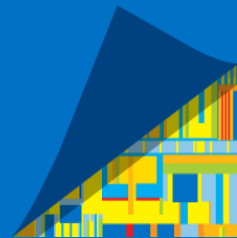
大数据的十个技术前沿

吴甘沙

英特尔中国研究院



Legal Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

The cost reduction scenarios described in this document are intended to enable you to get a better understanding of how the purchase of a given Intel product, combined with a number of situation-specific variables, might affect your future cost and savings. Circumstances will vary and there may be unaccounted-for costs related to the use and deployment of a given product. Nothing in this document should be interpreted as either a promise of or contract for a given level of costs.

Legal Disclaimers

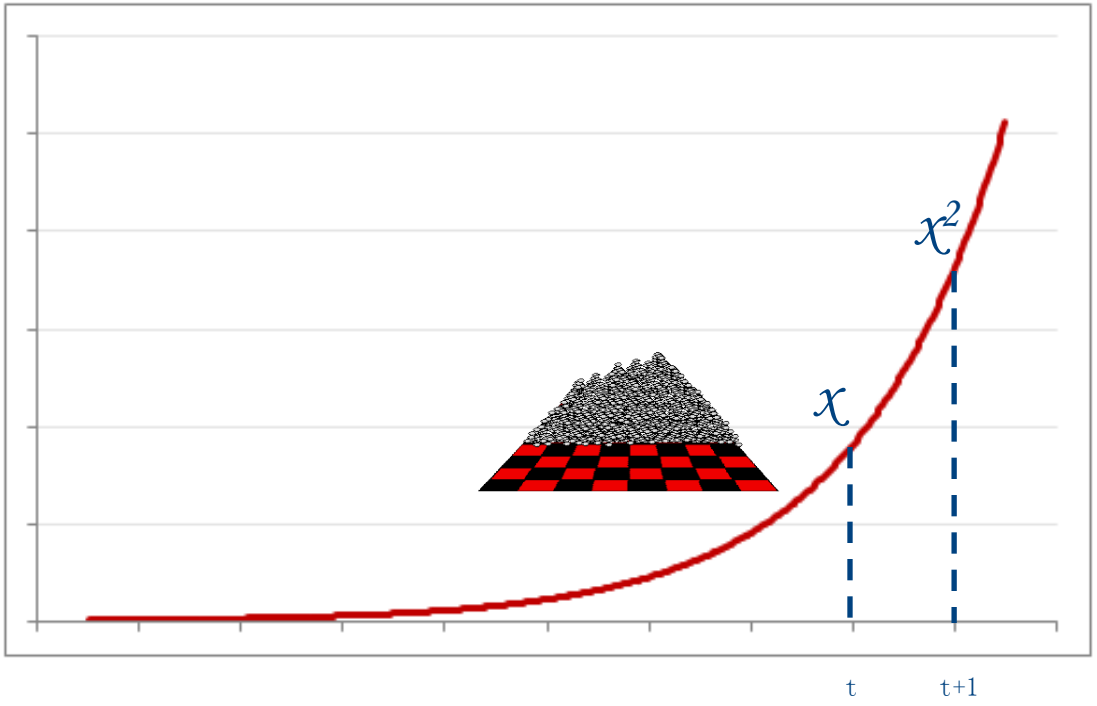
Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families: Go to: Learn About Intel® Processor Numbers

Intel, Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

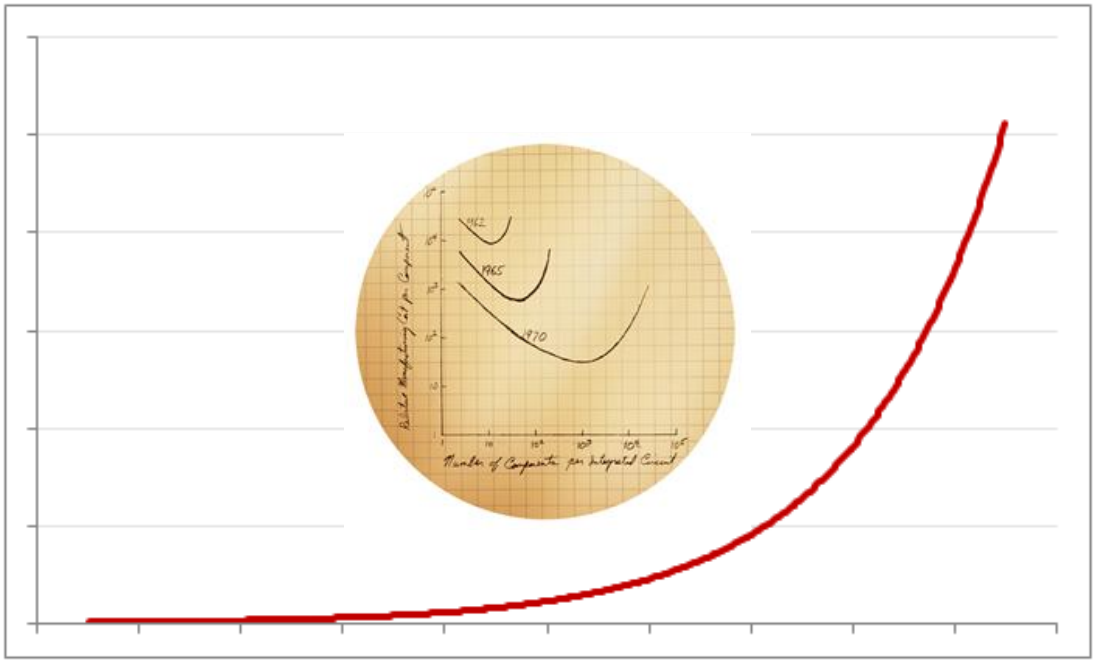
Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

*Other names and brands may be claimed as the property of others.

