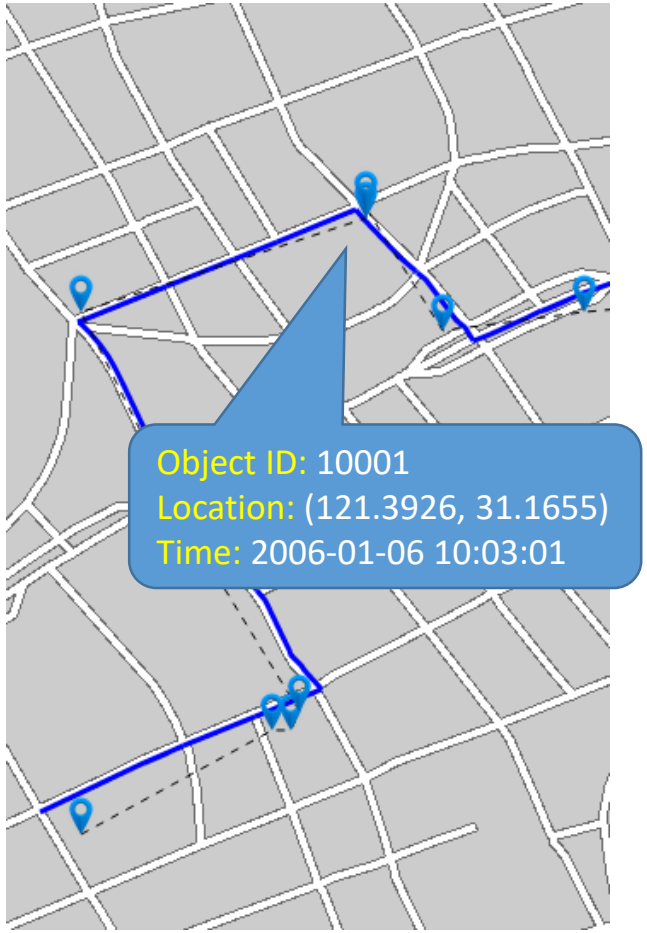
Analysis of Vast Amount Of Trajectories And Roads

Contents

INTRODUCTION
Trajectory

- The path that a moving object follows through space as a function of time
- Typical spatio-temporal data
- Collected from location-tracking devices, e.g., Global Positioning System (GPS)
- Formatted as sample points, or
- Formatted as road segments



INTRODUCTION

Trajectory Data of Taxis

	# of Taxis	Time Period	# of Records	Storage Size	Sampling Rate
Beijing	59,700	Jun – Aug, 2013	10 billion	853.7 GB	60 s
Shanghai	4,000	Jan 2006 – Nov 2007	1.8 billion	115.2 GB	60 s
Shenzhen	12,000	Sep 2013 – Jan 2014	3 billion	80 GB	15 s

Beijing Bus

- Jun 2013 - Aug 2013
- ~ 2.4 billion of records
- ~ 120 GB
- ~ 60 seconds sampling rate

Mobile

- Mar 2008 - Dec 2008
- ~ 170 GB

Miscellaneous

- Traffic videos
- Contravention snapshots
- Foursquare check-in data
- Yelp check-in data
- Sina Weibo / Twitter tweets
- Tmall user traces

INTRODUCTION

Trajectory Data of Taxis

Raw Data Structure

TID, TIME, LNG, LAT, SPEED, ANGLE, OCCUPY, N/A

```
B041C2,2009-09-01 00:00:23,114.10764,22.57017,  0,  0,  0,  0
B041C2,2009-09-01 00:00:43,114.10772,22.57015,  0,  0,  0,  0
B041C2,2009-09-01 00:01:03,114.10780,22.57013,  0,  0,  0,  0
B041C2,2009-09-01 00:01:43,114.10775,22.57013,  0,  0,  0,  0
B041C2,2009-09-01 00:02:03,114.10775,22.57007,  0,  0,  0,  0
B041C2,2009-09-01 00:02:23,114.10780,22.57003,  0,  0,  0,  0
B041C2,2009-09-01 00:02:43,114.10768,22.57013,  0,  0,  0,  0
B041C2,2009-09-01 00:03:03,114.10765,22.57013,  0,  0,  0,  0
B041C2,2009-09-01 00:03:23,114.10777,22.57015,  0,  0,  0,  0
B041C2,2009-09-01 00:03:43,114.10770,22.57013,  0,  0,  0,  0
```

