



BIG DATA: Gain Value with SAP HANA

邬学宁

首席专家 , SAP APJ

June 7, 2012

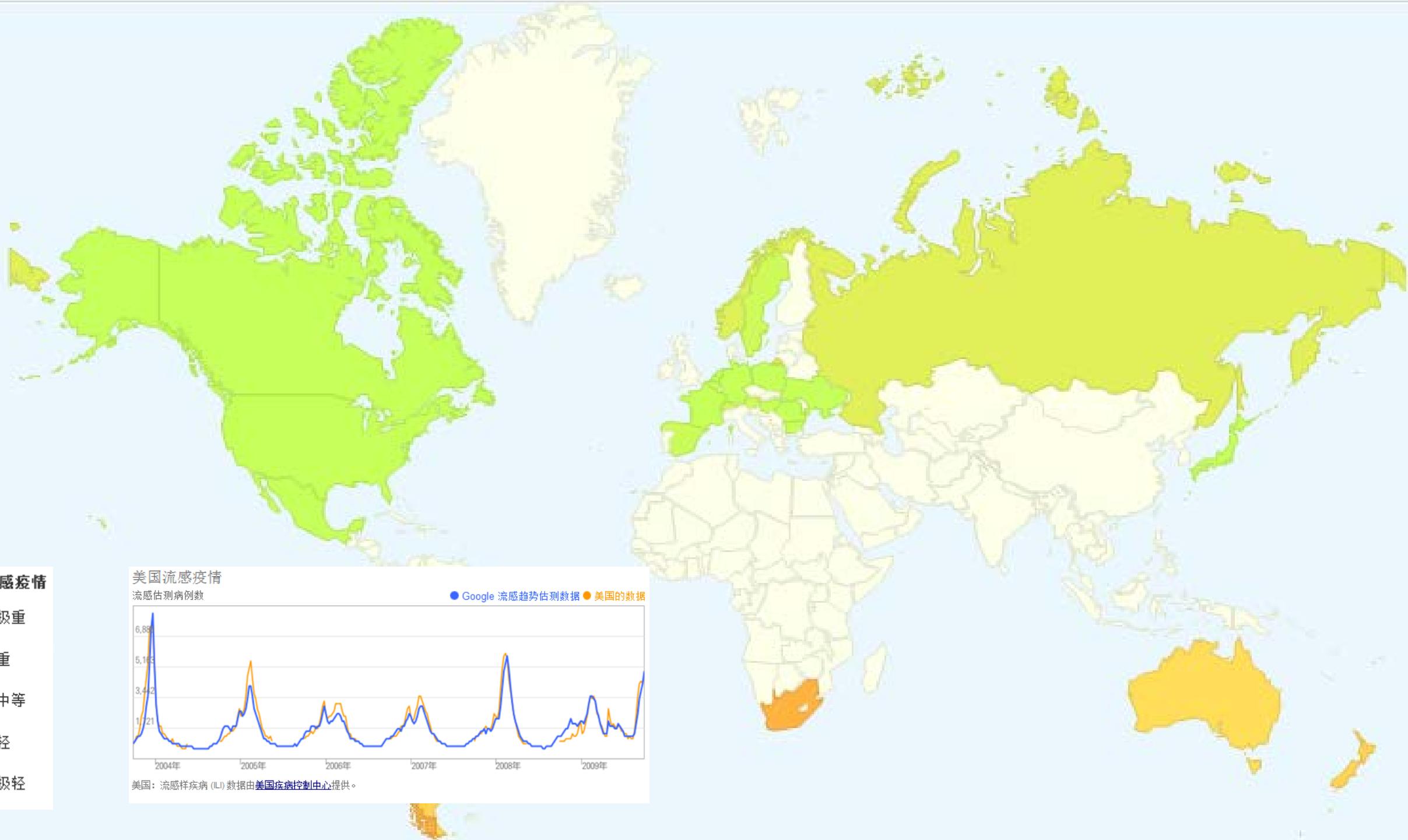
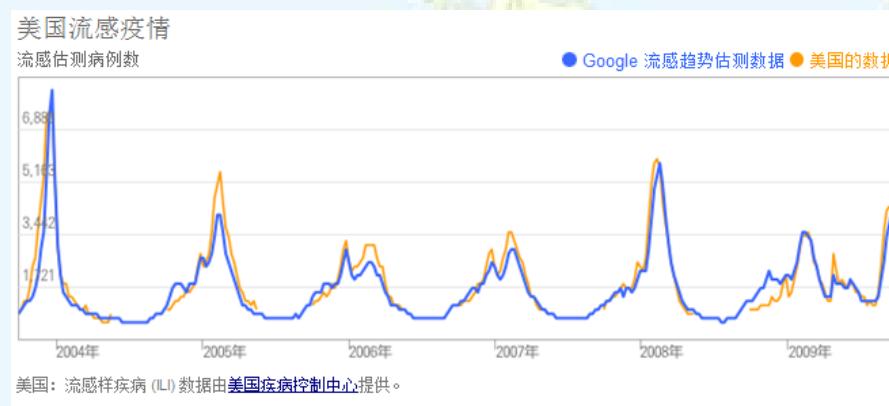


®



流感疫情

极重
重
中等
轻
极轻





flights to new york



TRAVEL FLIGHTS HOTELS

玩紐約省錢推薦-滙客來 - www.taiwanesehouse.url.tw

广告

背包客好評推薦！平價優質住宿環境，位交通便利皇后區，另有好玩行程可代客報名！

San Francisco, CA (SFO) 至 New York, NY (JFK)

週五, 06/14 - 週五, 06/21 · 1 成人 · 经济舱 · [更改搜索](#)

↑ 提示: Buy · 票价上涨 \$50+ · 80%+ 可信度

评级 \$548 ————— \$915

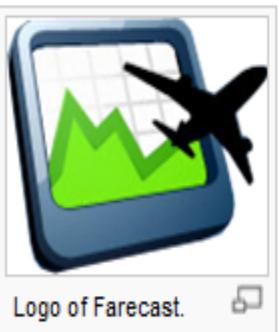
经停 直飞 \$666 停留 1 次 \$548 至少停留 2 次 \$716

[更多筛选器 >](#)

Los Angeles, CA (LAX) 至 New York, NY (JFK)

週五, 06/14 - 週五, 06/21 · 1 成人 · 经济舱 · [更改搜索](#)

↓ 提示: Wait · 票价下降 \$50+ · 78% 可信度



Forecast was founded four years before its launch in 2003; according to Forecast it has collected over 175 billion airfare observations as of 2007.^[2] Forecast's team of data miners use these airfare observations to build algorithms to predict future airfare price movements.

In April 2008, Forecast was bought by Microsoft at a price believed to be around \$75 million to 115 million.^[3] Microsoft officially integrated it as part of its Live Search group of tools in May 2008. On June 3, 2009, Microsoft officially rebranded Live Search Forecast as *Bing Travel* as part of its efforts to create a new search identity.

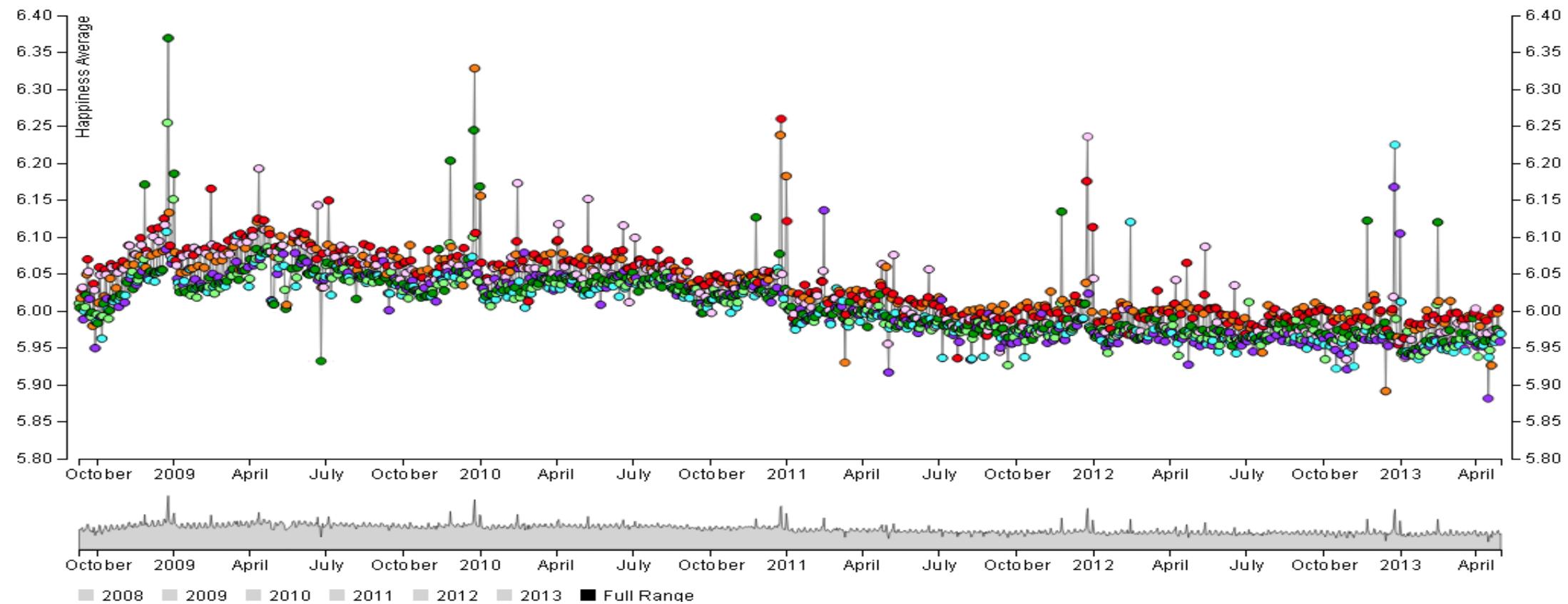


世界快乐指数

Daily Happiness Averages for Twitter, September 2008 to present

Hover over circles for more information

● Sun ● Mon ● Tue ● Wed ● Thu ● Fri ● Sat



“To possess facts is knowledge, to use is wisdom, to choose is education.

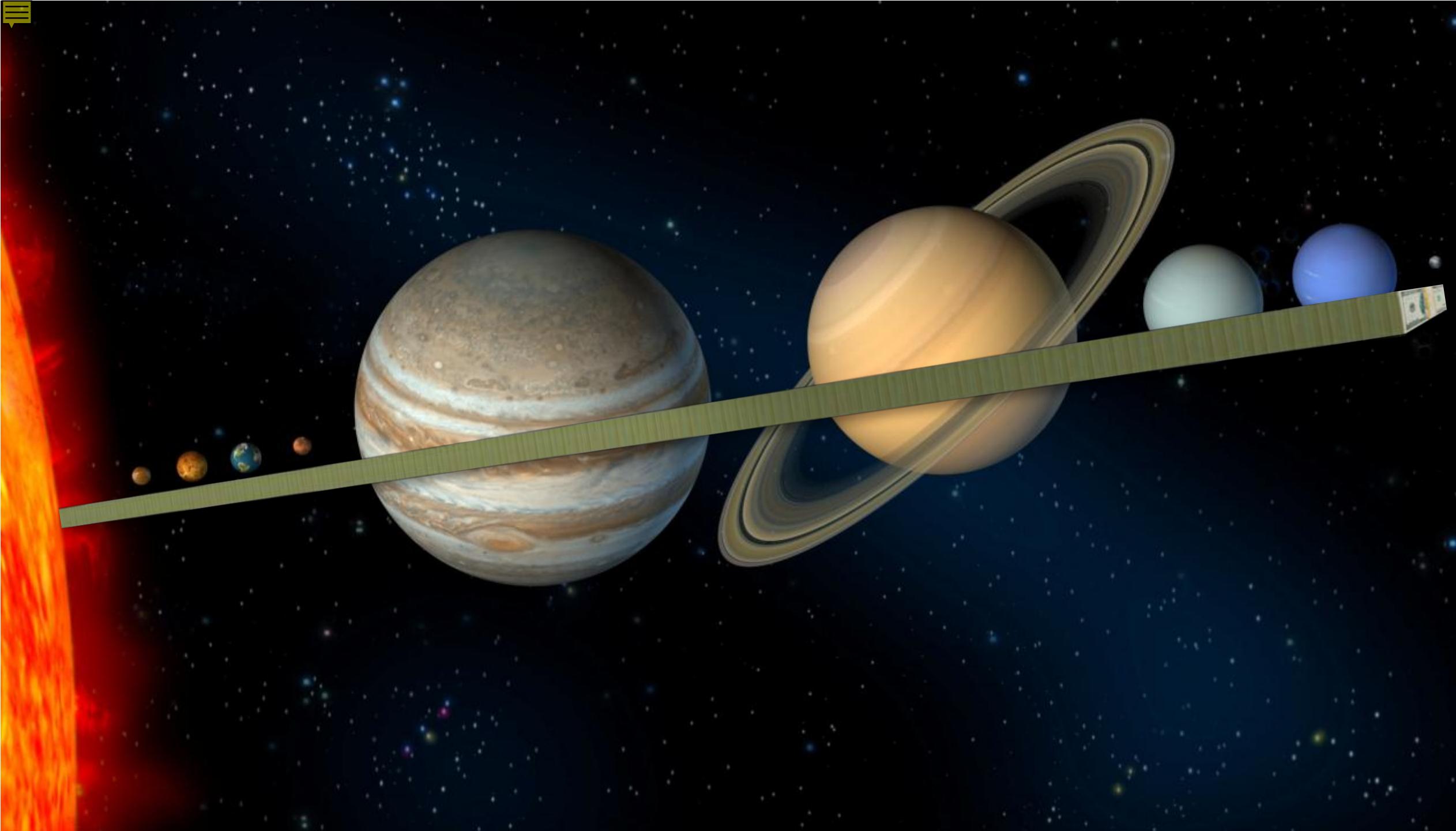
Knowledge is not power but riches, only have value when spending”

