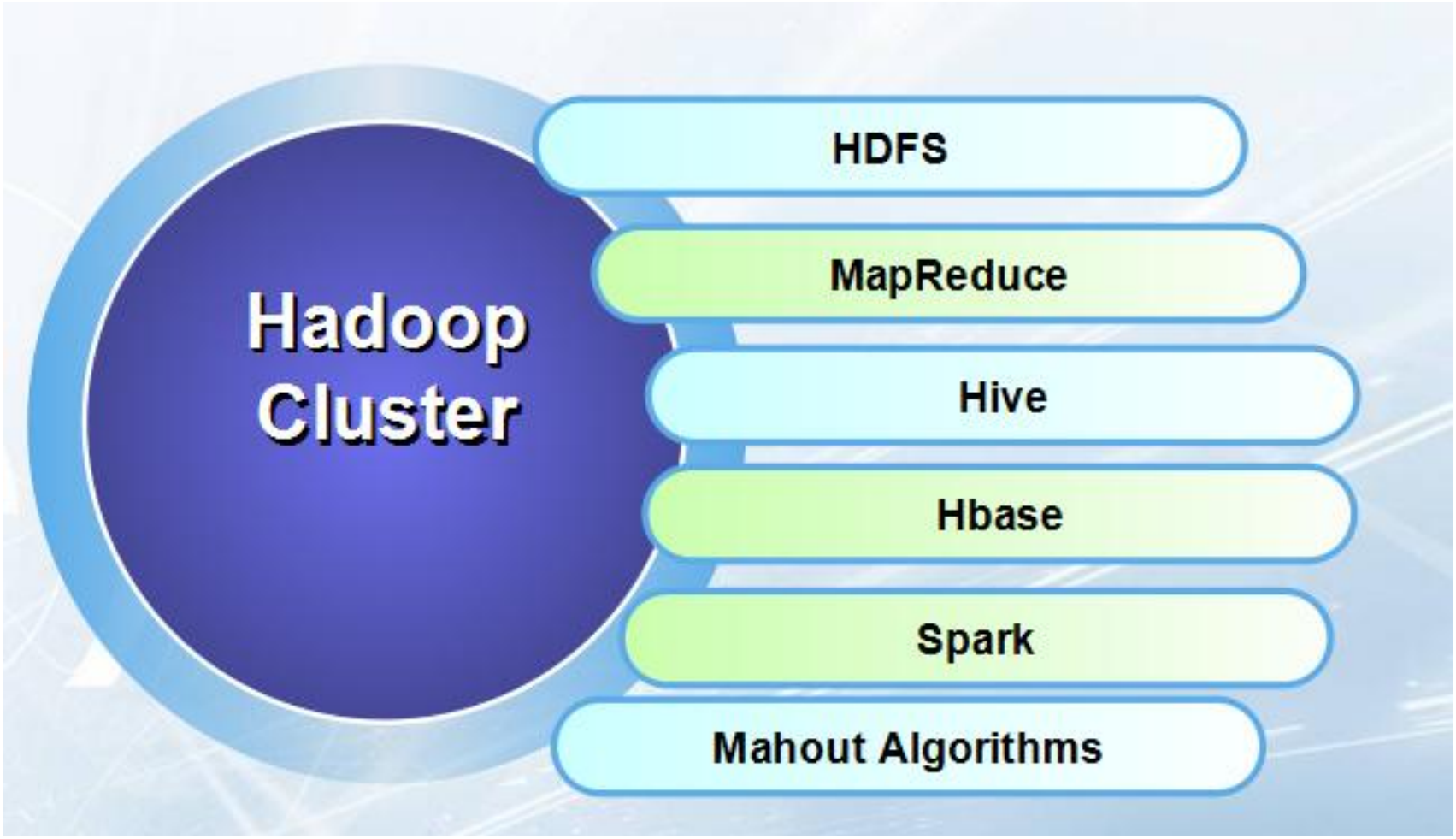
With big data implementation

Not serial but parallel example: Hadoop cluster

How to Deal with Big Data?

- Hadoop
- MapReduce
- Rational Database Management System(RDBMS)
- RDBMS vs MapReduce
- HDFS

How to Deal with Big Data?

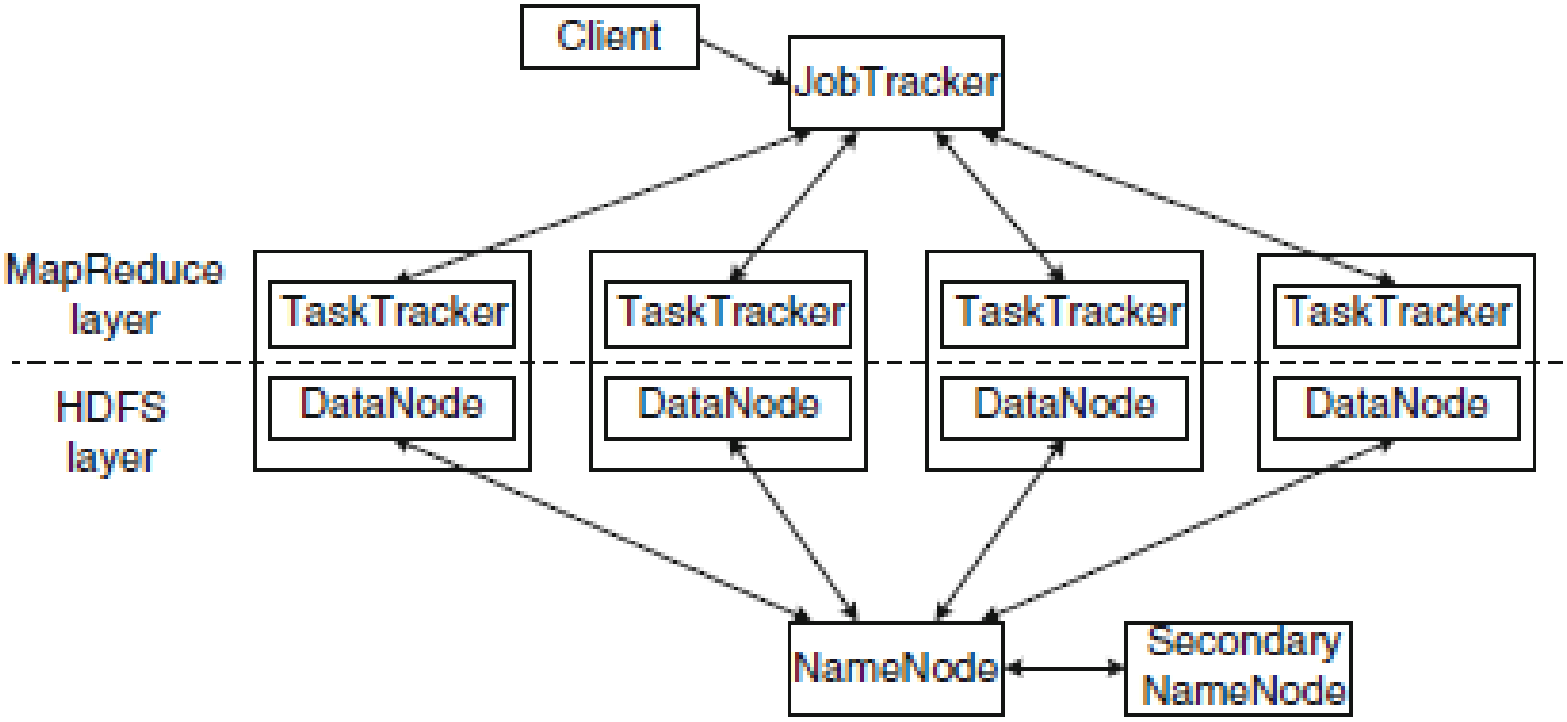


- Hadoop is an open-source implementation of MapReduce, and without doubt, the most popular MapReduce variant currently in use in an increasing number of prominent companies with large user bases, including companies such as Yahoo! and Facebook.

Topics (Outline)

- 1. What is Big Data?
- 2. Application of Big Data
- 3. Big Data opportunities and challenges
- 4. How to Deal with Big Data?
- 5. What's Hadoop/MapReduce?
- 6. Big data players/Software Tools/Platforms
- 7. Examples

Hadoop



Hadoop consists of two main parts: the Hadoop distributed file system (HDFS) and MapReduce for distributed processing. As illustrated in Fig., Hadoop consists of a number of different daemons/servers: NameNode, DataNode, and Secondary NameNode formanaging HDFS, and JobTracker and TaskTracker for performing MapReduce.

What is Hadoop

- Hadoop is a distributed computing framework with two main components:
  - a distributed file system and
  - a map-reduce implementation.

- Imagine you have a cluster of 100 computers. Hadoop's distributed file system makes it so you can put data "into Hadoop" and pretend that all the hard drives on your machines have coalesced into one gigantic drive.

- Under the hood, it breaks each file you give it into 64- or 128-MB chunks called blocks and sends them to different machines in the cluster, replicating each block three times along the way.

MapReduce

- MapReduce is a programming model for data processing. The model is simple, yet not too simple to express useful programs in.
- Most important, MapReduce programs are inherently parallel, thus putting very large-scale data analysis into the hands of anyone with enough machines at her disposal.
- Highly fault tolerant

  nodes are expected to fail
- Every data block (by default) replicated on 3 nodes (is also rack aware)

How does MapReduce work

MapReduce works by breaking the processing into two phases:

--- the map phase and the reduce phase

Each phase has key-value pairs as input and output, the types of which may be chosen by the programmer.

The programmer also specifies two functions:

--- the map function and the reduce function

Format and Types

- MapReduce model in detail, and, in particular, how data in various formats, from simple text to structured binary objects, can be used with this model
- MapReduce uses *key/value pairs*.

(Traditionally using rows and columns)

map: (K1, V1) → list(K2, V2)

reduce: (K2, list(V2)) → list(K3, V3)

Map

- MapReduce uses *key/value pairs*.
  (Traditionally using rows and columns)

Example: last name/chen
withdraw amount/20
transaction date/06-23-2013

Reduce

- all the intermediate values for a given output key are combined together into a list.

- The reduce() function then combines the intermediate values into one or more final values for the same key.

Example: A Weather Dataset

- For our example, we will write a program that mines weather data. Weather sensors collect data every hour at many locations across the globe and gather a large volume of log data, which is a good candidate for analysis with MapReduce because it is semistructured and record-oriented.
- The data we will use is from the National Climatic Data Center (NCDC, *http://www.ncdc.noaa.gov/*).

Example: A Weather Dataset

- Example: it is a small script to calculate the maximum temperature (Fahrenheit) for each year.
- Our map function is simple. We pull out the year and the air temperature because these are the only fields we are interested in. In this case,
- the map function is just a data preparation phase, setting up the data in such a way that the reducer function can do its work on it: finding the maximum temperature for each year.
- The reduce function is also a good place to drop bad records: here we filter out temperatures that are missing, suspect, or erroneous.

Example: A Weather Dataset

- To visualize the way the map works, consider the following sample lines of input data (some unused columns have been dropped to fit the page, indicated by ellipses):

```
0067011990999991950051507004...9999999N9+00001+99999999999...
0043011990999991950051512004...9999999N9+00221+99999999999...
0043011990999991950051518004...9999999N9-00111+99999999999...
0043012650999991949032412004...0500001N9+01111+99999999999...
0043012650999991949032418004...0500001N9+00781+99999999999...
```

These lines are presented to the map function as the key-value pairs:

```
(0, 0067011990999991950051507004...9999999N9+00001+99999999999...)
(106, 0043011990999991950051512004...9999999N9+00221+99999999999...)
(212, 0043011990999991950051518004...9999999N9-00111+99999999999...)
(318, 0043012650999991949032412004...0500001N9+01111+99999999999...)
(424, 0043012650999991949032418004...0500001N9+00781+99999999999...)
```

Example: A Weather Dataset

The keys are the line offsets within the file, which we ignore in our map function. The map function merely extracts the year and the air temperature (indicated in bold text), and emits them as its output (the temperature values have been interpreted as integers):

```
(1950, 0)
(1950, 22)
(1950, -11)
(1949, 111)
(1949, 78)
```

The output from the map function is processed by the MapReduce framework before being sent to the reduce function. This processing sorts and groups the key-value pairs by key. So, continuing the example, our reduce function sees the following input:

```
(1949, [111, 78])
(1950, [0, 22, -11])
```

Each year appears with a list of all its air temperature readings. All the reduce function has to do now is iterate through the list and pick up the maximum reading:

```
(1949, 111)
(1950, 22)
```

This is the final output: the maximum global temperature recorded in each year.

Fahrenheit

MapReduce logical data flow

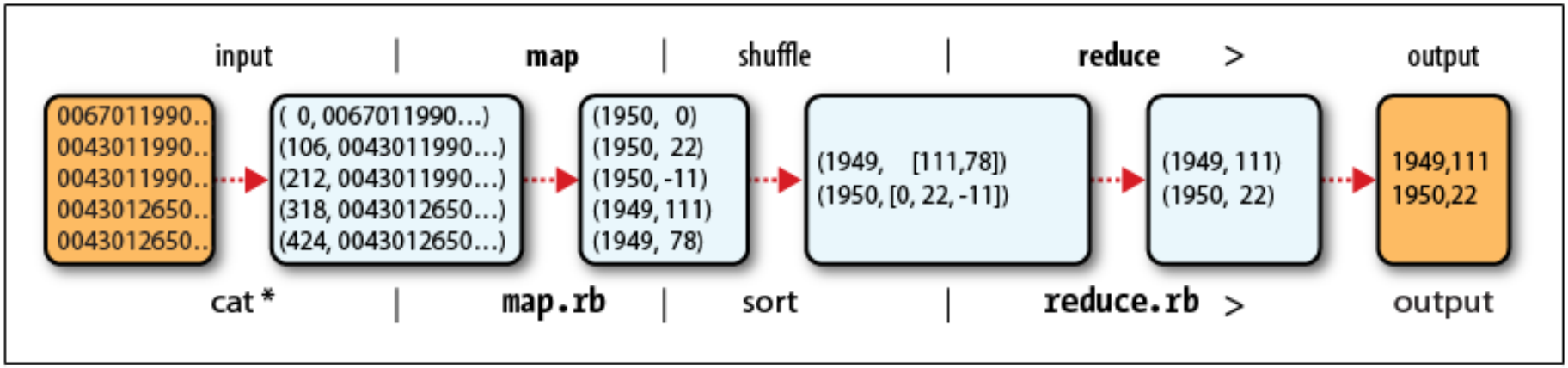


Figure. MapReduce logical data flow

The whole data flow is illustrated in Figure. At the bottom of the diagram is a Unix pipeline, which mimics the whole MapReduce flow and which we will see again later in this chapter when we look at Hadoop Streaming.

Mapreduce Special Feature

- Counter
- Sorting
- Joins
- Shuffle---merge / sort

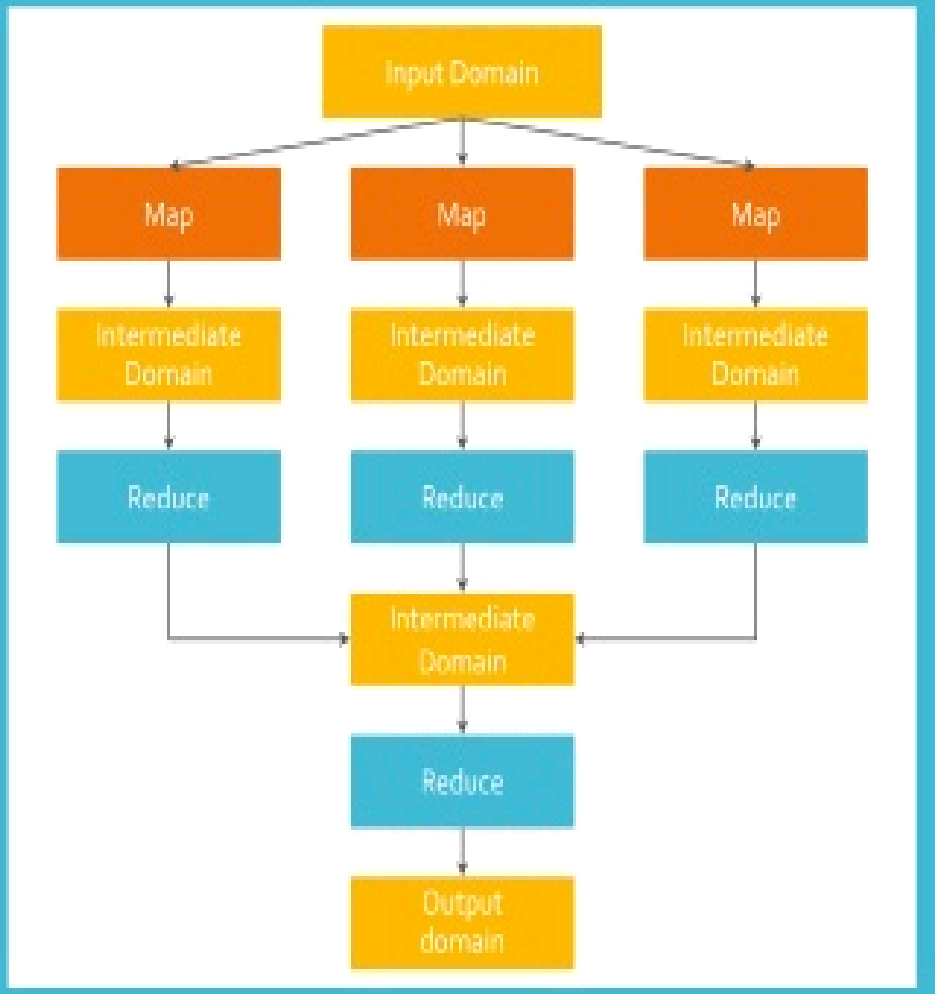
MapReduce guarantees that the input to every reducer is sorted by key. The process by which the system performs the sort—and transfers the map outputs to the reducers as inputs – Shuffle

MapReduce – Workflow

A MapReduce *job* usually splits the input data-set into independent chunks which are processed by the *map tasks* in a completely parallel manner

The framework sorts the outputs of the maps, which are then input to the *reduce* tasks

The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks



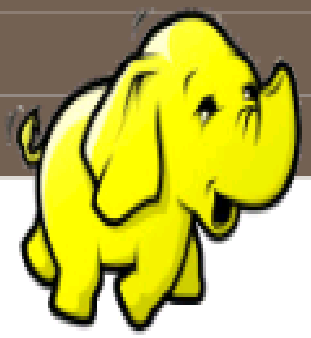
RDBMS

- fixed-schema, row-oriented databases with ACID properties and a sophisticated SQL query engine.
- The emphasis is on strong consistency, referential integrity, abstraction from the physical layer, and complex queries through the SQL language.
- easily create secondary indexes, perform complex inner and outer joins, count, sum, sort, group, and page your data across a number of tables, rows, and columns.

