



Predicting Crowds and Travel Times on Public Transportation

Marianne Winslett & Zhenjie Zhang
Advanced Digital Sciences Center &
University of Illinois at Urbana-Champaign



***ADSC is a
University of
Illinois
research
center
in Singapore.***



Faculty from
Computer Science
Computer Engineering



*technology
transfer*

Analytics for Video, Audio, & Text
Cyberphysical Systems
Security



Agency for
Science, Technology
and Research

\$95M 2009-2017



Why is ADSC in Singapore?



- Enhance Singapore's position as a hub for cutting-edge research in information technology.
- Provide Illinois with unique collaborative and funding opportunities for research with high potential impact.

What impact has ADSC had?

as of 30 Apr 2013

Best paper award nominations, other honors
Conference papers accepted last year
Journal papers accepted last year

15

87

43

Scientific Impact

ADSC technologies licensed
Startups facilitated in Singapore
Patent, trade secret applications

5

5 +1

3 +1

Economic Impact

(completed + underway)

Interns
BS holders
PhD holders

>100

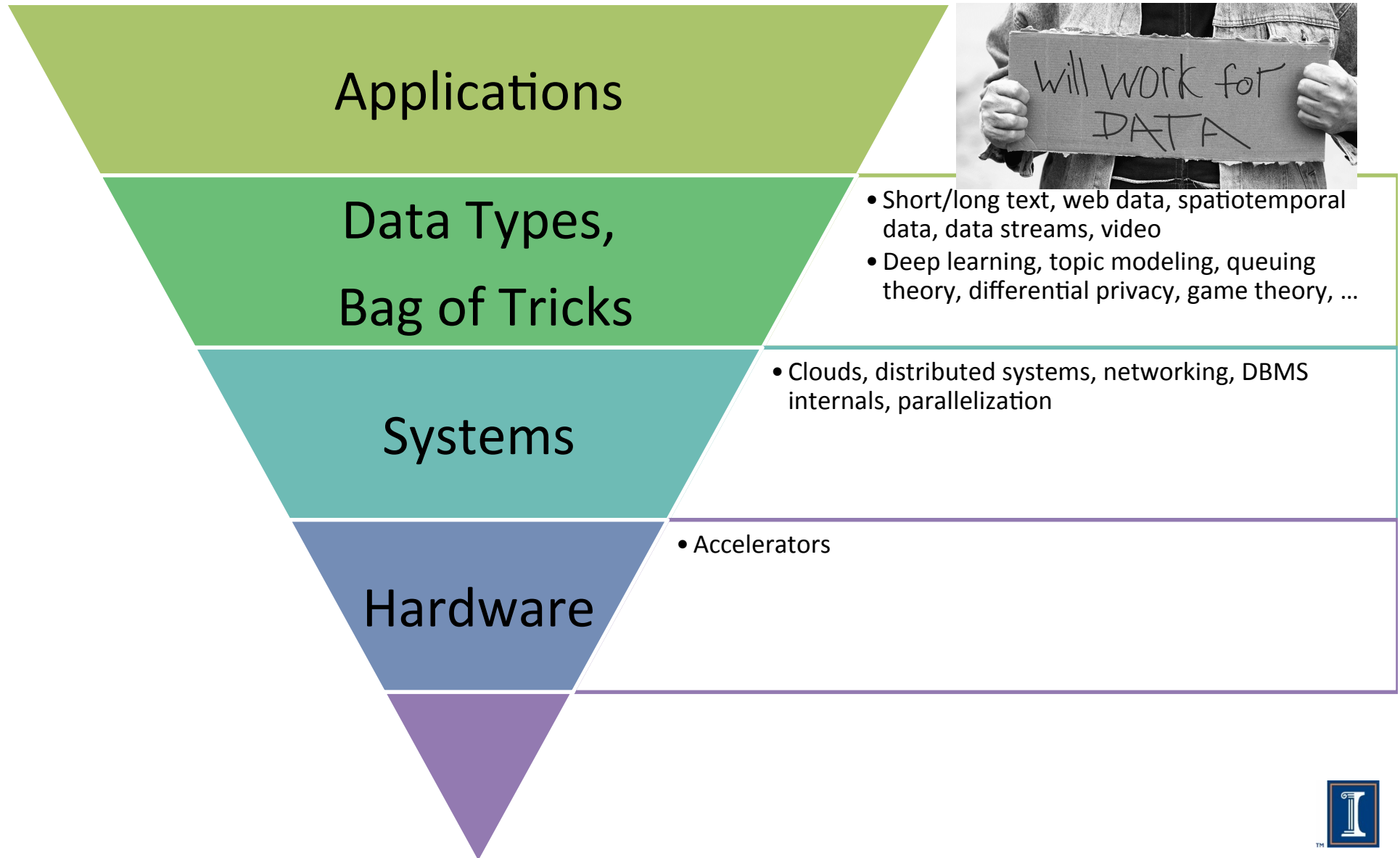
>60

>40

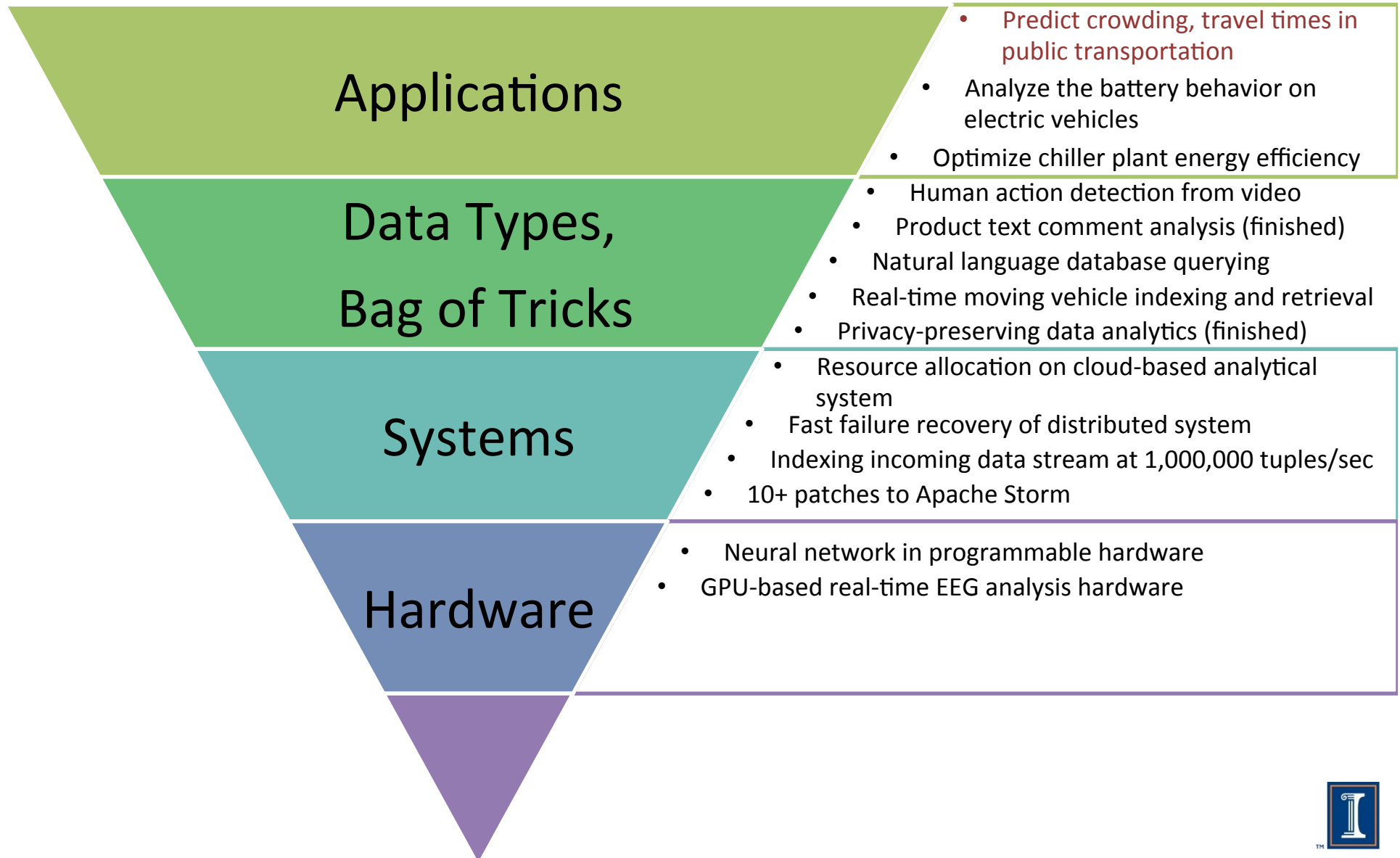
Workforce Development



Our group at ADSC does *analytics*.



Example recent projects



SHANGHAI

MERCURY



Available on the App Store



ANDROID APP ON Google Play

MERCURY: CURRENT PROGRESS

prototype version is published.
tube:

<https://www.youtube.com/watch?v=LatqguUa4oQ>



Check the APP

24 Apr



Data



metro, bus, taxi, transportation
card, weather, road condition

APP



crowdedness prediction for metro,
bus and taxi (iOS, Android supported)

System



distributed streaming system
(Spark, Storm)

Model



deep learning model for prediction
(recurrent neural network)

Commercialization



location-based recommendation

Optimization



improvement of model accuracy &
user experience

How can I avoid ?