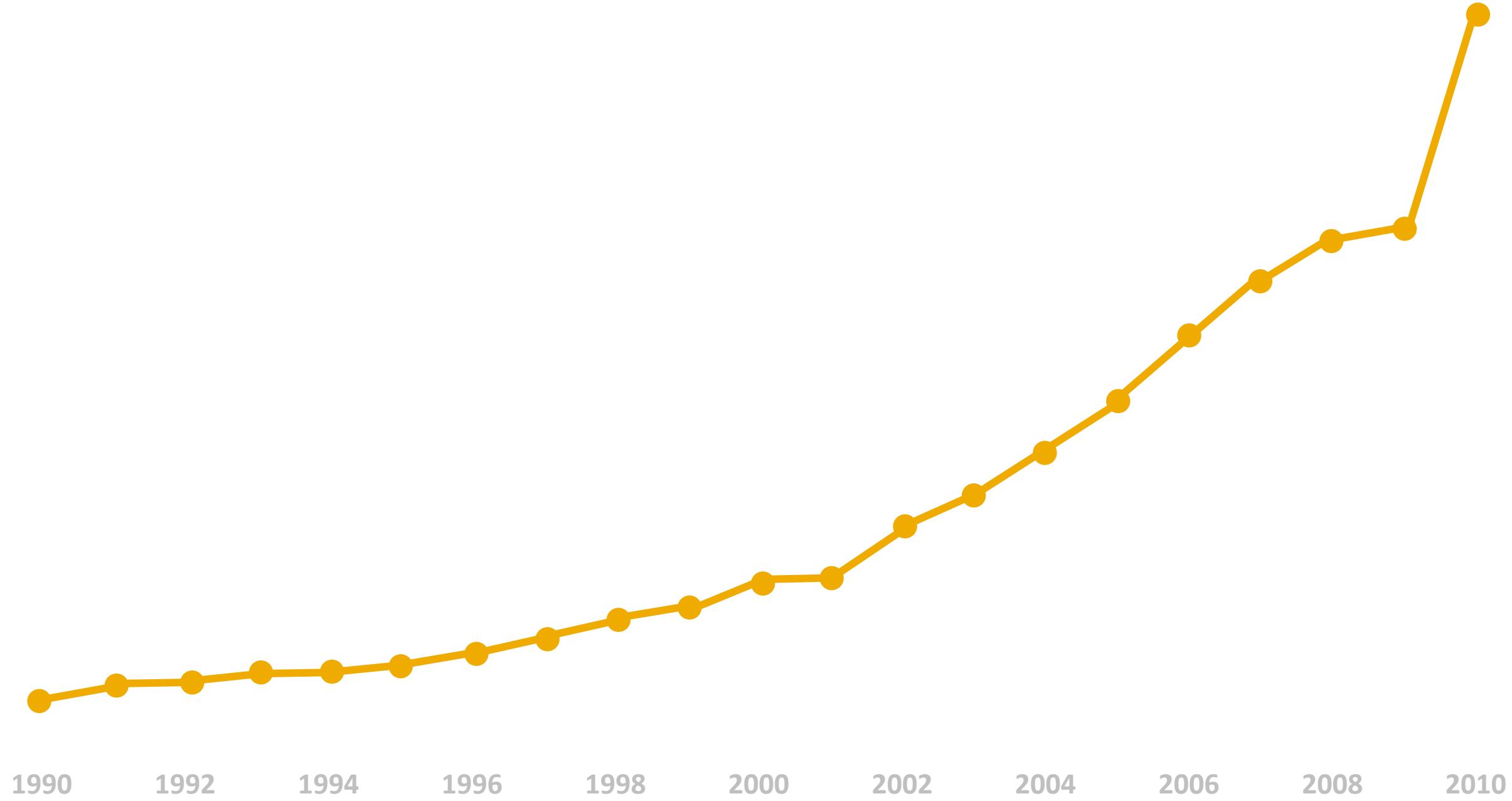
- Thomas Jefferson



2.5 PB

7.9zb

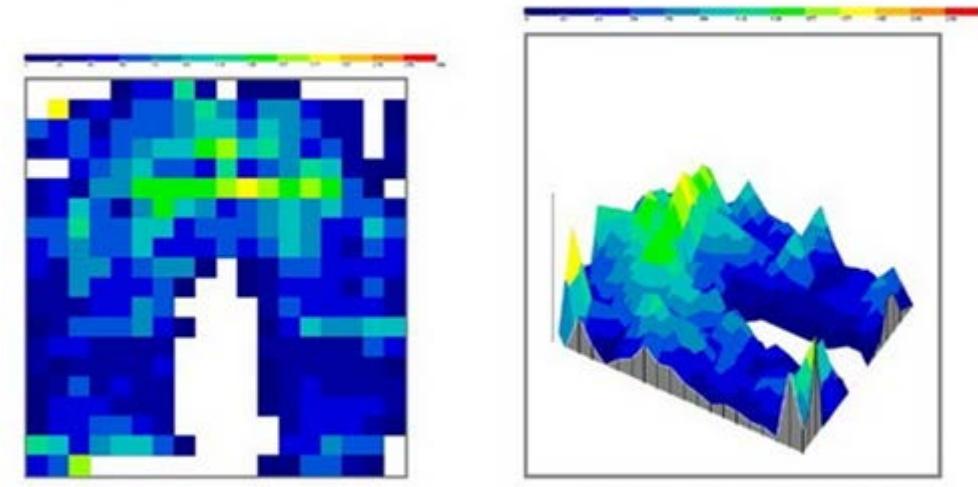




臀部大数据：360个传感器汽车防盗，防止疲劳驾驶！



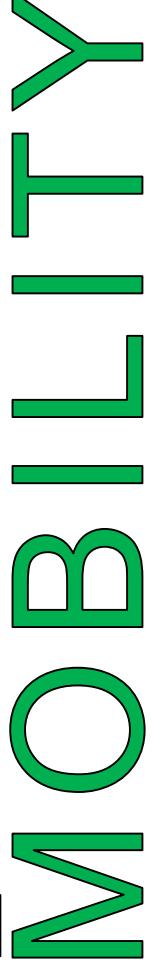
運転座席での座面圧力データ



39個の特微量を抽出し、マハラノビス距離を計算する

Source: <http://www.nownews.com/2011/12/29/11622-2771090.htm#ixzz2ViQqGjC2> 哦

BIG DATA VIRTUALIZATION **BYOD** CLOUD **ANALYTICS** **CLOUD** **BPM** SOCIAL **Real-Time** **PREDICTIVE** VIRTUALIZATION

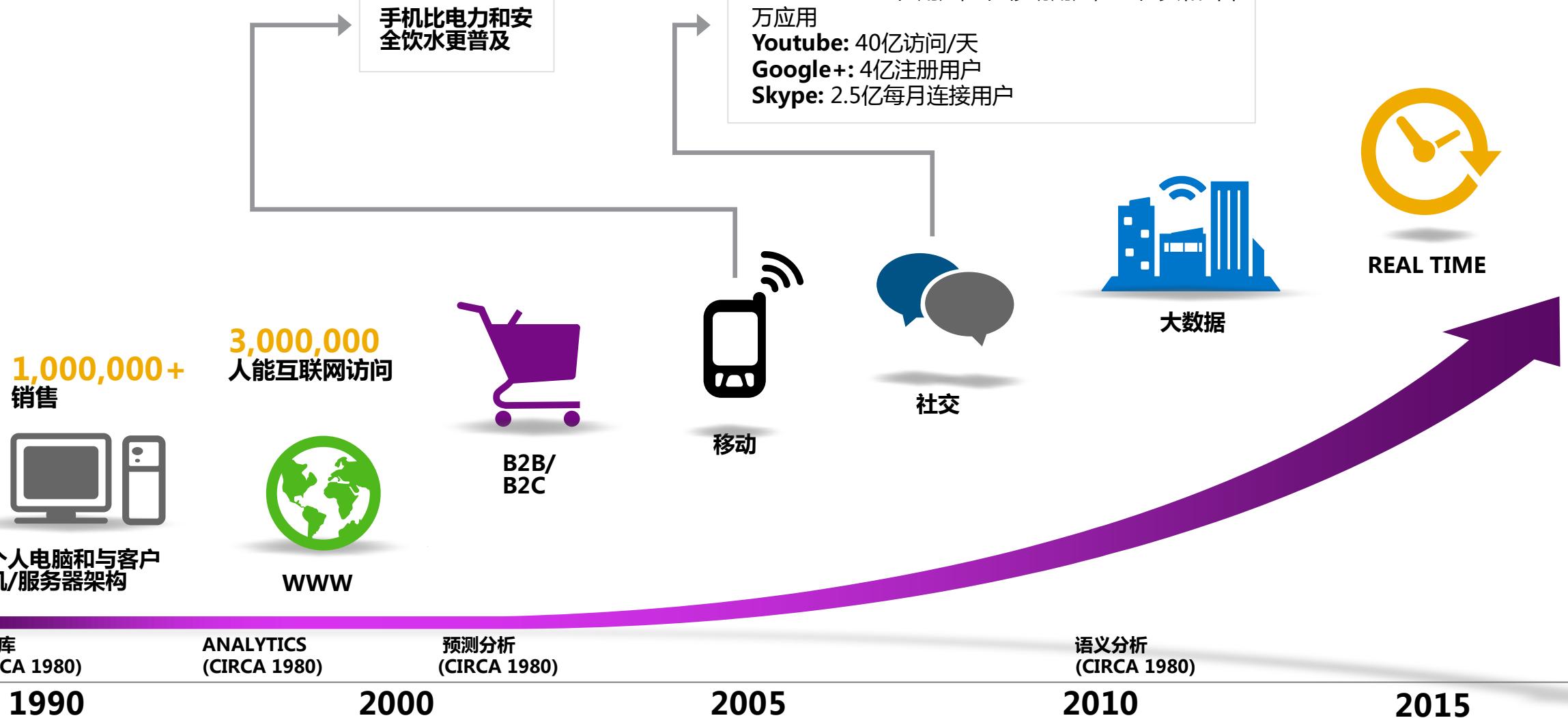




BIG DATA REAL TIME PREDICTIVE



我们如何发展到今天?

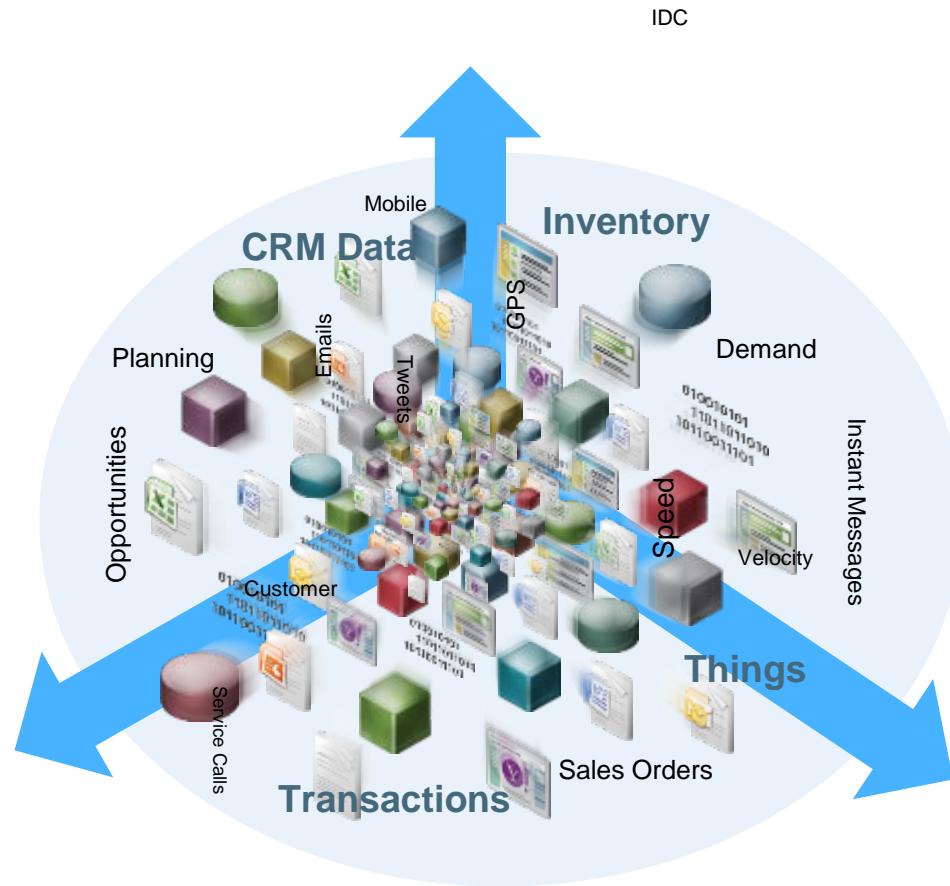




VOLUME

In 2005, humankind created 150 exabytes of information. In 2011, **1,200 exabytes will be created.**

The Economist



VELOCITY

Worldwide digital content will double in 18 months, and every 18 months thereafter.