指数时代：颠覆成为常态



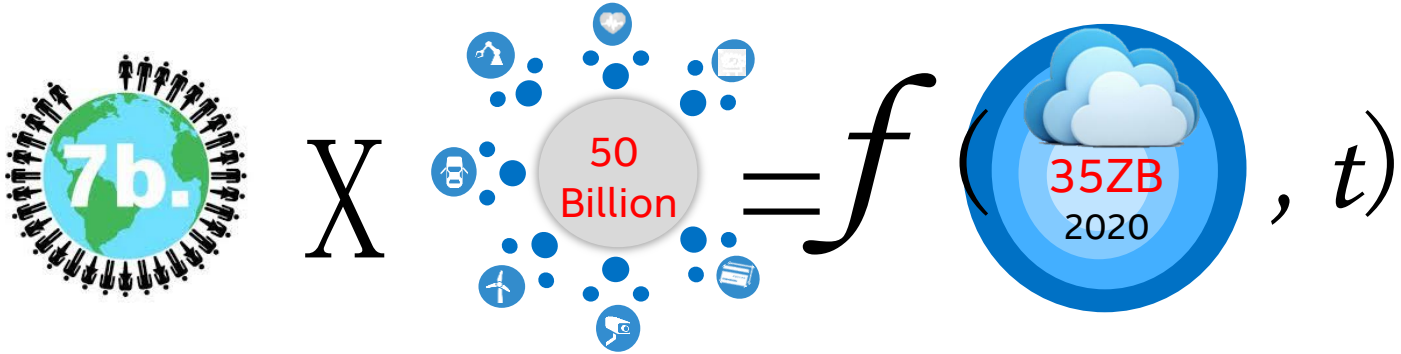
摩尔定律：指数社会的“基因”



晶体管数
性能
成本
功耗
存储
带宽
像素
… …

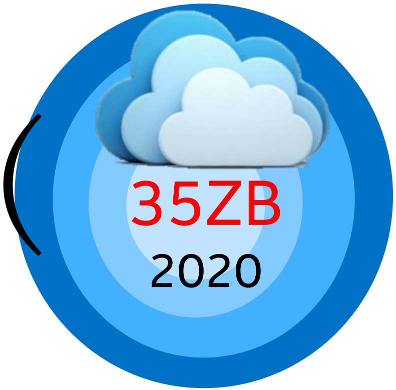
数据

大数据：指数社会的“蛋白质”