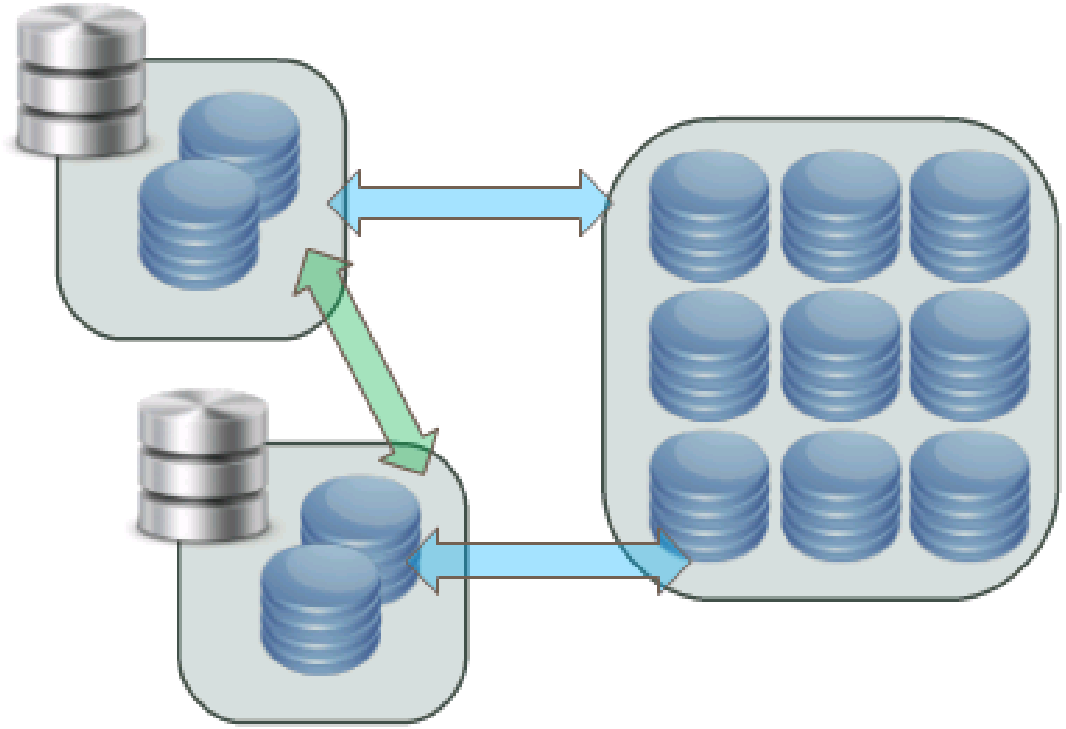
RDBMS vs MapReduce

	Traditional RDBMS	MapReduce
Data Size	Gigabytes *(Terabytes)*	Petabytes *(Hexabytes)*
Access	Interactive and Batch	Batch
Updates	Read / Write many times	Write once, Read many times
Structure	Static Schema	Dynamic Schema
Integrity	High (ACID)	Low
Scaling	Nonlinear	Linear
DBA Ratio	1:40	1:3000

Reference: Tom White's Hadoop: The Definiti



Comparing RDBMS and MapReduce



Traditional RDBMS: Move Data to Compute

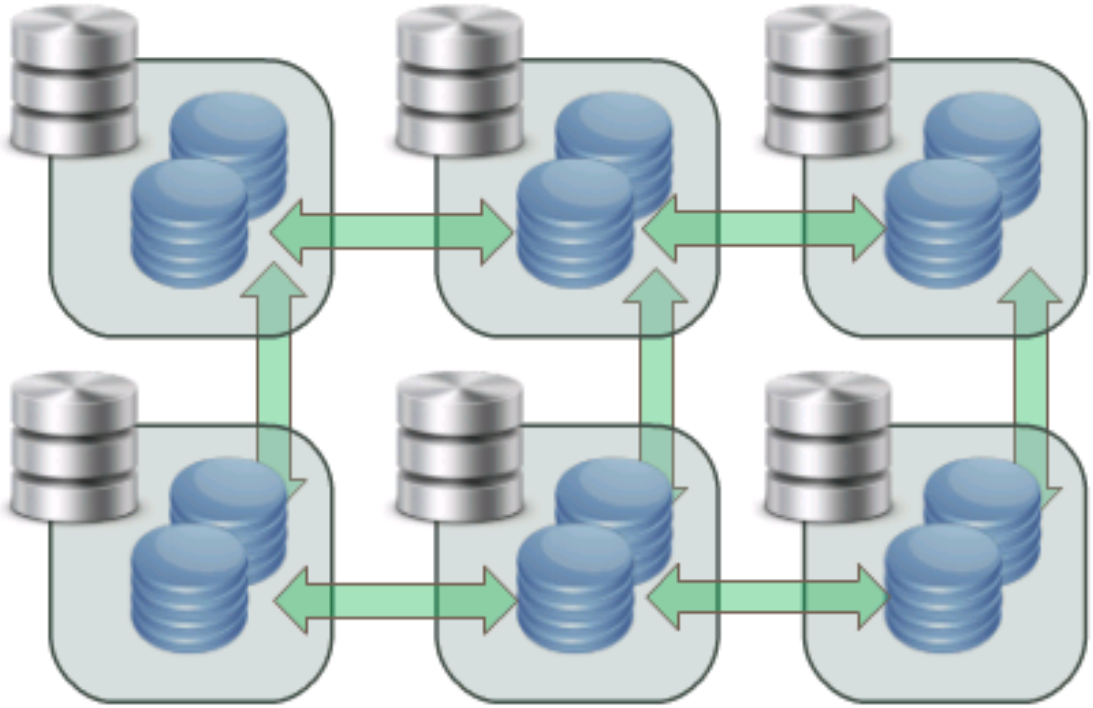
As you process more and more data, and you want interactive response

- Typically need more expensive hardware
- Failures at the points of disk and network can be quite problematic

It's all about ACID: atomicity, consistency, isolation, durability

Can work around this problem with more expensive HW and systems

- Though distribution problem becomes harder to do



Hadoop: Move Compute to the Data

Hadoop (and NoSQL in general) follows the Map Reduce framework

- Developed initially by Google -> Map Reduce and Google File system
- Embraced by community to develop MapReduce algorithms that are very robust
- Built Hadoop Distributed File System (HDFS) to auto-replicate data to multiple nodes
- And execute a single MR task on all/many nodes available on HDFS

Use commodity HW: no need for specialized and expensive network and disk

Not so much ACID, but BASE (basically available, soft state, eventually consistent)

Architechures

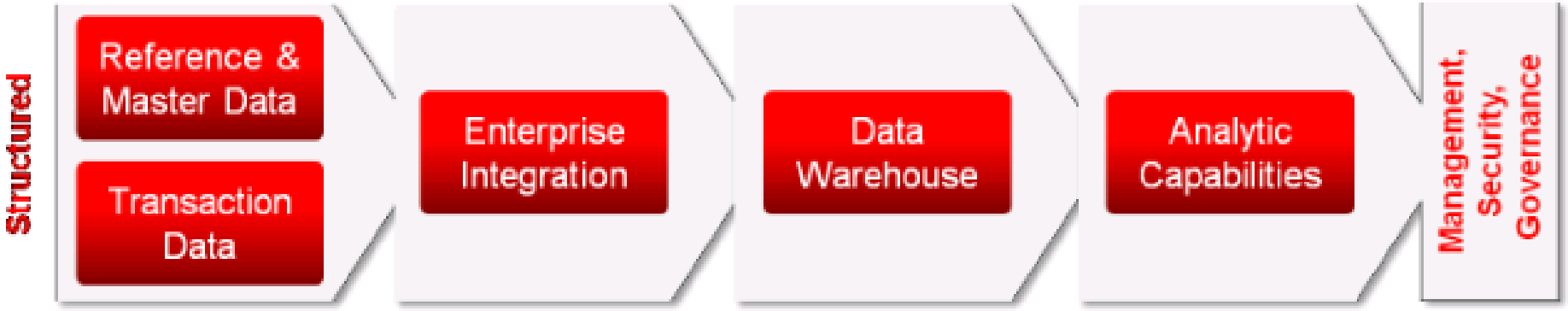


Figure 1: Traditional Information Architecture Capabilities

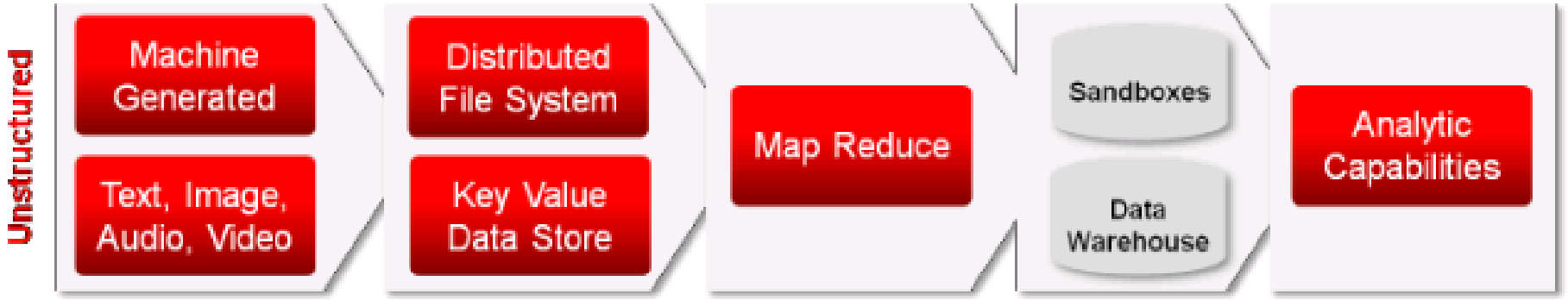
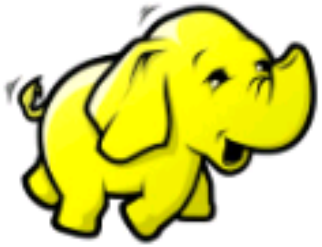


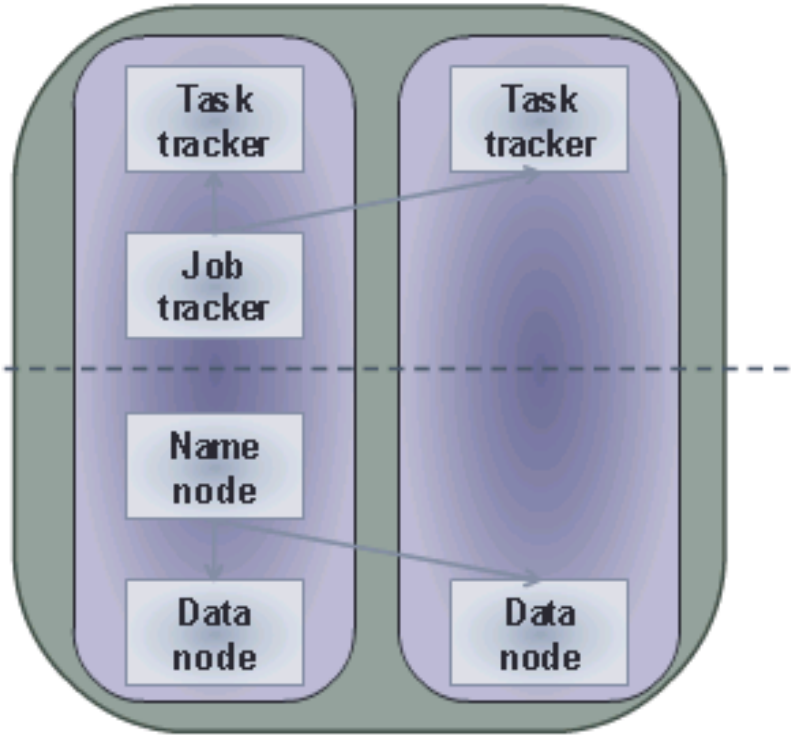
Figure 2: Big Data Information Architecture Capabilities

贵州科学院
Guizhou Academy of Sciences



Map Reduce Layer

HDFS Layer



Reference: http://en.wikipedia.org/wiki/File:Hadoop_1.png

What is Hadoop?

- Synonymous with the Big Data movement
- Infrastructure to automatically distribute and replicate data across multiple nodes and execute and track map reduce jobs across all of those nodes
- Inspired by Google's Map Reduce and GFS papers
- Components are: Hadoop Distributed File System (HDFS), Map Reduce, Job Tracker, and Task Tracker
- *Based on the Yahoo! "Nutch" project in 2003, became Hadoop in 2005 named after Doug Cutting's son's toy elephant*

Design of HDFS

- Namenodes (The Master)

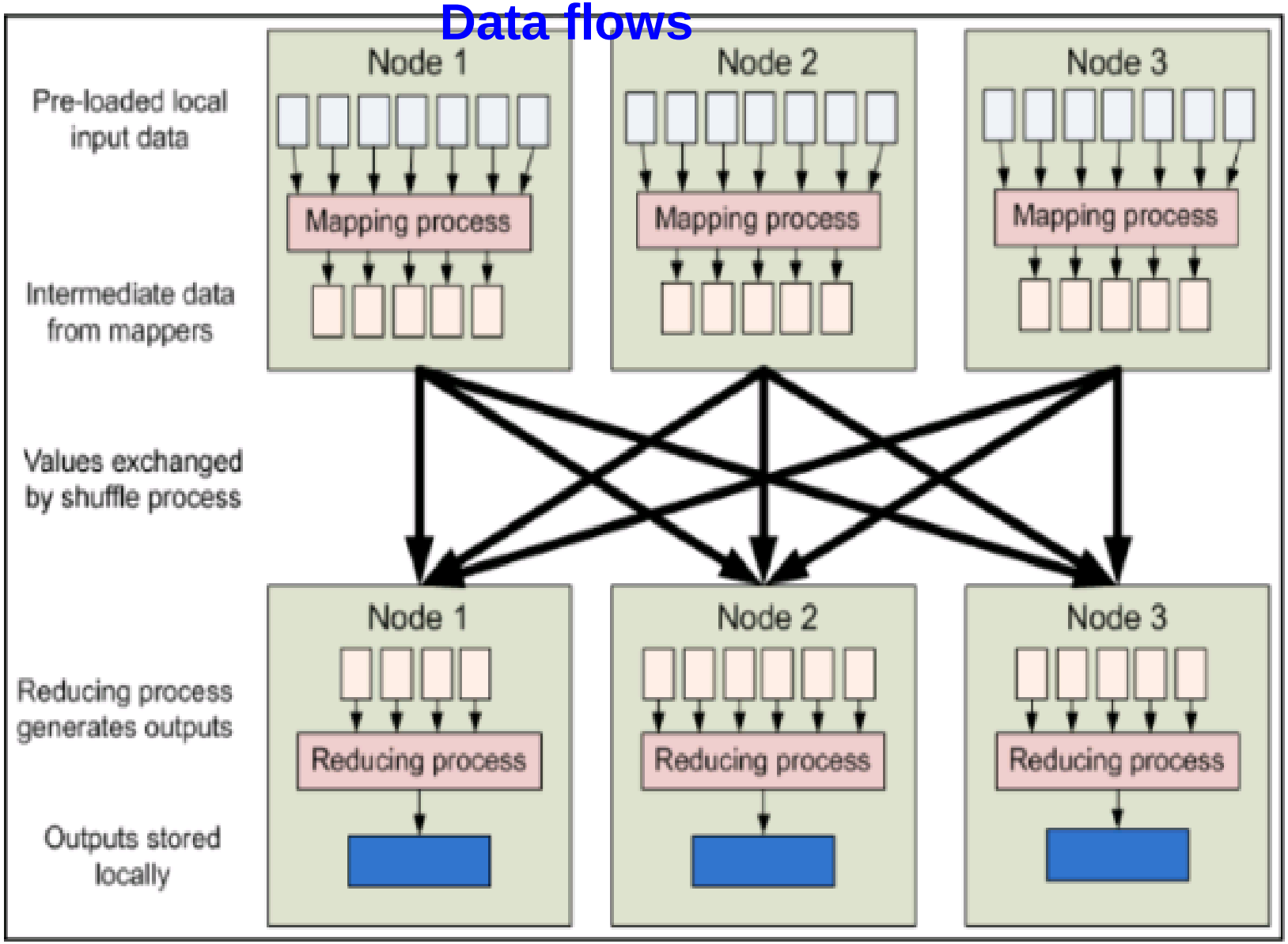
  Manage metadata/file trees
- Datanodes (Workers)

  store/retrieve data block
- Datanodes do not use RAID disk.