PROBLEM: EVERY MORNING

We have to face:



crowded metro trains



unavailable taxis



slowly moving buses



too many private cars



SOLUTION:
EARLY PREDICTION

Real-time Crowdedness Prediction in 24 Hours

Plan a day well

24 Apr



availability of taxis

driving time of buses

number of
passengers on trains

APP can answer my question



real-time stream data for
Short-term trend prediction



historical big data for
Deep learning model training

**DATA REQUIRED:
REAL TIME + HISTORY**

Subscribe the Prediction of Crowdedness

SUBSCRIPTION

METRO:

metro line

Will be crowd at 8am?



metro station

prediction time

metro map



Push the Notification of Crowdedness

map of metro station

It's time to go !



number of passengers

- ✓ in station
- ✓ on platform
- ✓ on train

METRO:
AVOID CROWDEDNESS



trend in 1 hour

METRO: DATA BEHIND

How to get the
numbers ?



Raw Data

transportation card
(only access records)

metro operation data



Route Planning

infer traveling path in-
between any stations



estimate traveling time
(including waiting and
interchange)



passengers in station
(easy)



passengers on platform
(hard)



passengers on train
(very hard)

METRO: PROBLEMS IN DATA

Ratio of Error-prone Data

So many problems !



passengers with incomplete
access records

6.85%



passengers in closed
hours

0.28%



passengers come in / out
at the same station

0.31%

TAXI:

Year	Percentage of people who have ever been in a romantic relationship
1990	85%
1995	85%
2000	82%
2005	78%
2010	75%

Carrier 1:12 AM

订制出租车空车预报

等车位置: 浦东新区, 世纪大道, 1220号

预报时间: 今天 12:00 PM

银城路 裕景大饭店 浦东大道 源深路 上海 浦东南路 南泉北路 地铁2号线 世纪大道 东昌路 世界广场 市政大楼 福山路 梅园 商业路 源深体育中心 世纪大道 江苏大厦 华都大厦 八百伴 浦电路 杨高路 峨山路 南泉路 京城大厦 香园 500米 蓝村路

Can I find a taxi ?



city map

Push the Notification of Availability

map of taxi stand



I should wait for
a while

prediction of
taxi availability

trend in 1 hour

TAXI:
FIND A TAXI

METRO: DATA BEHIND

Raw Data

taxi data
(Only GPS, indication)

Inference of Taxi Status



build a city grid at
granularity of 300m



recognize operating
& off-service taxi



estimate velocity,
heading and distance

How to locate a taxi ?



Statistics



of passengers picked
up in each grid



of passenger dropped
off in each grid



of available taxis in
each grid



average velocity in
each grid

TAXI: PROBLEMS IN DATA



Status of available / busy taxi is incorrect



incorrect Status of off-service taxi



Incorrect velocity



Sparsity of available taxis in suburbs

Need data cleaning



Subscribe the Prediction of Driving Time

SUBSCRIPTION

Bus:

bus line



起点

direction

How can I go home ?

departure & destination

终点

target time



Push the Notification of Driving Time

bus route



driving time between departure and destination

It's too slow !



trend in 1 hour

BUS:
MANAGE MY TIME

BUS: DATA BEHIND

Raw Data

ous data
(records of arriving &
eaving stop)

Inference of Bus Operation



correct bus route
by arriving time



calculate driving time
in-between stops



correct the records of
skipping stops

Statistics

How to know the
traveling time?



compute driving time
during off-peak hour



compute driving time
during peak hour



BUS: PROBLEMS IN DATA



skipped stops by the buses



incomplete bus arriving records



incorrect direction of bus operation



encoding error of Chinese

Bus data is complex



Real-time Onsite Recommendation

onsite recommendation



advertisement is information.