数据化生存　　感知-互联-智能　　价值为大

大数据的技术挑战

$f($  $, t)$

人与工具　　见微知著　　全集>采样　　实时和全时

十个技术前沿

膨胀宇宙

巴别之难

数据有价

软硬兼施

多快好省

天下三分

分久必合

精益求精

人机消长

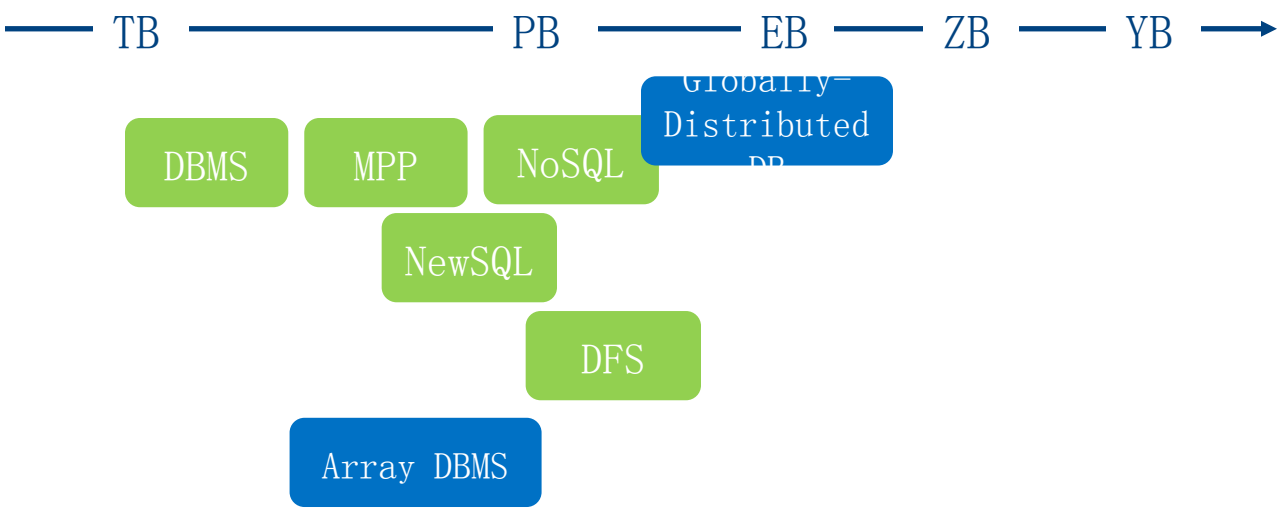
智能之争

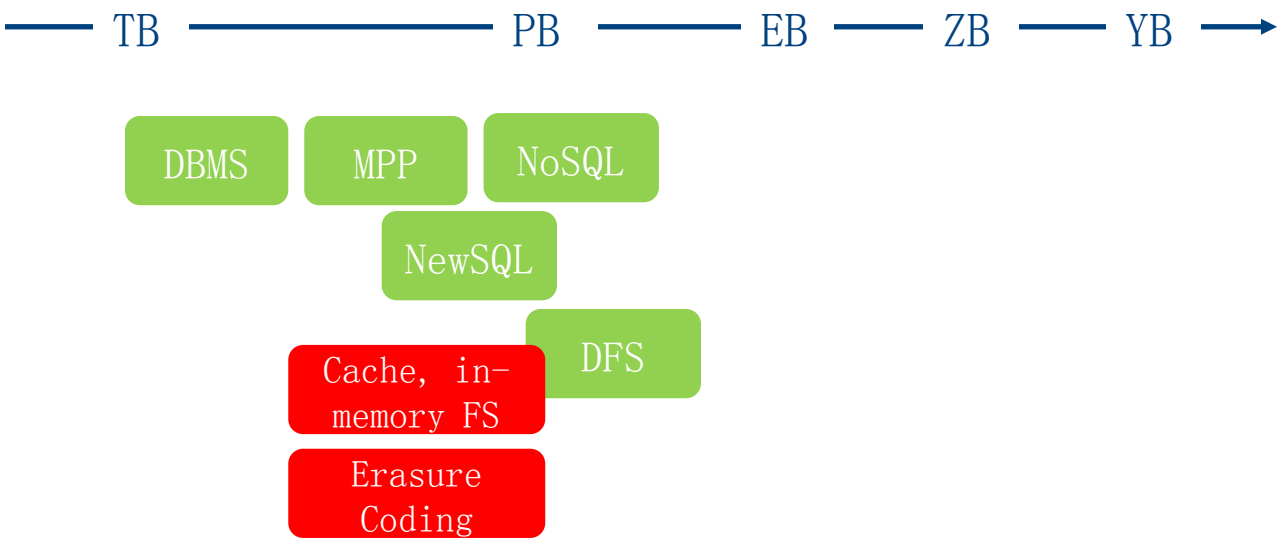
数据

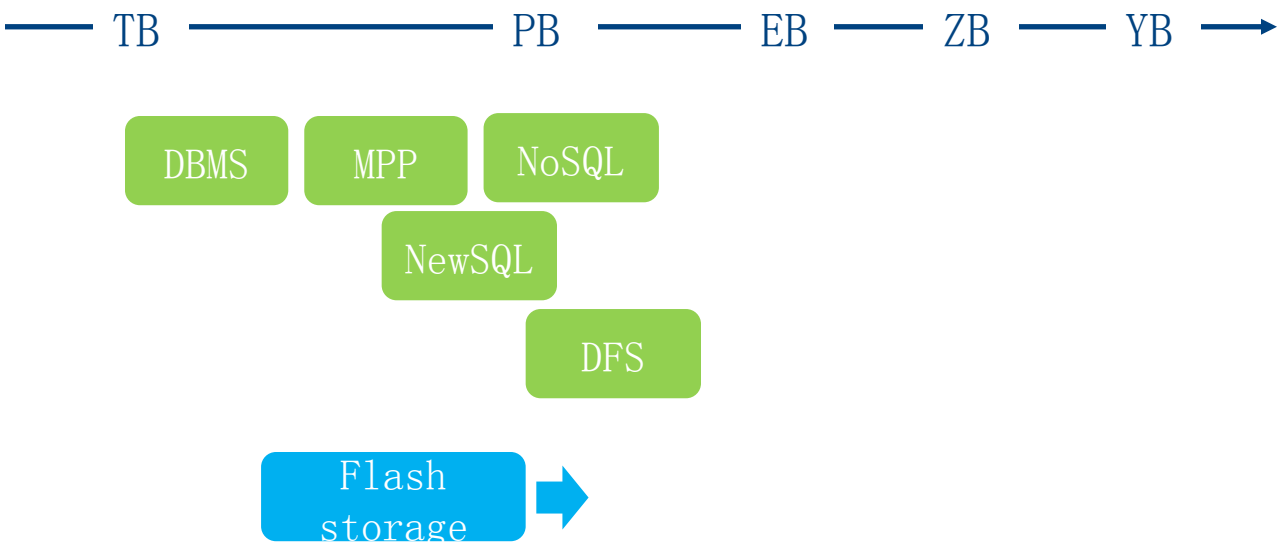
计算

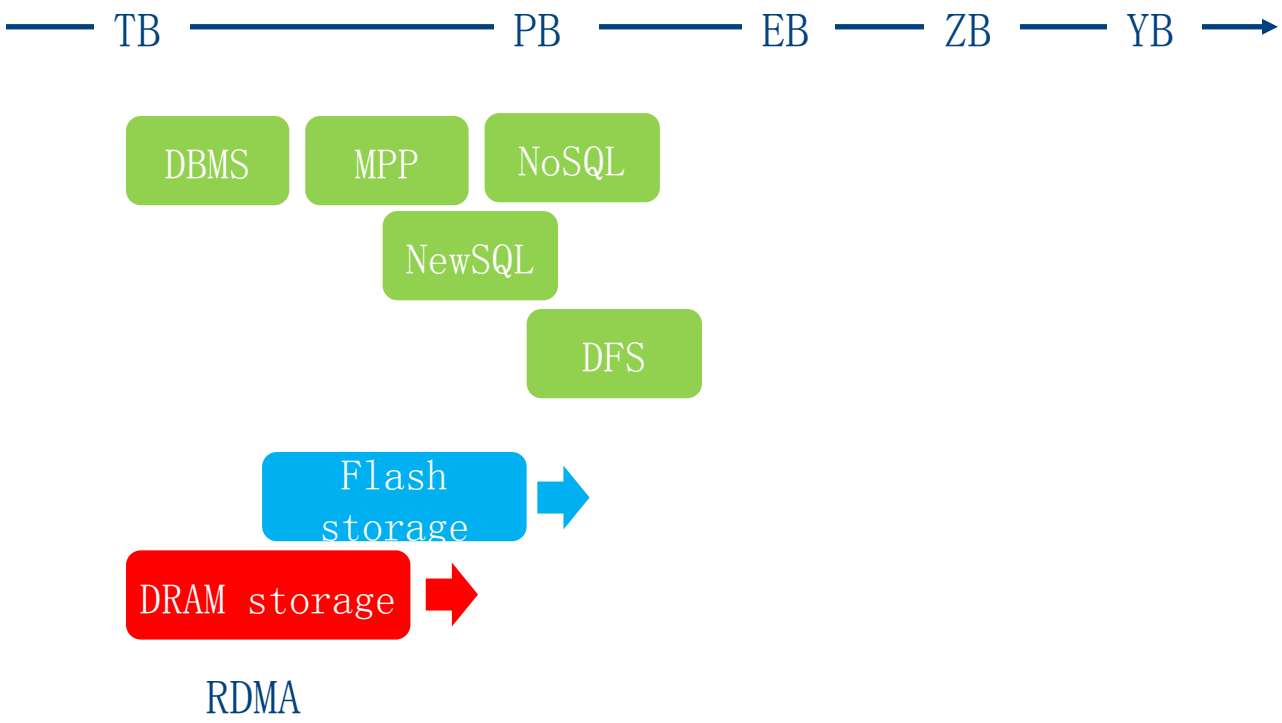
分析

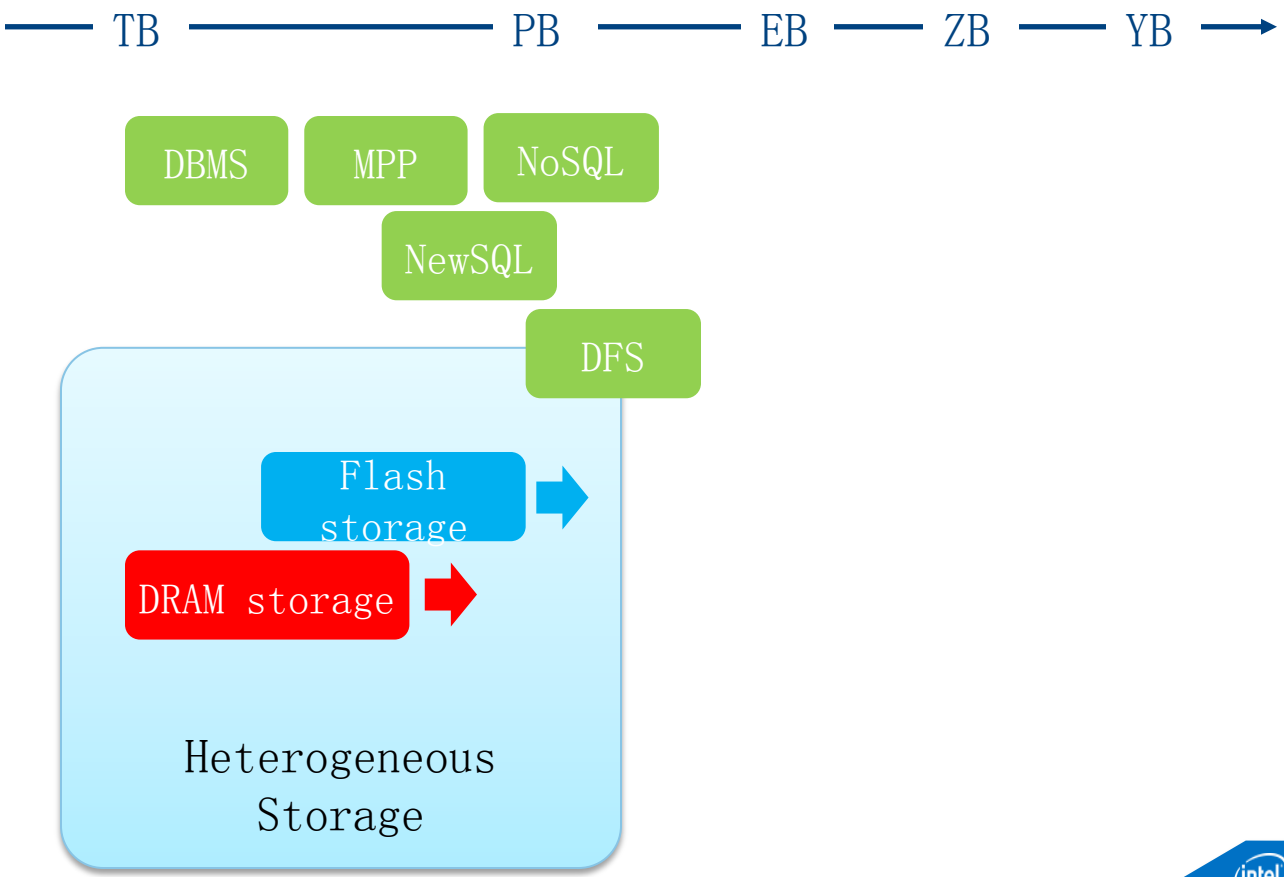
