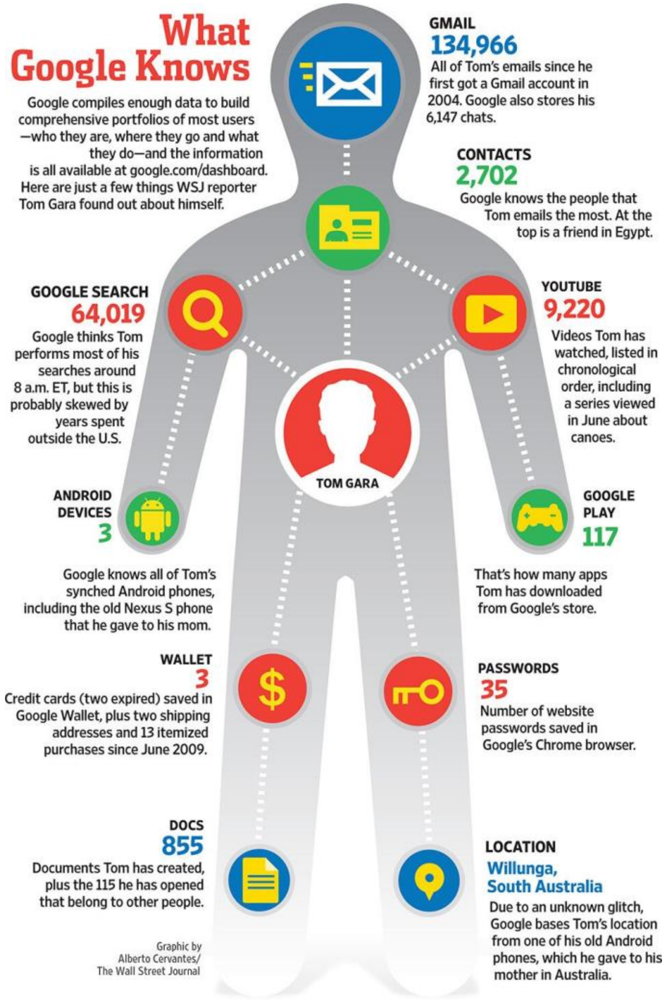
35th ASIAN RACING CONFERENCE
HONG KONG 2014

Making Big Profits from Big Data

Opportunities and Challenges in Business

Google Knows You Better Than You Do



What Google Knows

Google compiles enough data to build comprehensive portfolios of most users —who they are, where they go and what they do—and the information is all available at google.com/dashboard. Here are just a few things WSJ reporter Tom Gara found out about himself.

GMAIL
134,966
All of Tom's emails since he first got a Gmail account in 2004. Google also stores his 6,147 chats.

CONTACTS
2,702
Google knows the people that Tom emails the most. At the top is a friend in Egypt.

GOOGLE SEARCH
64,019
Google thinks Tom performs most of his searches around 8 a.m. ET, but this is probably skewed by years spent outside the U.S.

YOUTUBE
9,220
Videos Tom has watched, listed in chronological order, including a series viewed in June about canoes.

TOM GARA

ANDROID DEVICES
3
Google knows all of Tom's synched Android phones, including the old Nexus S phone that he gave to his mom.

GOOGLE PLAY
117
That's how many apps Tom has downloaded from Google's store.

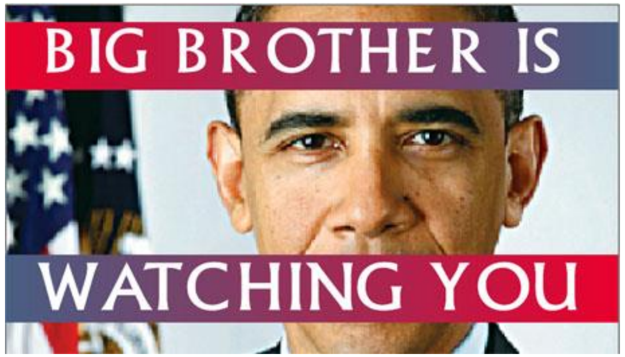
WALLET
3
Credit cards (two expired) saved in Google Wallet, plus two shipping addresses and 13 itemized purchases since June 2009.

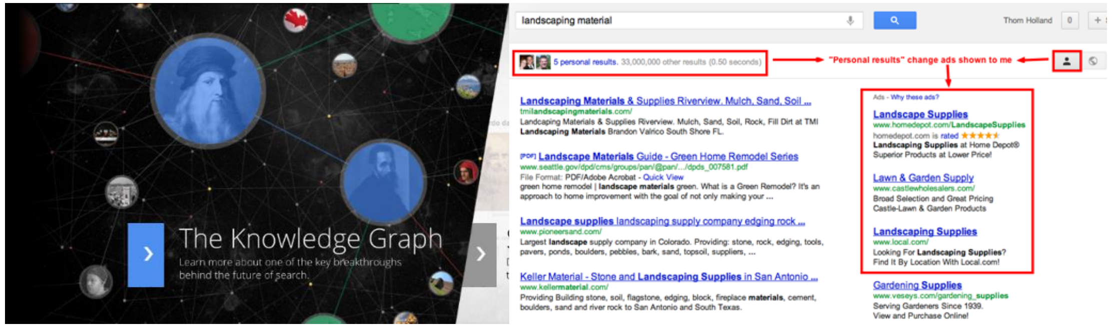
PASSWORDS
35
Number of website passwords saved in Google's Chrome browser.

DOCS
855
Documents Tom has created, plus the 115 he has opened that belong to other people.

LOCATION
Willunga, South Australia
Due to an unknown glitch, Google bases Tom's location from one of his old Android phones, which he gave to his mother in Australia.

Graphic by Alberto Cervantes/ The Wall Street Journal





The Knowledge Graph
Learn more about one of the key breakthroughs behind the future of search.

landscaping material

Thom Holland 0 +

5 personal results. 33,000,000 other results (0.50 seconds)

"Personal results" change ads shown to me

Landscaping Materials & Supplies Riverview. Mulch, Sand, Soil ...
tmilandscapingmaterials.com/
Landcaping Materials & Supplies Riverview. Mulch, Sand, Soil, Rock, Fill Dirt at TMI Landscaping Materials Brandon Valrico South Shore FL.

[PDF] Landscape Materials Guide - Green Home Remodel Series
www.seattle.gov/dpd/cms/groups/pan/@pan/.../dpds_007581.pdf
File Format: PDF/Adobe Acrobat - Quick View
green home remodel | landscape materials green. What is a Green Remodel? It's an approach to home improvement with the goal of not only making your ...

Landscape supplies landscaping supply company edging rock ...
www.pioneersand.com/
Largest landscape supply company in Colorado. Providing: stone, rock, edging, tools, pavers, ponds, boulders, pebbles, bark, sand, topsoil, suppliers, ...

Keller Material - Stone and Landscaping Supplies in San Antonio ...
www.kellermaterial.com/
Providing Building stone, soil, flagstone, edging, block, fireplace materials, cement, boulders, sand and river rock to San Antonio and South Texas.