  HDFS round-robins HDFS blocks between all disks. RAID limited by the slowest disk on the array.
- Block

64 MB/128MB (normal disk block 512 KB).

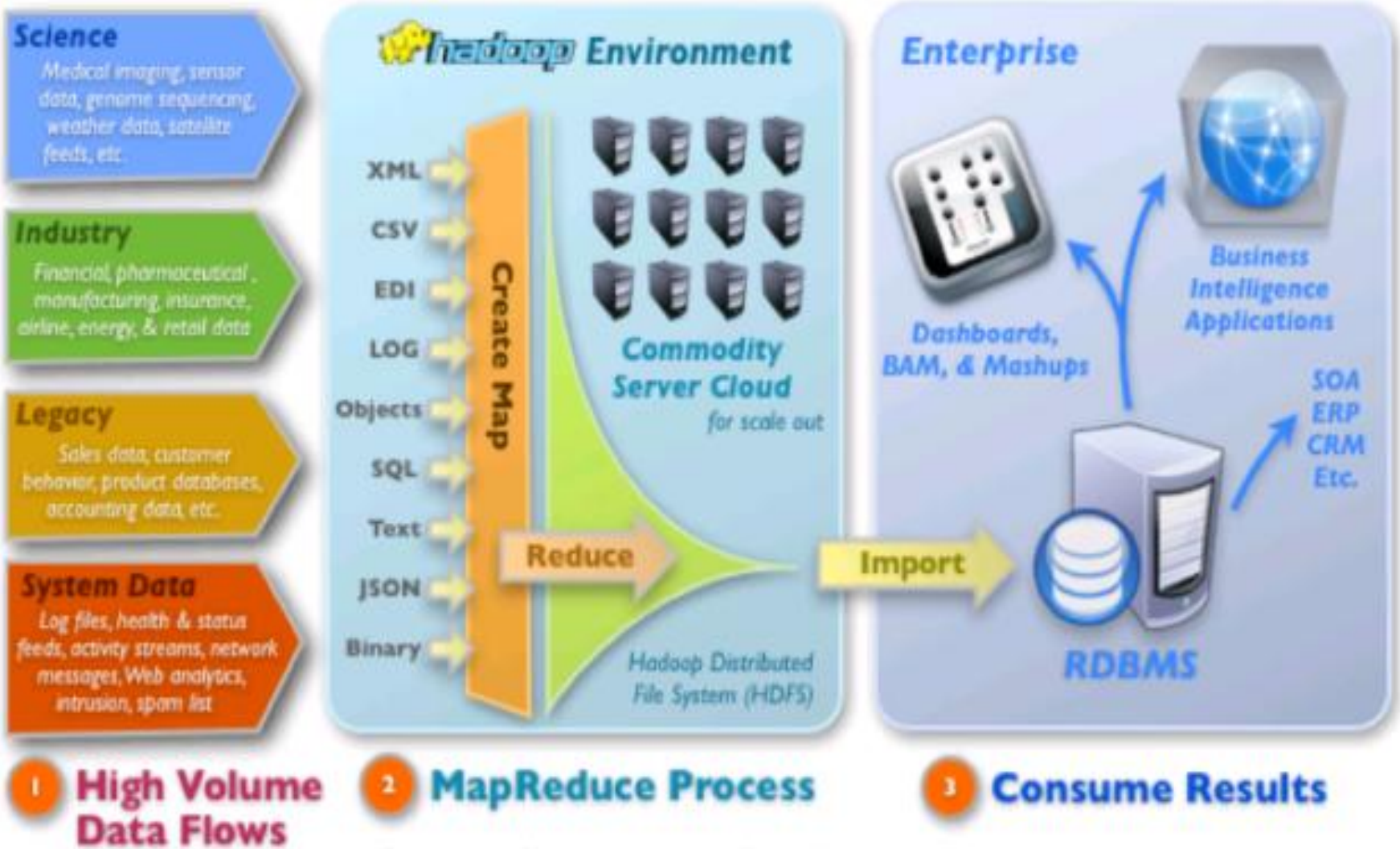
Data flows



Redundancy **Fault tolerant**

贵州科学院 Guizhou Academy of Sciences

Using Hadoop in the Enterprise



Topics (Outline)

- **1. What is Big Data?**
- **2. Application of Big Data**
- **3. Big Data opportunities and challenges**
- **4. How to Deal with Big Data?**
- **5. What's Hadoop/MapReduce?**
- **6. Big data players/Software Tools/Platforms**
- **7. Examples**

MapReduce logical data flow

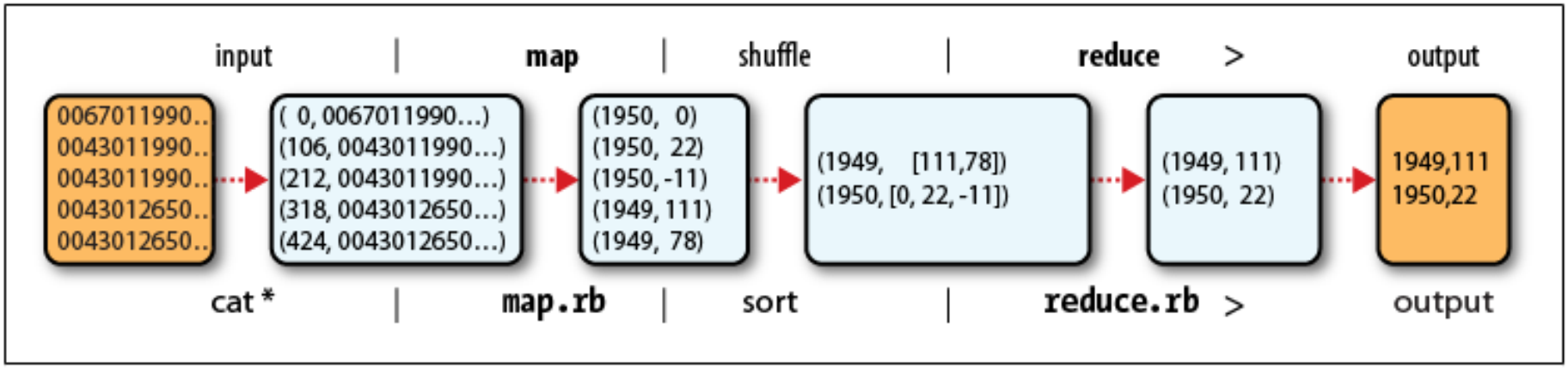
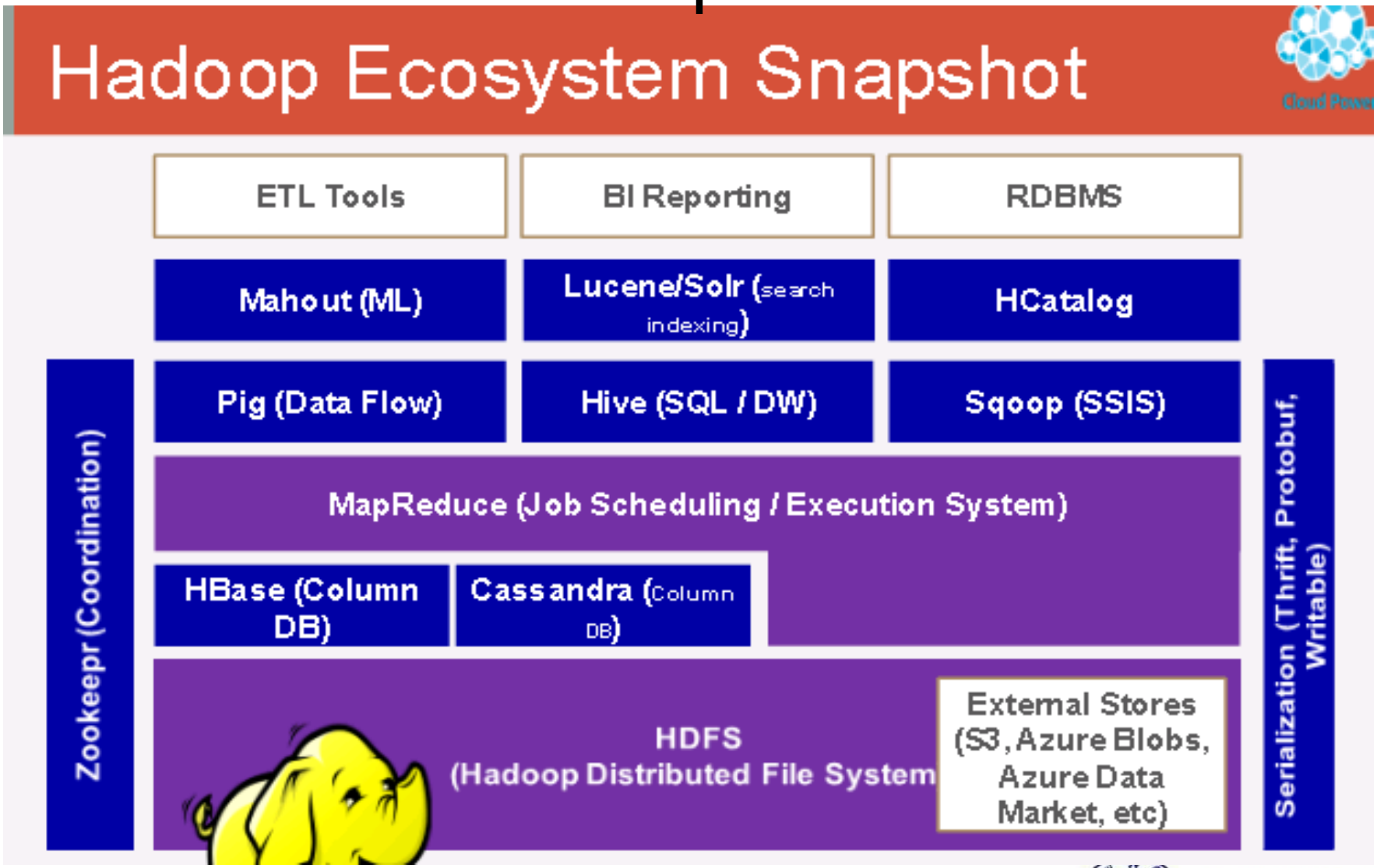


Figure. MapReduce logical data flow

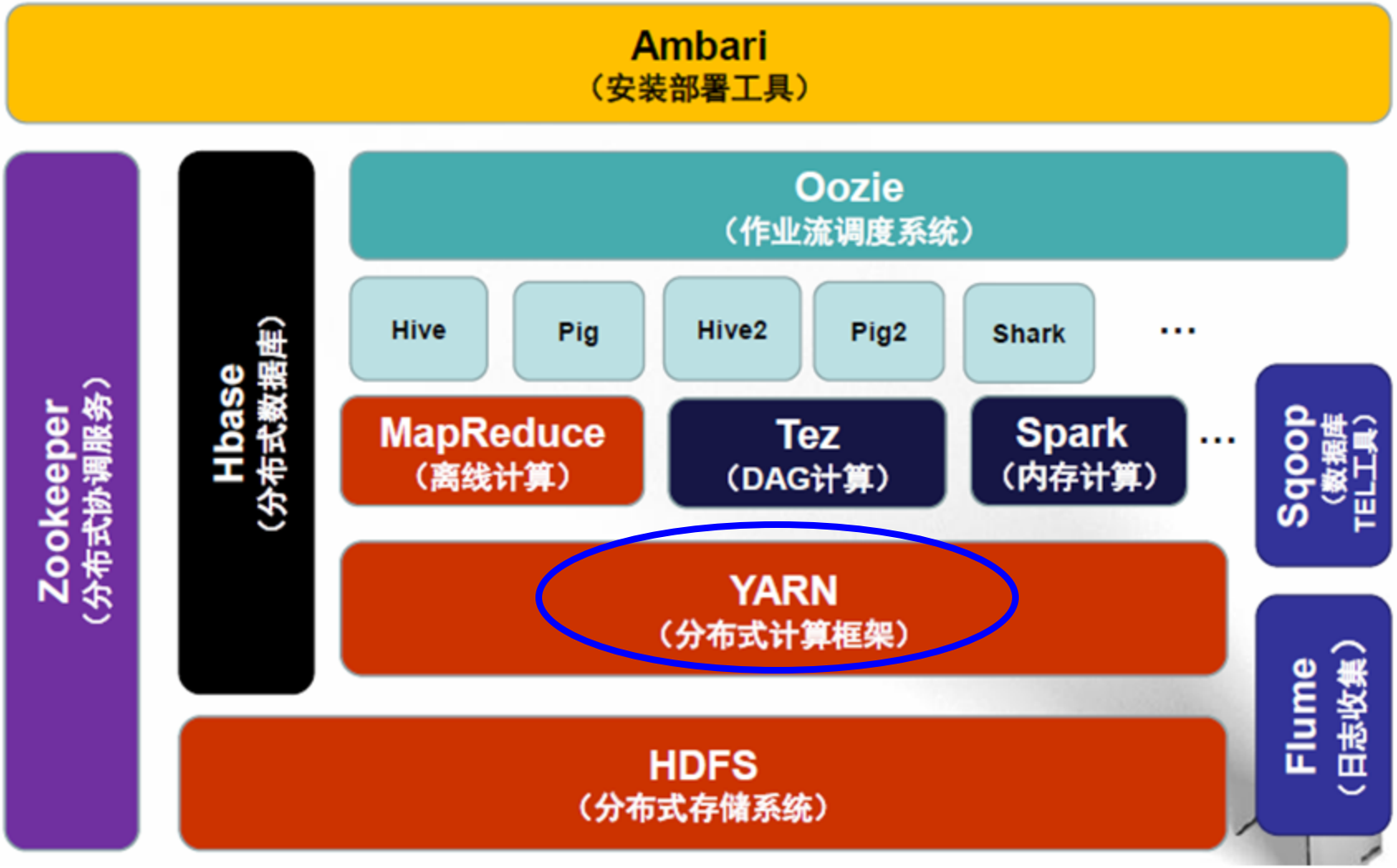
The whole data flow is illustrated in Figure. At the bottom of the diagram is a Unix pipeline, which mimics the whole MapReduce flow and which we will see again later in this chapter when we look at Hadoop Streaming.