1. 膨胀宇宙

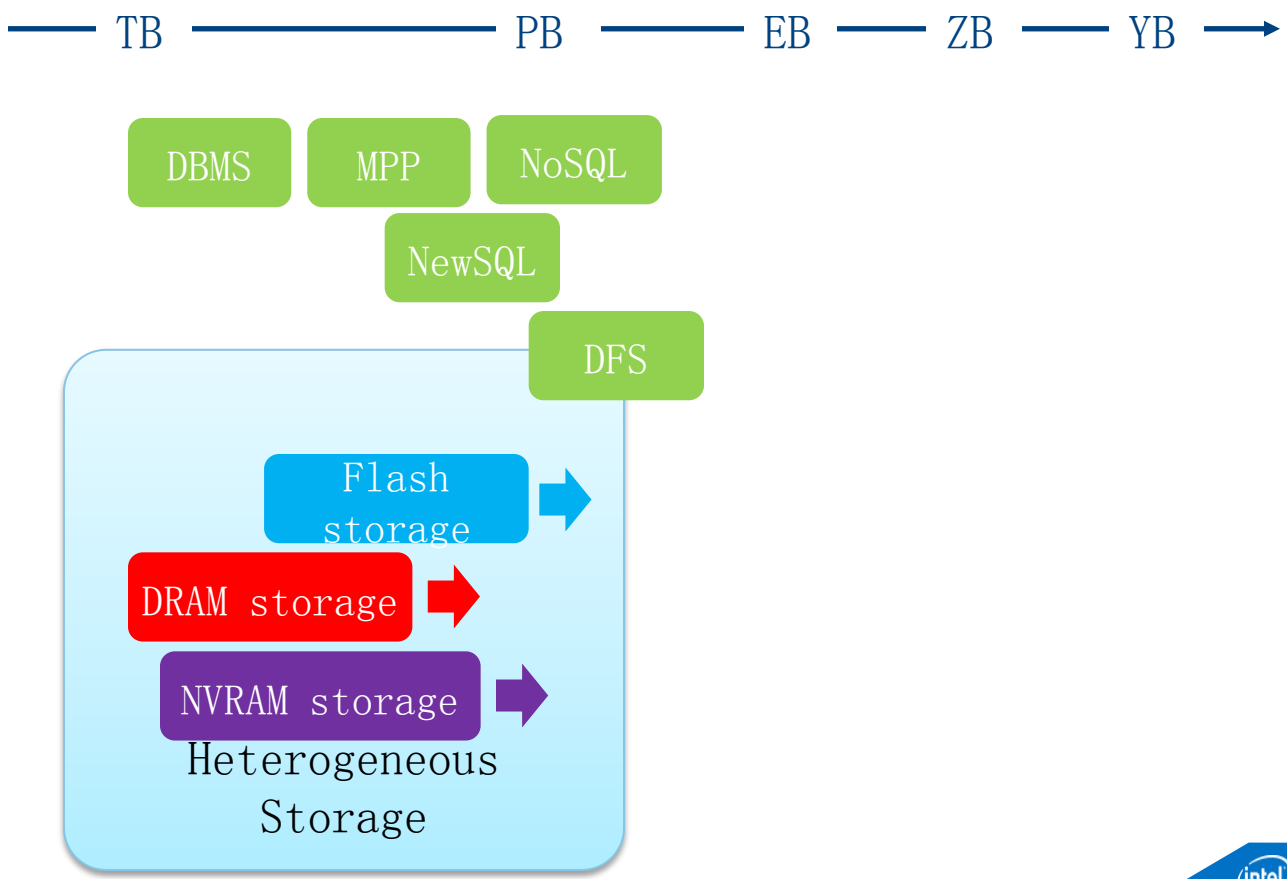

TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
Distributed
NewSQL
DFS
Array DBMS

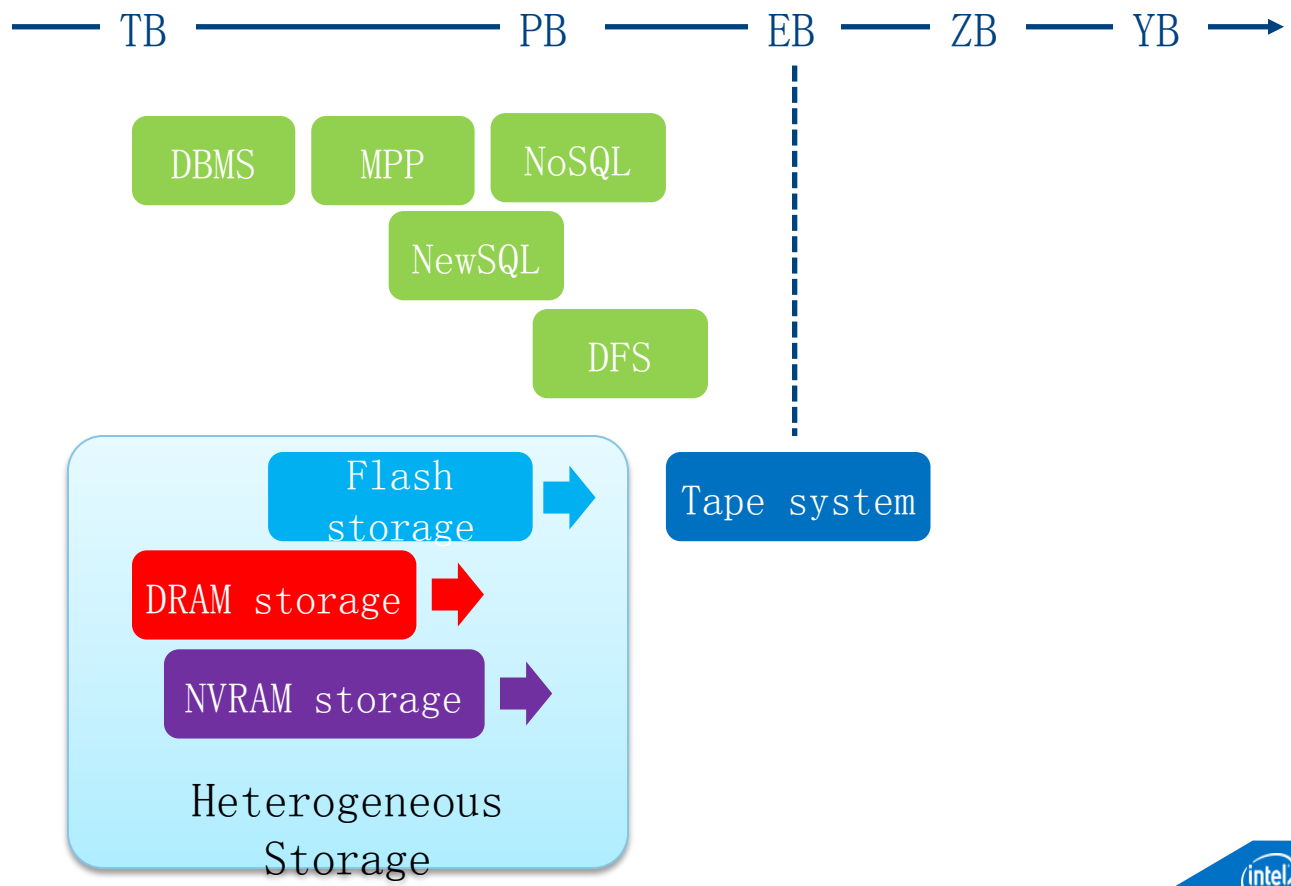

TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Cache, in-memory FS
Erasure Coding


TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Flash storage


TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Flash storage
DRAM storage
RDMA

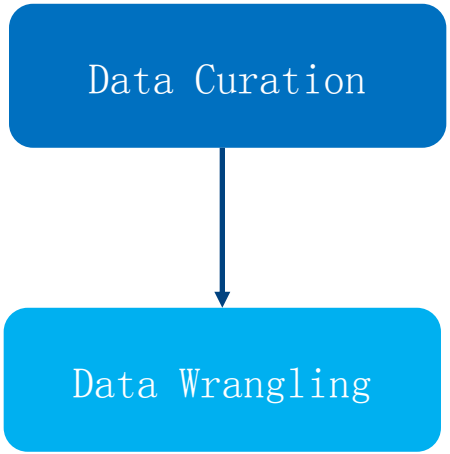

TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Flash storage
DRAM storage
Heterogeneous Storage


TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Flash storage
DRAM storage
NVRAM storage
Heterogeneous Storage