IDC

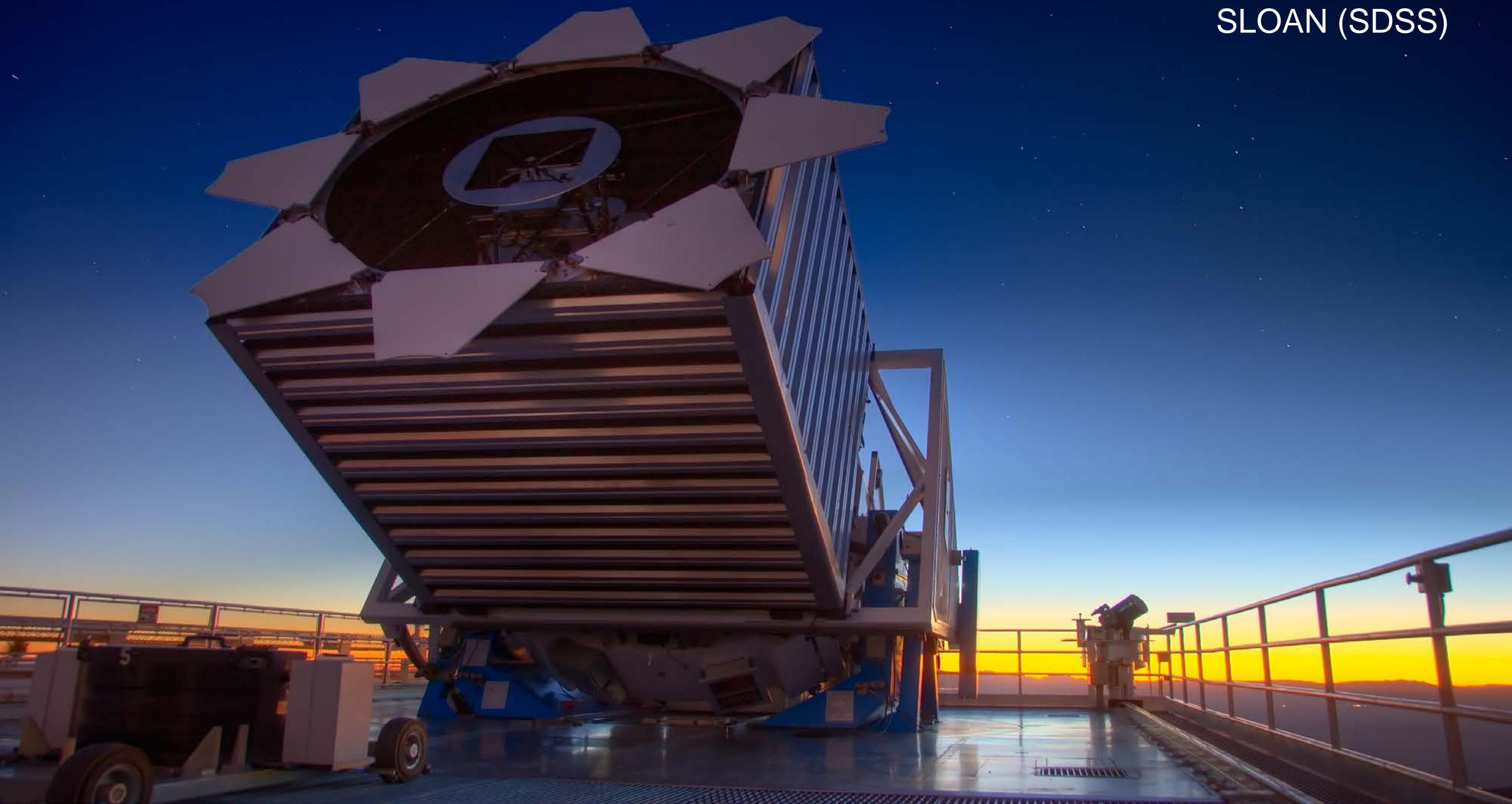
VARIETY

80% of enterprise data will be unstructured, spanning traditional and non traditional sources.

Gartner



SLOAN (SDSS)





Google books Ngram Viewer

Graph these case-sensitive comma-separated phrases: causality,correlation

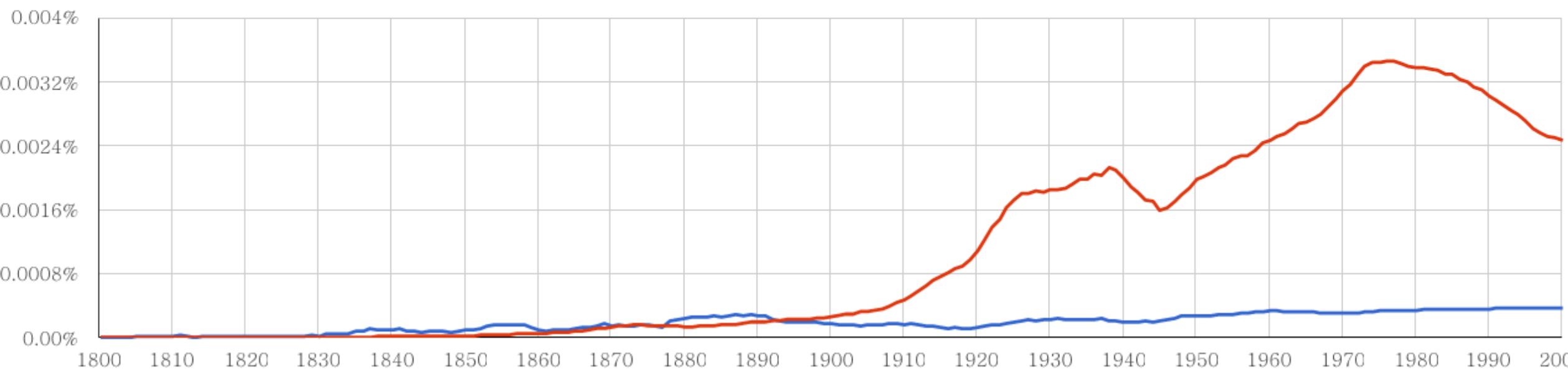


between 1800 and 2000 from the corpus English with smoothing of 3.



Search lots of books

causality correlation



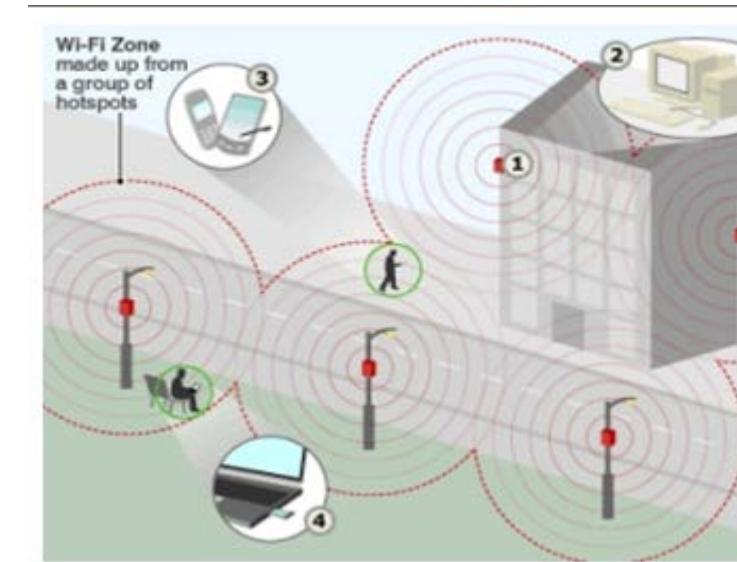
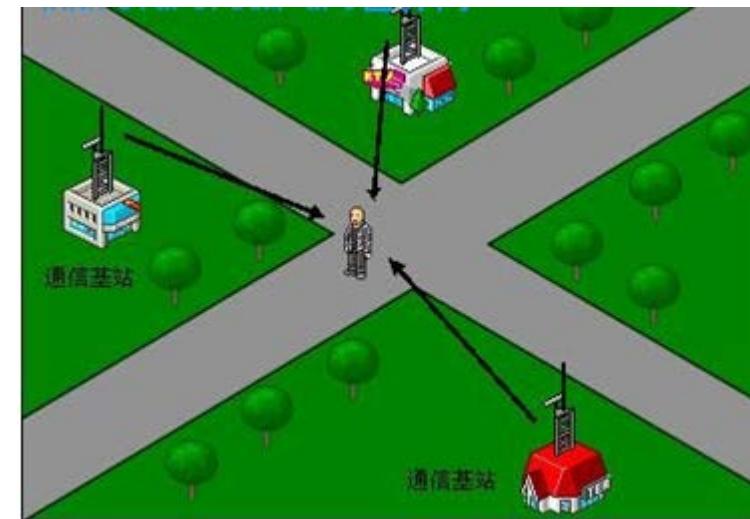
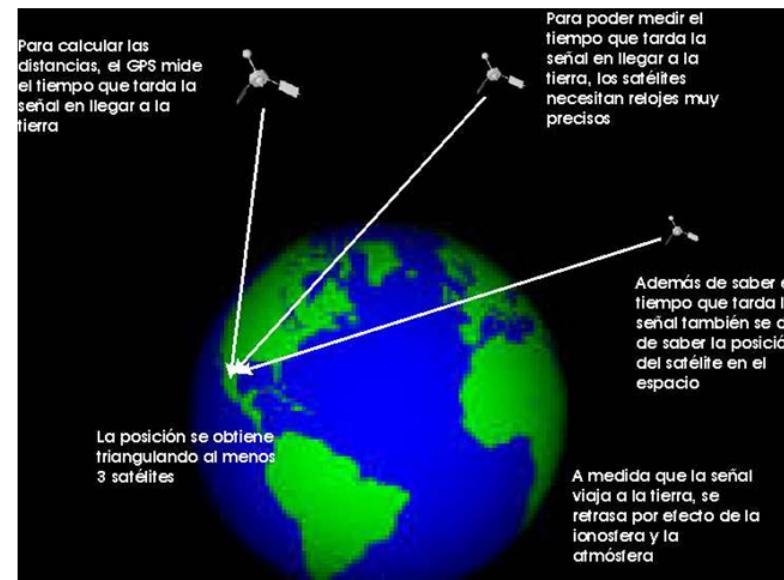