Hadoop 1.0

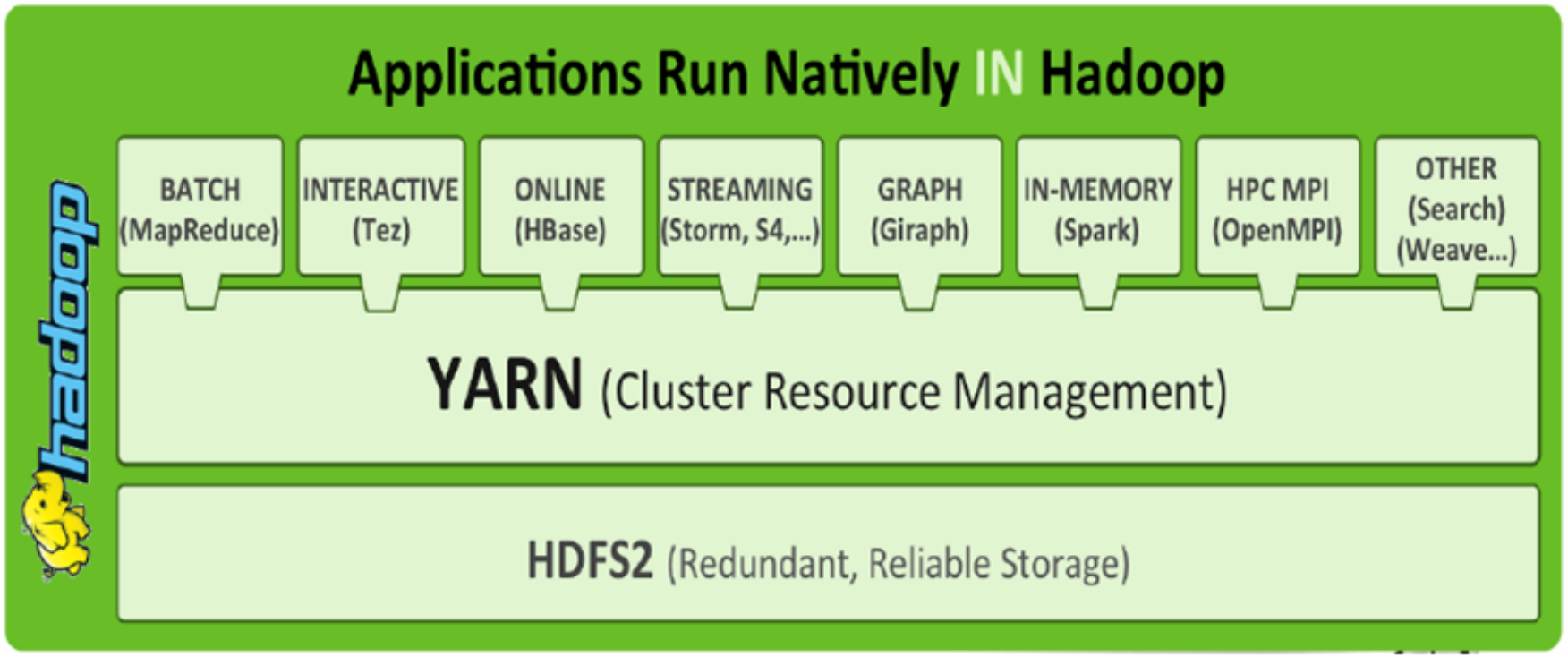


Hierarchy chart

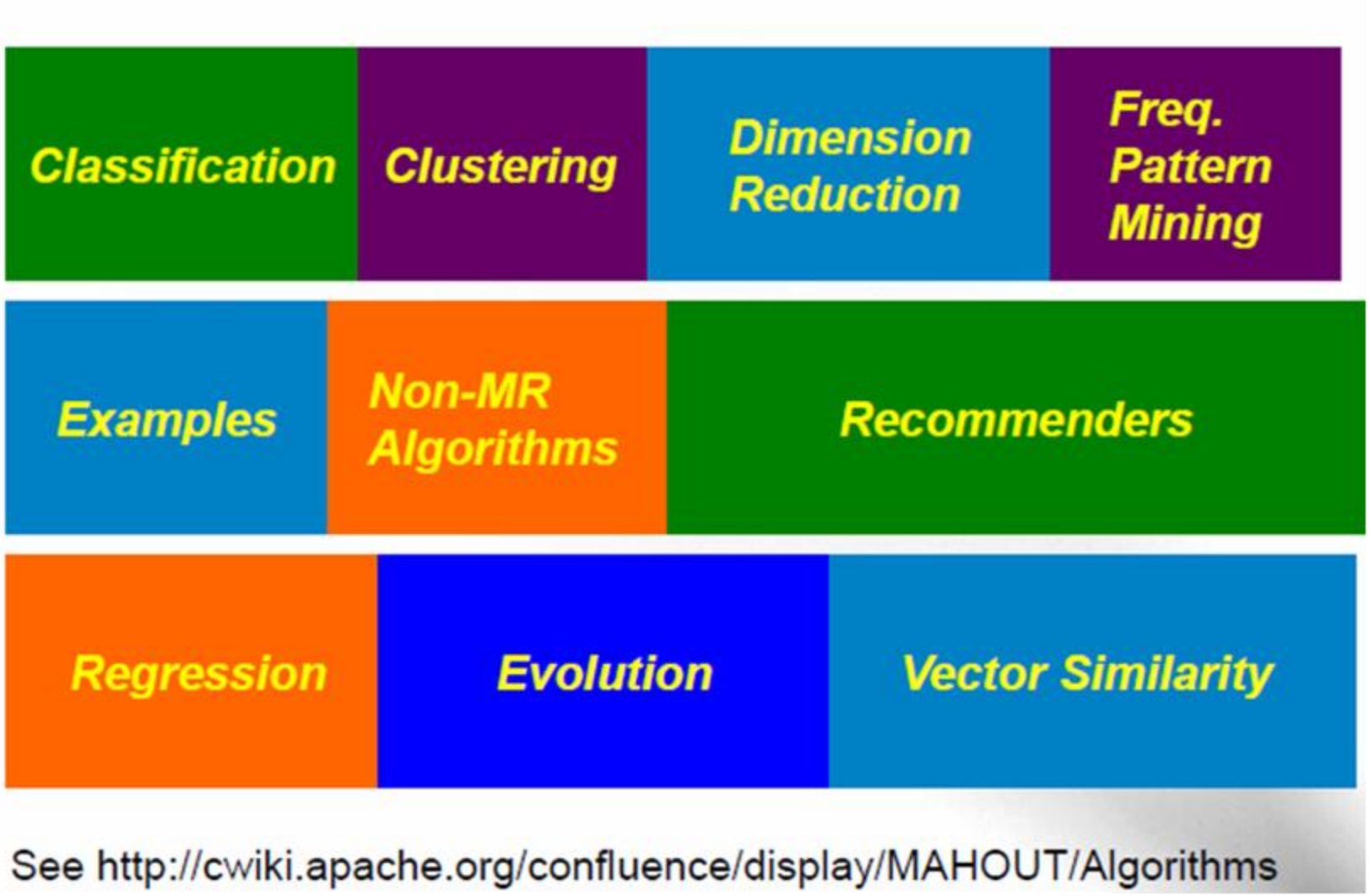
Hadoop 2.0



Hierarchy chart



Hierarchy chart



Machine learning algorithm

Install Hadoop

- % cd /usr/local
- % sudo tar xzf hadoop-x.y.z.tar.gz
- change the owner of the Hadoop files to be the hadoop user and group:
- % sudo chown -R hadoop: hadoop hadoop-x.y.z

- http://hadoop.apache.org/

Run on Linux operating system

Topics (Outline)

- **1. What is Big Data?**
- **2. Application of Big Data**
- **3. Big Data opportunities and challenges**
- **4. How to Deal with Big Data?**
- **5. What's Hadoop/MapReduce?**
- **6. Big data players/Software Tools/Platforms**
- **7. Examples**

1. We are starting the Big Data Projects

2. Examples

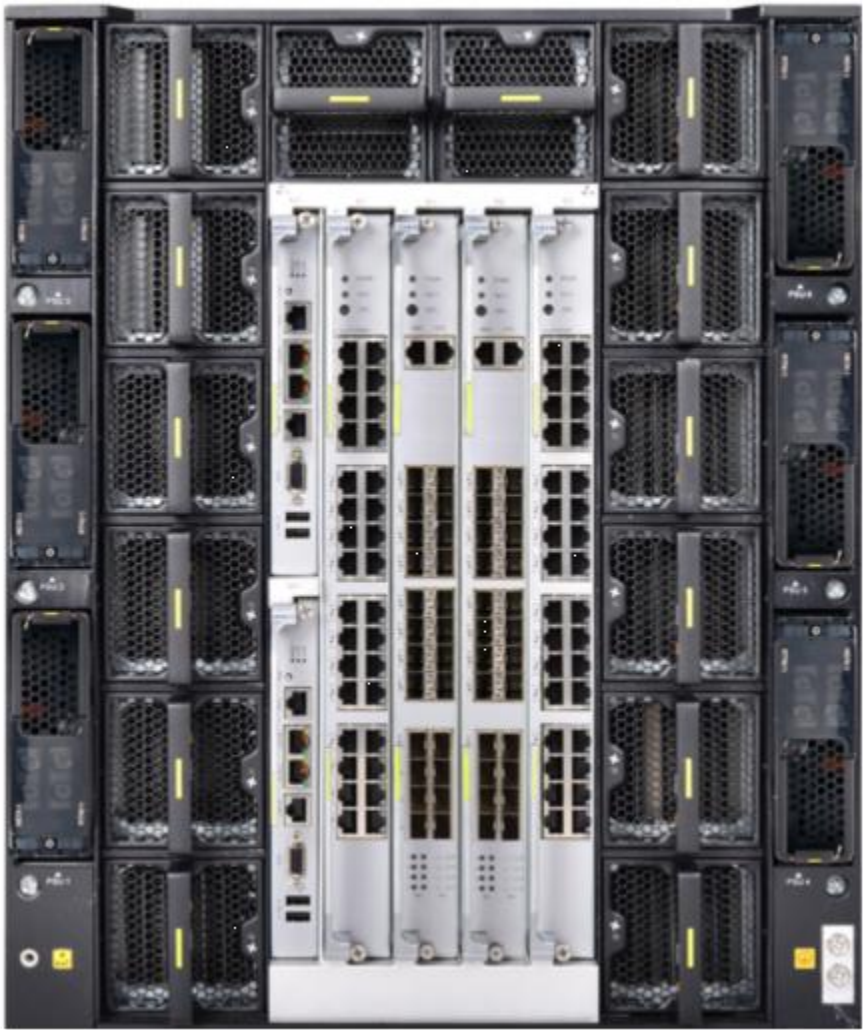
- Real-Time Urban Monitoring Using Cell Phones
- Mobile phone usage in complex urban systems: a space–time, aggregated human activity
- Unveiling patterns of international communities in a global city using mobile phone data
- NextCell Predicting Location Using Social Interplay from Cell Phone Traces

We are starting the Big Data Projects

- **Considering the infrastructure used, all the experiments were run at the atlas research group cluster. This cluster is composed of 16 nodes, each with two Intel E5-2630 microprocessors (at 2.30 GHz, 128 GB cache) and 2 TB of main memory, connected with 1 Gb/s ethernet. All of them work under Linux CentOS 6.4.**

- **The cluster is configured with Hadoop and Mahout. One of the nodes is configured as name-node and job-tracker, and the remaining nodes are both datanodes and trask-trackers.The Hadoop version used is 2.0 (Cloudera CDH5.2.0) and the Mahout version is 0.8.**

Platform



Huawei E9000 The total value is ￥1,700,000

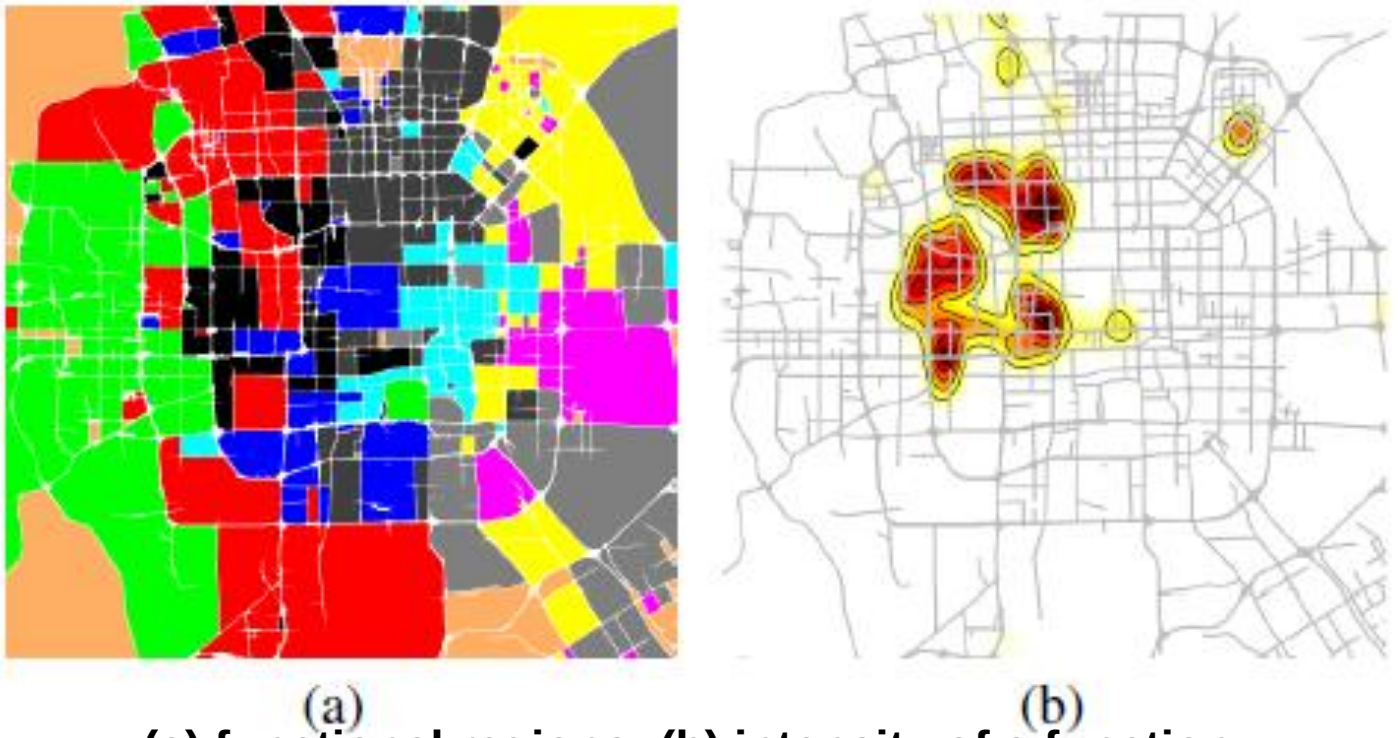
整机硬件

设备(当前配置数目/满配数目)

CPU(2/2)	
CPU 1	Intel(R) Xeon(R) CPU E5-2630 @ 2.30GHz
CPU 2	Intel(R) Xeon(R) CPU E5-2630 @ 2.30GHz
内存(8/24)	
DIMM000	DIMM000,Samsung, 16384 MB, 1600 MHz
DIMM010	DIMM010,Samsung, 16384 MB, 1600 MHz
DIMM020	DIMM020,Samsung, 16384 MB, 1600 MHz
DIMM030	DIMM030,Samsung, 16384 MB, 1600 MHz
DIMM100	DIMM100,Samsung, 16384 MB, 1600 MHz
DIMM110	DIMM110,Samsung, 16384 MB, 1600 MHz
DIMM120	DIMM120,Samsung, 16384 MB, 1600 MHz
DIMM130	DIMM130,Samsung, 16384 MB, 1600 MHz
硬盘(2/2)	
RAID卡(1/1)	
RAID卡	LSI SAS2308
网卡(4/12)	
网卡 1	MAC: E0-97-96-02-13-41
网卡 2	MAC: E0-97-96-02-13-42
网卡 3	MAC: E0-97-96-02-13-43
网卡 4	MAC: E0-97-96-02-13-44
Mezz卡(1/2)	

Hardware parameters for each server

The objectives of the Big Data Projects

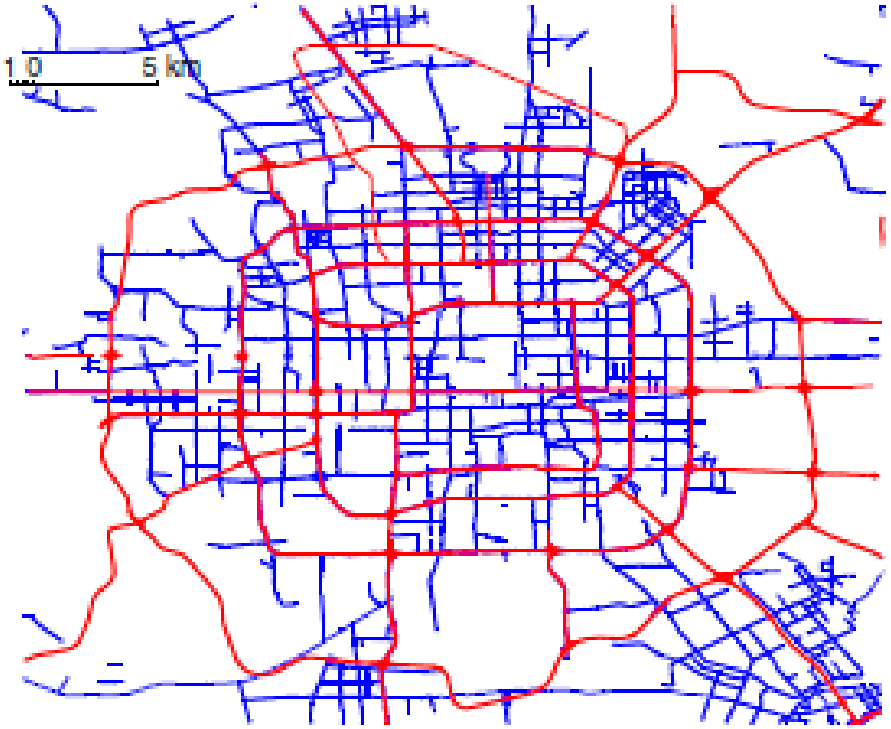


(a) functional regions; (b) intensity of a function

The step of urbanization and modern civilization leads to different functional regions in a city, e.g., residential areas, business districts, and educational areas, which support different needs of people's urban lives and serve as a valuable organizing technique for framing detailed knowledge of a metropolitan. These regions may be artificially designed by urban planners, or naturally formulated according to people's actual lifestyle, and would change functions and territories with the development of a city.

Discovering regions of different functions can enable a variety of valuable applications.

- **First, it can provide people with a quick understanding of a complex city** (like New York City, Tokyo, Beijing, and Paris) and social recommendations. For example, tourists can easily differentiate some scenic areas from business districts given these functional regions, thereby reducing effort for trip planning.
- **Second,** these functional regions can **calibrate the urban planning of a city and contribute to the future planning to some extent**. It is not surprising that a city did not evolve as its original planning, given the complexity of urban planning itself and the difficulty in predicting the development of a city.
- **Third,** these functional regions would also **benefit location choosing for a business and advertisement.** For instance, when building a supermarket we need to consider the distance to the residential areas, and the advertisement for a training course could be better put considering the geospatial intensity of the educational function.