based on user's
subscriptions and current
traffic conditions

Have a cup of
coffee first



COMMERCIALIZATION:
LB - RECOMMENDATION

A Solution for City Management

 reward passengers who travel during off-peak hours



I can earn rewards



**REWARDS:
TAGGED SHIFTS**

Dataset for Training

1 Apr. 2015 – 23 Apr. 2015

SYSTEM: DATA MINING

Data Cleaning

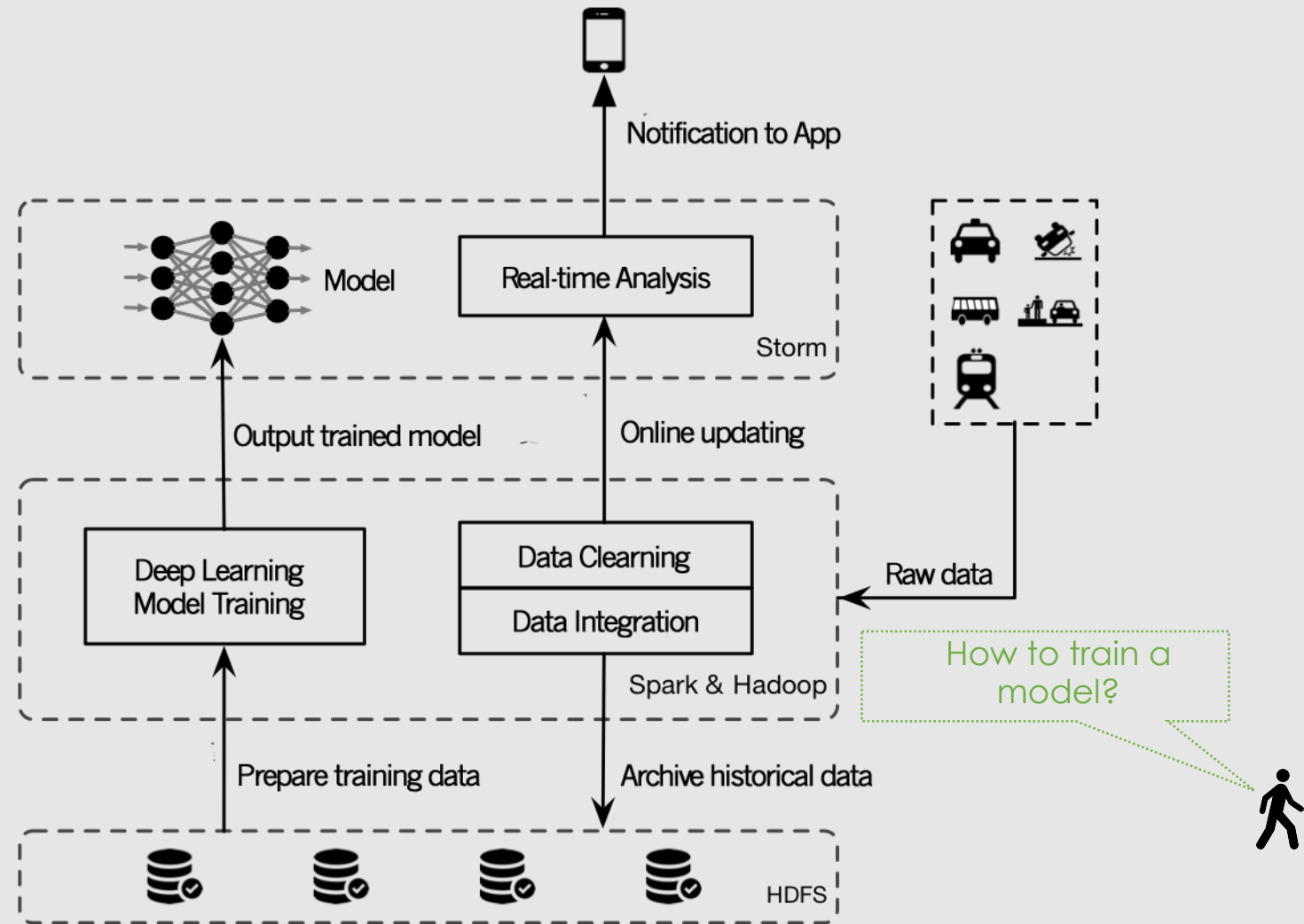
pre-process massive data on Hadoop

Model Training

train deep learning model offline on Spark

Real-time Predicting

compute near-future trend online via Storm



SYSTEM: REAL-TIME PREDICTION

Dataset for Testing

2015.4.24 - 2015.4.30

Real-time

response time < 1ms

High Concurrency

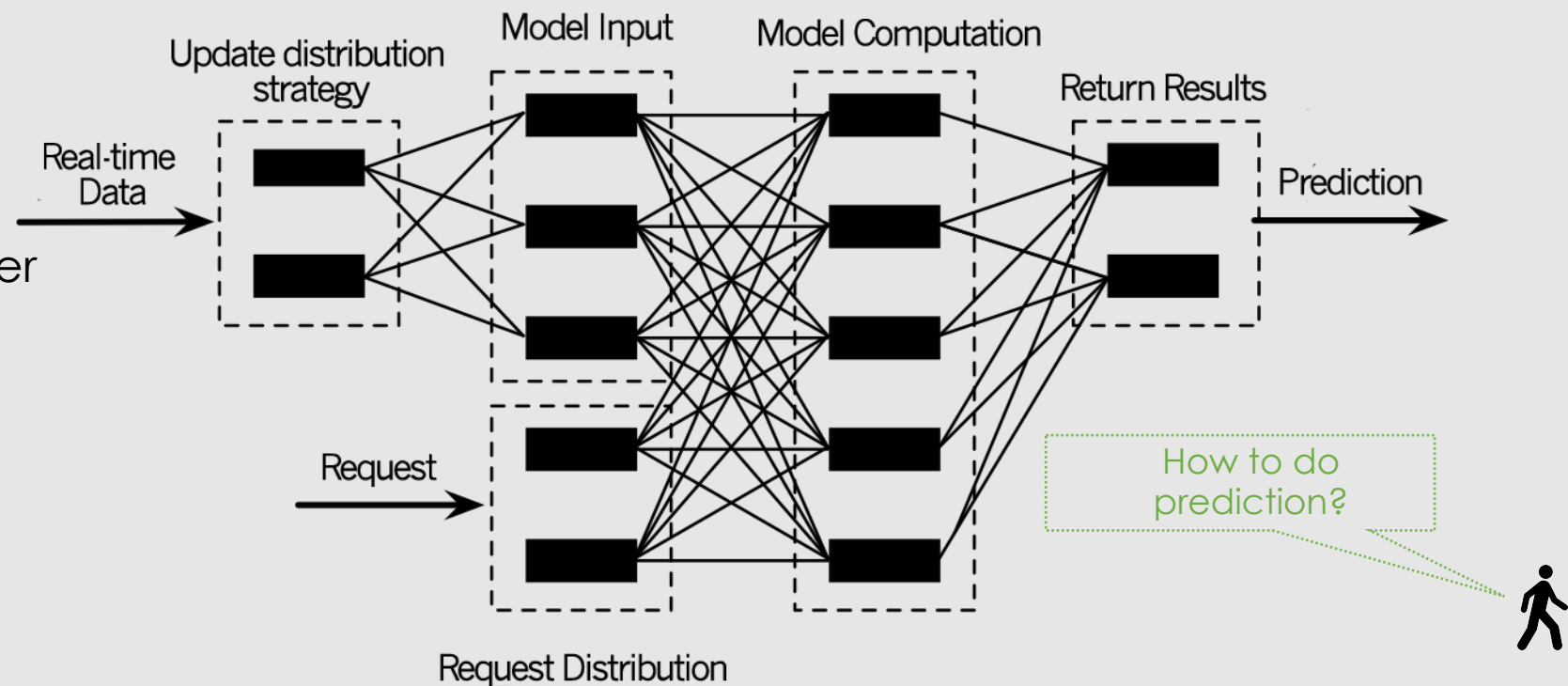
150K times prediction per second per node

Elastic Computation

increase / decrease nodes on demand

Highly Usable

distributed fault tolerance



Integration of Road and Weather Information to Improve Model Accuracy

-  weather
-  accident on road
-  traffic index of city road