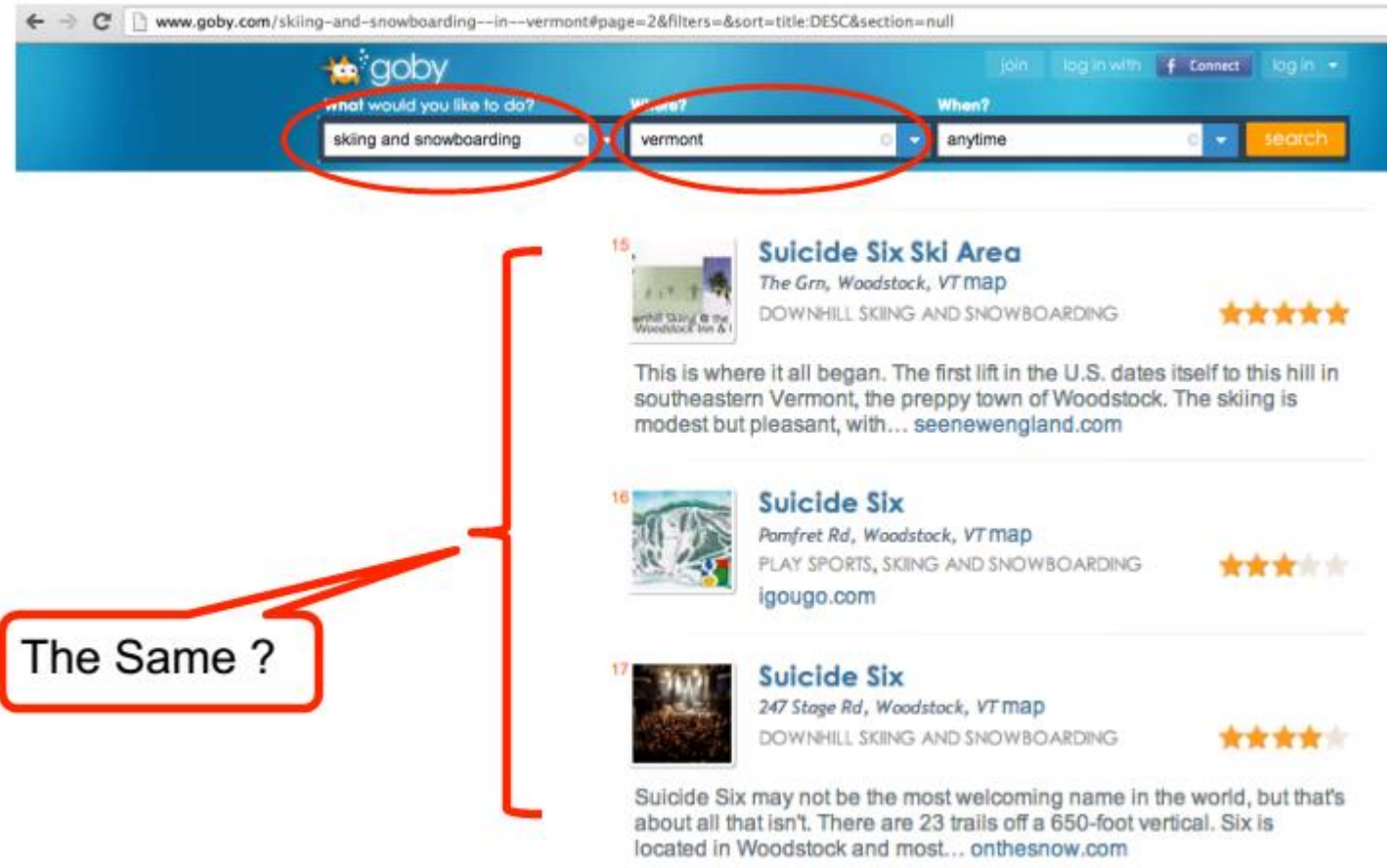

TB
PB
EB
ZB
YB
DBMS
MPP
NoSQL
NewSQL
DFS
Flash storage
DRAM storage
NVRAM storage
Heterogeneous Storage
Tape system

2. 巴别之难

Data talks, but in different languages



DB-hard



DataTamer

Courtesy: Michael Stonebraker

Courtesy: Michael Stonebraker

DETAILS FROM WEB

dining rating: 4.00/5
helicopter: No
peak elevation: 650 Ft / 198 M
service rating: 5.00/5
snow conditions: Yes
snowcast: No
terrain rating: 5.00/5
trail map: Yes
value rating: 5.00/5
vertical (ft): 650'
webcam: No

advanced trails %: 30%
beginner trails %: 30%
expert trails %: 0%
intermediate trails %: 40%
lift count: 3 - 2 Doubles; 1 Surface Lift
snowphone: (802) 457-1622
snowmaking: 50%
trail acreage: 100 skiable acres
trail count: 22
vertical drop (feets): 650 feet

DETAILS FROM WEB

lift tickets: Adult Junior Senior Weekday Full Day: $36 $30 $30 Weekday Half Day: $30 $23 $23 Weekend Full Day: $55 $40 $40 Weekend Half Day:...
state: Vermont
advanced runs: 30
average annual snowfall: 90"
base elevation: 550ft
beginner runs: 30
expert runs: 0
intermediate runs: 40
number of double chairs: 2
number of high speed quads: 0
number of high speed sixes: 0

number of regular quads: 0
number of surface lifts: 1
total lifts: 3
number of gondolas and trams: 0
number of triple chairs: 0
visitor recommendations: Single/Newlyweds: 101%, Beginner: 101%, Families: 126%, Intermediate: 126%, Empty Nesters: 101%, Advanced: 76%
summit elevation: 1200ft
vertical drop: 650ft
years open: 79
event category: Kids Activities Recreation Ski Area

Courtesy: Michael Stonebraker

Active work

Data interpretation via automated learning

- Discovery of structure
- Entity resolution
- Transformation

From (semi-)structured to unstructured

Data and metadata organization

- Apache UIMA (Unstructured Information Management Architecture)