Ads - Why these ads?

Landscape Supplies
www.homedepot.com/LandscapeSupplies
homedepot.com is rated ★★★★★
Landscaping Supplies at Home Depot®
Superior Products at Lower Price!

Lawn & Garden Supply
www.castlewholesalers.com/
Broad Selection and Great Pricing
Castle-Lawn & Garden Products

Landscaping Supplies
www.local.com/
Looking For Landscaping Supplies?
Find It By Location With Local.com!

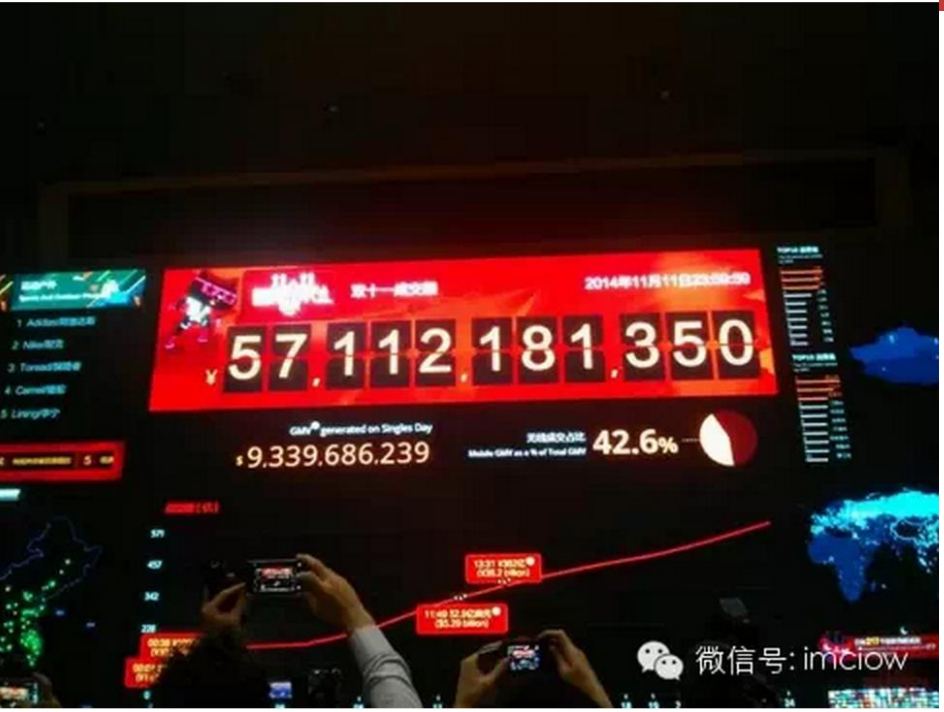
Gardening Supplies
www.veseys.com/gardening_supplies
Serving Gardeners Since 1939.
View and Purchase Online!

Personalized ads based on YOUR data

Google AdWords
Advertise your business on Goo
No matter what your budget, you can display your ads on
G$$gle
Adsense

Enhanced click-through rate = MORE profits!

Alibaba One Day Transaction (Nov. 11, 2014)



境外消费TOP 10

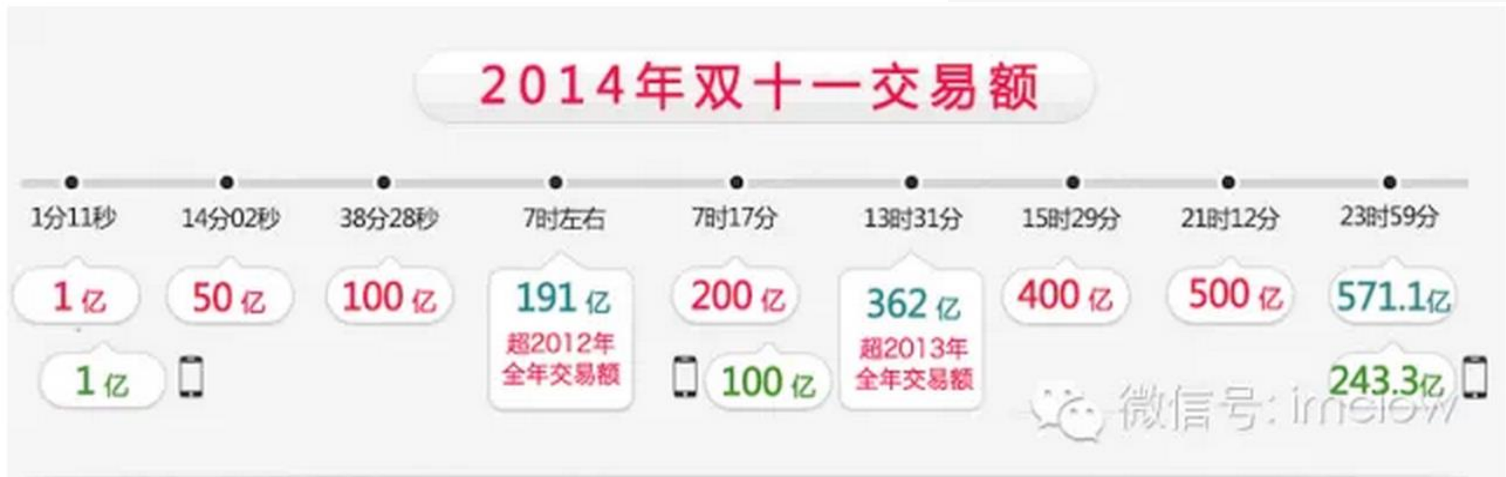
1.香港
2.俄罗斯
3.美国
4.中国台北
5.澳大利亚
6.新加坡
7.加拿大
8.澳门
9.巴西
10.马来西亚

（*地区排名与地图非对应关系）

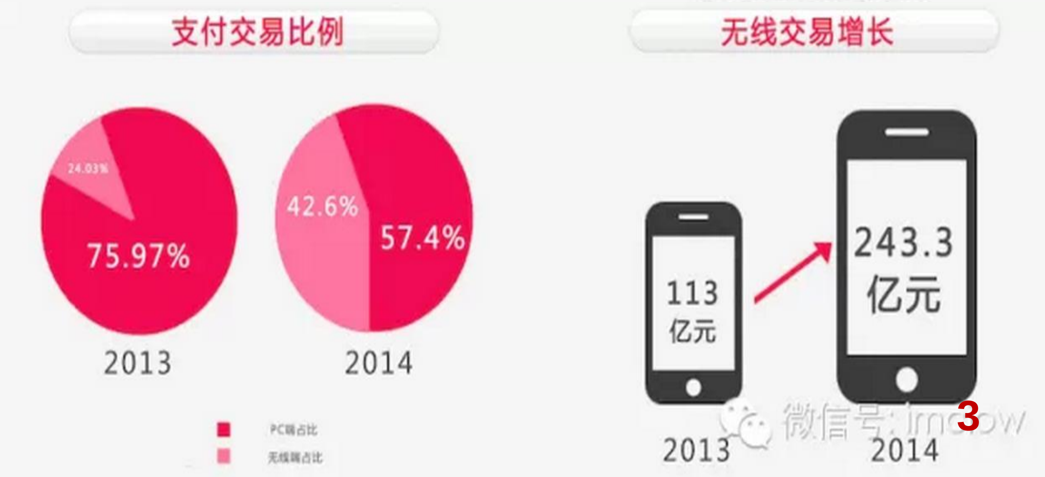
海淘热门商品TOP 5

手机　奶粉　箱包　女裙　面膜

(*海淘数据来自好奇心日报)