ID, TID, BEGIN_DATE, END_DATE, END_PRICE, TASK_DISTANCE ,
RECORD_TIME, TASK_AMOUNT, FREE_DISTANCE, COMPANY_ID, TAXI_TYPE

```
32532516,B45U23,2009-9-1 00:05:40,2009-9-1 00:16:26,2.88,4.92,79,18,1.34,7753,2
32532829,B97S42,2009-9-1 00:12:32,2009-9-1 00:17:42,2.88,4.03,5,15,11.35,7753,2
32535829,B45U23,2009-9-1 00:29:10,2009-9-1 00:33:05,2.88,1.82,47,9,5.65,7753,2
32537800,B67E96,2009-9-1 00:00:00,2009-9-1 00:04:00,3.12,2.85,20,16.1,41.69,7745,1
32539264,B13K97,2009-9-1 00:05:00,2009-9-1 00:10:00,3.12,2.74,25,16.1,1.36,7741,1
32539806,BQH042,2009-9-1 00:00:05,2009-9-1 00:03:20,3.12,1.303,35,16.1,5.208,7746,1
32539814,B19E52,2009-9-1 00:07:00,2009-9-1 00:16:00,3.12,5.94,60,26.9,6.07,7761,1
32540019,B916D5,2009-9-1 00:03:01,2009-9-1 00:06:45,13.12,1.594,66,16.1,0.544,8267,1
32540068,B61E52,2009-9-1 00:02:08,2009-9-1 00:06:48,3.12,0.32,168,16.1,18.89,7745,1
```

INTRODUCTION
Taxi

- The source of our trajectory data
- Can be considered as mobile sensors
- Unique specifications:
  - Additional information (vs. private vehicles)
    - Equipped with taximeters
  - Two different driving behaviors (vs. other public transportations)
    - Vacant (slow, redundant) & occupied (high speed, efficient)
  - Privacy (vs. other trajectory data, e.g., mobile data)
    - Data is collected by the government

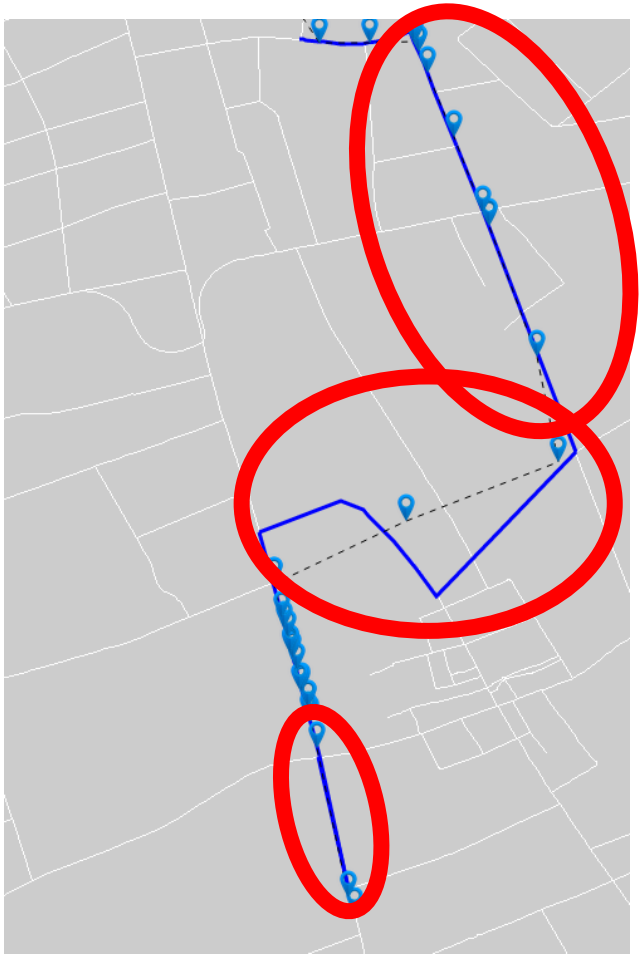
INTRODUCTION
Taxis in China

- Large number (2011) [1]
  - Beijing: 63k, Shanghai: 50k, Shenzhen: 15k
- Stopped growing [2][3]
  - Low passenger load factor: many taxis are vacant
  - Bad scheduling: difficult to hire taxis even there are many vacant taxis
- Bad service facilities
  - Few waiting lanes / parking places