MOBILE



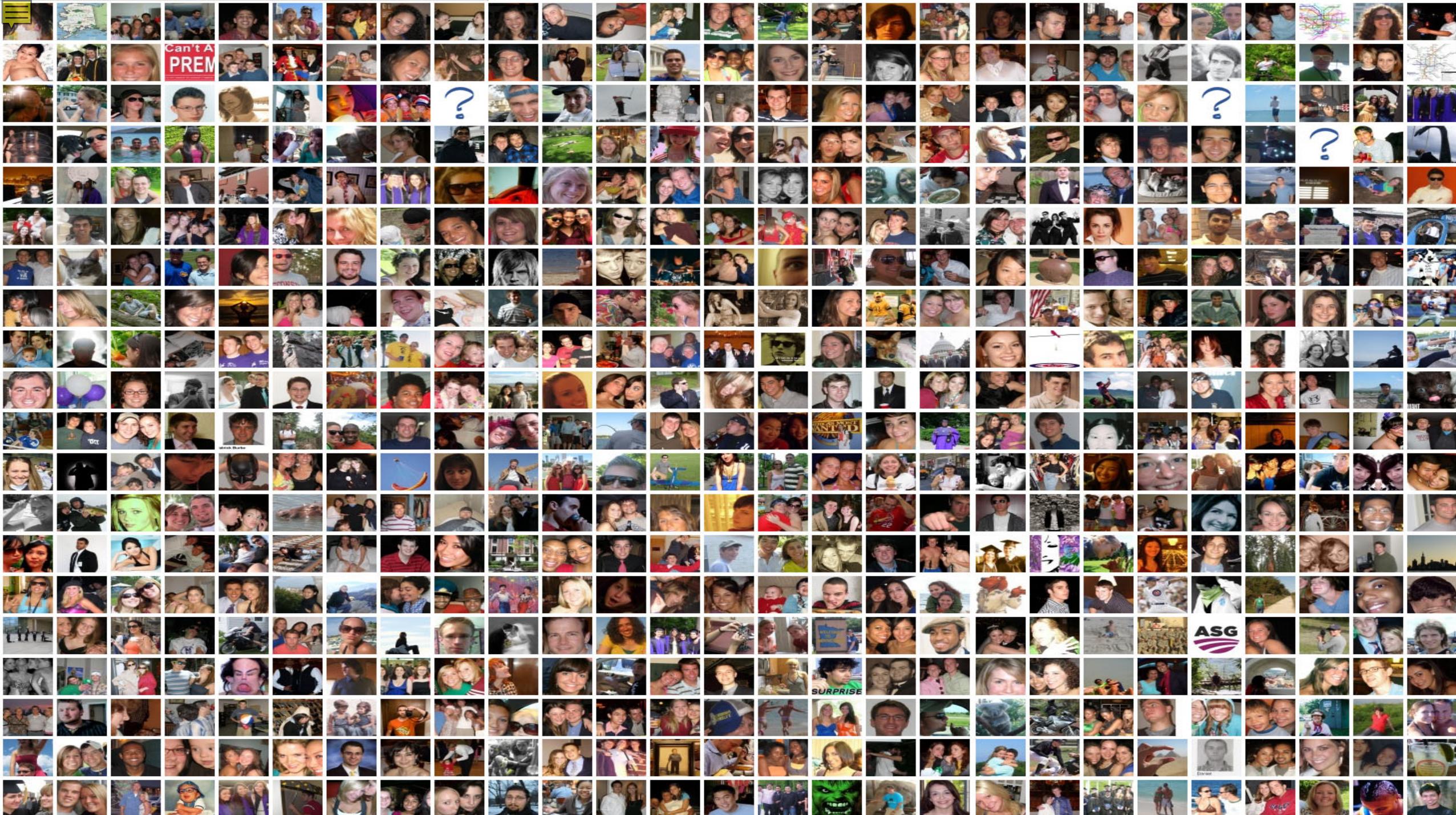


LBS



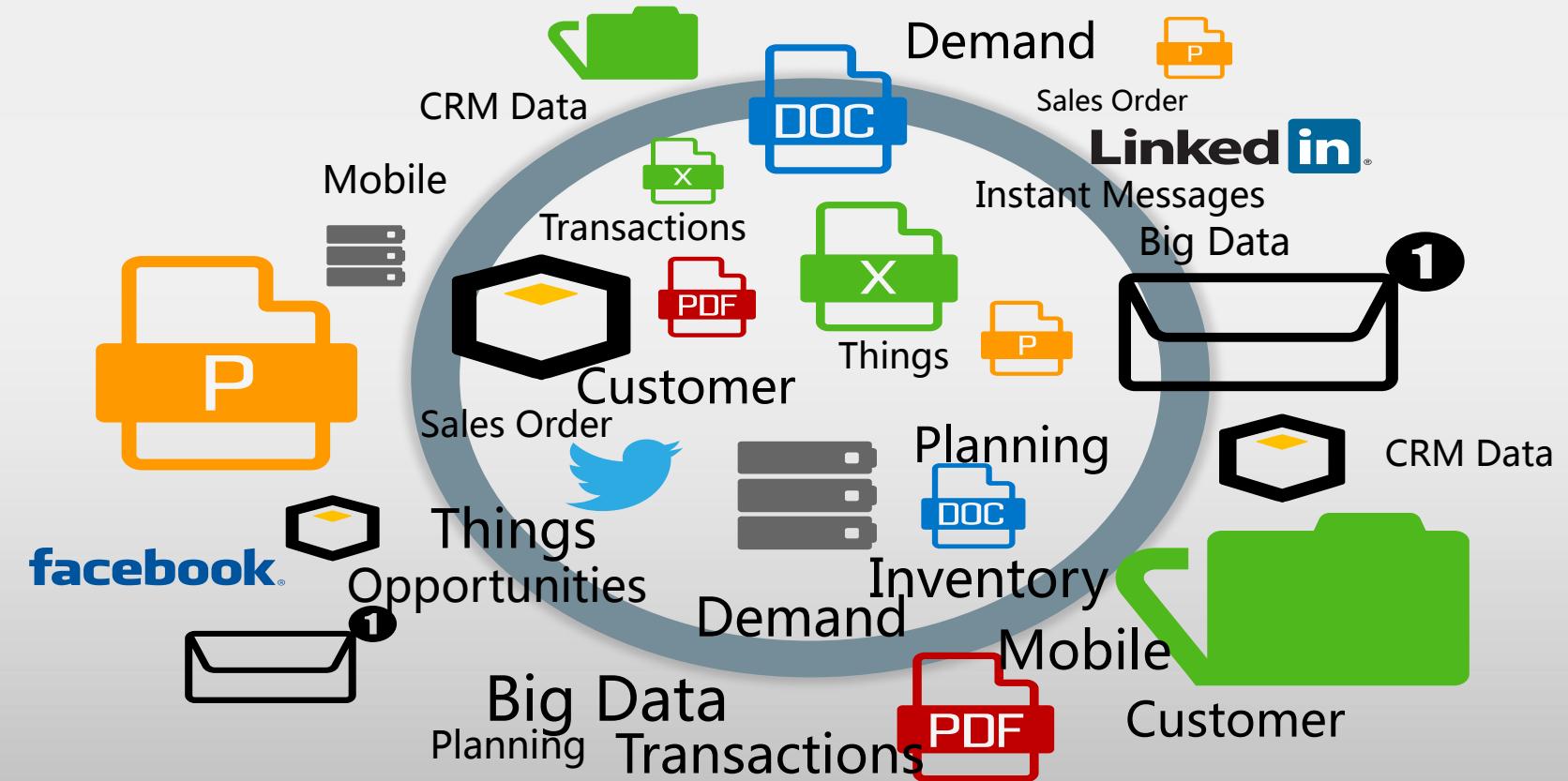
A top-down photograph of four people working together at a white desk. In the center, a man in a pink shirt and a woman in a grey shirt are looking at a black tablet. To the left, a woman in a red shirt is looking up at them. In the bottom right, a person's hands are on a white keyboard. A computer monitor is visible in the top right corner. The desk is cluttered with papers, a smartphone, a pair of glasses, and a telephone.

SOCIAL

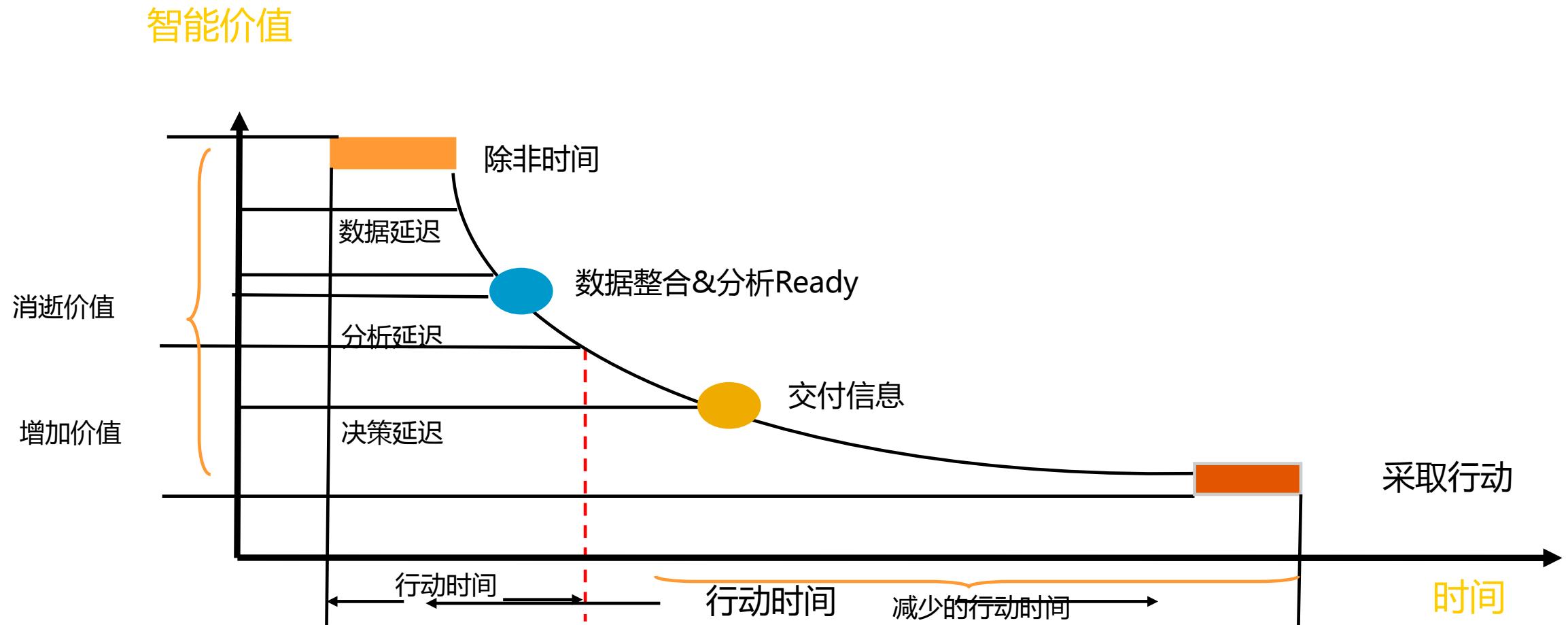


数据的体量和类型

在急剧增加



减少延迟 / 增加价值



来源: Dr. Richard Hackathorn, Bolder Technologies Inc.



实时:巨量数据集





CLOUD

就是现在！处理交易的内存数据库

摩尔定律：

IT 性能价格每18个月翻倍 →
每十年翻倍 6+ 次 →
~每十年大约100倍

• 比屋顶还高





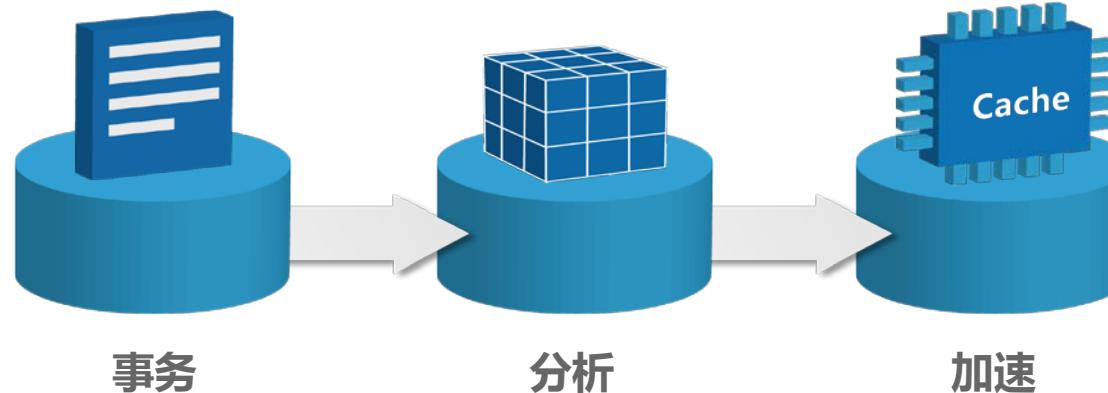
SAP的内存数据管理创新

提供企业应用实时平台

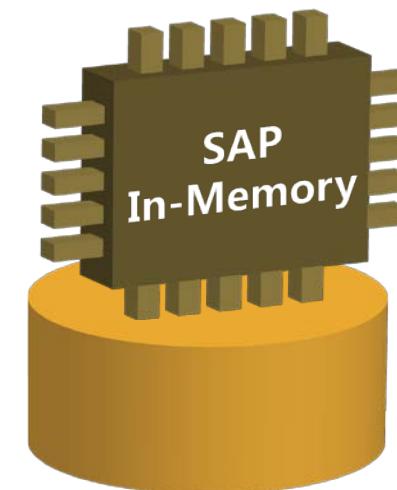


一个使用内存计算的列式数据库整合了OLTP和OLAP

Hasso Plattner



VS

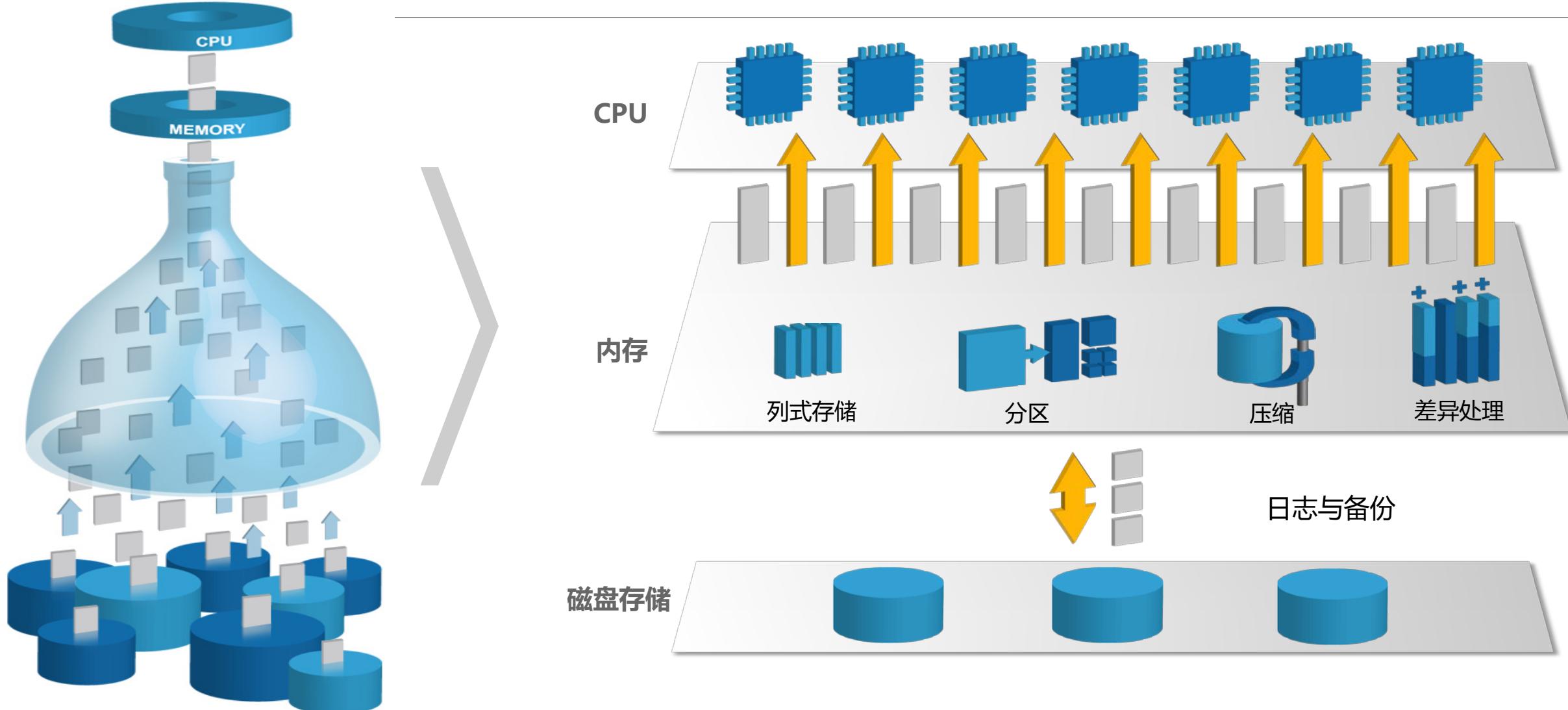


在内存中直接进行

通过创新和专注与客户进行转型

整合事务与分析

内存技术驱动的低延迟计算



IN-MEMORY



FASTER





SMARTER



BIG DATA



ANALYTICS



CLOUD



MOBILE



SOCIAL

HANA REAL-TIME PLATFORM



呈现

Big Data Applications?