Beijing road network. red: level-0/1; blue: level-2

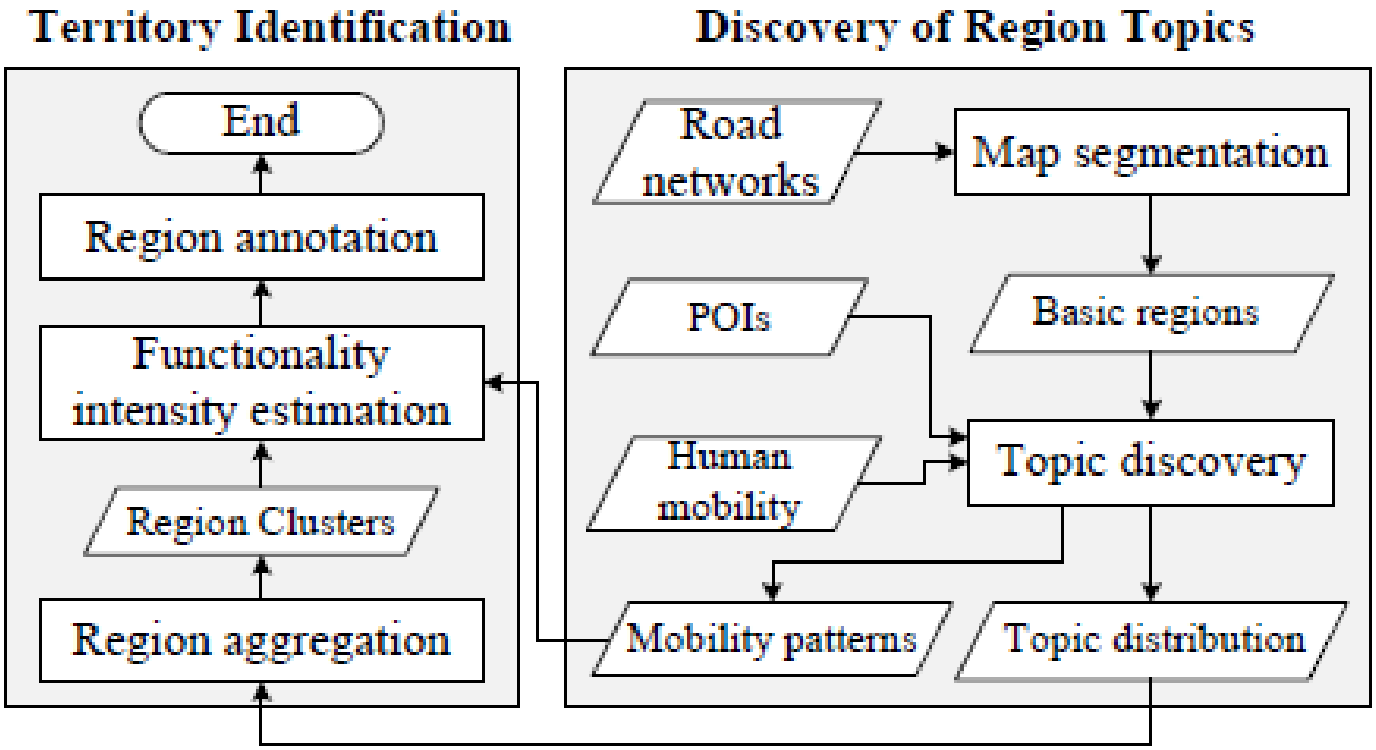
A road network is usually comprised of some major roads like highways and ring roads, which naturally partition a city into regions.

For example, as shown in Figure, the red segments denote freeways and city expressways in Beijing, and blue segments represent urban arterial roads. The three kinds of roads are associated with a road level 0, 1, and 2 respectively (in a road network database), forming a natural segmentation of the urban area of Beijing.

STEP

- Map Segmentatio
- Topic Discovery

Framework



Framework for discovering the functional regions

Table : POI category taxonomy

code	POI category	code	POI category
1	car service	16	banking and insurance service
2	car sales	17	corporate business
3	car repair	18	street furniture
4	motorcycle service	19	entrance/bridge
5	Café/Tea Bar	20	public utilities
6	sports/stationery shop	21	Chinese restaurant
7	living service	22	foreign restaurant
8	sports	23	fastfood restaurant
9	hospital	24	shopping mall
10	hotel	25	convenience store
11	scenic spot	26	electronic products store
12	residence	27	supermarket
13	governmental agencies and public organizations	28	furniture building materials market
14	science and education	29	pub/bar
15	transportation facilities	30	theaters

Discovers Regions of different Functions in a city by using mobile phone data

- 中国联通每天新增1500亿条上网行为记录
- 每条上网行为记录由26个字段组成

China Unicom operators
150 billion generated every day

Real-Time Urban Monitoring Using Cell Phones

13003759599	8551	55401	128390081 39491	304	2013/11/12 7:25:26	2013/11/12 7:30:36	311	1277	104	1381	0	10.76.254.10	140.206.160.215	0		UNINET	460015358901400	116
13003763810	55082	27612	865521012428430	304	2013/11/12 7:25:29	2013/11/12 7:25:29	1	375	0	375	1	10.51.233.6	140.207.54.36	0		3gnet	460017896601560	220
13003763810	55082	27612	865521012428430	304	2013/11/12 7:25:29	2013/11/12 7:29:30	240	112	316	428	1	10.51.233.6	140.207.54.36	0		UNINET	460015850933188	220
13003765732	55082	34171	866166010129677	304	2013/11/12 7:25:26	2013/11/12 7:25:26	1	56	64	120	1	10.60.181.115	140.206.160.215	0		3gnet	460017575470080	220
13004621608	17941	61666	860312024090890	304	2013/11/12 7:25:28	2013/11/12 7:27:38	130	862	1216	2078	2	10.60.168.125	140.206.160.215	0		3gnet	460011909263237	220
13004621608	17941	61666	860312024090890	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	68	0	68	2	10.60.168.125	140.206.160.215	0		3gnet	460012212758054	116
13004621608	17941	61666	860312024090890	304	2013/11/12 7:25:28	2013/11/12 7:25:28	;	68	0	68	2	10.60.168.125	140.206.160.215	0		3GNET	460017635914516	220
13004621608	17941	61666	860312024090890	304	2013/11/12 7:25:29	2013/11/12 7:25:29	1	68	0	68	2	10.60.168.125	140.206.160.215	0		3GNET	460010905271381	116
13004719068	55116	3336	356551040012500	304	2013/11/12 7:25:27	2013/11/12 7:29:29	241	52	52	104	1	10.73.141.131	140.206.160.215	0		3gnet	460018925400428	220
13004719068	55116	3336	356551040012500	304	2013/11/12 7:25:29	2013/11/12 7:25:29	1	52	0	52	1	10.73.141.131	140.206.160.215	0		3gnet	460016925039910	220
13004777799	55144	14323	354965057423220	304	2013/11/12 7:25:27	2013/11/12 7:25:27	1	52	128	180	1	10.73.217.102	140.207.54.36	0		3gnet	460015626061273	220
13008929535	55076	17001	353723055111833	304	2013/11/12 7:25:26	2013/11/12 7:25:26	1	56	0	56	1	10.47.158.250	140.206.160.215	0		3gnet	460015866080333	220
13008933263	55082	31101	860308025151310	304	2013/11/12 7:25:30	2013/11/12 7:27:17	108	1152	1408	2560	1	10.60.13.138	140.206.160.215	0		3gnet	460014633022036	220
13008943055	55100	26751	13468003734950	304	2013/11/12 7:25:30	2013/11/12 7:25:30	1	0	1204	1204	1	10.85.155.64	140.207.54.36	0		3gnet	460018275304019	220
13008944728	42257	41688	13354006821460	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	56	0	56	1	10.84.40.190	140.207.54.36	0		3gnet	460017405614367	220
13008955205	55082	41893	13417001487430	304	2013/11/12 7:25:26	2013/11/12 7:31:26	360	1418	1060	2478	1	10.76.237.87	140.207.54.36	0		3gnet	460016170904269	116

13056704641	16000	56140	13189002898501	304	2013/11/12 7:25:53	2013/11/12 7:30:12	259	1000	52	1052	2	10.85.92.83	140.207.54.36	0		3gnet	460016955004060	220
13002601383	16930	12662	12966005976782	304	2013/11/12 7:25:28	2013/11/12 7:28:03	154	1221	31167	32388	2	10.76.173.65	140.207.54.36	0		3gnet	460015870653290	116
13002601383	16930	12662	12966005976782	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	0	52	52	2	10.76.173.65	140.207.54.36	0		3gnet	460015741979456	220
13002601383	16930	12662	12966005976782	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	744	0	744	2	10.76.173.65	140.207.54.36	0		3gnet	460016630606833	116
13002601383	16930	12662	12966005976782	304	2013/11/12 7:25:29	2013/11/12 7:28:02	153	1136	0	1136	2	10.76.173.65	140.207.54.36	0		uniwap	460015958963313	220
13002601383	16930	12662	12966005976782	304	2013/11/12 7:25:30	2013/11/12 7:25:30	1	454	0	454	2	10.76.173.65	140.207.54.36	0		3gnet	460012212758530	116
13002627555	16011	4393	355637054708130	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	56	0	56	2	10.99.206.187	140.207.54.36	0		3GNET	460016071201697	116
13002631930	55220	22443	13183003933891	304	2013/11/12 7:25:30	2013/11/12 7:27:08	98	586	64	650	1	10.98.130.232	140.207.54.36	0		3gnet	460016620650430	116
13002631930	55220	22443	13183003933891	304	2013/11/12 7:25:30	2013/11/12 7:25:30	1	0	52	52	1	10.98.130.232	140.207.54.36	0		uniwap	460015958963313	220
13002631930	55220	22443	13183003933891	304	2013/11/12 7:25:30	2013/11/12 7:25:30	1	56	0	56	1	10.98.130.232	140.207.54.36	0		UNINET	460015756083879	220
13002662838	55136	32392	862966021080617	304	2013/11/12 7:25:29	2013/11/12 7:25:29	1	52	0	52	1	10.98.51.199	140.206.160.215	0		3gnet	460015850695993	220
13002665566	55142	38003	353638057787620	304	2013/11/12 7:25:26	2013/11/12 7:25:26	1	52	0	52	1	10.99.51.242	101.226.76.145	0		3gnet	460013740065529	220
13002682781	18210	31558	352343051868640	304	2013/11/12 7:25:28	2013/11/12 7:25:28	1	0	64	64	2	10.98.231.129	101.227.131.106	0		3gnet	460015800685630	220
13003700875	55080	36601	12958007001551	304	2013/11/12 7:25:27	2013/11/12 7:28:27	180	636	0	636	1	10.85.182.210	140.206.160.215	0		3gnet	460018915400870	220
13003701870	55084	28544	357376058551900	304	2013/11/12 7:25:26	2013/11/12 7:29:59	273	112	0	112	1	10.85.205.28	140.207.54.36	0		3gnet	460017975053326	116

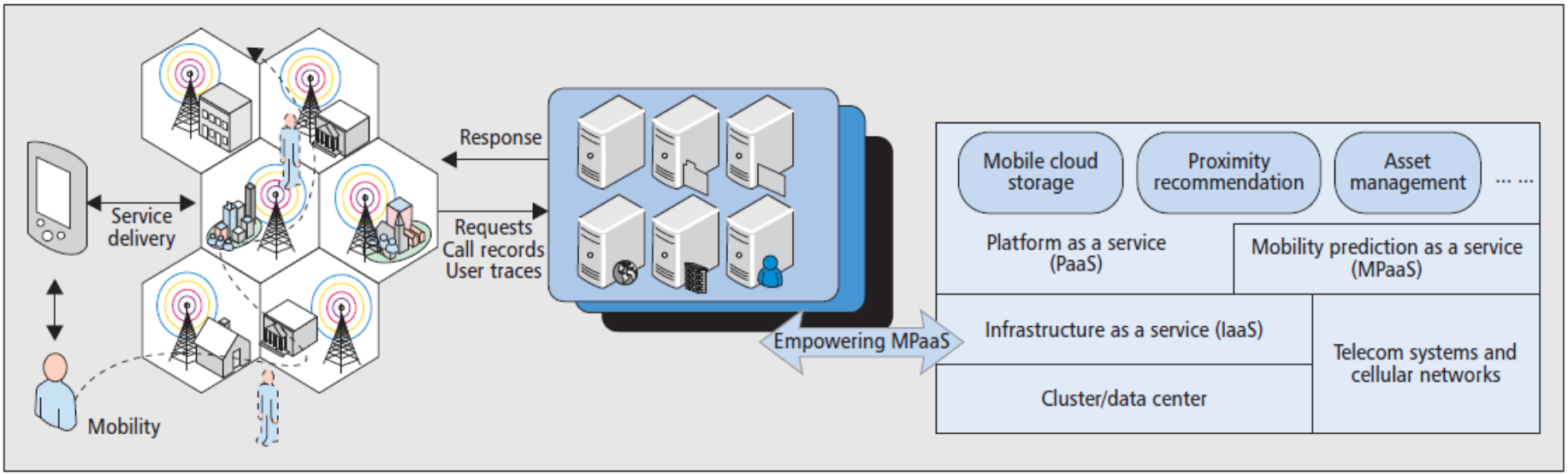
Each record has 26 fields.

Query of Internet access record

18658758639|8360|10443|868439014912357|304|2013-11-12 03:30:29.0110500|2013-11-12 03:30:29.0110500|1|246|0|246|0|10.11.85.36|140.206.160.215|0||3gnet|
460018755603960|220.206.151.134|116.79.206.4||34248|8080|0|0|1|0
15657770623|55050|18421|358967045670080|304|2013-11-12 03:30:29.0637750|2013-11-12 03:30:29.0637750|1|0|52|52|1|10.72.4.199|140.206.160.215|0||3gnet|460016640343299|
116.79.206.81|116.79.206.8||42124|8080|0|159|0|1
18667375068|16917|14061|351661058166274|304|2013-11-12 03:30:29.0702330|2013-11-12 03:30:29.0702330|1|0|64|64|2|10.13.4.189|112.64.237.188|0||3GNET|460017376601177|
116.79.206.20|116.79.206.4||52927|8080|0|159|0|1
13136330113|16027|4145|866323012043937|304|2013-11-12 03:30:29.1246130|2013-11-12 03:30:29.1246130|1|56|0|56|2|10.11.239.216|140.206.160.215|0||3gnet|
460015550947907|220.206.180.93|116.79.206.2||52533|8080|0|0|1|0
13018805053|55216|10483|864048019181607|304|2013-11-12 03:30:29.1377260|2013-11-12 03:30:29.1377260|1|0|52|52|1|10.81.98.244|140.207.54.36|0||3GNET|460012630009660|
220.206.180.105|116.79.206.6||34511|8080|0|212|0|1
18658721389|41212|22821|866240011336460|304|2013-11-12 03:30:29.1494870|2013-11-12 03:30:29.3592160|1|306|0|306|1|10.14.255.193|140.206.160.215|0||3gnet|
460018705601822|220.206.165.37|116.79.206.3||52888|8080|0|0|2|0
18605754887|55218|11511|013471004496100|304|2013-11-12 03:30:29.1528900|2013-11-12 03:30:29.2828910|1|321|0|321|1|10.11.148.6|140.207.54.36|0||3GNET|460015850648130|
220.206.180.111|116.79.206.4||57364|8080|0|0|2|0
13095659618|17934|62526|012839005919781|304|2013-11-12 03:30:29.1550490|2013-11-12 03:30:29.1550490|1|320|0|320|2|10.12.81.141|140.207.54.36|0||3GNET|
460015679032729|220.206.180.103|116.79.206.2||61642|8080|0|138|1|0
13067565029|16930|21475|861113028918867|304|2013-11-12 03:30:29.2034840|2013-11-12 03:30:29.2034840|1|56|0|56|2|10.14.11.158|140.206.160.215|0||3GWAP|
460017566001366|116.79.206.25|116.79.206.35||40368|8080|0|0|1|0
13065679093|55090|31051|354963051394820|304|2013-11-12 03:30:29.2225930|2013-11-12 03:30:29.2225930|1|304|0|304|1|10.72.118.148|140.206.160.215|0||3gnet|
460015676002276|220.206.180.116|116.79.206.35||39473|8080|0|138|1|0
13065805189|16008|3783|867203010564830|304|2013-11-12 03:30:29.2332890|2013-11-12 03:30:29.2332890|1|56|0|56|2|10.3.162.106|140.206.160.215|0||uninet|
460015656018239|220.206.180.96|116.79.206.35||37847|8080|0|0|1|0
13065679093|55090|31051|354963051394820|304|2013-11-12 03:30:29.2453560|2013-11-12 03:30:29.3209720|1|56|316|372|1|10.72.118.148|140.206.160.215|0||3gnet|
460015676002276|220.206.180.116|116.79.206.35||39473|8080|0|0|1|1
18668313272|55200|50813|861235013005597|304|2013-11-12 03:30:29.3096630|2013-11-12 03:30:29.3096630|1|56|0|56|1|10.72.0.144|140.206.160.215|0||3GNET|460018316604722|
116.79.206.20|116.79.206.4||47058|8080|0|0|1|0
13056876956|55076|17001|356788041512610|304|2013-11-12 03:30:29.3108870|2013-11-12 03:30:29.3108870|1|56|0|56|1|10.14.147.55|163.177.71.160|0||3gnet|460016875001011|
220.206.180.116|116.79.206.3||53022|8080|0|0|1|0
13221381628|16955|19491|352000044325760|304|2013-11-12 03:30:29.3924230|2013-11-12 03:30:29.3924230|1|56|0|56|2|10.11.110.17|140.207.54.36|0||3GNET|460017696007045|
116.79.206.20|116.79.206.2||51481|8080|0|0|1|0
13095977360|37380|40991|351979047837170|304|2013-11-12 03:30:29.4102780|2013-11-12 03:30:29.4102780|1|52|0|52|2|10.72.20.145|140.206.160.215|0||3gnet|
460015979001146|220.206.136.137|116.79.206.2||55155|8080|0|207|1|0
13164923766|18205|30294|359341033288142|304|2013-11-12 03:30:29.4251260|2013-11-12 03:32:28.5112640|119|280|0|280|2|10.97.81.102|140.206.160.215|0||3gnet|
460014926901081|116.79.206.28|116.79.206.36||59237|8080|0|0|2|0
13221380749|16917|21714|860814024858937|304|2013-11-12 03:30:29.4278260|2013-11-12 03:30:29.4278260|1|56|0|56|2|10.81.135.167|140.206.160.215|0||3GNET|
460011342229250|116.79.206.22|116.79.206.6||51314|8080|0|0|1|0
18667891694|54563|16672|351979046003591|304|2013-11-12 03:30:29.4422740|2013-11-12 03:30:29.4422740|1|56|0|56|1|10.13.243.7|112.64.237.188|0||3gnet|460017896602678|
220.206.143.33|116.79.206.8||55167|8080|0|0|1|0
13221395086|55206|58831|351526045573692|304|2013-11-12 03:30:29.4892470|2013-11-12 03:30:29.4892470|1|56|0|56|1|10.83.26.183|140.206.160.215|0||3GNET|
460011392200285|116.79.206.20|116.79.206.6||45772|8080|0|0|1|0

Query of Internet access record

- 1. build the relevant test database
- 2. tables and various business inquiries of the database
- 3. case test/experiment test
- 4. visualization of data ---R, Matlab, Python and other graphics tools



Telecom cloud architecture empowered by cloud-based mobility prediction.