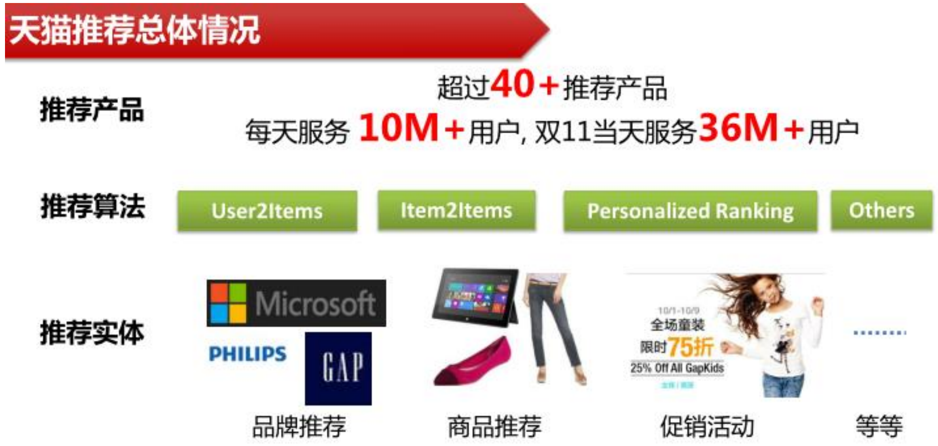

2014年双十一交易额

1分11秒	14分02秒	38分28秒	7时左右	7时17分	13时31分	15时29分	21时12分	23时59分
1亿	50亿	100亿	191亿 超2012年全年交易额	200亿	362亿 超2013年全年交易额	400亿	500亿	571.1亿
1亿 📱				📱 100亿				243.3亿 📱


支付交易比例

无线交易增长

Brand Recommendation in Tmall.com (Alibaba)



天猫推荐总体情况

推荐产品 超过40+推荐产品
每天服务 10M+用户, 双11当天服务36M+用户

推荐算法 User2Items Item2Items Personalized Ranking Others

推荐实体 品牌推荐 商品推荐 促销活动 等等

Microsoft
PHILIPS
GAP
10/1-10/9
全场童装
限时75折
25% Off All GapKids
....

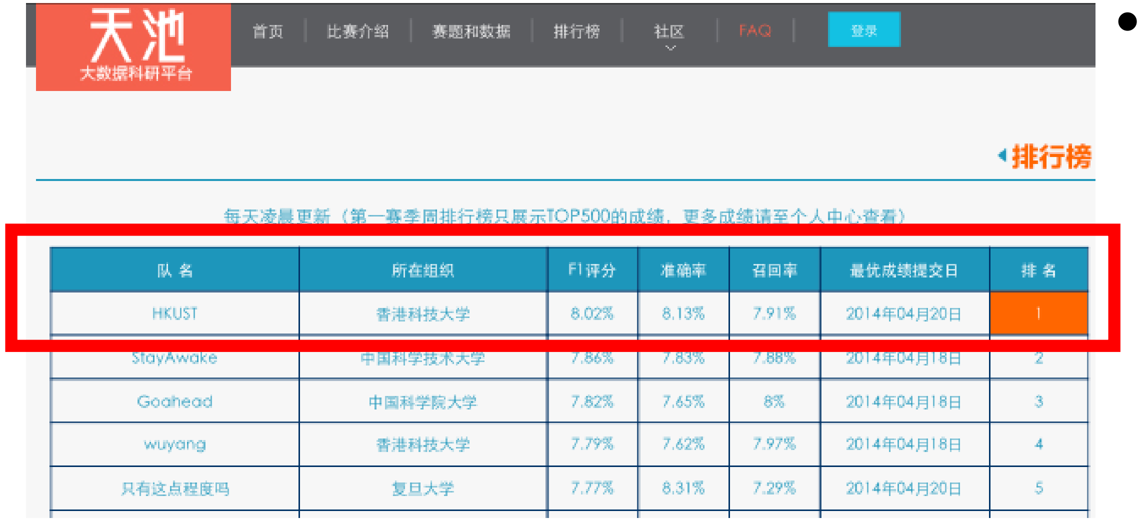
- Serving 10 million+ users every day, 36 million+ users at the busiest day
- 40+ recommended commodities for each user
- Predicting user preferences from behaviors data

Due to better models, the revenue of Tmall.com increased by 20%+

Alibaba Big Data Competition

User behavior data from Tmall.com (Alibaba)

User	Brand	Date	Behavior
Alice	Lenovo	2014-04-18	Click
Bob	Sony	2014-04-20	Click
Bob	Sony	2014-04-20	Buy
…	…	…	…



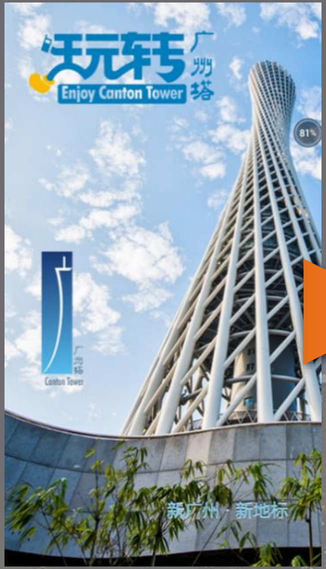
每天凌晨更新（第一赛季周排行榜只展示TOP500的成绩，更多成绩请至个人中心查看）

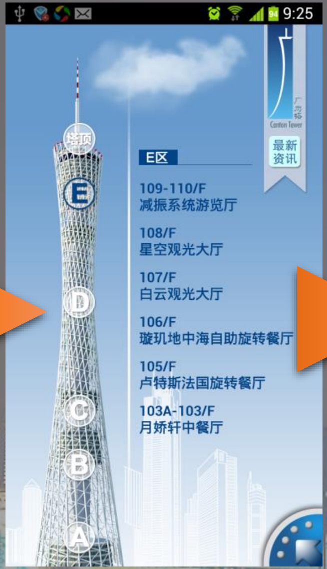
队 名	所在组织	F1评分	准确率	召回率	最优成绩提交日	排 名
HKUST	香港科技大学	8.02%	8.13%	7.91%	2014年04月20日	1
StayAwake	中国科学技术大学	7.86%	7.83%	7.88%	2014年04月18日	2
Goahead	中国科学院大学	7.82%	7.65%	8%	2014年04月18日	3
wuyang	香港科技大学	7.79%	7.62%	7.97%	2014年04月18日	4
只有这点程度吗	复旦大学	7.77%	8.31%	7.29%	2014年04月20日	5

- Alibaba organizes a competition to look for even better prediction models
- Given the log of user behaviors (including CLICK, BUY, BOOKMARK, and ADD-TO-CART) of a certain period
- Predict which users will buy which brands at a later time

We are the No.1 out of 7000+ teams in season 1.