处理

Azkaban Oozie Pig Hive

Hadoop MapReduce S4 Storm

存储

Voldemort Cassandra Hbase

摄入

Kafka Flume Scribe



小心！再次落入陷阱！

复杂技术架构. 令人困惑的供应商选择



需要全面“端到端”的方法

呈现

处理

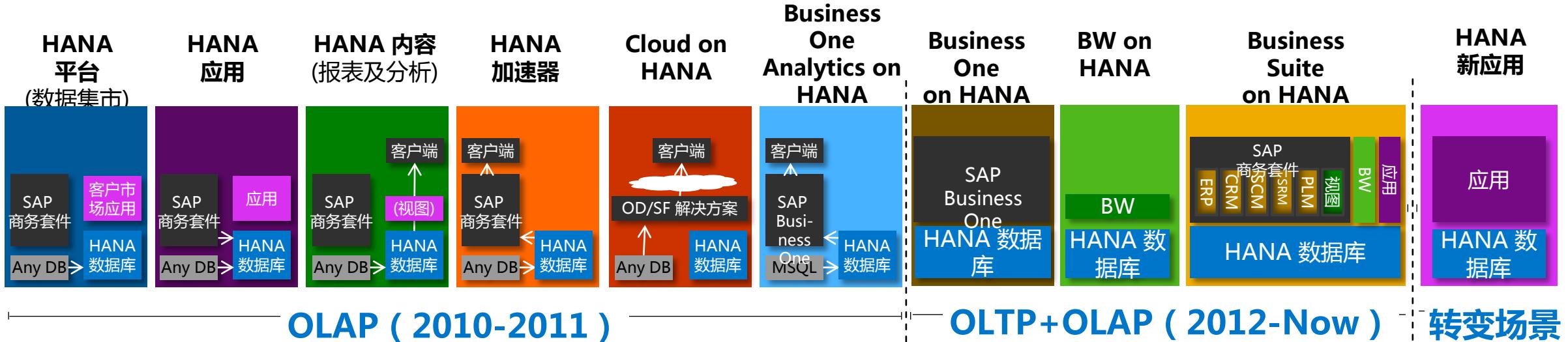
存储

摄入



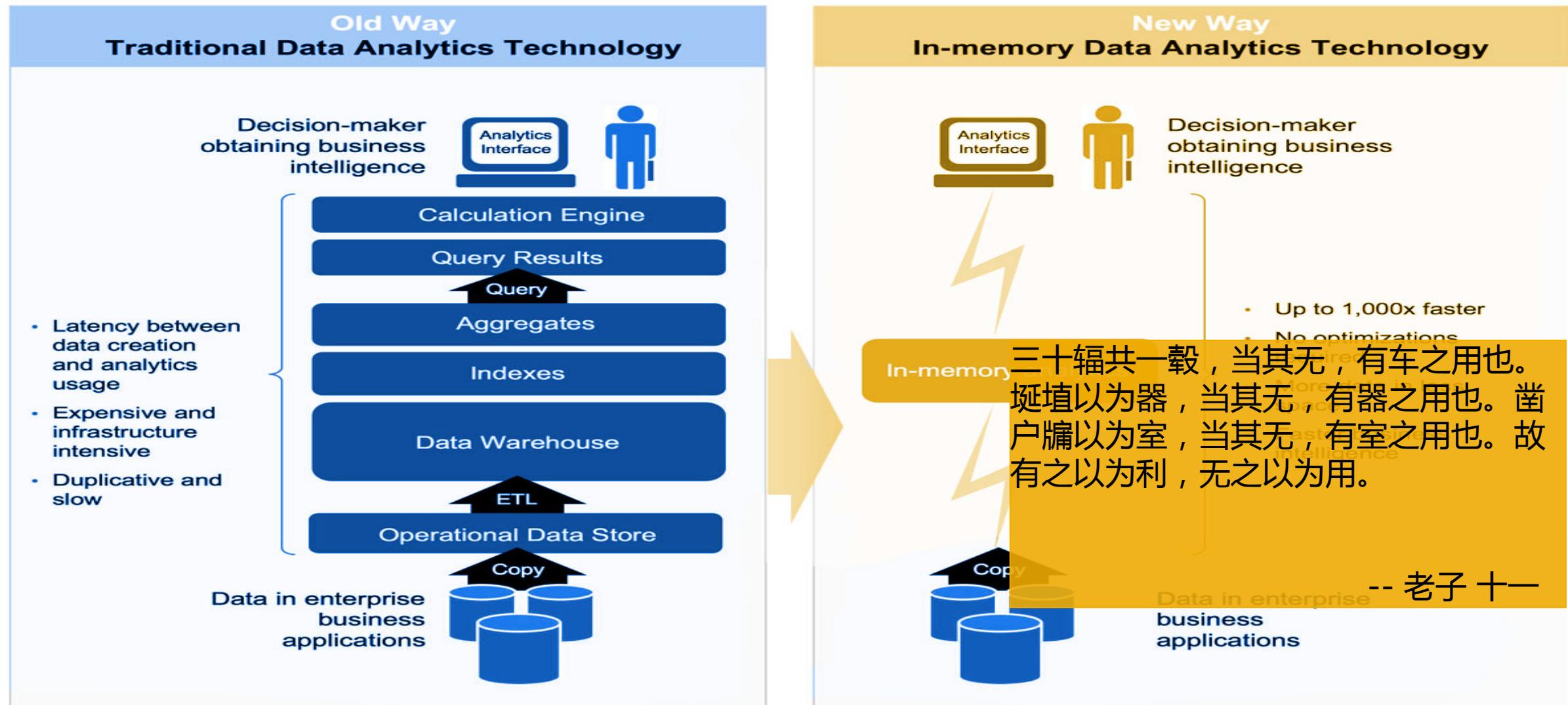
SAP HANA

不断创新



- 从OLAP到OLTP
- 从Side-by-Side到主数据库
- 从BI到预测
- 从结构化到非结构化

HANA – 简化BI与分析



SAP 巨量数据处理框架



呈现

移动

巨量数据分析

巨量数据应用

处理

Sybase ESP

事件与流处理

Sybase ASE

交易处理

SAP HANA

内存技术

Sybase IQ

分析网格

Hadoop

MapReduce
Batch Compute Framework

存储

DB 引擎

DB 引擎

DB 引擎

Hive/HDFS

摄入

Sybase ESP

监控 / 过滤事件流

Sybase Replication Server, SAP BusinessObjects Data Services

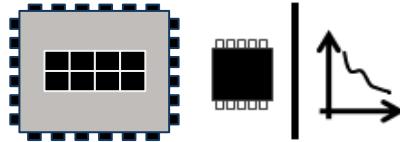
(跨越不同部署选项整合/ 同步数据)

半结构化数据

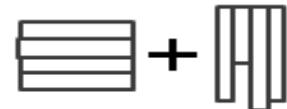
结构化数据

非结构化数据

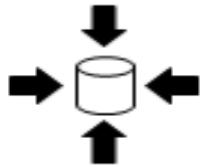
HANA – 软件和硬件的共同创新



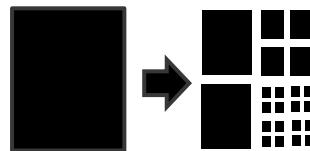
多核CPU
大内存



行与列存储



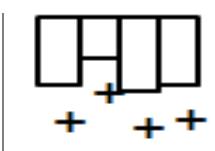
压缩



分区



无汇总表
无含数据的视图



实时复制
只插入差异数据

计算能力：
比硬盘访问快 $1M \times$
1 TB 服务器, 64-80 核

列式存储 = 快速查询

5x 压缩
1 TB 数据, ~ 200GB 内存

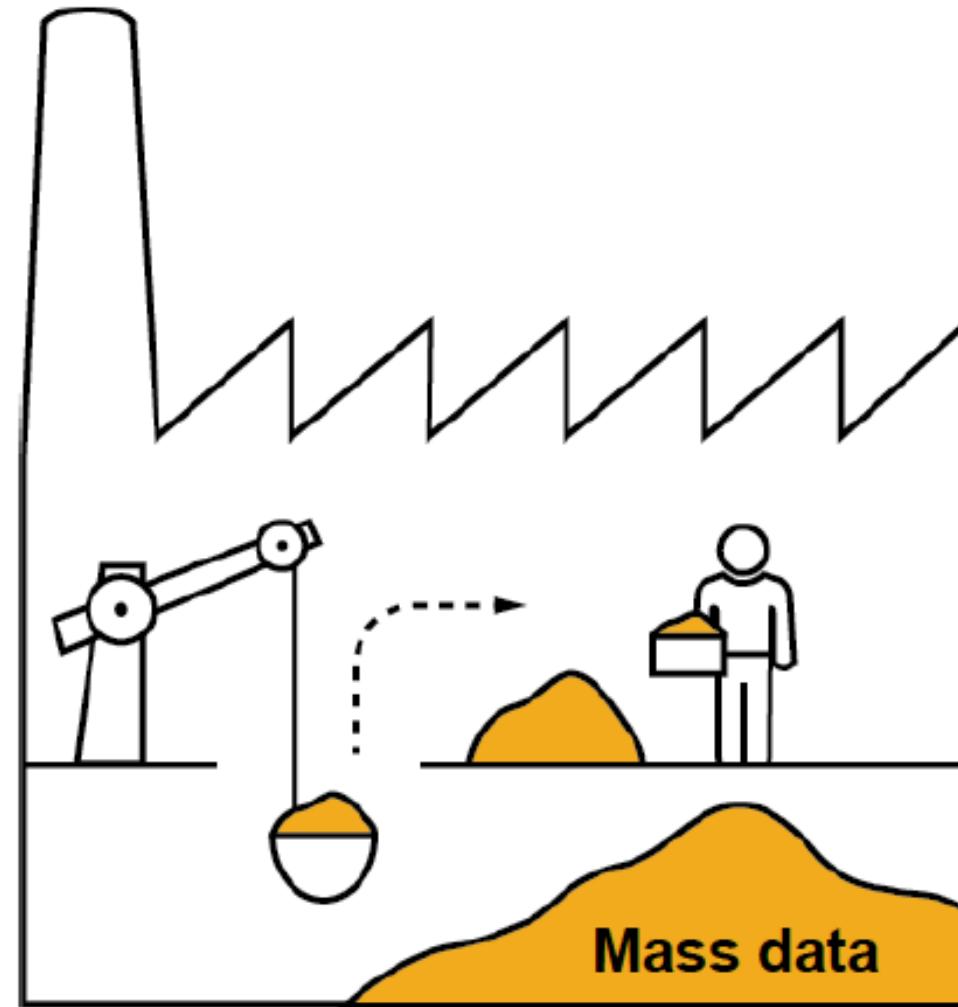
分析巨量数据集
复杂计算

灵活建模
无需数据复制

快速数据 Loading



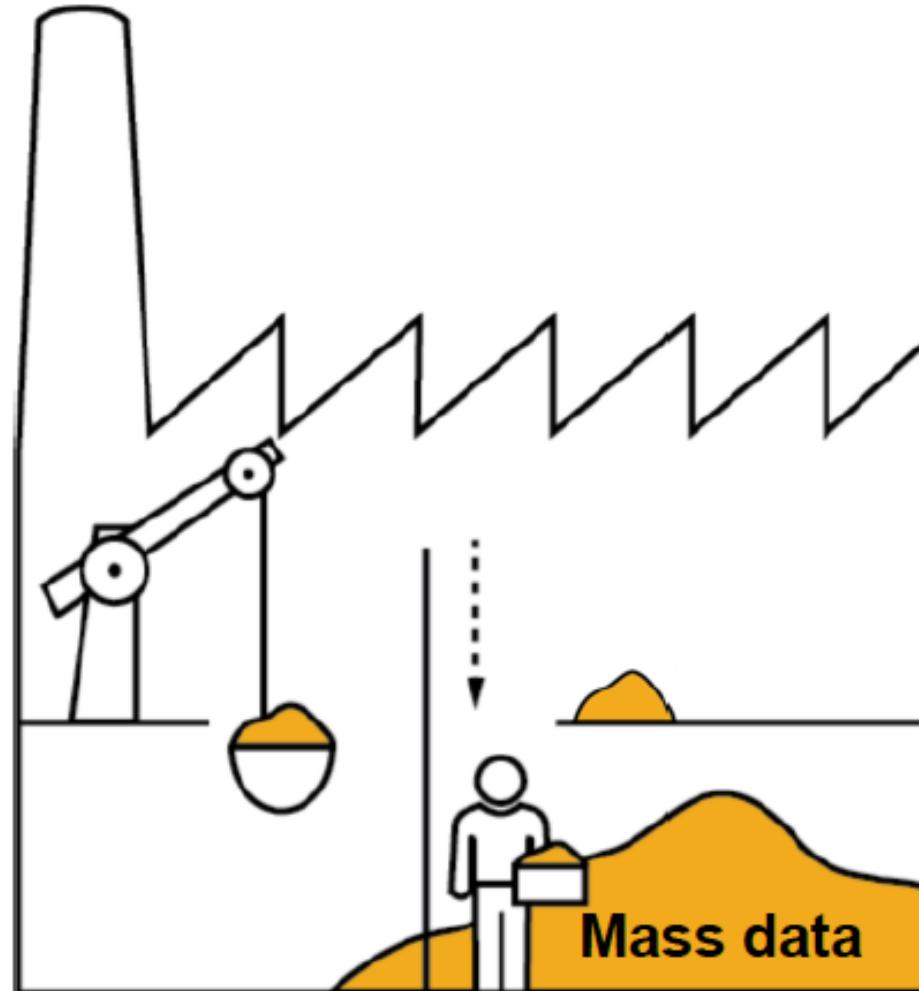
在应用层处理数据



Application

Database

Push Down 业务和代码



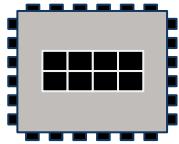
Application

Database

Mass data

内存计算 – 就是现在！

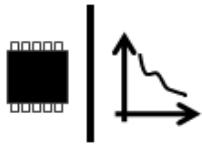
HW Technology Innovations



Multi-Core Architecture (8 x 8core CPU per blade)