-  closure of ramp

**MODEL:
MULTIPLE FACTORS**

Consider other factors?

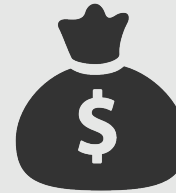




{ Comfortable Travel }



{ Smart Transportation }



{ Efficient Business }



{ Green City }

VALUABLE



{ Safe Driving }

A new day coming





SMART SENTOSA



IMS
Intelligent Mobile Solutions

ADSC
Illinois at Singapore Pte Ltd

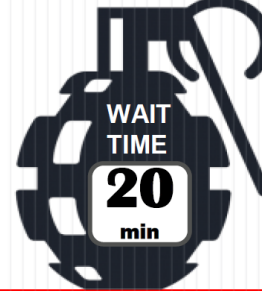
XJER

PROBLEM ANALYSIS

Unable To Capture Tram & Bus Ridership Information

Two earlier considerations made were **not workable**:

- ☐ People counting mats: guests front or back door making col
- ☐ RFID tags/wearable: no prac are free



Using Queue Marker As Proxy For Ridership Information

- ☐ Duty personnel **manually observes** which **queue** marker the last person in the queue is at **via CCTV**;
- ☐ **Waiting time is estimated** by the last queue marker's position
- ☐ Data is **manually entered** into an excel file and analysed for average waiting time
- ☐ Cycle is repeated every 10 minutes;