Smart Tour in Canton Tower





9:25

塔顶

E区

109-110/F
减振系统游览厅

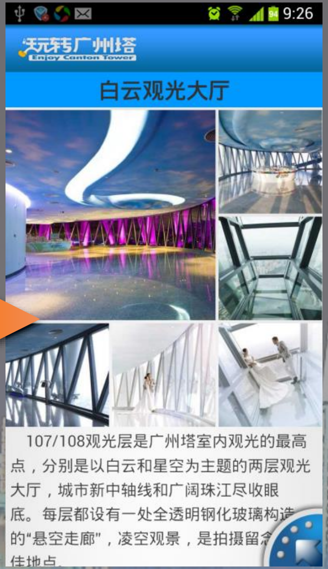
108/F
星空观光大厅

107/F
白云观光大厅

106/F
璇玑地中海自助旋转餐厅

105/F
卢特斯法国旋转餐厅

103A-103/F
月娇轩中餐厅

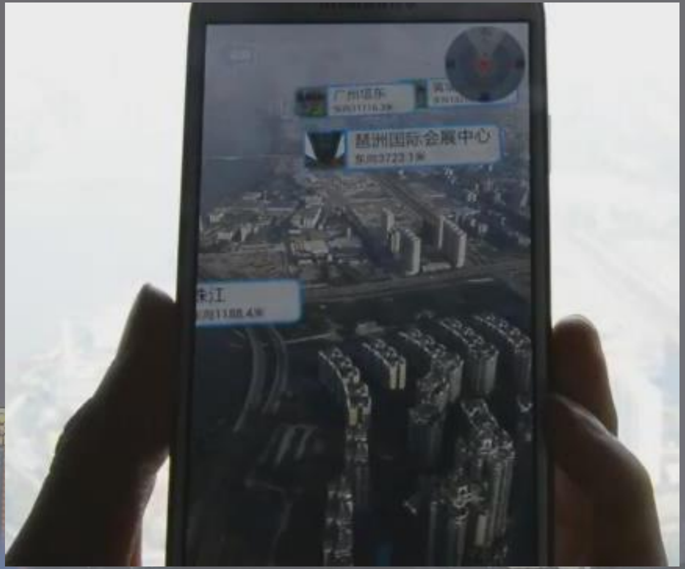


9:26

玩转广州塔 Enjoy Canton Tower

白云观光大厅

107/108观光层是广州塔室内观光的最高点，分别是以白云和星空为主题的两层观光大厅，城市新中轴线和广阔珠江尽收眼底。每层都设有一处全透明钢化玻璃构造的“悬空走廊”，凌空观景，是拍摄留念佳地点。

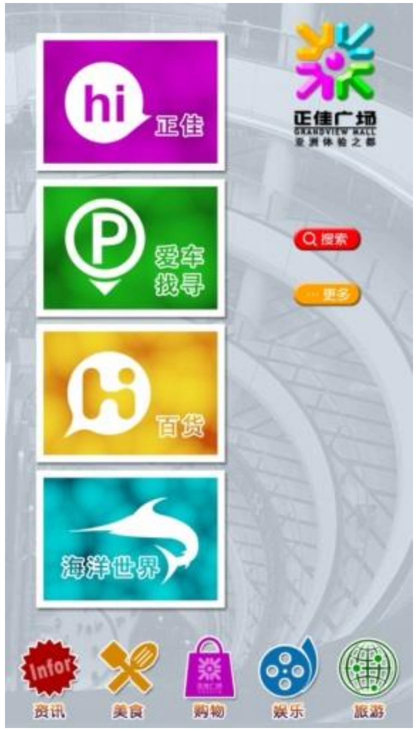


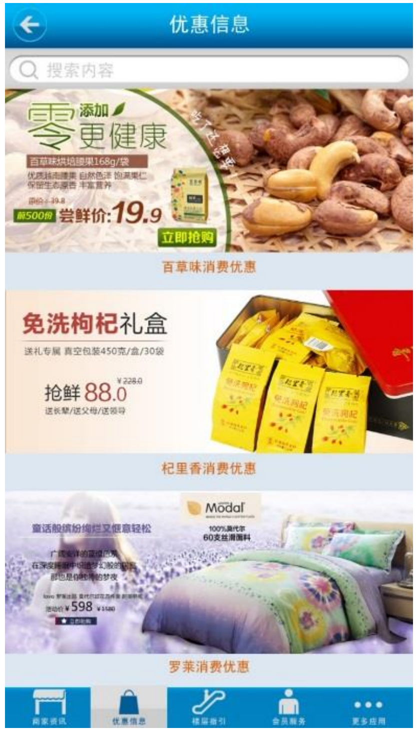
A mobile phone app with augmented reality

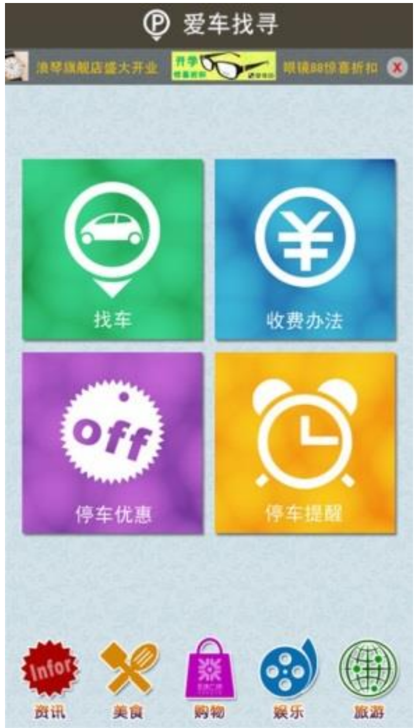
Smart Shopping Mall

Your smartphone will accompany you go shopping

- Indoor navigation
- Personalized and location-based recommendation
- Finding your cars in big parking lots







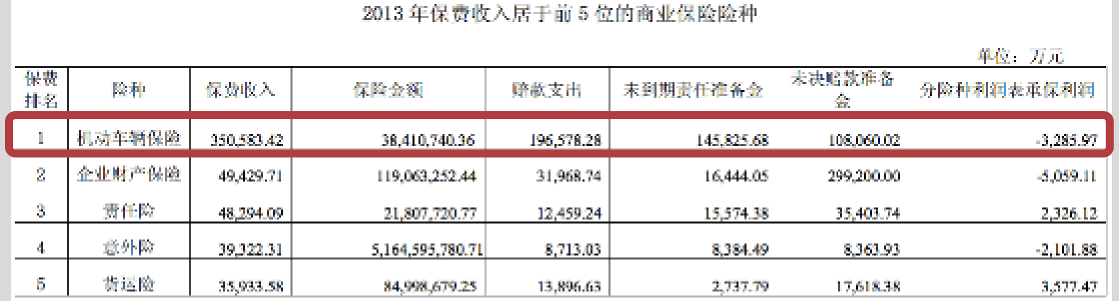
Better shopping experience with very low cost!