Massive parallel scaling with many blades

One blade ~\$50.000 = 1 Enterprise Class Server

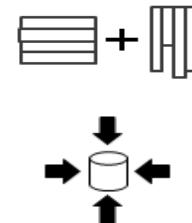


64bit address space – 2TB in current servers

100GB/s data throughput

Dramatic decline in price/performance

SAP SW Technology Innovations

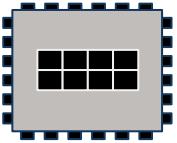


Row and Column Store

Compression

内存计算 – 就是现在！

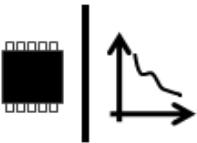
HW Technology Innovations



Multi-Core Architecture (8 x 8core CPU per blade)

Massive parallel scaling with many blades

One blade ~\$50.000 = 1 Enterprise Class Server

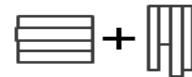


64bit address space – 2TB in current servers

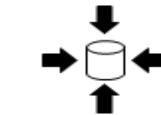
100GB/s data throughput

Dramatic decline in price/performance

SAP SW Technology Innovations



Row and Column Store

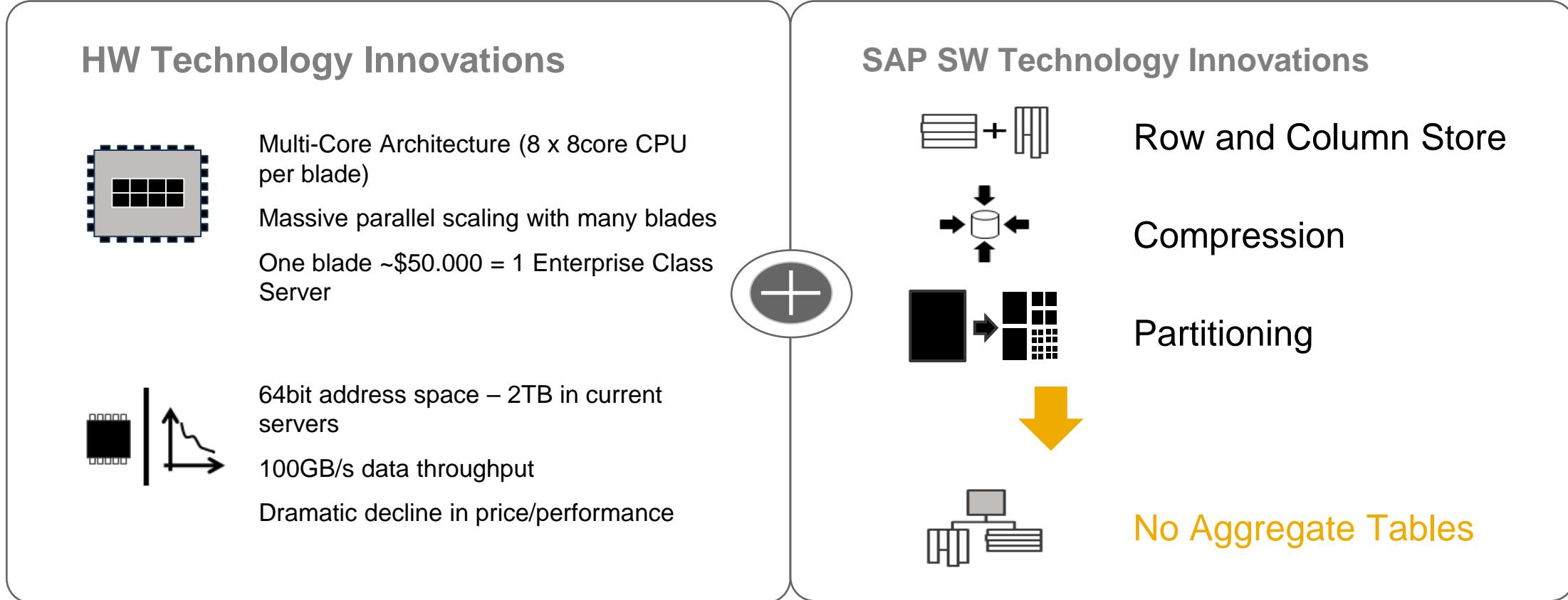


Compression



Partitioning

内存计算 – 就是现在！



快速！为内存计算进行软件优化

概念视图

A	10	€
B	35	\$
C	2	€
D	40	€
E	12	\$

传统数据库按行存储

按列存储使快速内存操作（例如汇总）成为可能

- 列布局支持顺序内存访问
- 简单汇总行扫描一次即可

映射至内存

按行组织

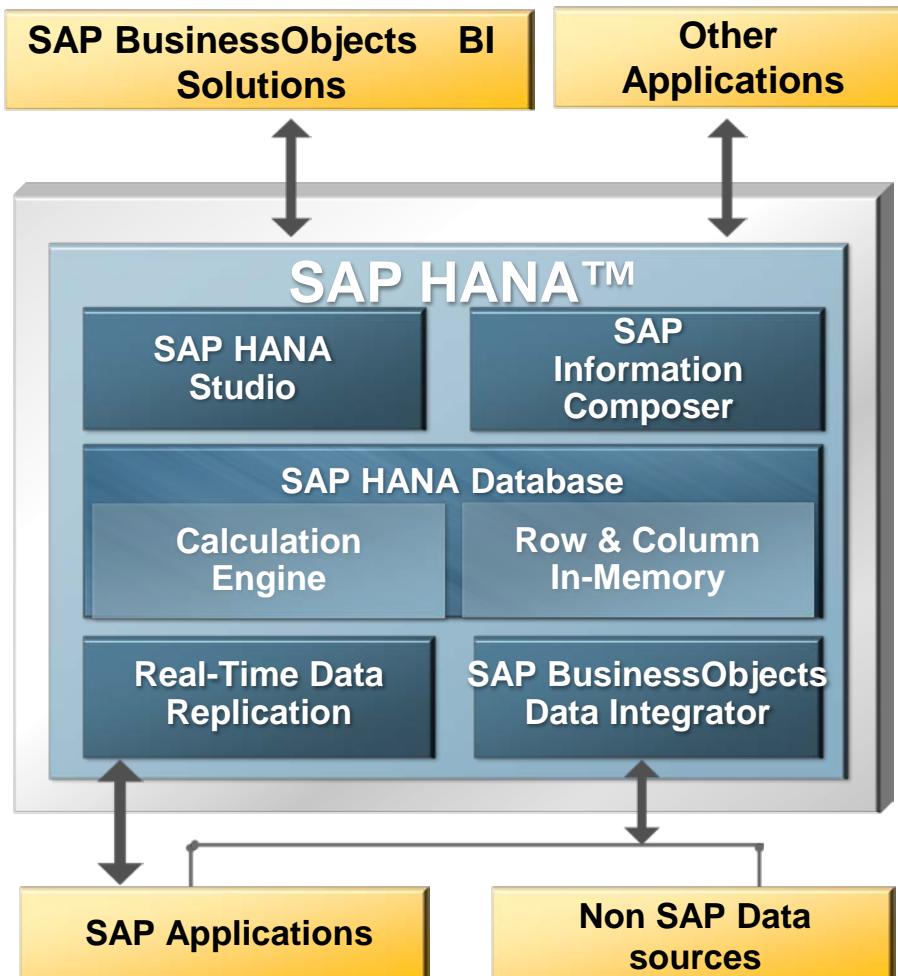


按列组织



memory address

SAP In-Memory Appliance (SAP HANA™)



SAP HANA™

- In-Memory software + hardware (HP, IBM, Fujitsu, Cisco, Dell)
- Data Modeling and Data Management
- Real-time Data Replication
- SAP BusinessObjects Data Services for ETL capabilities from SAP Business Suite, SAP NetWeaver Business Warehouse (SAP NetWeaver BW), and 3rd Party Systems

Capabilities Enabled

- **Analyze information in real-time** at unprecedented speeds on large volumes of non-aggregated data
- **Create flexible analytic models** based on real-time and historic business data
- **Foundation for new category of applications** (e.g., planning, simulation) to significantly outperform current applications in category



个性化癌症治疗

IMITSUI KNOWLEDGE INDUSTRY

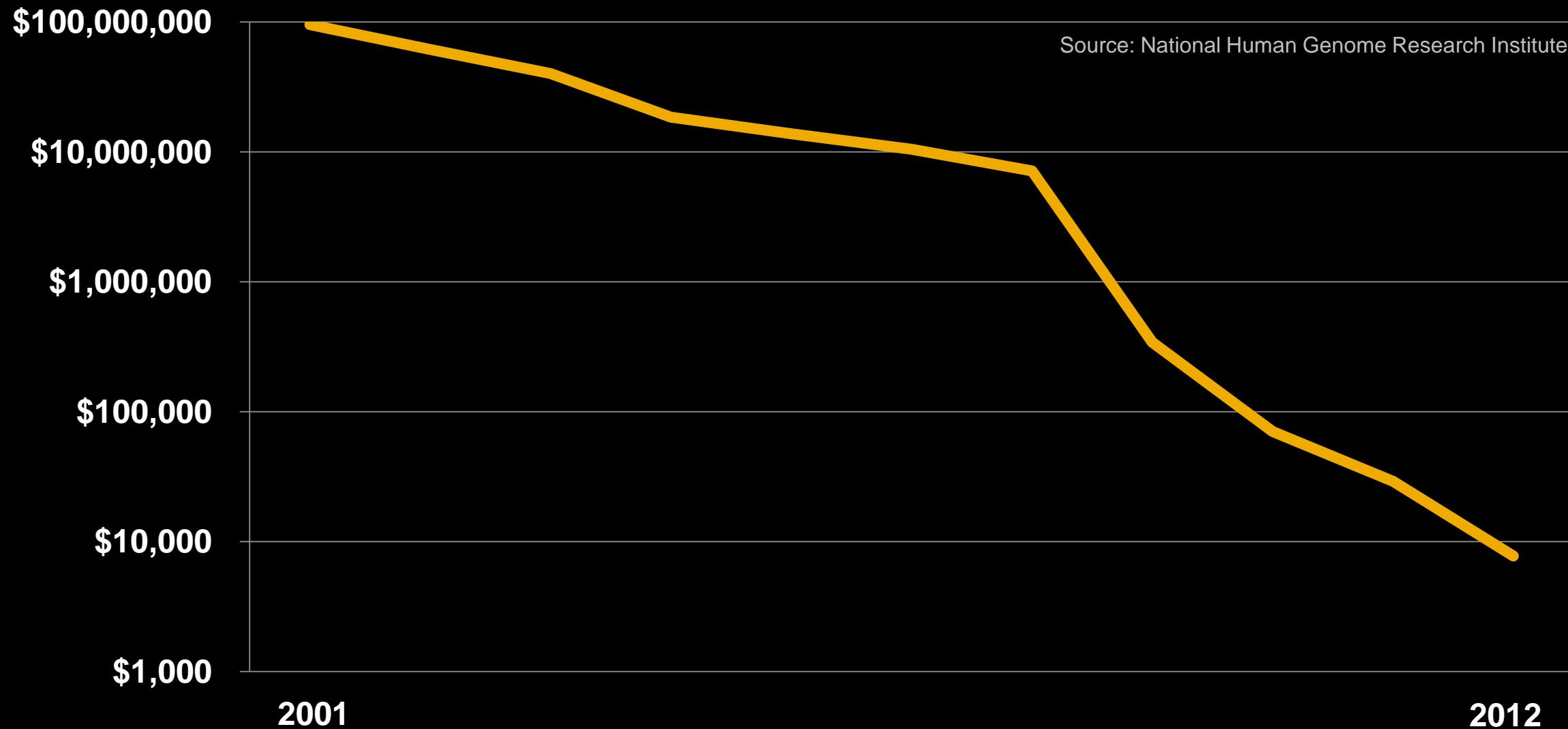


SOME PEOPLE WOULD
EASILY RECOGNIZE THEM





基因组序列分析成本





Preprocess

- Alignment of patient's DNA sequence with a normal sequence



Data analysis

- Variant calling against preprocessed sequences



Annotation

- List of actionable mutated genes and related medicines
- Create predictive model (prognosis, driver mutation, etc.)

SAP HANA + Hadoop + R



408,000X faster than
traditional disk-based
systems in PoC



216X faster DNA
analysis results - from
2-3 days to 20 minutes

Benefits

- Reduces time to detect variant DNA
- In-memory accelerates predictive & correlation analysis
- Optimized treatment plans based on DNA mutations
- Long-term study of DNA-based cancer treatment

“ ”

Genomic DNA analysis in real-time will transform how we enable comprehensive patient care to fight against cancer. SAP HANA will be the mission critical and reliable data platform to make real-time cancer analytics into a reality. Separately, our internal technical comparison demonstrated that SAP HANA outperforms a traditional disk-based system by factor of 408,000 when performing other types of data analysis.

Yukihisa Kato, Director & Executive Officer, CTO, Research and Development Center, MITSUI KNOWLEDGE INDUSTRY CO.,LTD.