[1] http://www.hnczc.com/news_show.aspx?id=1751
[2] http://politics.people.com.cn/n/2012/0816/c1001-18753269.html
[3] http://www.voc.com.cn/Topic/article/201109/201109061730026480.html

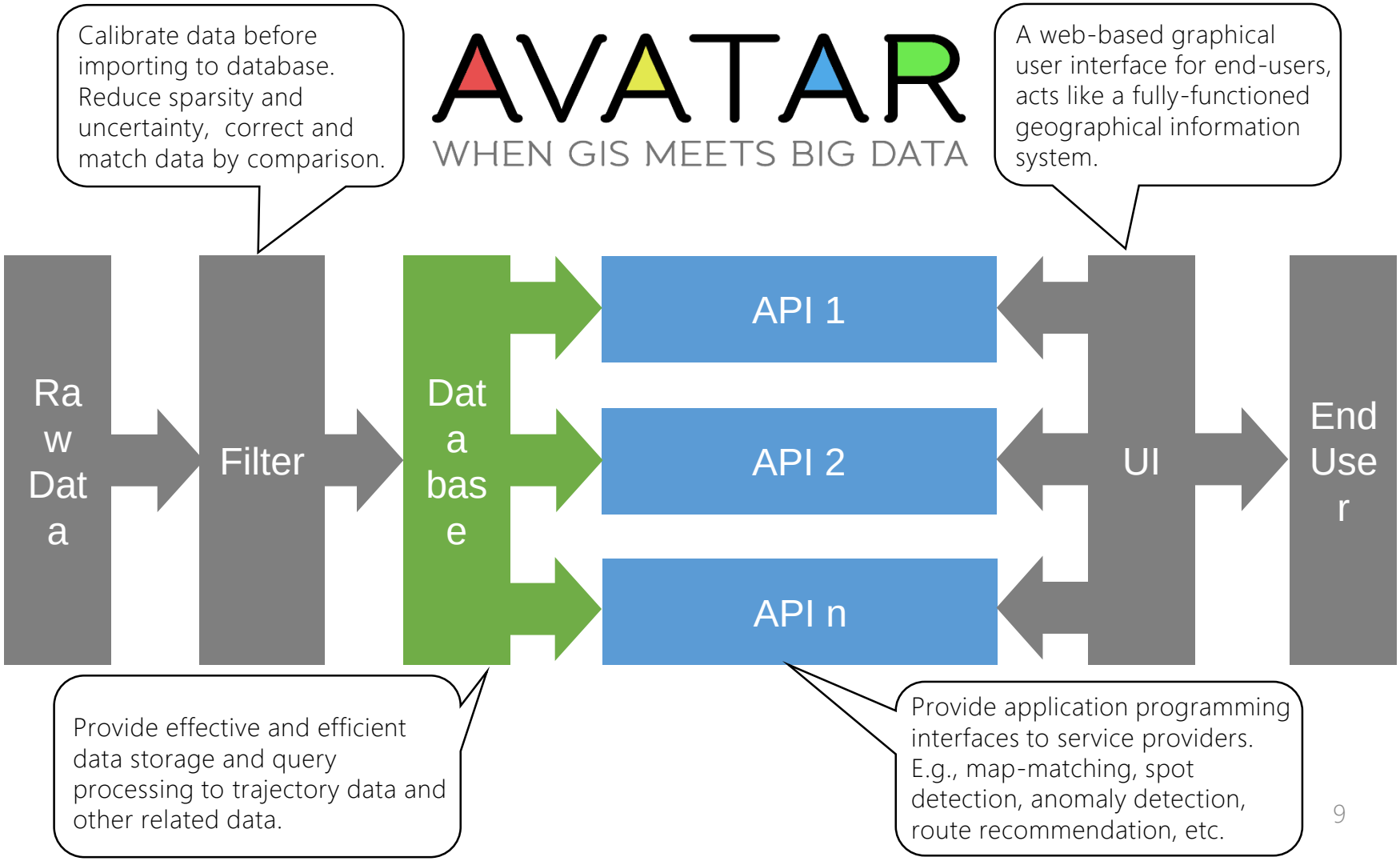
INTRODUCTION
Challenges

- Sparse
  - Low sampling rate
  - Missing data
- Uncertain
  - GPS device is not accurate
- Non-uniform
  - Sampling rate is not stable
  - Heterogeneous data