Fraud Detection for Automobile Insurance

- It is estimated that there are more than 2500 false insurance claims each year in UK.
- Each false claim could cause a loss of up to $300,000 HKD.
- In the market, there are old and expensive anti-fraud solutions by FICO, SPSS, etc.
- Our solution uses new machine learning technologies.

Our solution detects more frauds while cuts the cost by more than a half.

Fraud Detection for Automobile Insurance

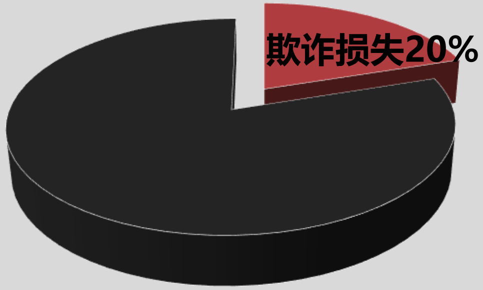


2013 年保费收入居于前 5 位的商业保险险种

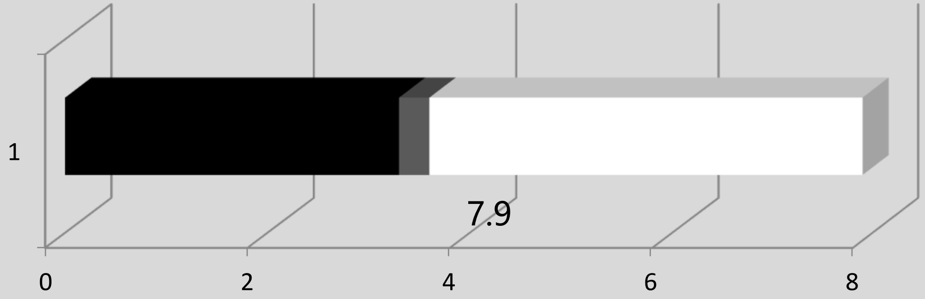
单位：万元

保费排名	险种	保费收入	保险金额	赔款支出	未到期责任准备金	未决赔款准备金	分险种利润表承保利润
1	机动车辆保险	350,583.42	38,410,740.36	196,578.28	145,825.68	108,060.02	-3,285.97
2	企业财产保险	49,429.71	119,063,252.44	31,968.74	16,444.05	299,200.00	-5,059.11
3	责任险	48,294.09	21,807,720.77	12,459.24	15,574.38	35,403.74	2,326.12
4	意外险	39,322.31	5,164,595,780.71	8,713.03	8,384.49	8,363.93	-2,101.88
5	货运险	35,933.58	84,998,679.25	13,896.63	2,737.79	17,618.38	3,577.47

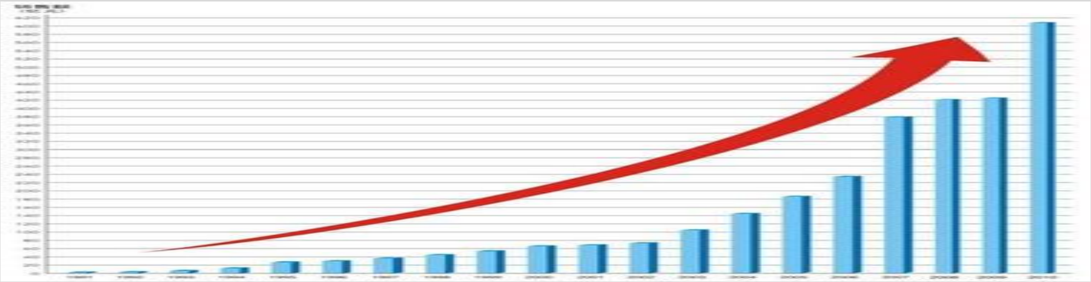
An auto insurance company had compensation expenses 1.97B, with net loss 33M in 2013