Choose a Patient

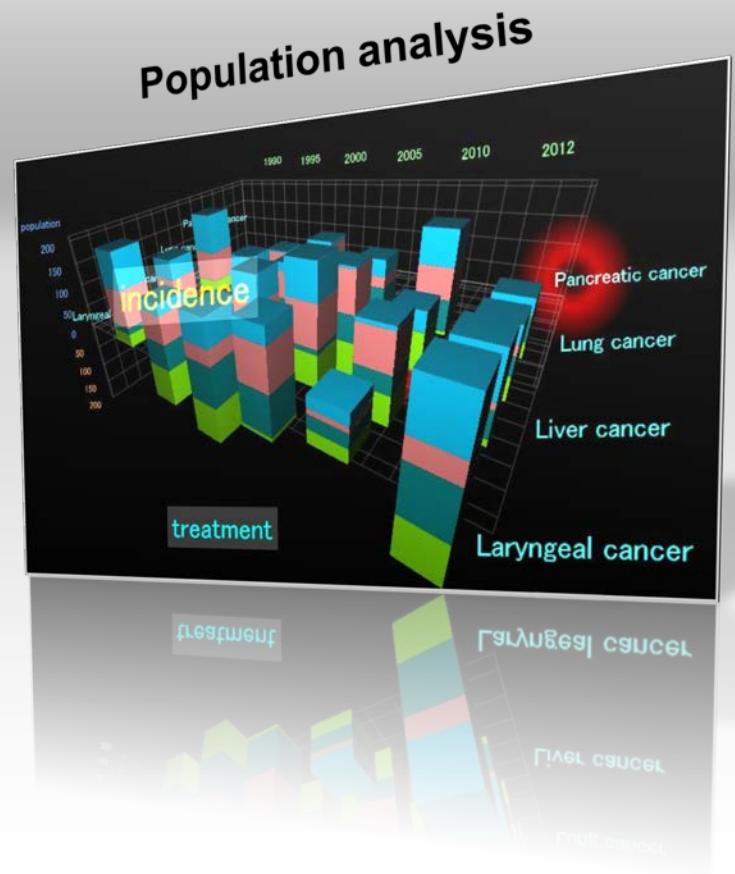
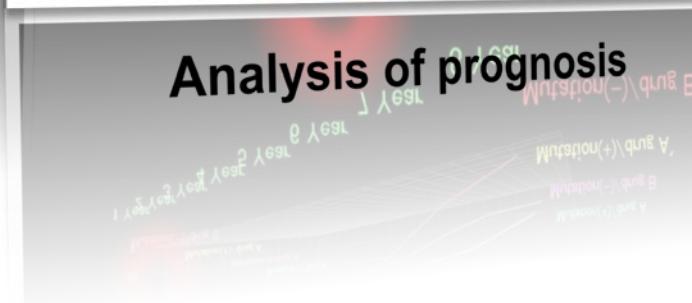
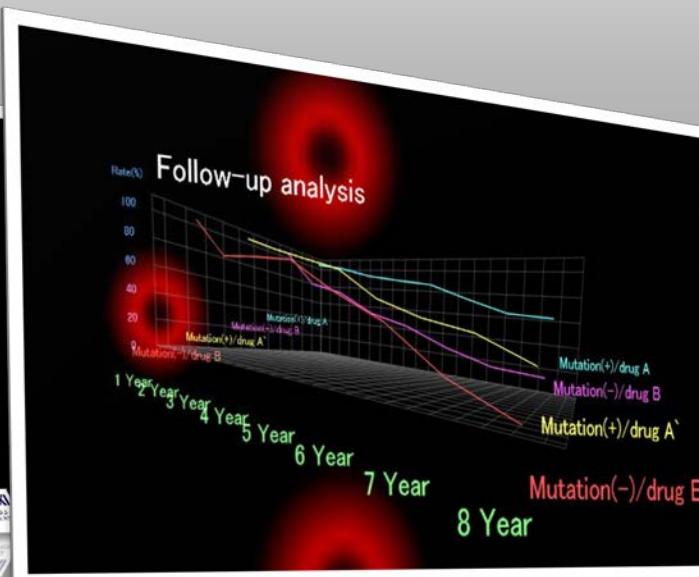
No.002 sex:F blood:O birth: Jun.21, 1968 Elizabeth Ford

No.001 sex:M blood:A birth: Apr. 6, 1974 Michel Bennet

Powered by

Patients list

Powered by MKI





传统组织绩效评价方法 都有几个世纪了

- 财务报表
- 资产负债表
- 损益表

传统财务报表
是向后看的！



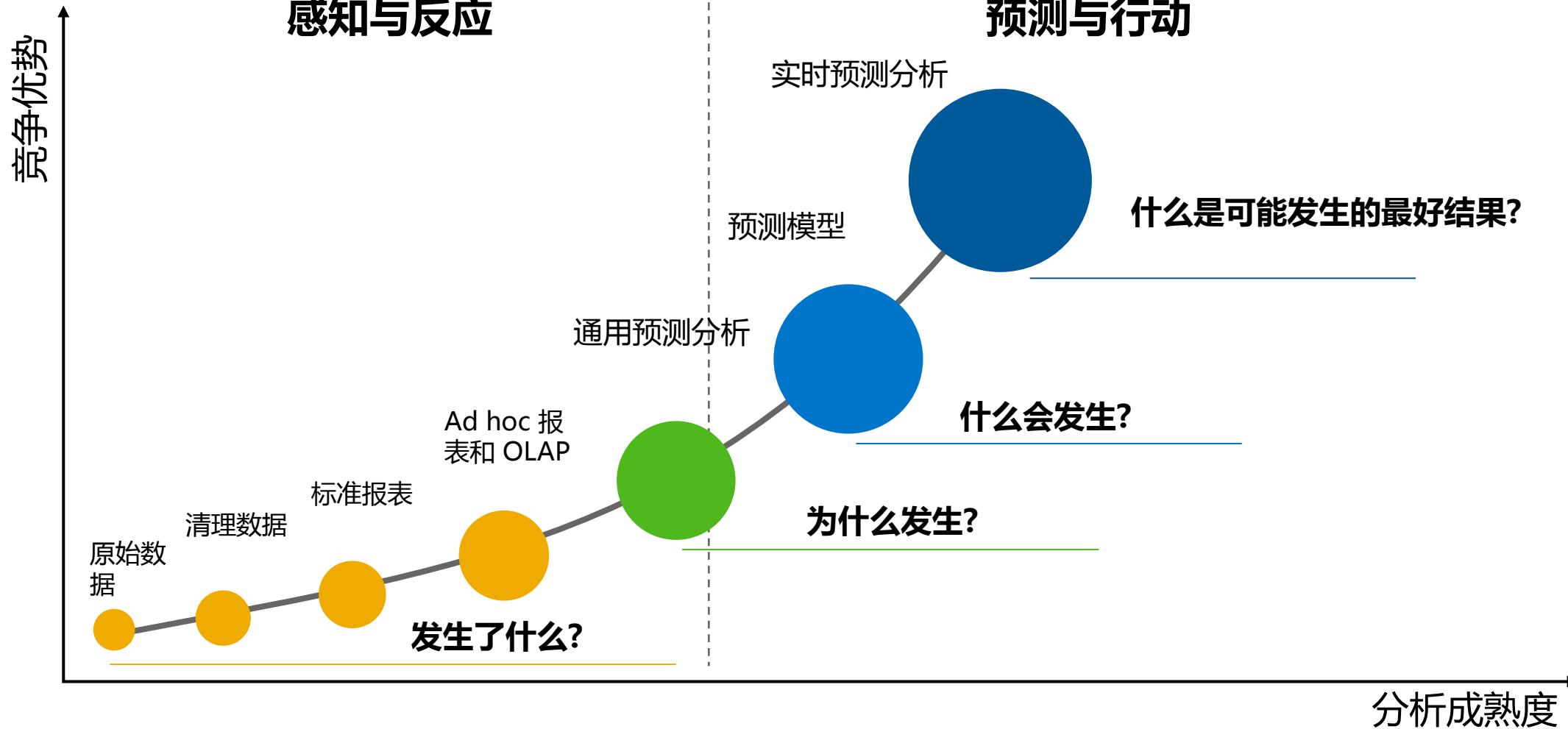
Most
Established
KPIs too



我们需要
向前看！



拓展您的分析



Fresh Direct



**“Our Food is Fresh.
Our Customers Are Spoiled”**

“As a working mom of three, I stock up on your pizzas & other ready-to-heat & serve family meals so I know what my family is eating.”

- Linda R., Working Mom



Food Made To Order

From the thickness of your steaks to the grind of your coffee, we prepare your order to perfection.



Freshly Baked Daily

We've combed the market on artisanal breads and pastries, handcrafted everyday by our own bakers. Pop our baguettes or croissants in your oven and the scent will remind you that hand-made is better.



Healthy & Delish

Shop online from our broad selection of **organic** and **all-natural** products as well as fresh dairy and meat that's raised without antibiotics.

Local Farmers. Less Travel Time!

Visit your Private Online Farmers' Market. Best of all, we cut out the middleman and pass on the freshness and savings directly to you.



Cool Meals For Kids

Each one of these delicious meals provides a complete serving of vegetables, provides great sources of the vitamins and nutrients kids need.



Easy Meals For You

Let our expert chefs give a hand with freshly-made, restaurant-quality entrees and sides, wholesome meals for kids and irresistible party platters.



More Value Everyday

Our smart web circular, "President's Picks®" brings you up to 50% off groceries every week. Get yours by email or at our website.

Buy Big & Save

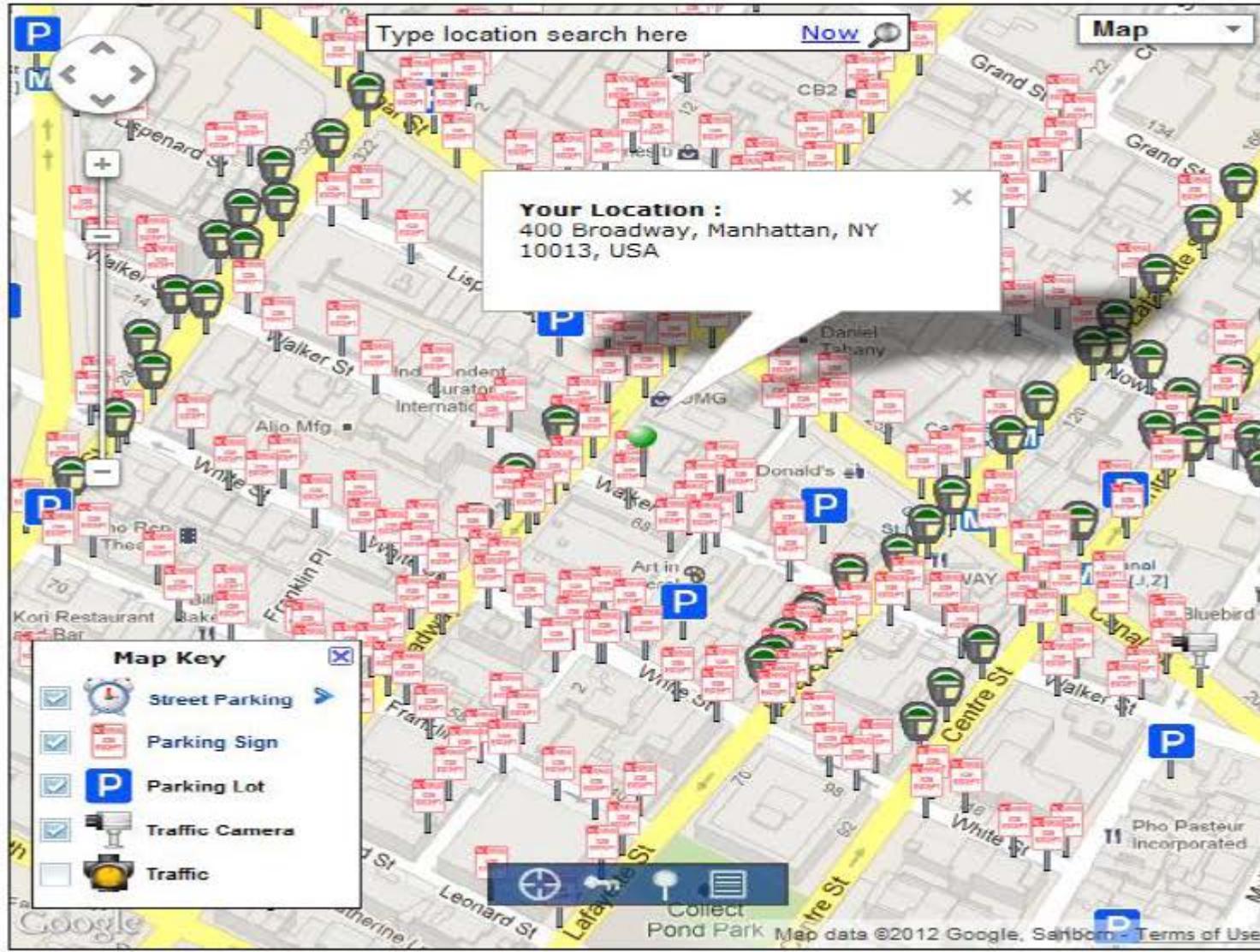
We deliver hundreds of groceries by the case and multipack, including beverages, Tide®, Bounty®, Charmin®, diapers, pet food and coffee.