INTRODUCTION

System Structure



Contents

- Introduction
- Database Layer
- API Layer
- Conclusion

DATABASE LAYER
Spatio-Temporal Database

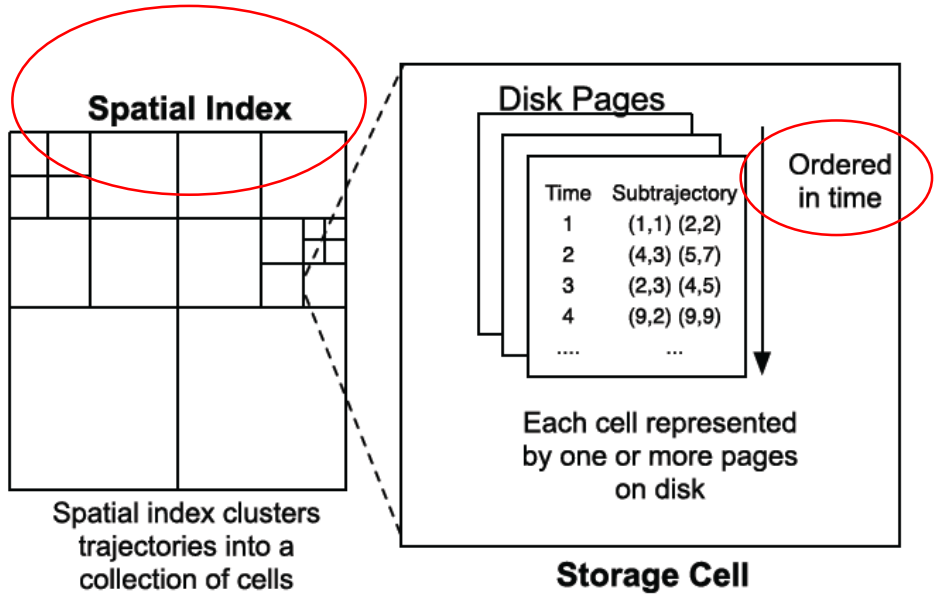
- A database that manages both space and time information
- An extension of spatial databases
- Captures spatial and temporal aspects of data and deals with:
  - Geometry changing over time, and / or
  - Location of objects moving over invariant geometry (known variously as moving objects databases or real-time locating systems)
- Have both spatial and temporal indices

DATABASE LAYER
Spatio-Temporal Database

[ICDE'10] TrajStore: An Adaptive Storage System for Very Large Trajectory Data Sets

- Spatio-temporal database
- Designed for trajectories
- Fully functioned
- Adaptive index
- Faster than traditional databases

R-Tree, K-D Tree, Quadtree, etc.



DATABASE LAYER
Problems of Spatio-Temporal Database

- Distributed: query can be processed in distributed machines simultaneously via MapReduce
- Adaptive: how to determine the size of the spatial index based on historical queries?
- Streaming: modify spatial and temporal index while inserting data
- Real-time: execute queries in real-time
- Compressed: compress trajectories via both spatial and temporal aspects

DATABASE LAYER
Queries of Spatio-Temporal Database

- Basic spatio-temporal range query
  - Show the trajectories passed by a specific area within a specific time period
- Queries require additional calculations
  - Path-finding query (shortest path, most frequent path, etc.)
  - Statistical query (traffic, energy-consumption, gather location, etc.)
  - Predictive query (vehicle location, arriving time, etc.)
- NoDB Queries
  - Instantly show statistics / benchmarks of raw data
  - Query from compressed data

Contents

API LAYER
Characteristics of Trajectories

- Map-matching
- Traffic volume prediction
- Energy consumption prediction
- Arriving time prediction
- Route recommendation
- Anomaly detection
- Event / spot detection (hot spot, gather location, etc.)
- Driving pattern learning from different type of vehicles