CIRC statistics show that about 20% of claims are fraudulent



Save up to 78M from the loss of 390M fraud

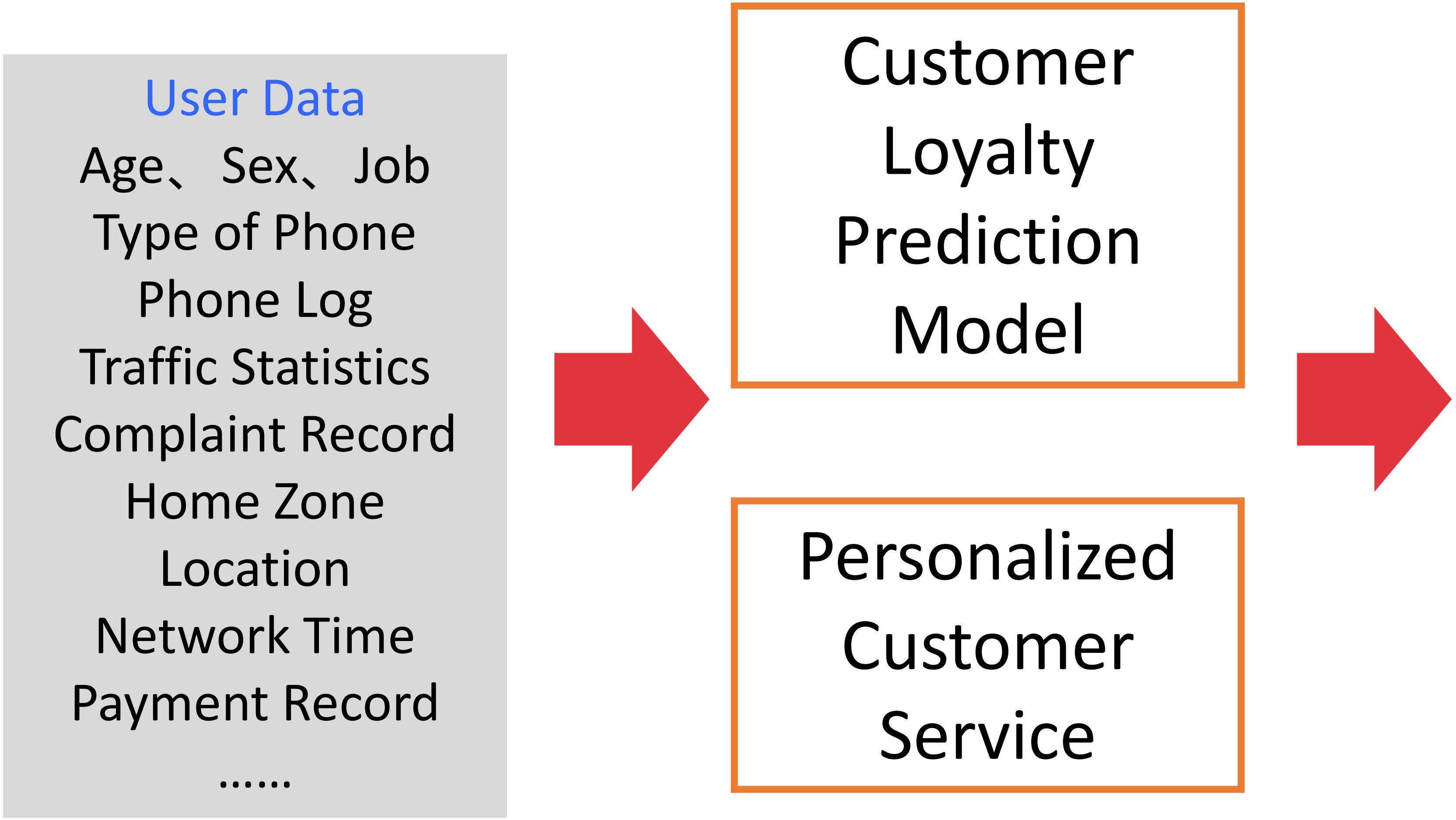


With 3M system cost, the company made profit of 42M

Fraud detection rate increased by 20%, you can change from a loss of 9.4 ‰ to 12.2 ‰ profit

Benefitted from Big Data, the company becomes profitable!

T-Mobile Uses Big Data to Retain Customers



User Data

Age、 Sex、 Job

Type of Phone

Phone Log

Traffic Statistics

Complaint Record

Home Zone

Location

Network Time

Payment Record

……

Customer Loyalty Prediction Model

Personalized Customer Service

In Q1 2011, customer drop out rate reduced by 50% in the US

T-Mobile Uses Big Data to Retain Customers

Mobile Internet allows real-time communication with the Internet

- Always connected
- Greater user stickiness
- Longer access time
- Lower cost of participation

Significantly change to user behavior

Data-as-a-Service:
Revitalizing Payment Transaction Log

FB: $16B

>$10B

New IT Business

Data Center Infrastructure

Smart Shopping Mall

Microfinance Companies

Data-based Precise Marketing

Credit Scoring Service

It requires new information technologies

Big Data Collection

The Knowledge Graph
Learn more about one of the key breakthroughs behind the future of search.

Large-scale Data Mining

Advanced Machine Learning

Data

Knowledge

Insight

The point of Big Data is to make sense of it

Challenges of Harnessing Big Data for Business Value

Data Collection

- Extracting siloed data
- Understanding metadata
- Mixed structured and unstructured data
- Data cleaning
- Data calibration
- Data integration
- …

Data Processing

- Data loading
- Building large indexes
- Parallel algorithms design
- Fault-tolerance
- Adapt to new hardware
- Data compression
- Real-time response
- …

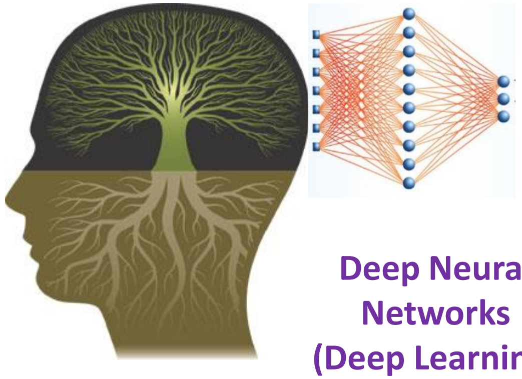
Data Mining

- Statistical analysis
- Data clustering
- Predictive modeling
- Ensemble of models
- Abnormal detection
- Unsupervised learning
- Data visualization
- …

“Big Data is a Revolution that Will Transform How We Live, Work, and Think”

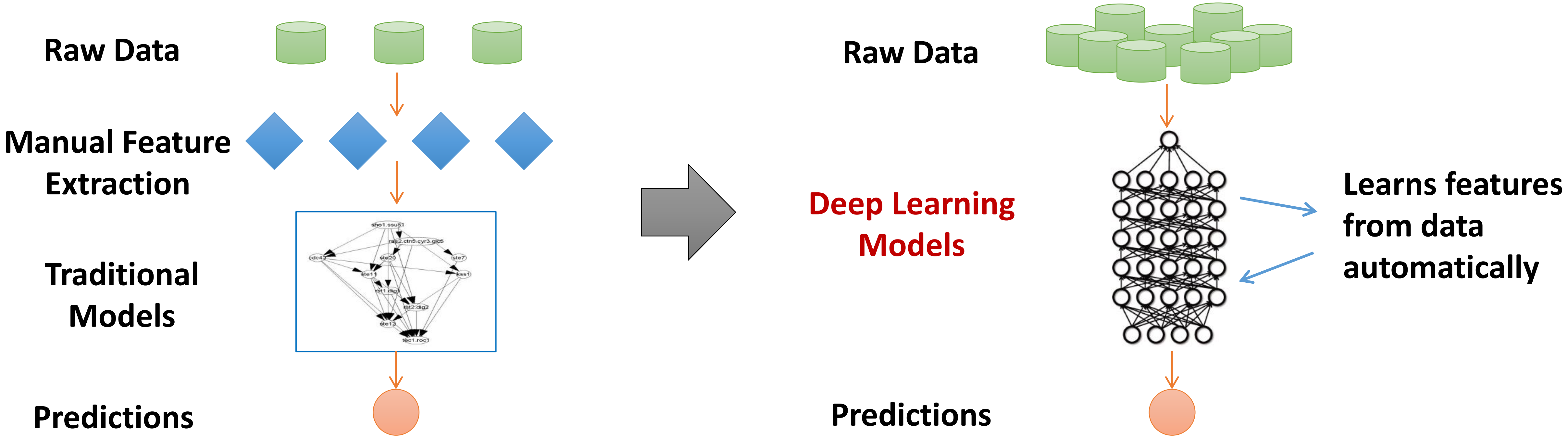
Winners in the Big Data era would be those who:

- Have as much as possible data
- Have a creative mind of what the data can do (the value of the data)
- Know how to extract knowledge and gain insight from data
- Have the computation resources and know how to process huge amounts of data



Deep Neural Networks (Deep Learning)

Deep Learning:
A Frontier of Today's Artificial Intelligence



Raw Data

Manual Feature Extraction

Traditional Models

Predictions

Raw Data

Deep Learning Models

Learns features from data automatically

Predictions

Amazing accuracy (> 85%) for image recognition; yet many works remain to be done to apply it for business data.