At first, we analysis data structure

- 中国联通每天新增1500亿条上网行为记录
- 每条上网行为记录由26个字段组成

150 billion generated every day

在 Hive 准备使用的表结构，请大家提建议。

上网行为记录组成 —— 中国联通

序号	字段名称	英文字段名称	数据类型	必填	说明
1	手机号码	PhoneId	STRING	是	手机上网的唯一标识码（主键）
2	位置区域编码	Lac	INT	是	用于标识不同的位置区
3	小区编码	CellId	INT	是	识别小区（基站号）
4	终端编码	IMEI	STRING	是	手机的唯一识别号码
5	流量类型	NType	INT	是	2G、3G、4G、5G 网络
6	起始时间	STime	STRING	是	
7	结束时间	ETime	STRING	是	
8	上网时长	MS	DOUBLE	是	
9	上行流量	Ub	DOUBLE	是	Upbit: 本机向 inter 网发送的字节数
10	下行流量	Db	DOUBLE	是	Downbit:从网络中下载的字节数
11	总流量	Sb	DOUBLE	是	Sumbit: 网络总量
12	网络类型	NetType	BOOLEAN	是	NetMA: 手机网络模式的含义包括：双模 1---GSM 模式和 2---CDMA 模式
13	终端 IP	TerIP	STRING	是	TerminalIP
14	访问 IP	AccIP	STRING	是	Access the IP
15	状态码	Sc	INT	是	status code(一般标识上网的状态的标识如用 0 代替)
16	终端 UserAgent	UA	STRING	是	数据中全为NULL
17	APN	Apn	STRING	是	[redacted]
18	IMSI	Imsi	STRING	是	国际移动用户识别码
19	SGSN IP	SgsnIp	STRING	是	
20	GGSN IP	GgsnIp	STRING	是	
21	Content Type	CType	STRING	是	数据中全为NULL
22	源端口	Sport	STRING	是	Source Port
23	目标端口	Dport	STRING	是	Destination port: 8080 端口同 80 端口，是用于 WWW 代理服务的
24	记录标识	Rid	STRING	是	Record ID
25	合并记录数	Nrecords	DOUBLE	是	With the number of records
26	URL （网址）	Url1	STRING	是	网址
27	??	Url2	STRING	是	数据中 多出了字段？？

Communication base transceiver stations (BTSs) hone-number

```
hive> select * from yuqing_9 limit 100;
OK
55114   40296   1083
55116   5424    879
55100   53081   853
55120   18419   839
55084   28561   801
55110   3192    799
55074   13112   797
55084   28281   786
55222   21621   755
55090   41883   710
55106   56983   710
55080   22163   656
55100   26822   648
55216   13423   648
55084   35273   647
55090   31051   646
55216   12501   645
55082   25442   638
55120   6084    636
55075   51363   633
55080   22171   627
55080   22282   609
13841   6617    608
55206   58341   606
55084   28562   604
55148   16462   603
55110   16591   598
55120   18324   594
55075   37361   590
55040   153     590
55054   25381   586
55148   16292   585
55074   13113   585
55148   16561   576
55206   58313   573
55082   25692   570
```

```
hive> select * from yuqing_10 limit
OK
55114   40296   1677
55074   13442   1098
55120   18419   1031
55074   13112   1010
55222   21621   882
55054   25381   881
55090   41883   809
55148   16462   793
55138   12691   770
55080   22282   768
55106   56983   756
55074   49093   740
55156   15773   732
55100   53081   724
55074   13563   718
55084   13852   710
17000   60151   709
55120   5774    692
55092   40233   691
55106   10671   685
55075   37361   683
55084   28281   678
55084   28503   669
55110   3192    666
55176   3052    663
```

```
hive> select * from yuqing_11 limit 100;
OK
58657   3810    1845
55114   40296   1801
55120   18419   1374
55074   13442   1245
55072   10173   1172
55074   13112   1111
55100   26822   1107
55076   16421   1066
55100   53081   1012
55074   49093   1011
55090   41883   1003
55120   18664   986
55106   10671   975
55074   13732   936
55084   13852   926
55084   28561   915
55116   5424    908
55074   13272   907
55110   48602   904
55148   16561   886
16014   4023    879
55106   48509   876
55078   19212   862
17000   60151   845
55156   15781   838
55106   56983   829
55084   35273   828
55082   25692   824
55156   15881   822
16047   1535    822
55074   13563   820
55206   58312   813
55222   21621   811
55110   58102   810
55054   25381   795
55120   10764   784
55078   27051   784
55084   28281   783
55078   19211   781
55154   31543   778
55220   20081   777
55206   58852   766
55074   13733   755
```

```
hive>  select * from yuqing_12 limit 100;
OK
55074   13442   1538
55114   40296   1401
55106   10671   1237
55076   16421   1140
55084   28561   1132
55074   13701   1101
55072   10173   1087
55074   13272   1085
55120   2344    1051
55106   48509   1012
55078   19211   972
55156   15881   963
55120   18419   946
55206   58313   933
55110   48602   932
55074   13732   931
55074   13112   928
55082   25692   927
55200   51011   927
55074   49093   915
55078   19212   913
55074   13563   907
55120   10764   906
16047   1535    905
55222   21621   895
55120   18664   893
55110   3192    874
55206   58191   870
55074   13733   866
55042   6251    866
55054   25381   852
55138   12693   847
55200   50513   846
55100   26822   843
55090   41883   840
55156   17492   828
55136   10952   816
55086   31393   810
55074   13113   806
55074   13661   804
55084   35343   799
55084   35273   795
55075   37361   785
```

```
select * from yuqing_13 lim
OK
55054   25381   1446
55074   13442   1384
55084   28561   1300
55074   13701   1179
55074   13112   1166
55222   21621   1141
55066   41613   1071
55210   57381   1064
55074   13301   1059
55114   40296   1057
55042   5732    1002
55076   16426   964
55064   48642   958
55206   58191   941
55106   10671   926
55050   18561   922
55074   13272   916
55054   25423   907
55120   18419   900
55078   19211   897
55072   10232   890
55050   16661   877
55106   48509   877
55074   13732   870
55076   16424   868
55072   37155   865
55066   44113   852
55072   10173   850
55074   13563   847
55078   19291   832
13844   57139   828
55092   40233   821
55074   13271   816
16047   1556    813
55092   40232   812
55076   16421   804
55200   51011   801
55082   25692   801
55066   44162   801
55084   28503   798
55200   50513   797
55072   37153   788
```

```
hive> select * from yuqing_
OK
55074   13442   1919
55206   58191   1398
55074   13112   1386
55084   28561   1356
55076   16424   1349
55074   13661   1325
55074   13732   1323
55074   49093   1300
55210   57381   1263
55074   13701   1252
55076   16426   1243
55120   18419   1179
55106   10671   1152
16047   1535    1150
55114   40296   1127
55072   37155   1110
55074   13272   1107
55078   19211   1105
55200   50513   1097
55106   48509   1094
55106   56983   1088
55103   16344   1088
55074   13261   1087
55110   48602   1033
55078   19291   1020
55082   25692   1014
55200   51011   981
55084   28281   979
55076   16421   974
55074   13563   973
55072   37153   973
55100   26822   964
55110   3192    937
55072   10232   930
55216   13451   928
55231   54573   927
55092   40233   921
55074   13733   919
55074   13271   913
16047   1556    898
55072   11265   873
55156   17392   859
```

```
hive>  select * from yuqing_15 limit 100;
OK
55074   13442   2145
55084   28561   1698
55074   13112   1657
55074   13701   1445
55082   25442   1390
55074   49093   1364
55120   18419   1362
55072   37155   1295
55114   40296   1278
55074   13732   1264
55206   58191   1255
55074   13272   1252
55076   16421   1197
55110   3192    1183
55110   48602   1178
55076   16424   1160
55074   13271   1156
55106   10671   1149
55078   19211   1140
55210   57381   1128
55072   37153   1127
55090   41883   1123
55084   28281   1089
55076   16426   1087
55072   10232   1078
55222   21621   1074
55106   48509   1069
55084   46133   1069
55120   2344    1047
55074   13113   1045
55074   13932   1044
55200   50513   1033
55072   10173   1022
55120   18664   1003
55078   19291   997
55106   56983   995
16047   1556    988
55084   29523   975
55074   13733   973
55084   13852   968
55200   51011   952
55072   49811   932
```

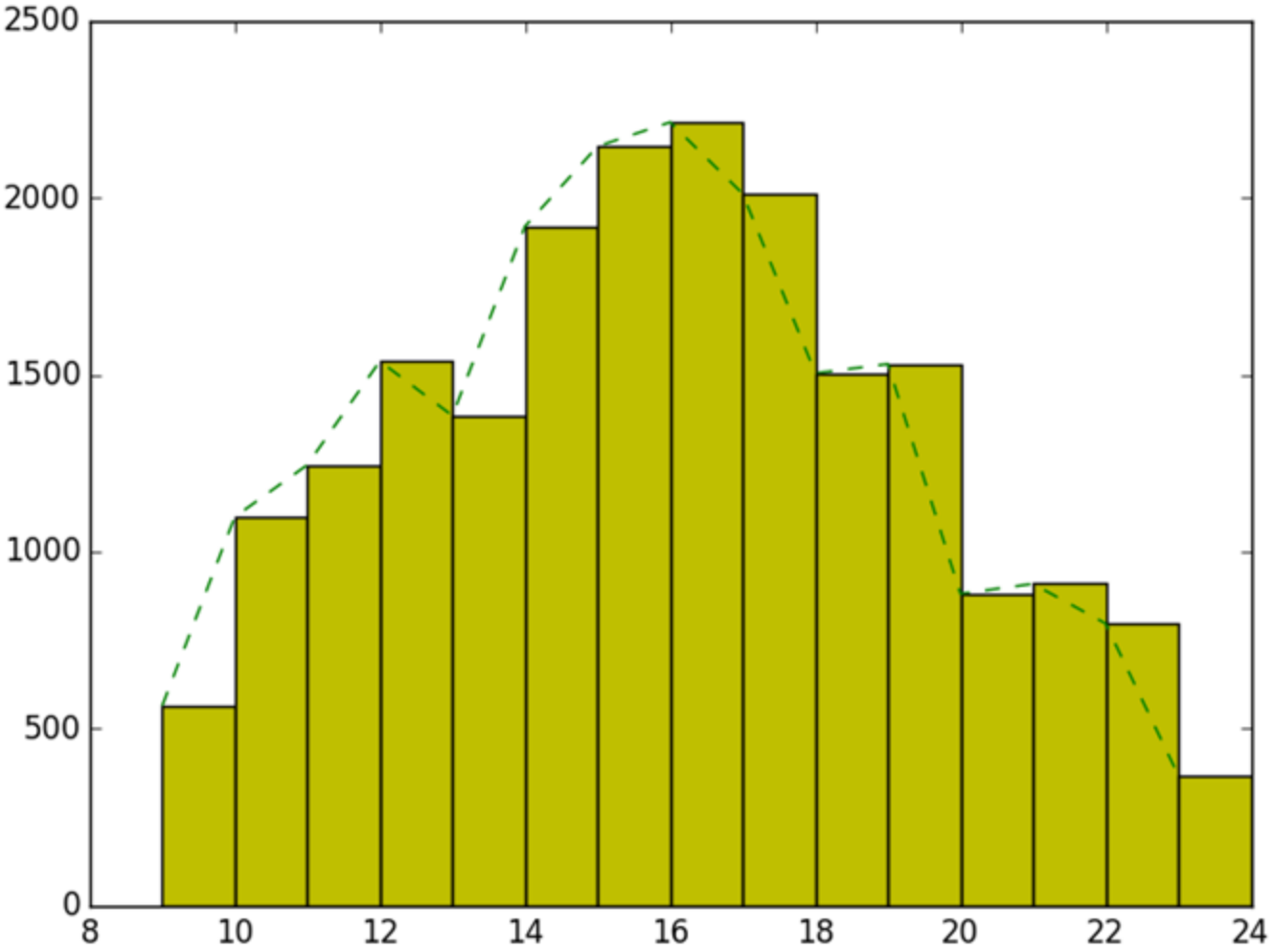
```
ive> select * from yuqing_16
K
5074    13442   2213
5084    28561   1747
5074    13112   1619
5072    10173   1341
5074    13272   1315
5078    19211   1287
5074    13732   1252
5074    13701   1236
5074    49093   1220
5074    13733   1168
5072    37155   1151
6047    1535    1125
5110    48602   1120
5090    41883   1114
5120    18419   1104
5072    49811   1081
5222    21621   1080
5084    28562   1044
5076    16421   1019
5114    40296   1013
5106    48509   1011
5206    58191   1006
5216    13451   1000
5200    51011   993
5074    13932   975
6047    1556    970
5078    32092   962
5084    29523   920
5106    56983   908
5156    17492   897
5086    31393   884
5210    57381   879
5120    10764   869
5220    20082   865
5200    50513   856
5080    22282   854
5080    22383   839
5074    13261   835
5082    25692   835
5078    19292   828
5114    40322   827
5138    12712   827
```

```
hive> select * from yuqing_17
OK
55074   13442   2011
55084   28561   1431
55074   13112   1431
55074   13272   1353
55074   13701   1262
55066   44113   1214
55222   21621   1179
55074   13732   1178
55054   25381   1169
55072   49811   1162
55120   18419   1136
55074   49093   1127
55072   10173   1116
55200   50513   1110
55206   58191   1108
55106   56983   1104
55084   28562   1088
55074   13733   1066
55084   35273   1037
55106   10671   995
55106   48509   979
55114   40296   976
55084   13852   976
55043   19572   968
55216   13451   961
55090   41883   954
55120   10764   949
55090   37533   948
55084   28281   912
55056   38531   912
55138   12712   906
55110   3192    889
55076   16421   886
55214   61037   885
55048   13082   881
55080   22282   878
55054   25423   870
55074   13932   866
55078   19211   862
55200   51011   845
16047   1535    843
55064   48642   834
```

```
hive>  select * from yuqing_18 limit 100;
OK
55074   13442   1504
55084   28561   1321
55074   13732   1225
55074   13112   1149
55072   10173   1118
55090   41883   1096
55220   20081   1046
55074   13272   1040
55082   25692   996
55072   10101   984
55084   29523   983
55110   3192    978
55084   28562   975
55074   13733   975
55074   49093   970
55106   56983   949
55084   28392   934
55120   18419   932
55222   21621   924
55076   16421   916
55226   35032   909
55084   13852   901
55200   50513   897
55084   28281   892
55074   13711   879
55090   37581   875
55080   22282   875
55080   22383   867
55080   23463   860
55078   32092   858
55206   58191   855
55050   16661   855
55206   58312   846
55156   17491   844
55066   44113   827
55200   50692   825
55074   13661   822
55074   13712   818
55072   10103   807
55075   15721   804
55120   10764   802
55072   10171   796
```

```
hive>  select * from yuqing_22 limit 100;
OK
55074   13032   1696
55084   28561   1561
55054   25381   1447
55084   35273   1243
55156   51511   1154
55075   40512   1117
55100   53081   1102
55086   31393   1048
55216   10083   1007
13844   7438    981
55216   13872   947
55220   20081   913
55064   38353   908
55216   10173   896
55094   12221   883
55150   19433   881
55200   50513   879
55080   22282   876
55120   18419   874
55084   28503   871
55090   41883   860
55222   21621   848
55054   25423   838
55138   51852   836
55103   16101   824
55075   40513   824
55216   13451   820
55050   16661   820
55090   37581   809
55156   17492   807
55138   12661   806
55072   10161   803
55074   13442   797
55104   62141   793
55280   12373   793
55202   41021   790
55074   53403   780
55072   49813   755
55075   42531   750
55076   16421   744
55043   19572   734
55040   693     734
55202   31251   726
```

```
 select * from yuqing_23 limit 100;
OK
55074   13032   1135
55216   10173   860
55086   31393   759
55216   13872   698
55100   53081   685
55086   31243   668
55084   28561   651
55220   20081   635
55084   28503   635
55216   13451   630
55072   10161   603
55202   41021   602
55084   35273   591
55090   31051   590
55090   41883   588
55202   52101   585
55072   10101   578
13844   7438    575
55208   53131   570
55106   56983   566
55156   15823   564
55208   35392   564
55050   18561   555
55100   26762   553
55094   12221   552
55106   19553   550
55094   33712   541
55114   19743   541
55058   32112   539
55216   50761   538
55156   17492   537
55072   10173   531
55202   31251   528
55058   31982   524
55076   51813   521
55136   11001   515
55138   12712   514
55280   10431   513
55216   13793   511
55090   37581   510
55054   25381   509
55150   19223   508
55138   12442   506
```