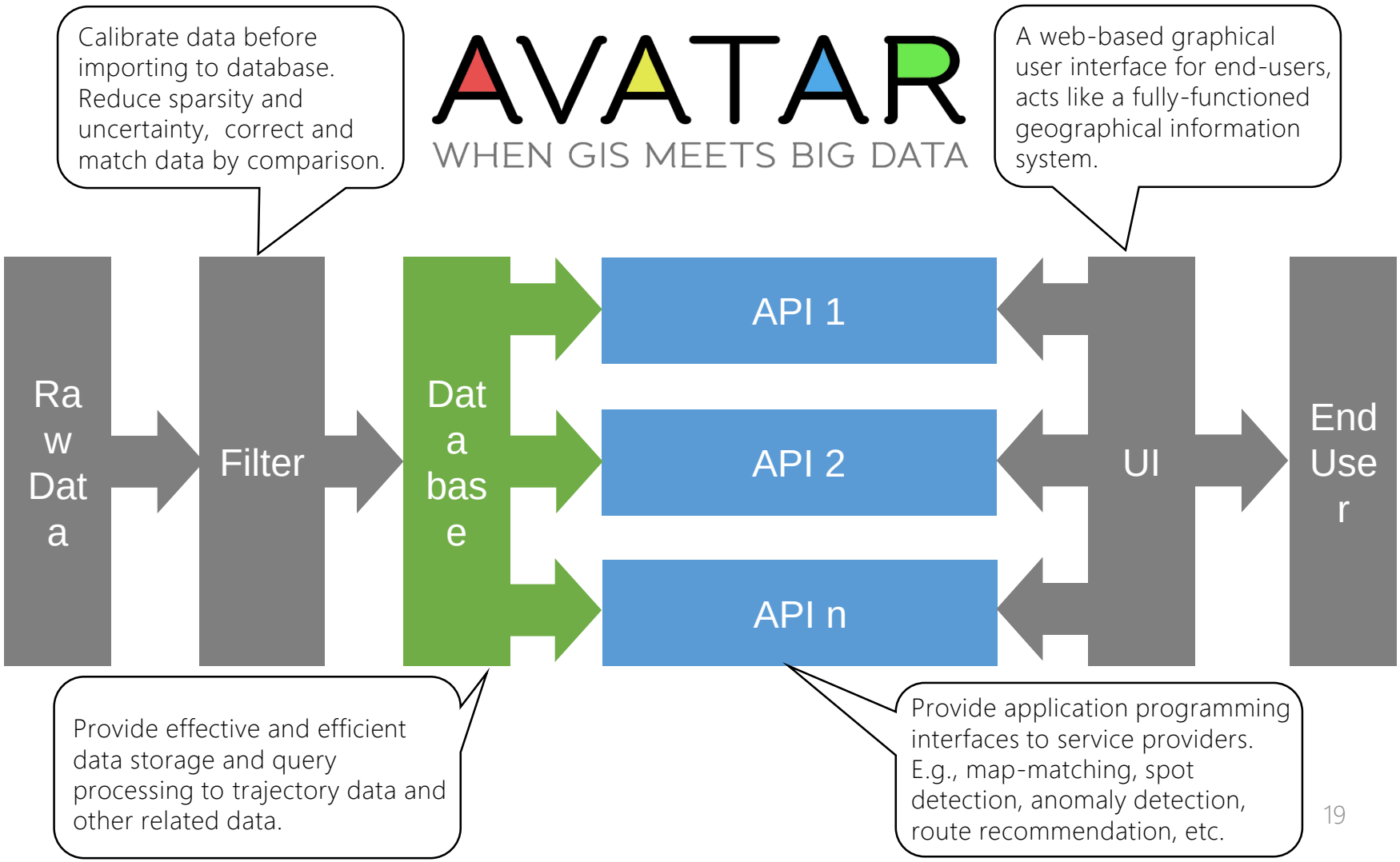
API LAYER
Urban Computing / Urban Sensing

- Urban planning flaw detection
- Traffic control / scheduling
- Estimation of the efficiency of public transportations
- Map enhancement (road type detection, no left turn detection, etc.)
- Region recognition
- POI (point-of-interest) placement (e.g., gas station)

API LAYER
Combining Different Data Sources

- PM 2.5 detection
- Weather / climate sensitive region detection
- Combine location-based social network services (LBS)
- Combine O2O (online-to-offline) services (e.g., traveling services)
- Combine / compare the data of different cities

Conclusion
System Structure



Thanks!