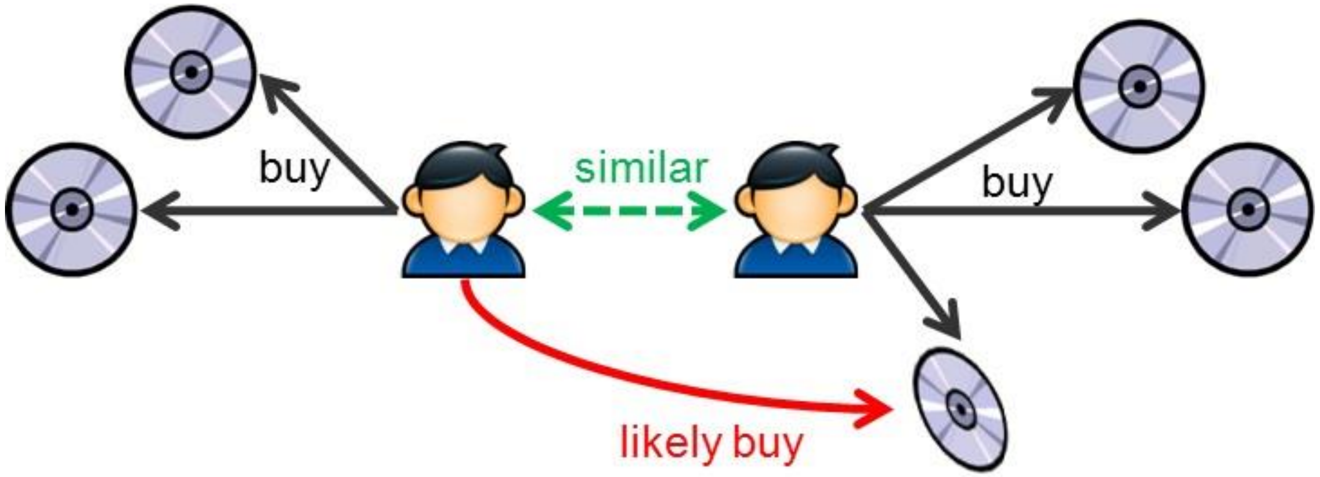
Alibaba Big Data Competition Revisited

Models are no secret

- They are all textbook knowledge.
- Alibaba even let the players know what models they use.

The point is how to well use them

- How similar is similar?
- How many similar users/brands to use?
- How to tune the parameters?
- …



Rocket science?
No! It's all art!

What Companies Should Do

To collect more data

- "Data will become more valuable than you thought when you collect it."

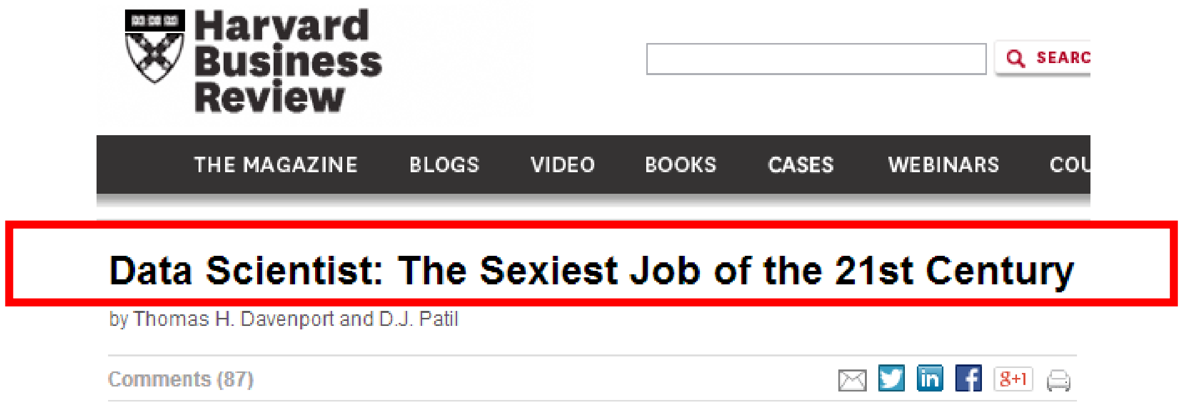
– Dr. Jian Wang, CTO of Alibaba Corp.

To profit from data

- Treating big data as profitable assets
- Figuring out how data will help the business

To hire experienced data scientists

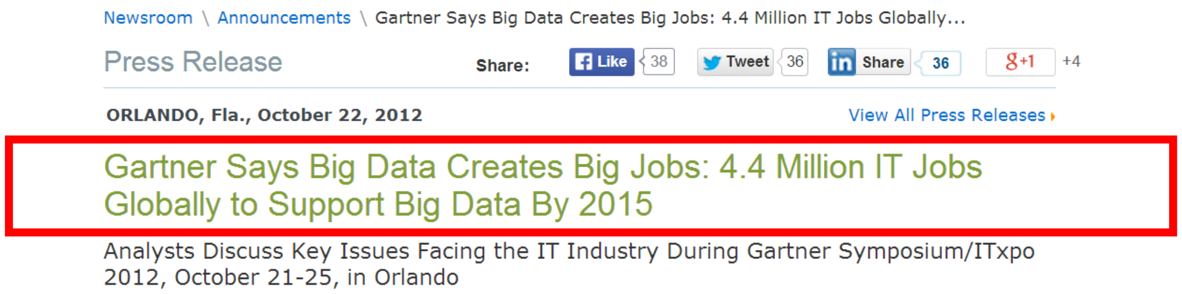
- Data science is more of an art than a science.



Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (87)



Newsroom \ Announcements \ Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally...

Press Release

ORLANDO, Fla., October 22, 2012

View All Press Releases

Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015

Analysts Discuss Key Issues Facing the IT Industry During Gartner Symposium/ITxpo 2012, October 21-25, in Orlando

What Government Should Do

Build data center

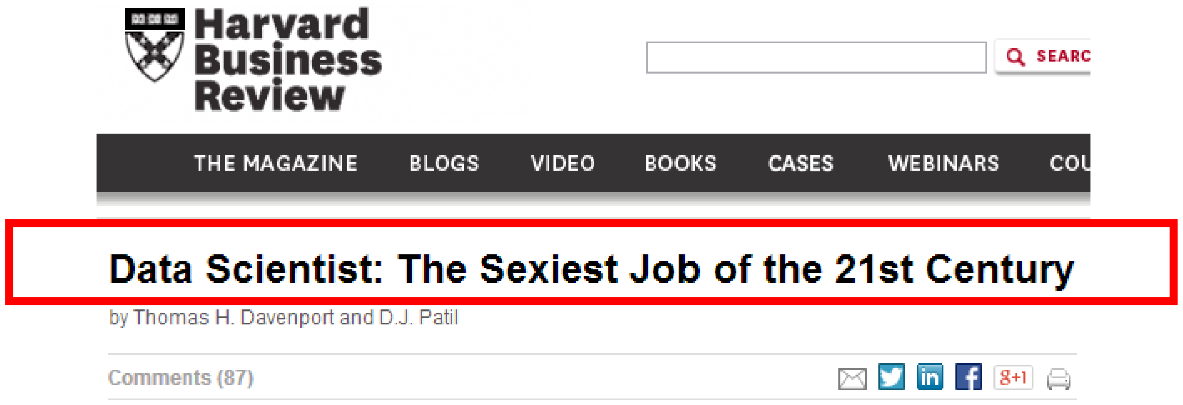
- Not affordable to SMEs.