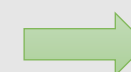
Problem

Crowd Information Collection



Cause

Lack of Effective Perception



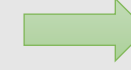
Technical Intuition

Real-time Video Analytics

Staff Management

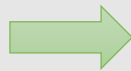


Lack of Prediction Capability

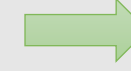


Predictive Analysis by integrating data

Guest Engagement

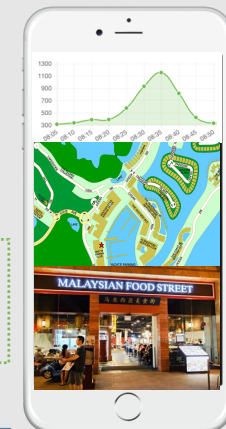
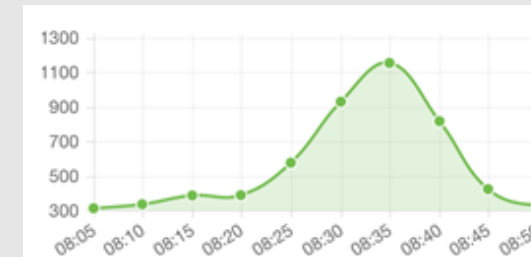
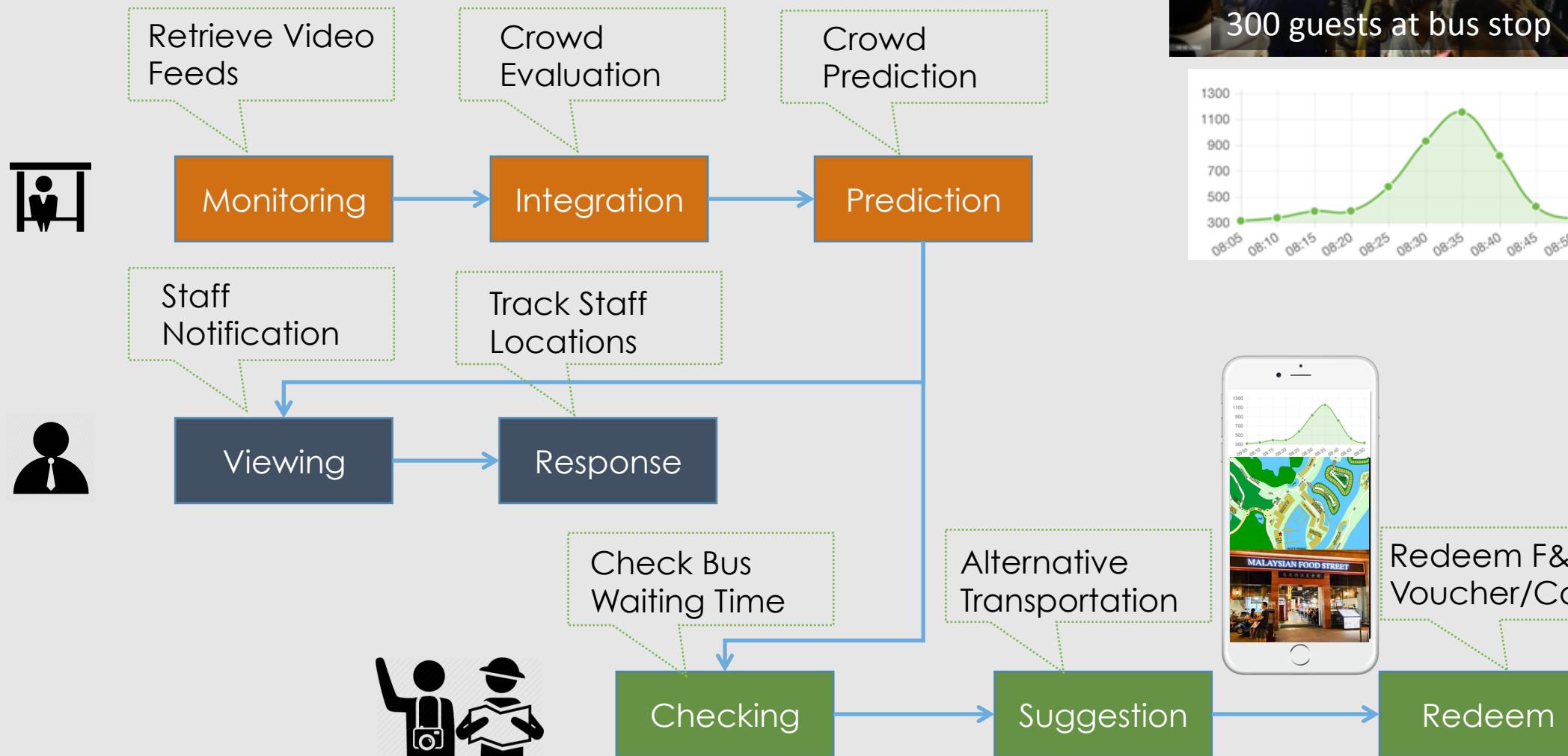