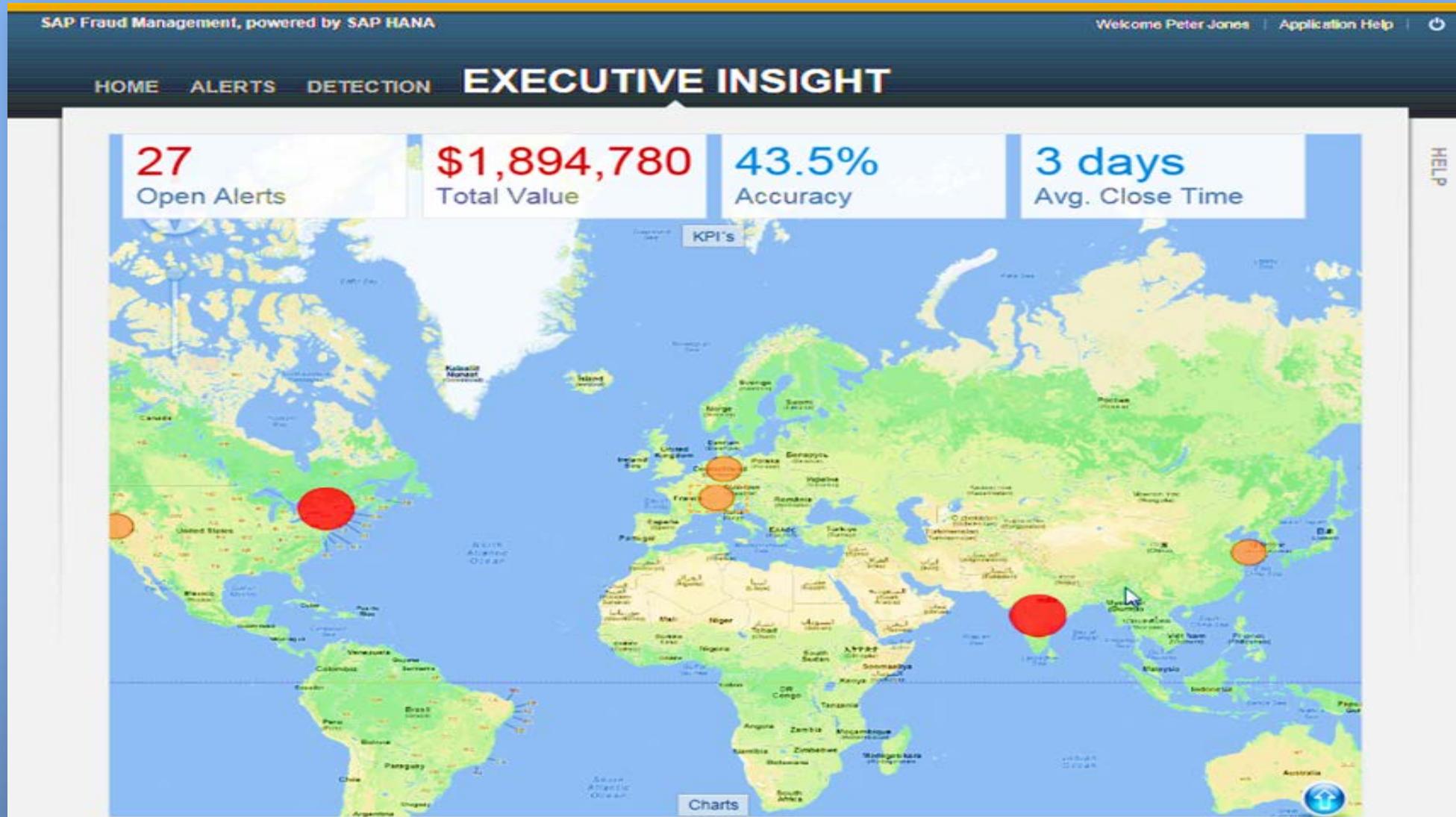
Daily Freshness Ratings

Our food experts rate all our produce and seafood every morning, so you'll always know what's the best and in season.

See how our team selects the best produce.



创新:SAP 反欺诈管理





120+ sensors per race car

Real-time visualization of
10kHz sensor readings
(10000/s) and 4Mbit/s data
speed

one thousand
simulations a minute
during a race

Real-time race
prediction & Strategy
Simulations

Detect in real-time
and avoid the
“preventable
retirement”



Real-Time Analytics of Telemetry Data



*BIG DATA: 30 years
of Historical Data*

*Data Analytics
speed up by
factor 14,000*

*Car & driver
performance
optimization over
multiple tracks &
seasons*

*Models and
simulations can
be used to work
out "what
would" have
happened*

*better predicting
the life and
boundaries of
components to
optimizing setups
& car design*

Car & Driver Performance Analysis



VODAFONE McLAREN MERCEDES

DRIVER

JENSON BUTTON

BUT

LAPTIME
ASSISTANTCOOLANT
LEVEL

0.000000

EMPTY ON
LAP 51

1

+1.078

CURRENT LAP
123.052BEST
122.565

TOP LAP SPEED

320 km/h

RACE
269 km/h

ENGINE

7

THROTTLE

AUTODROMO DI MONZA
LAP
31/53

WEATHER

Cloudy

25°C

WIND,
4.0 m/sTRACK TEMP
45°C

DRS ZONE



2

+1.078

CURRENT LAP
124.052BEST
123.781

TOP LAP SPEED

319 km/h

RACE
307 km/h

ENGINE

2

THROTTLE



LEWIS HAMILTON

HAM

LAPTIME
ASSISTANTCOOLANT
LEVEL

0.000000

EMPTY ON
LAP 48

TYPE AND LAPS RACED

TYRE PRESSURE AND LIFE



NEXT PIT WINDOW

36 - 39

ESTIMATED TIME
3.024sec

LAP 19

CURRENT TELEMETRY DATA

vCar	=259.3	=258.4	kph
Ngear	=7	=2	
rThrottlePedal	=100.0	=100	%
pBrakeR	=8.5	=0.0	bar
lBrakeR	=123	=123	

DATA STATISTICS

1	2	3
2	3	4
3	4	5
4	5	6

Transferring =3.4 MB/sec

NEXT PIT WINDOW

NO MORE PITS

ESTIMATED TIME
2.674sec

LAP 19

TYRE PRESSURE AND LIFE



TYPE AND LAPS RACED



PREDICTIVE TIMELINE

SAP



Powered by HANA



6:02

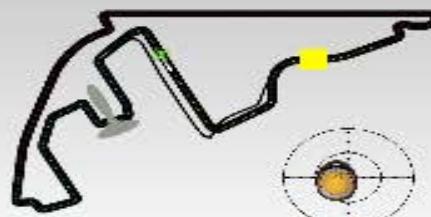
* 100%

MP4-27

ON



Lap: 2 - Laptime: 00:17:302
Speed: 240 kph



Leaders

Switch to Grid View

[Points](#) [Rebounds](#) [Assists](#) [Defense](#) [Playmaking](#) [Efficiency](#) [Clutch](#) [Scoring](#) [Players](#) [Teams](#) [Season ▾](#)

PLAYERS & TEAMS

PTS

TS% ▾

Carmelo Anthony
New York Knicks

29.0

57.2%

Kevin Durant
Oklahoma City Thunder

29.0

65.7%

LeBron James
Miami Heat

27.1

63.6%

Kobe Bryant
Los Angeles Lakers

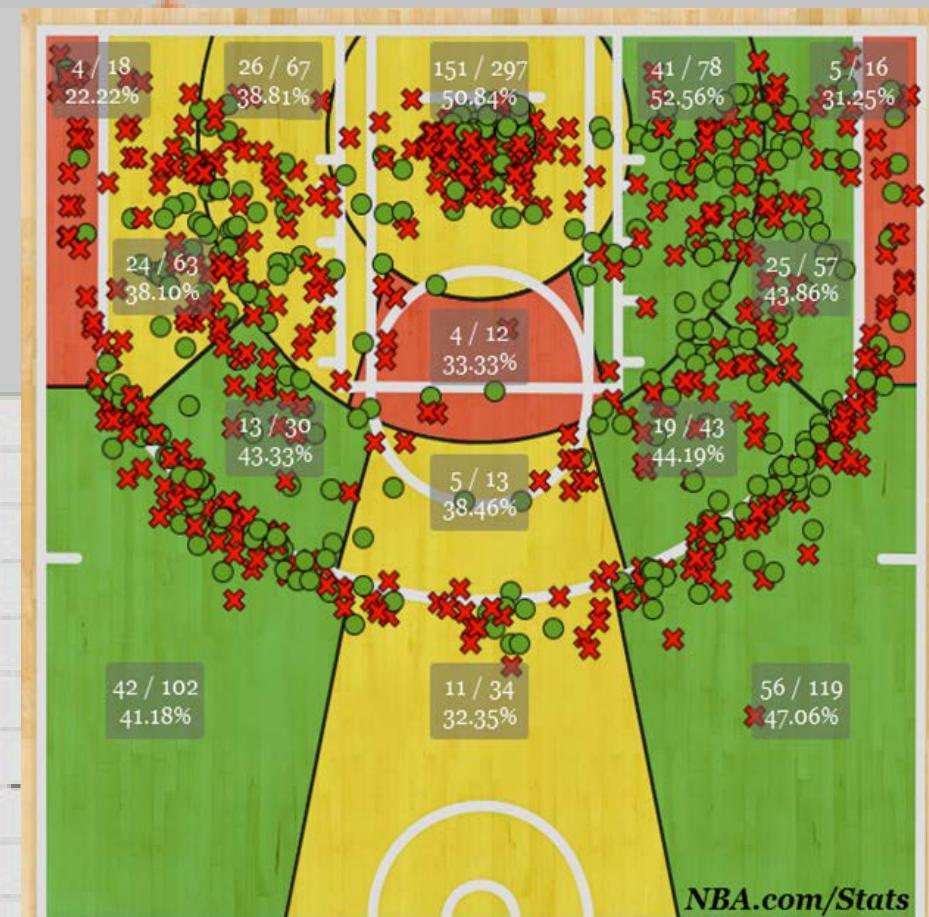
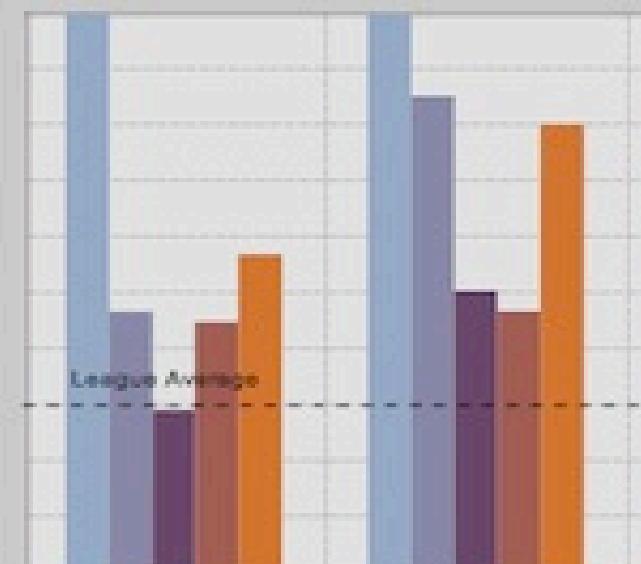
27.0

56.5%

James Harden
Houston Rockets

26.1

60.0%

[MORE](#)[MORE](#)

NBA.com/Stats

What is TS% ?

Calculates shooting percentage for a player or team adjusting for the value of free throws and three-point field goals. [Learn More](#)





100% ingame material

<https://www.youtube.com/watch?v=N6bUkkiZHoE>

SAP 实时数据平台：一个游戏改变者!



HITACHI

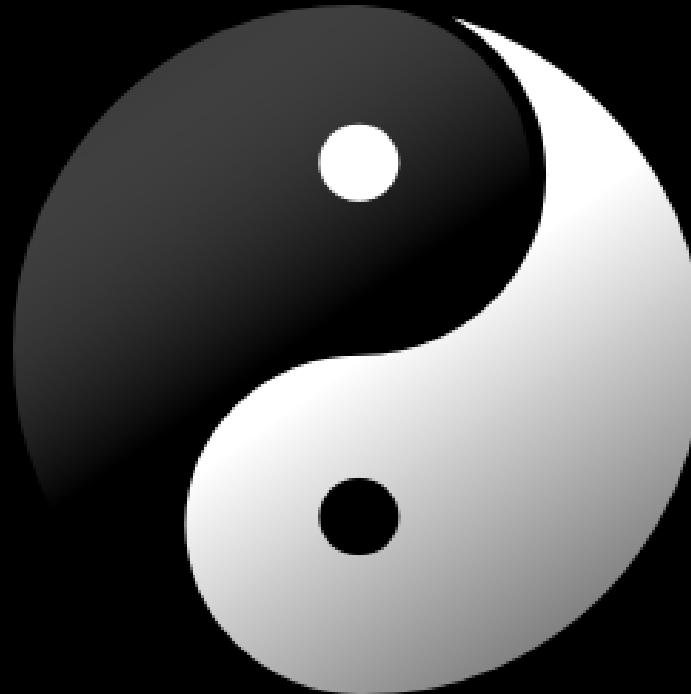




利用BI的分析建立大数据的方案

商业智能

Operations
Theorems (facts)
Statements
Instant answer
High value data
Highly structured
Rigid process



发现洞见

Research
Hypothesis
Questions
Indefinite result
Any data
Any structure
Flexible strategies

“Only one truth”

“Many paths to wisdom”



As we may
WORK





0 MEMBERS MERGUI ARCH

MERGUI ARCHIPELAGO



IN THE AREA, WHEN A STORM BROKE DOWN COMMUNICATIONS
THROUGHOUT THE REGION, WORKERS ARE SCRAMBLING TO REPAIR



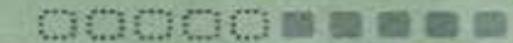
VANNEVAR BAY



VARIN VISHEL



ALAN ENGEBAK







THANK YOU!

邬学宁
首席专家，SAP APJ

E: x.wu@sap.com

T: 86-21-61085287