Collect more data

- Collect and publish public data.

Train more data scientists

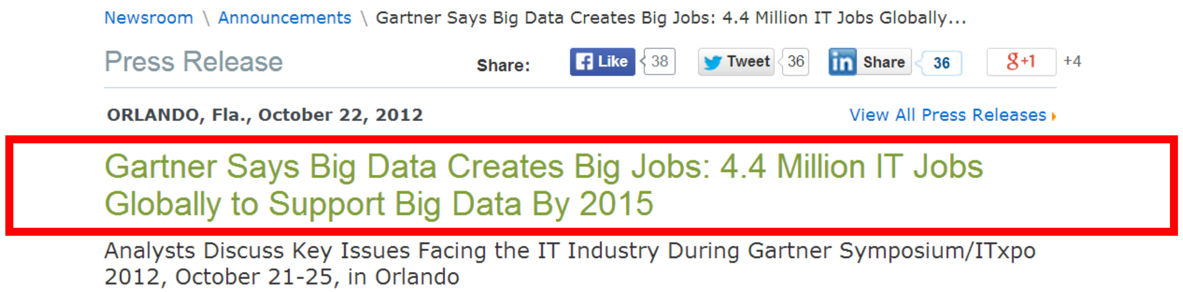
- Preferential policies for Big Data initiatives
- More research funding in Big Data



Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (87)


Newsroom \ Announcements \ Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally...
Share:
Like 38
Tweet 36
Share 36
g+1 +4
View All Press Releases

Press Release

ORLANDO, Fla., October 22, 2012

Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015

Analysts Discuss Key Issues Facing the IT Industry During Gartner Symposium/ITxpo 2012, October 21-25, in Orlando

Big Data is a New Type of Infrastructure

Flow of people
(e.g., octopus data)

Taxi and bus GPS data

Weather and
environment data

Mobile phone
base station data

Collecting and providing data as a public service

City data centers

Taking care of privacy issues

Business

Research

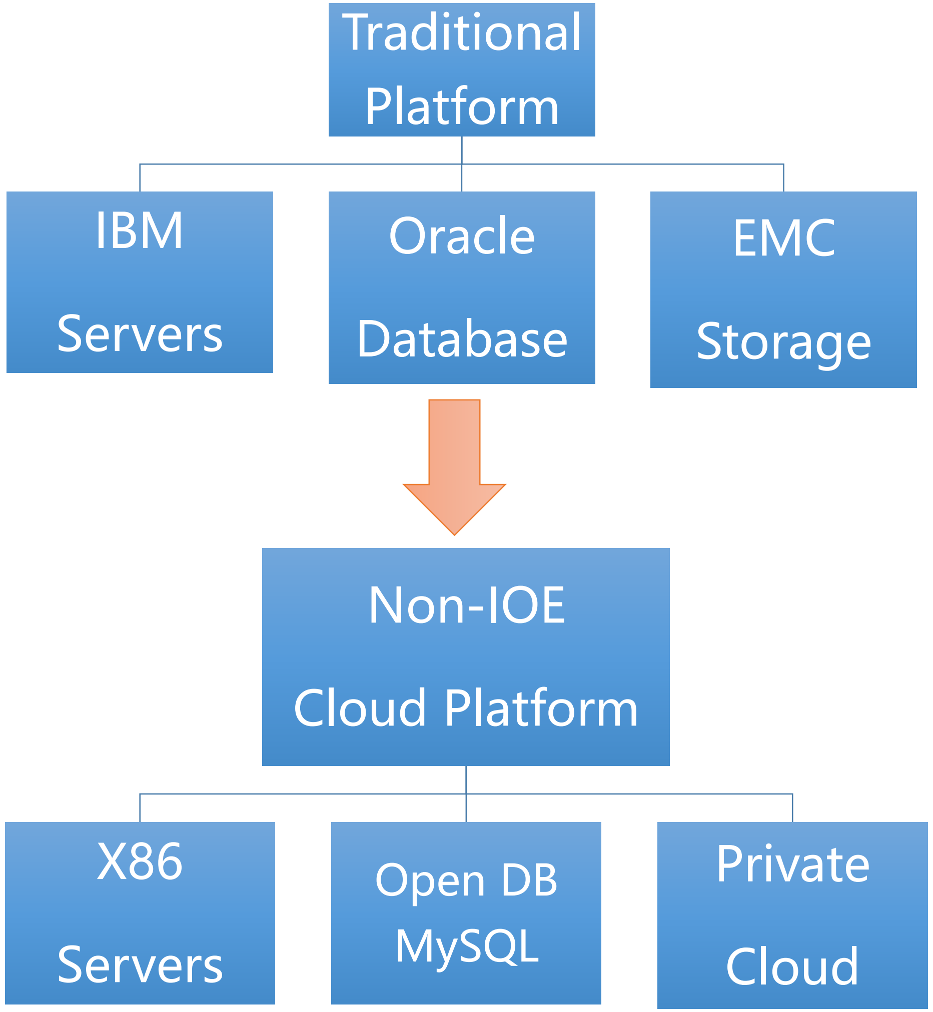
Public administration

Personal use

Traditional Enterprise Data Management: IOE

IBM servers, Oracle database, EMC storage systems.
IOE have dominated the hardware and software of enterprises for data management, especially those big enterprises, financial companies in China.

Road to Non-IOE Era



Traditional Platform

IBM Servers

Oracle Database

EMC Storage

Non-IOE Cloud Platform

X86 Servers

Open DB MySQL

Private Cloud

Data-intensive vs. Computation-intensive

In the past ten years, Big Data is mainly about data-intensive processing.

Easy tasks, e.g., simple queries, building inverted index

Petabytes of (text) data

Thousands of commodity (cheap) servers with lots of hard disks

Example: Google’s cluster for web search

More than 10,000 low-end servers

In the next ten years, Big Data may be more about computation-intensive processing.

Difficult tasks, e.g., training complex machine learning models

Relatively small amount of (feature) data extracted from raw data

Hundreds of powerful servers with lots of processing units/cores

Example: IBM Watson

Less than 100 high-end servers

Be Aware of the Traps and Pitfalls of Big Data

- Data are always biased, regardless of its size.
- In most cases, big data supports decision making; yet it is unwise or even dangerous to let data alone make decisions for us (human beings).
  - Correlation rather than causality
- Value big data, do not ignore “small data”.

Big Data: A young man, a conqueror, and a reaper, who still needs practicing, developing, receiving and understanding

- New technologies do bring other concerns, such as privacy. We should face them and solve them, other than escape from them.
- Technologies will move the world forward, and there is no way back.
- The history has told us that those companies who overlook the technology advances will be out of business, faster than you thought.

Do You Still Remember Kodak?

Big Data Will Change Business

In the near future:

**Big Data Should Be and Could Be
A Core Competency of Your Enterprise**

Thank You !