Lack of Contact Channel and Personalized Recommendation



Targeted Incentives on smart phones

OVERALL WORKFLOW



SYSTEM ARCHITECTURE

Predictive Analytics

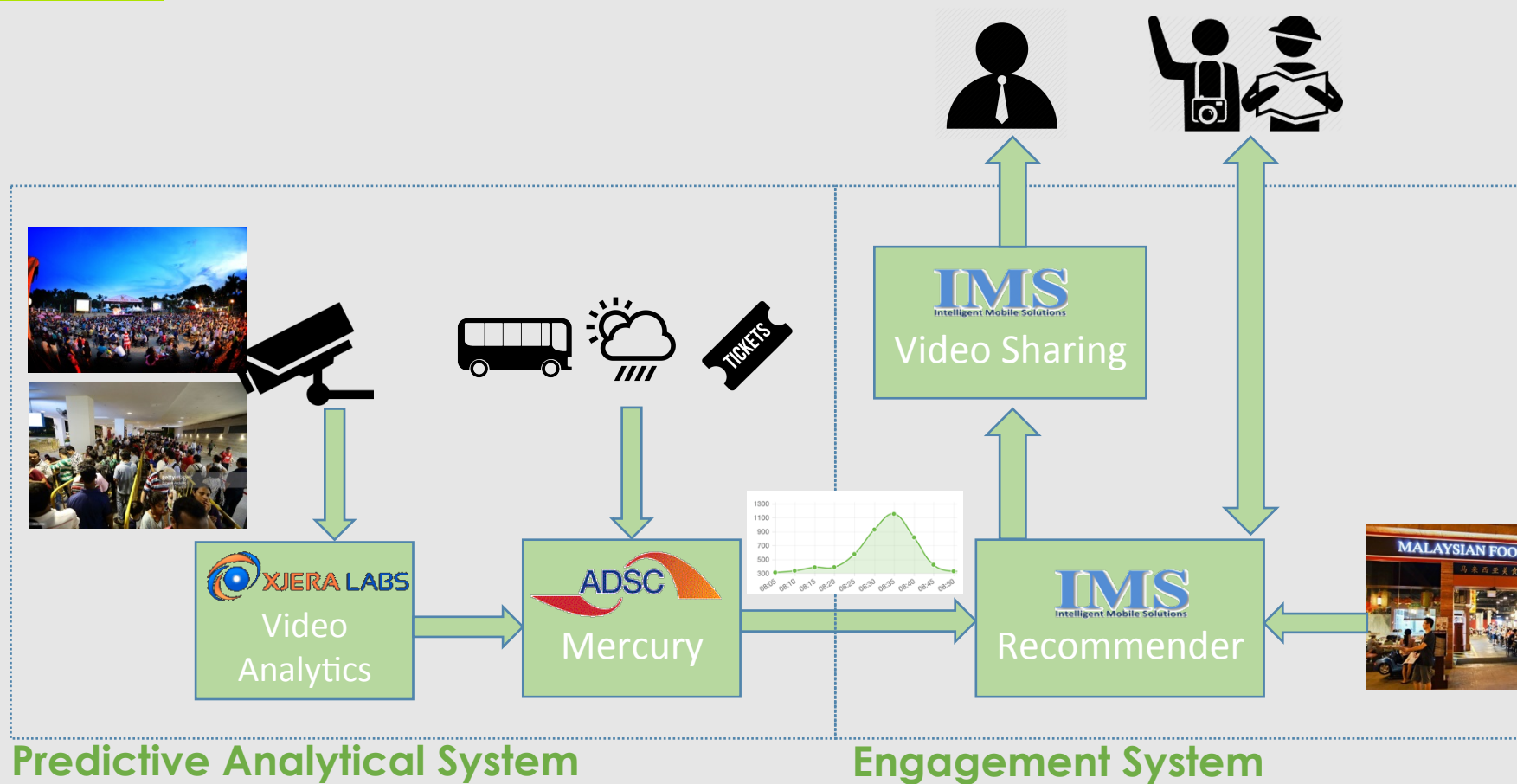
Video Analysis
Density Evaluation
Density Prediction