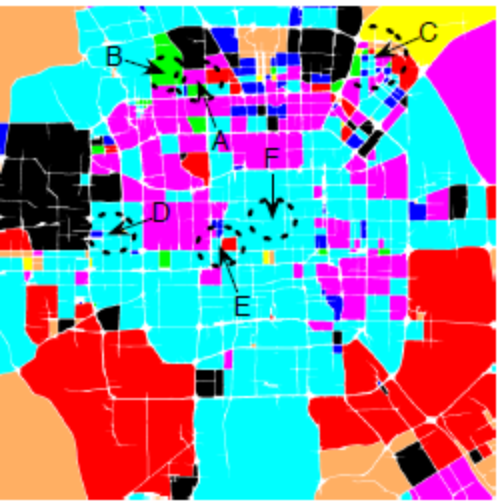


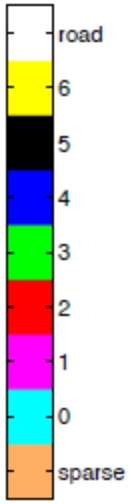
base transceiver stations (BTSs)

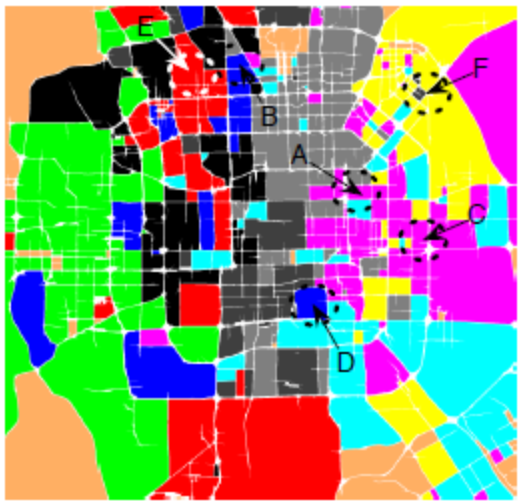
(a) functional region c_1

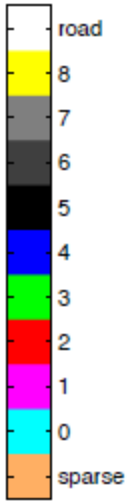
(b) functional region c_4

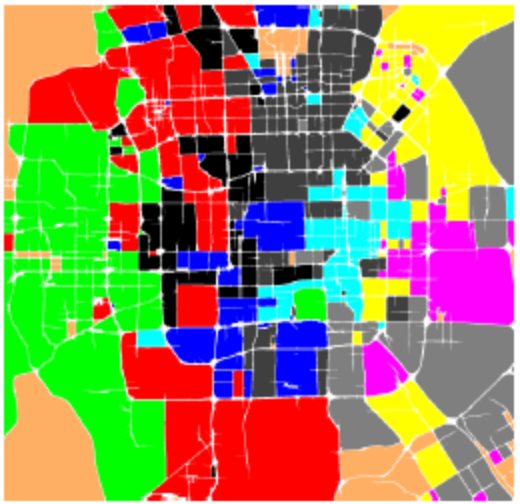
Functionality intensity of functional regions



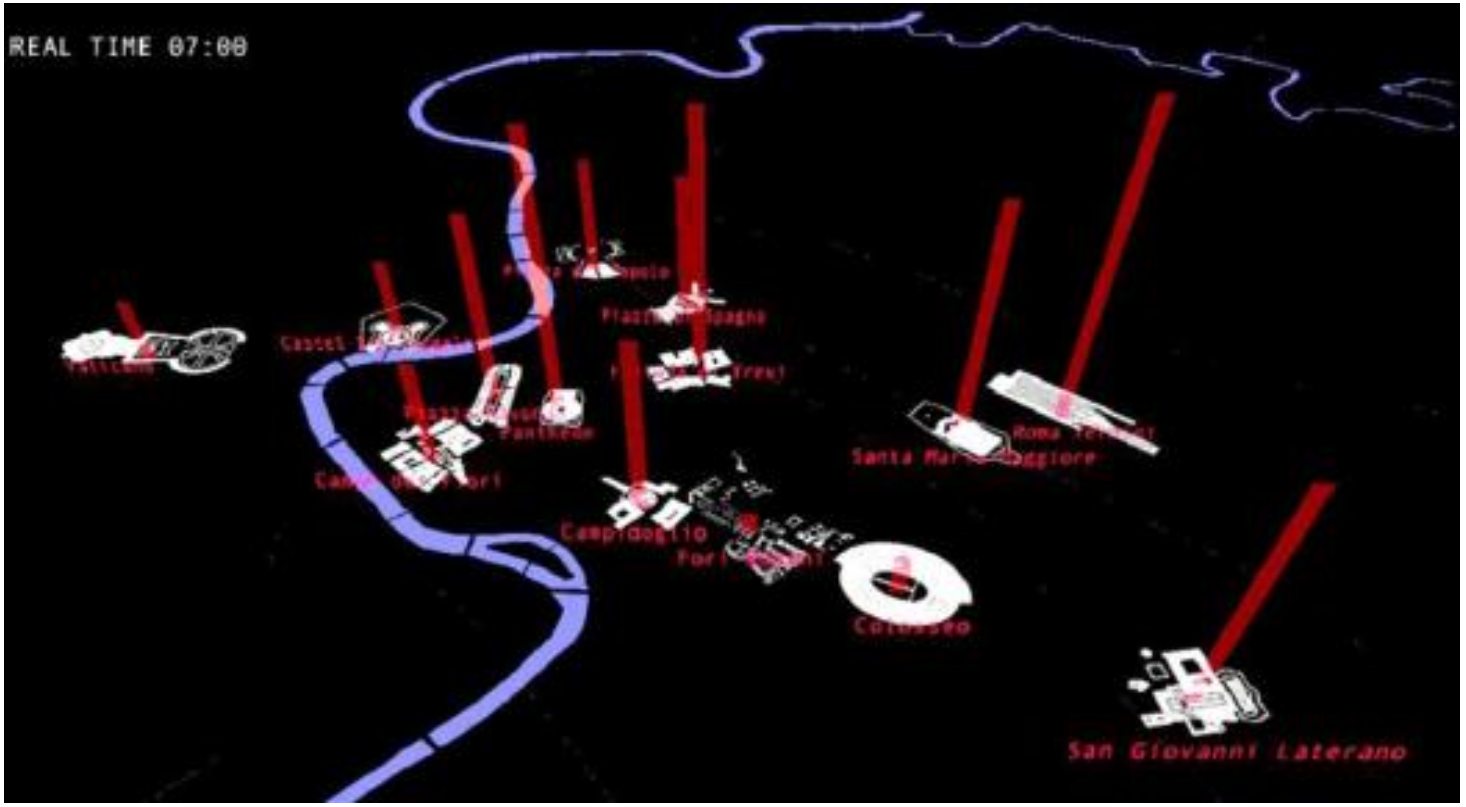








residential areas, business districts, and educational areas



Icons: Which landmarks in place attract more people

Visitors: Where are tourists congregating?

1. We are starting the Big Data Projects

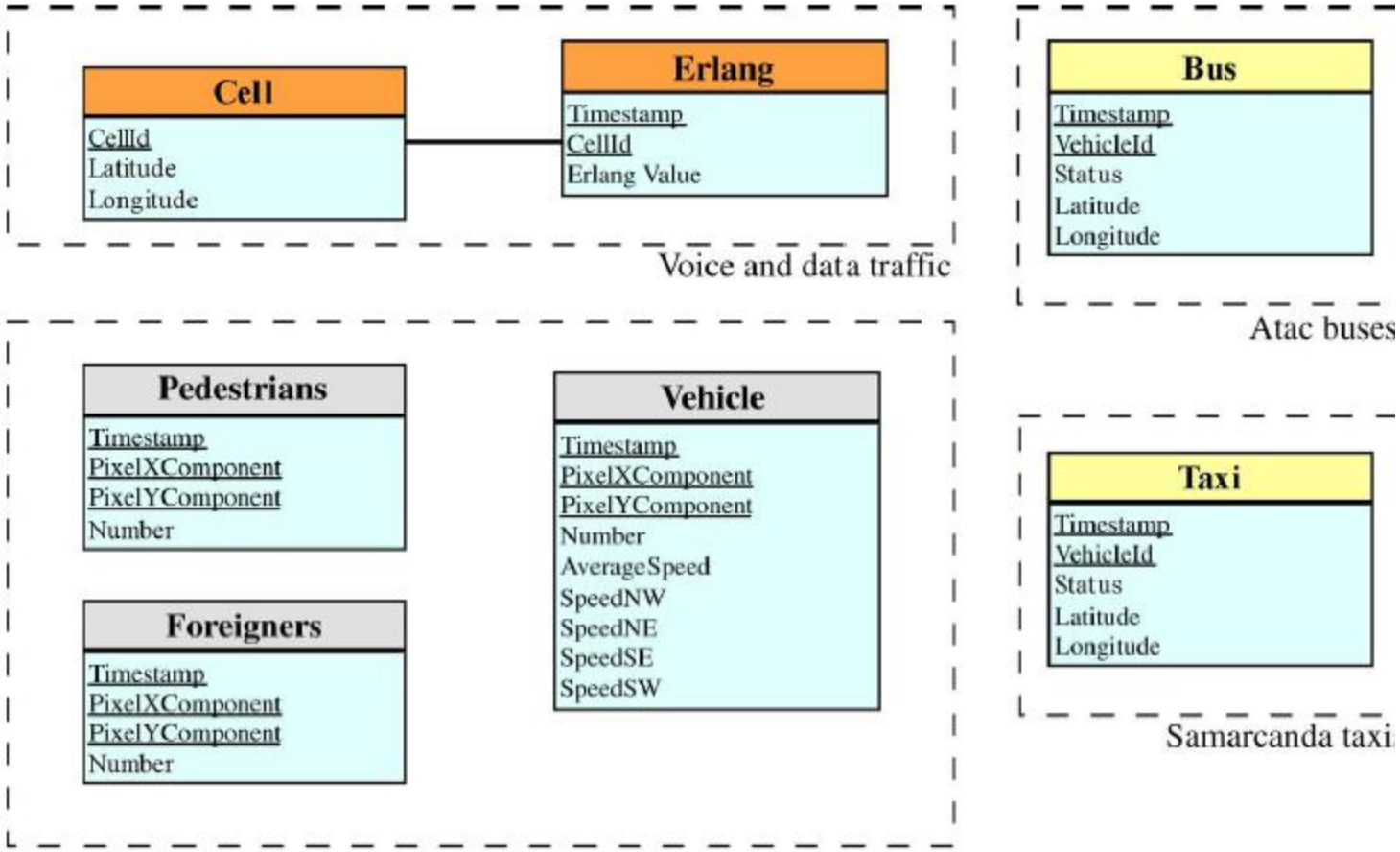
2. Examples

- Real-Time Urban Monitoring Using Cell Phones
- Mobile phone usage in complex urban systems: a space–time, aggregated human activity
- Unveiling patterns of international communities in a global city using mobile phone data
- NextCell Predicting Location Using Social Interplay from Cell Phone Traces

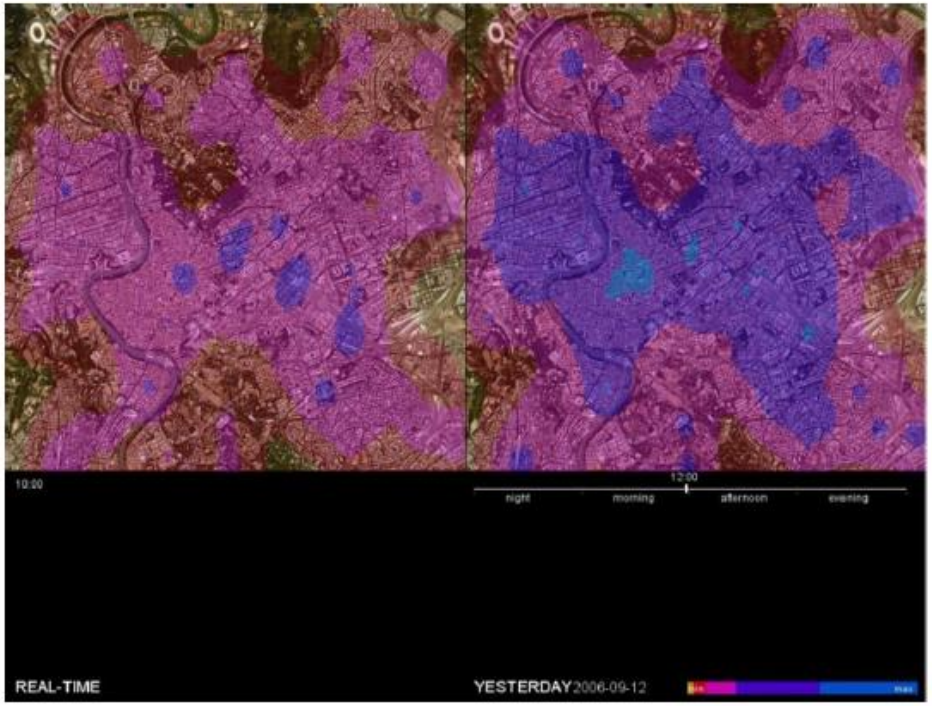
Real-Time Urban Monitoring Using Cell Phones

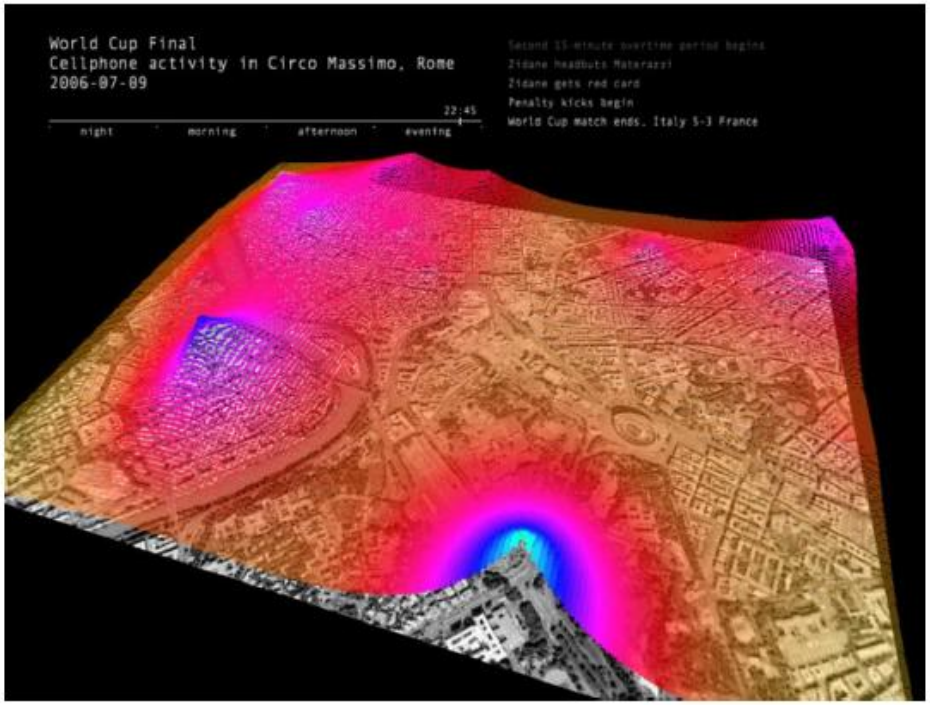
AN INTELLIGENT transportation system (ITS) is a transportation system that makes use of information and communication technology (ICT) to address and alleviate transportation and congestion problems. In general, an ITS relies on location-based information: It monitors and processes the location of a certain number of vehicles (used as probes) to obtain information on estimated travel time, driving conditions, and traffic incidents. Using a relatively large amount of probes, the early stages of bottlenecks can be detected, and traffic can be directed to other routes to mitigate congestion and provide more expedient and efficient itineraries to travelers. A variety of sensors can be used to obtain traffic information. These traditionally fall into two main categories: fixed sensors and Global Positioning System (GPS) receivers.

- Mobile phone positioning techniques generally provide less accuracy than the GPS, but the wide diffusion of mobile equipment (ME) in addition to the widespread installation of radio transmitters, or base transceiver stations (BTSs), in both urban and rural areas makes such positioning techniques very appealing.



Schematic of the database tables.