3. 数据有价

安全和定价

安全

- 系统的安全
- 数据的安全
- 数据使用的安全

定价：效用和稀缺性

系统的安全

Hadoop用户和服务鉴权（基于Kerberos）

HDFS文件和数据块权限控制

数据高度分布、去中心化场景下的安全：分布式安全架构

- 如blockchain

数据的安全

静态数据安全

- 加密：HADOOP-10150
- 访问控制：Apache Accumulo，Hbase

动态数据安全

- 动态审计能力：数据泄露防护（Data Leakage Prevention）

数据脱敏/匿名化

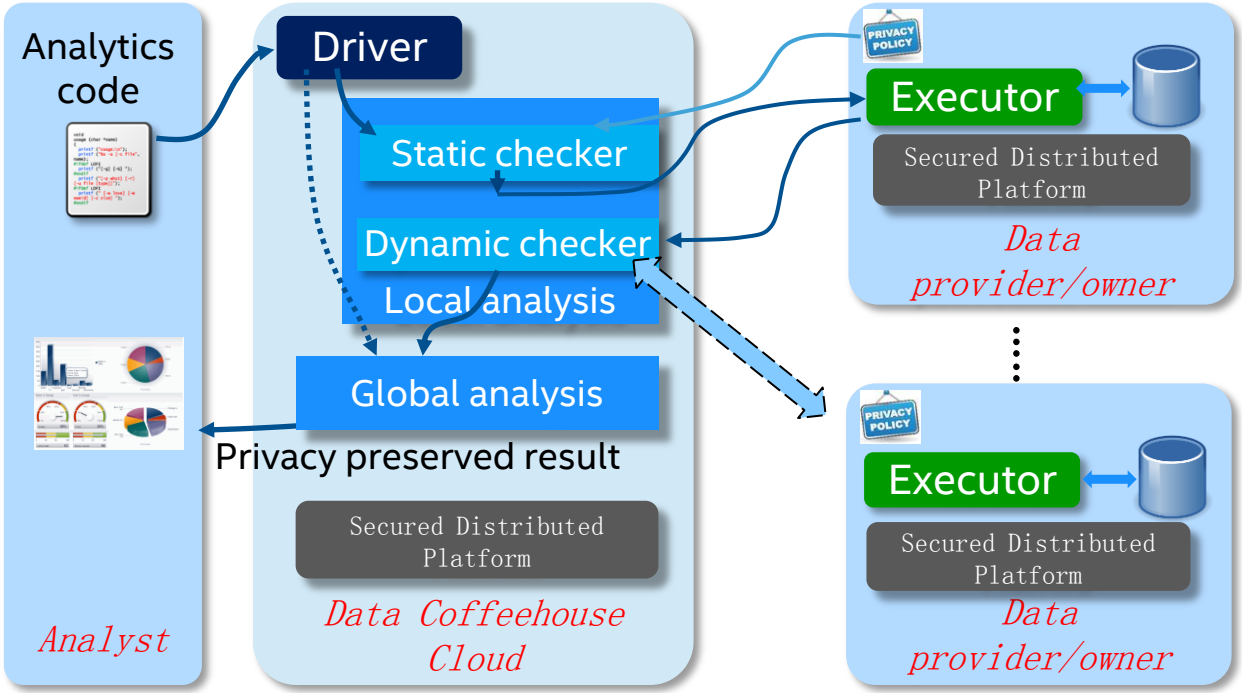
- 去标识符，但基于准标识符（quasi identifiers）仍能重新标识化
- k-anonymity、L-diversity、T-Closeness
- 差分隐私（differential privacy）
- 隐私安全性和数据可用性的平衡

数据使用的安全

可用不可见，相交不相识

- 同态加密数据库技术： CryptDB/Monomi
- 基于加密协议的多方安全计算
- 基于可信计算环境的多方安全计算：TXT/TPM/VT-d → SGX

Multi-source privacy preserving data analytics



4. 软硬兼施

选择更好的硬件架构——计算

大小核

- Brawny cores vs. wimpy cores

异构计算

- 集成异构多核，GPGPU，Xeon Phi，FPGA，ASIP（e.g. NPU）

内存与处理器更加靠近

- eDRAM，3D stacking，in-memory computing

选择更好的硬件架构——存储和互联

大内存服务器

- 内存容错：lineage，checkpointing，redundancy，log-based

SSD → PCIe SSD → flash storage

- 重构系统软件栈，智能数据迁移

NVRAM

- No checkpointing, no ser-de, no file…

互联

- Silicon photonics，RDMA

软件与硬件架构协同优化

针对硬件特点对软件栈优化

- 把硬件暴露给软件栈：NativeTask，Spark，线性代数库
- 重新设计软件栈：全闪存存储，NVRAM
- 一体机
- 云化：虚拟化和资源管理，HVE，YARN，Docker

Big Learning System：机器学习算法与底层系统的更好配合

- VW，GraphLab，DistBelief，Project Adam，Petuum

5. 多快好省

多快好省：未必能兼而得之

内存计算

降低空间、时间复杂度

并行化/分布式

内存计算

在硬件平台层：大内存，全闪存，NVRAM，RDMA

在数据管理和存储层：cache，堆外内存/in-memory FS

在计算处理层：in memory data grids，Spark

在数据分析和可视化层：

- 重新设计数据结构，原位（in-situ）分析和可视化
  - 分析：图（GraphLab vs. GraphChi）
  - 可视化：in-memory data cube（e.g. Nanocube）

降低空间、时间复杂度

空间（把大数据变小）

- 压缩（e.g. Apache Parquet）；缓存和多温度存储；稀疏结构
- Spark： hash-based shuffle → sort-based shuffle

时间

- 采样和近似：BlinkDB，Summingbird/Algebird 近似计算Monoid
- 更多数据 + 简单模型
- 简单模型的组合（ensemble）
- 降维和混合建模

并行化/分布式

ACID → BASE (Basically Available Soft-state services with Eventual-consistency)

Jacobi方法 vs. Gauss-Seidel方法

- 使用过时数据来打破迭代间数据依赖
- Parameter Server，Petuum/Stale Synchronous Parallel

数据并行 → 图并行（GraphLab），模型并行（Petuum）

减少通讯

- 缓存，一致性，本地性（locality），划分（cut）和调度
- 同步、半同步（e.g. Petuum）和异步，批量与个别，全部与变化部分（Spark bit torrent）

Alternating Direction Method of Multipliers (ADMM)

6. 天下三分

数据类型的分野

表格/K-V，数组/矩阵，图

- 关系查询，以线性代数为代表的复杂分析，图计算

计算范式的分野

计算图：数据依赖，无计算依赖

- 批量计算：数据不动，计算动
  - MapReduce：二阶段
  - BSP：三阶段
  - DAG和多迭代计算
- 流式计算：计算不动，数据恒动
  - Record-at-a-time vs. minibatch
  - 简单计算 → 流式/在线机器学习（e.g. SAMOA、Jubatus）