Staff Engagement

Video Sharing
Location Tracking

Guest Engagement

Coupon Recommend
Coupon Redeem



USE CASES: L-TIME TRAFFIC PREDICTION

Data Cleaning



pre-process
massive data
on Hadoop

Model Training

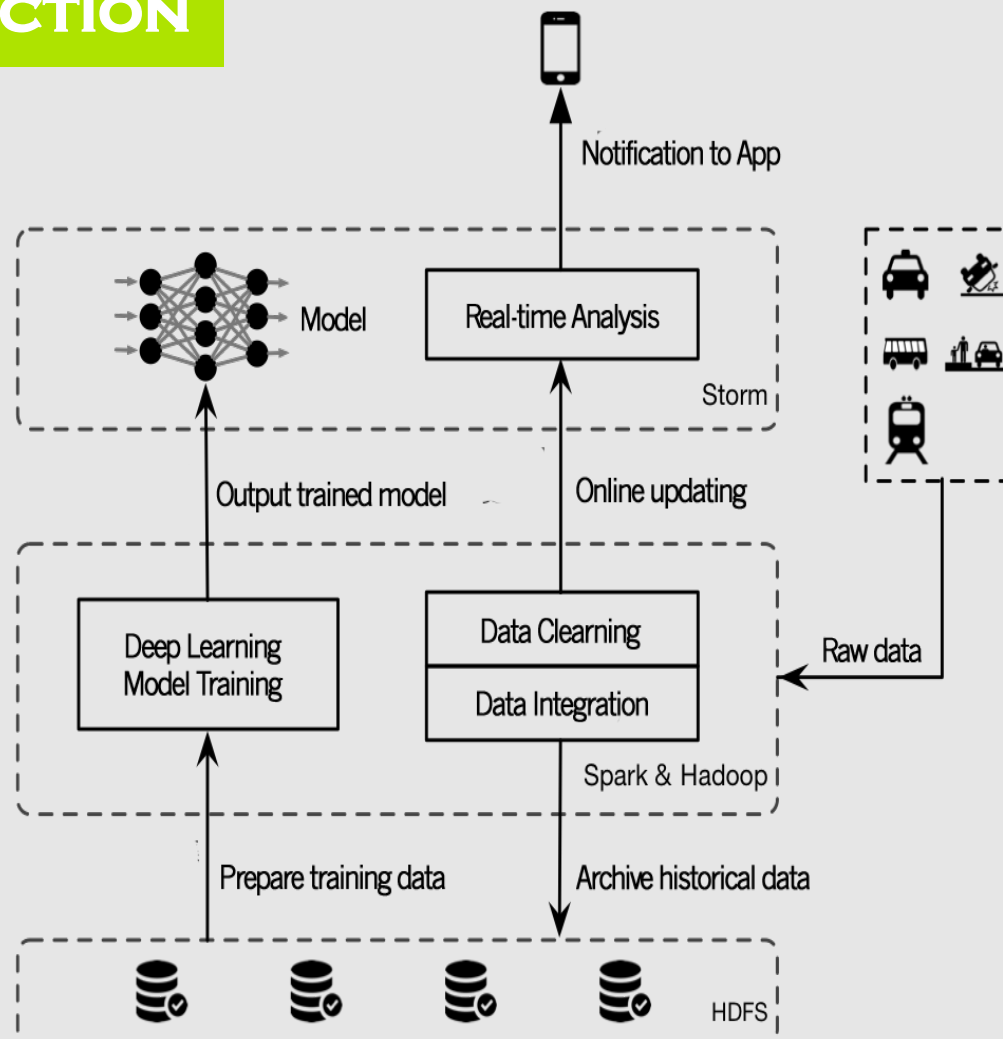


train deep
learning model
offline on Spark

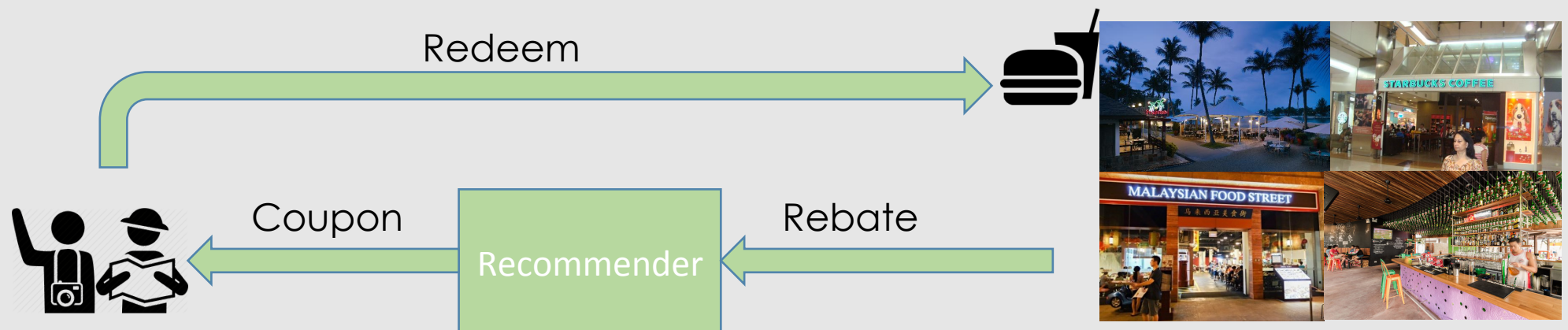
Real-time Predicting



compute near-
future trend
online via Storm



SUSTAINABLE BUSINESS MODEL



ECONOMIC ANALYSIS

One-Time Hardware Investment



Server for video processing, predictive analysis and app hosting



Software on analytics and mobile apps

Operational Cost



Electricity on computation



Software maintenance

Incomings



Each coupon charges the vendor dozens of cents