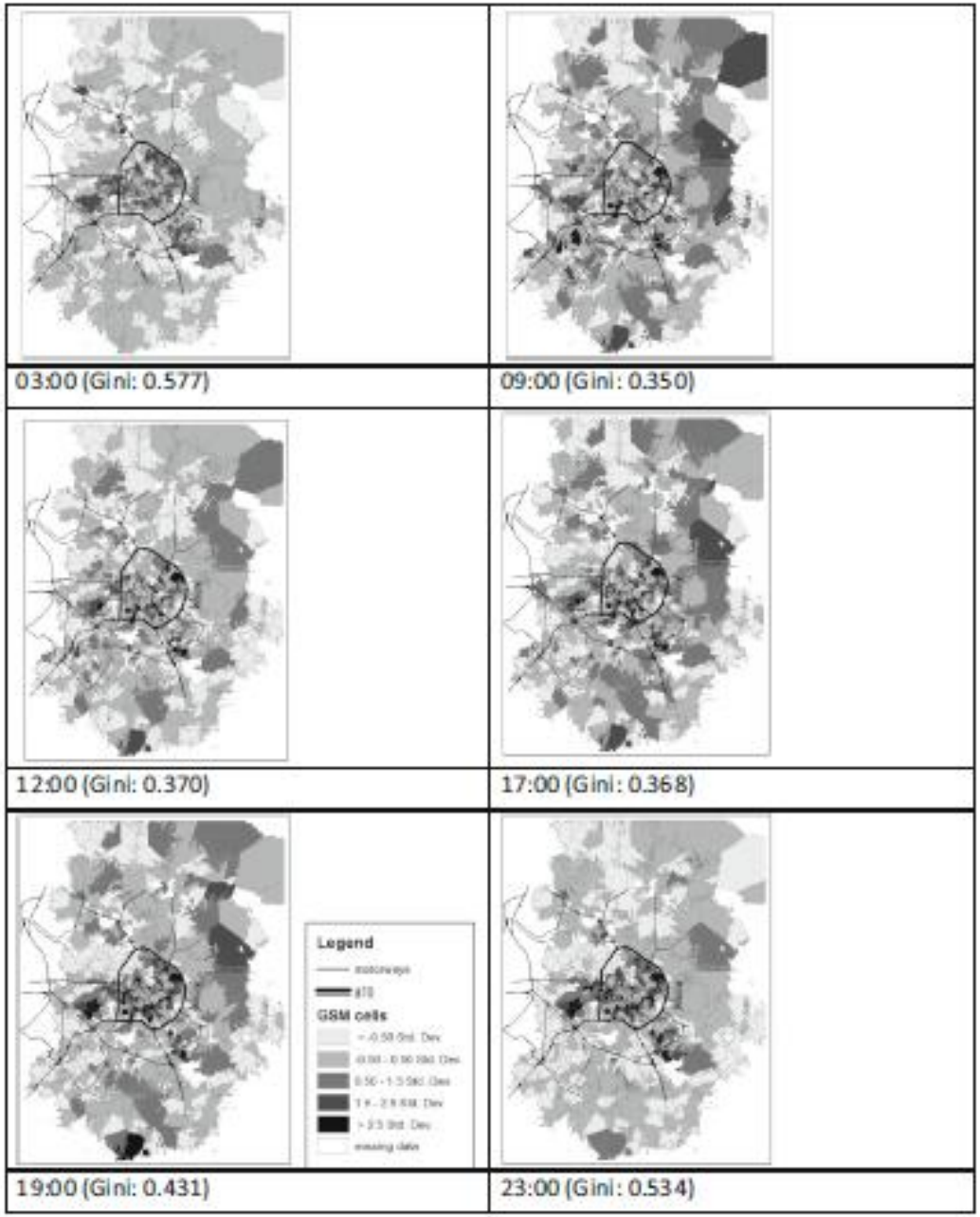


Pulse: What are the patterns of use in place

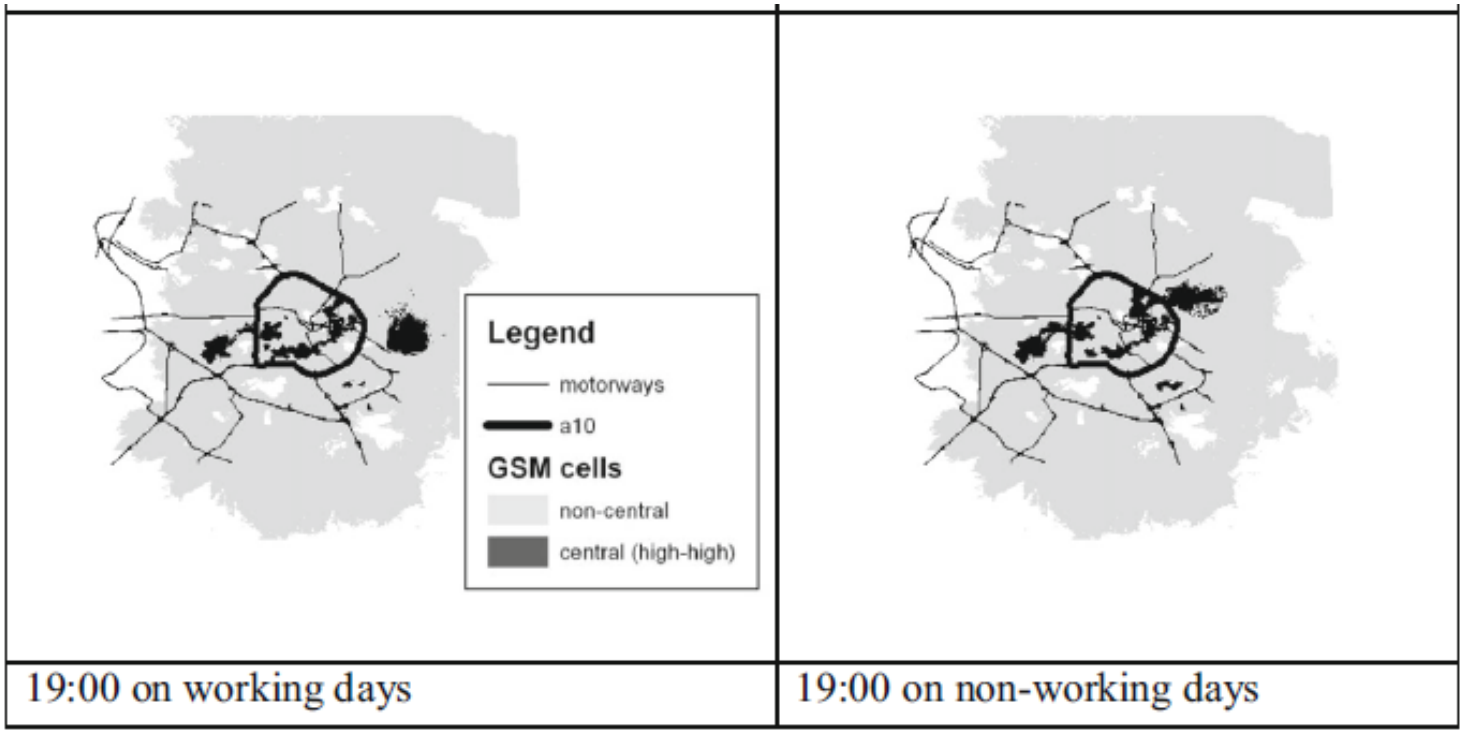
Gatherings: What does Rome look like during special events?

Mobile phone usage in complex urban systems: a space–time, aggregated human activity

- The present study aims to demonstrate the importance of digital data for investigating space–time dynamics of aggregated human activity in urban systems. Such dynamics can be monitored and modelled using data from mobile phone operators regarding mobile telephone usage.



Mobile phone usage during the day in some place. Note: Mobile phone usage measured as Erlangs per hour at the level of the cell.



19:00 on working days | 19:00 on non-working days

Central GSM zones—local indicators of spatial autocorrelation (LISA) high–clusters. Note: Mobile phone usage measured as Erlangs per hour at the level of the cell.

Unveiling patterns of international communities in a global city using mobile phone data

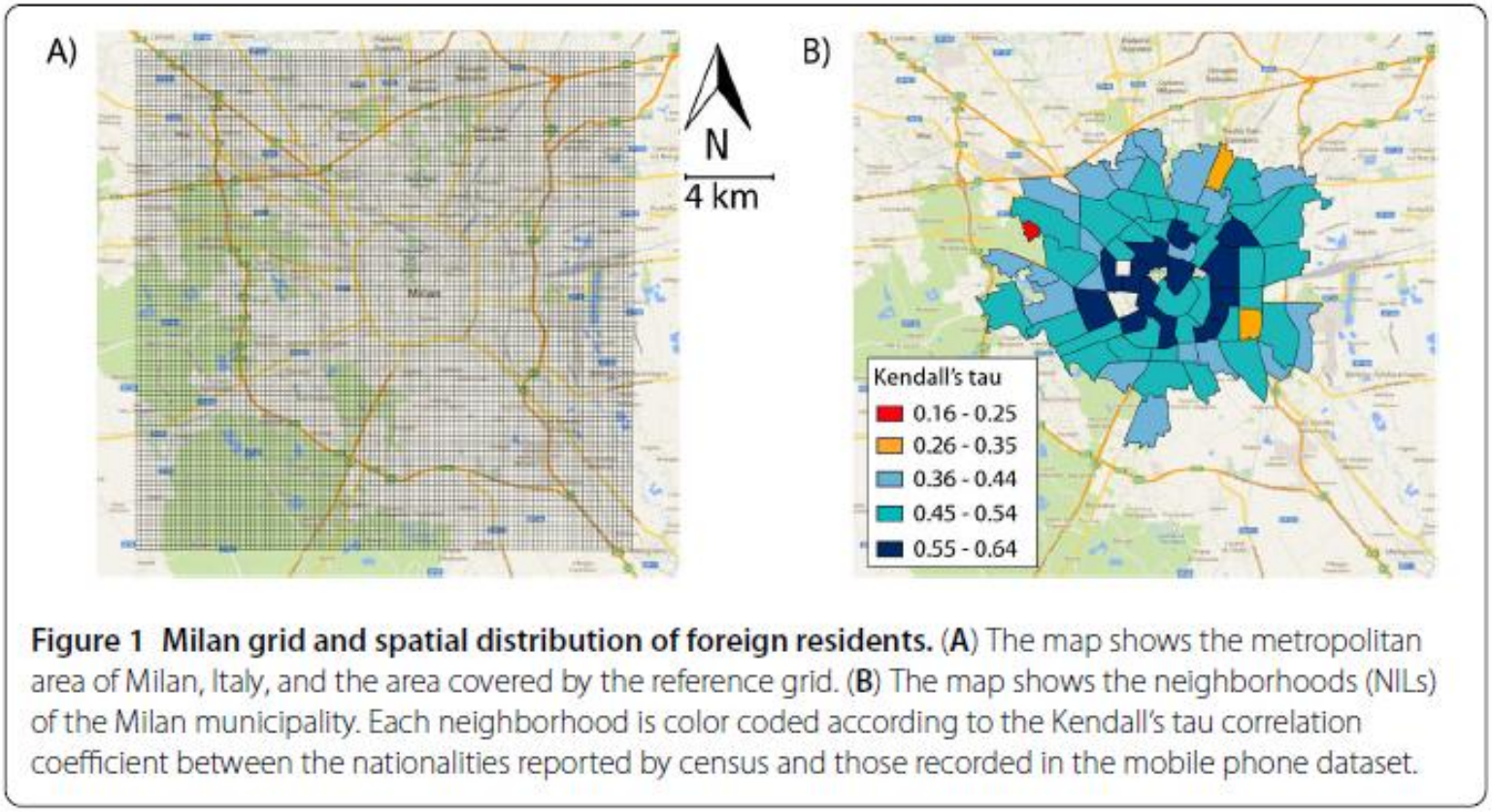
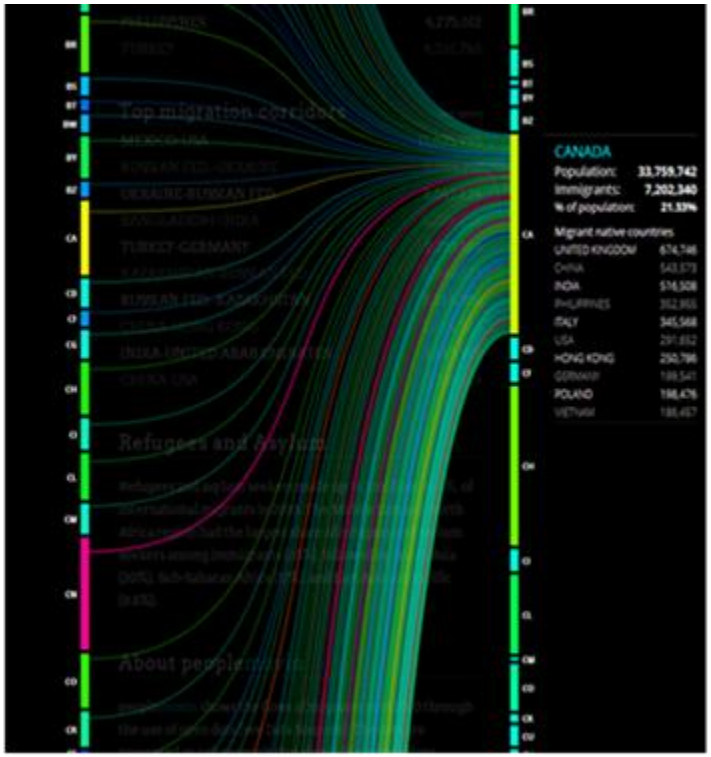
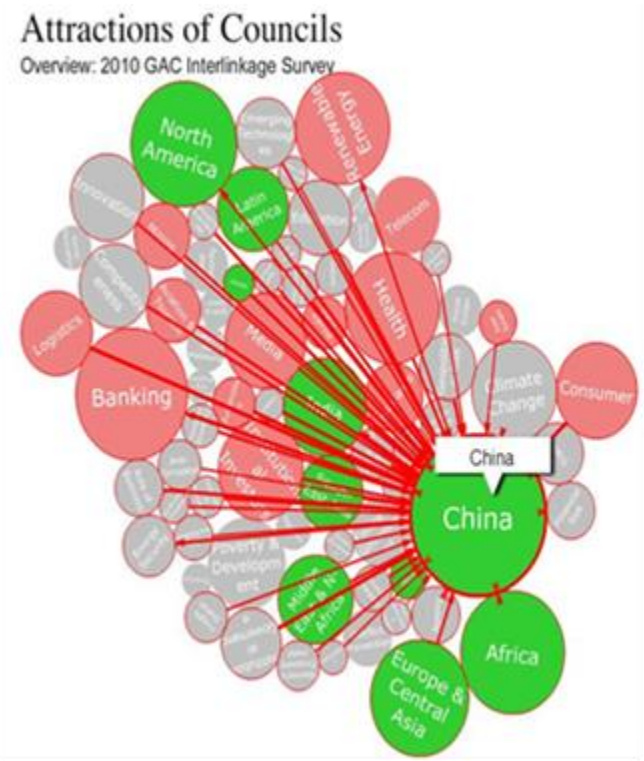


Figure 1 Milan grid and spatial distribution of foreign residents. (**A**) The map shows the metropolitan area of Milan, Italy, and the area covered by the reference grid. (**B**) The map shows the neighborhoods (NILs) of the Milan municipality. Each neighborhood is color coded according to the Kendall's tau correlation coefficient between the nationalities reported by census and those recorded in the mobile phone dataset.

Such as: visualization of data





贵州科学院
Guizhou Academy of Sciences

Web visualization of data

- 2014 Chinese "Guizhou on Cloud " big data business model contest starts in Guiyang.
- We participated in the contest.

http://contest.gzdata.com.cn/

贵州科学院
Guizhou Academy of Sciences

- 2014 Chinese "Guizhou on Cloud " big data business model contest starts in Guiyang.
- We participated in the contest.

https://www.gzdata.gov.cn/

We won the Outstanding Project

贵州科学院
Guizhou Academy of Sciences

2015 Guiyang International Big Data

数博会
BIG DATA EXPO
中国·贵阳 GUIYANG·CHINA

大数据，超乎我们的想象，超越我们的梦想！
Big data has gone beyond our imagination and our dreams.

http://www.bigdata-expo.org/cn/

贵州科学院
Guizhou Academy of Sciences

数博会
BIG DATA EXPO
中国·贵阳 GUIYANG·CHINA

大数据，超乎我们的想象，超越我们的梦想！
Big data has gone beyond our imagination and our dreams.

“互联网+”时代的数据安全与发展
Data Security and Development in the Era of “Internet Plus”

http://www.bigdata-expo.org/cn/

贵州科学院
Guizhou Academy of Sciences

2015贵阳大数据草根创新公开赛
BIG DATA
Competition
中国·贵阳 GUIYANG·CHINA

云上贵州·数聚贵阳
Guizhou above the Cloud·Data Gathering in Guiyang

http://guiyang.chinacloudapp.cn/

贵州科学院
Guizhou Academy of Sciences

食品安全营养信息化云平台

FSNIP（Food Safety and Nutrition Information Platform）

食品安全营养网

食安测

食品安全营养信息服务云平台

食品安全营养
知识信息系统

食品安全营养
测试信息系统

食品安全营养
监管信息系统

贵州科学院
Guizhou Academy of Sciences

Business Value

- Guiyang Global Big Data Exchange

Chinese President Xi Jinping visited the development of big data in guiyang, 2015.6

贵州科学院
Guizhou Academy of Sciences

Big Data References

Hadoop: The Definitive Guide by Tom White
SQL Server Sqoop http://bit.ly/rulsjX
JavaScript http://bit.ly/wdaTv6
Twitter https://twitter.com/#!/search/%23bigdata

Hive http://hive.apache.org
Excel to Hadoop via Hive ODBC http://tinyurl.com/7c4qjjj
Hadoop On Azure Videos http://tinyurl.com/6munnx2
Klout http://tinyurl.com/6qu9php

Microsoft Big Data http://microsoft.com/bigdata
Denny Lee http://dennyglee.com/category/bigdata/
Carl Nolan http://tinyurl.com/6wbfxy9
Cindy Gross http://tinyurl.com/SmallBitesBigData

Softwares

- http://hadoop.apache.org

- Oreilly: Hadoop. The definitiive Guide 3rd Edition. May 2012.pdf

- Oracle products and price

https://shop.oracle.com/pls/ostore/product?p1=OracleBigDataAppliance&p2=&p3=&p4=&sc=ocom_BigDataAppliance

- Microsoft

http://www.windowsazure.com/en-us/manage/services/hdinsight/

- IBM

http://www-01.ibm.com/software/data/bigdata/platform/product.html

- Cloudera

http://www.cloudera.com/content/cloudera/en/why-cloudera/hadoop-and-big-data.html

Q&A

Thank You !