图计算：数据和计算依赖

编程模型

数据并行，任务并行（流式），图并行

图结构与关系结构、数组结果的互操作

事件驱动

- reactive范式：从ErLang Actor到Scala Akka

概率编程模型

7. 分久必合

融合

Twitter Summingbird在**编程接口层面**融合

- 支持批量和流式

Lambda架构在**应用框架层面**融合

- 增量计算和批量缓存

Spark在**实现框架层面**完成融合

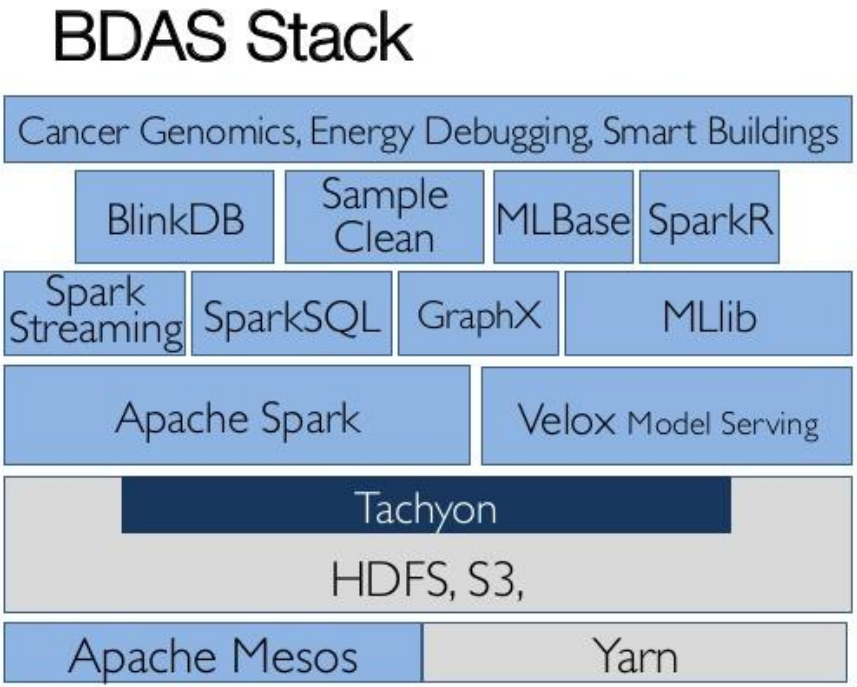
- 几乎支持所有计算范式

REEF通过**资源管理层**来支持多计算模型

Big Dawg普适编程模型

- 支持关系和线性代数、复杂数据模型、迭代计算和并行计算

Spark的融合


BDAS Stack
Cancer Genomics, Energy Debugging, Smart Buildings
BlinkDB
Sample Clean
MLBase
SparkR
Spark Streaming
SparkSQL
GraphX
MLlib
Apache Spark
Velox Model Serving
Tachyon
HDFS, S3,
Apache Mesos
Yarn

Spark的融合

流查询：Spark Streaming + Spark SQL

实时+批量（类Lambda结构）：Spark Streaming + Spark Core

流处理+机器学习： Spark Streaming + MLlib

图流水线：从MapReduce+图引擎到Spark Core/SQL + GraphX

- 图创建/ETL+图计算+后处理

即席查询+机器学习：Spark SQL + MLlib

8. 精益求精

机器学习精确度的提升

对更多数据的包容性

- 模型：复杂 > 简单，Deep Learning
- 混合模型
  - Non-parametric vs. Parametric
  - Linear vs. Non-Linear
  - Discriminative vs. Generative
- 优化算法：good enough > 复杂？

覆盖长尾 Exponential assumption vs. long tail

- PCA/LDA/pLSA vs. 分级训练、模型组合、概率图模型/DNN

在线/流式学习

- 增量训练，模型异步更新，快速部署

9. 人机消长

人和机器作用的变化

自动化分析和可视化

- MLBase和VizDeck

可视化库、框架和工具

- 文本、网络/图、时空数据、多维数据
- 交互式可视分析：多侧面、多尺度、多焦点交互

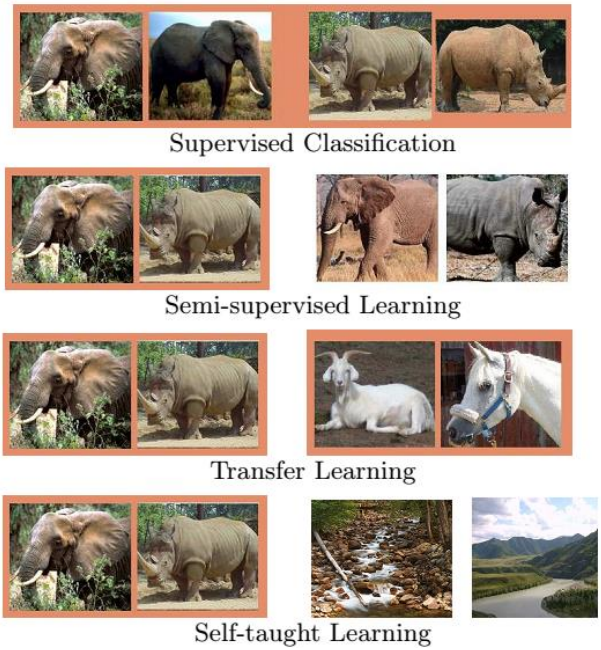
大规模协作分析

- CrowdDB，Kaggle，Duolingo

机器学习中人的角色

标记数据

- Unsupervised learning, semi-unsupervised learning, transfer learning, self-taught learning



机器学习中人的角色

标记数据

- Unsupervised learning, semi-unsupervised learning, transfer learning, self-taught learning

特征工程

- Deep Learning的unsupervised特征学习

方便易用性

- 全pipeline框架设计，模型管理

10. 智能之争

生物智能 vs. 机器智能

计算智能

模式匹配

人工神经网络

进化计算

模糊逻辑

人工免疫系统

群体智能

统计学习

当前在热烈讨论的问题

深度学习有没有可能包打天下

- 深度学习的理论基础，deep flaws，Hinton的Capsules theory

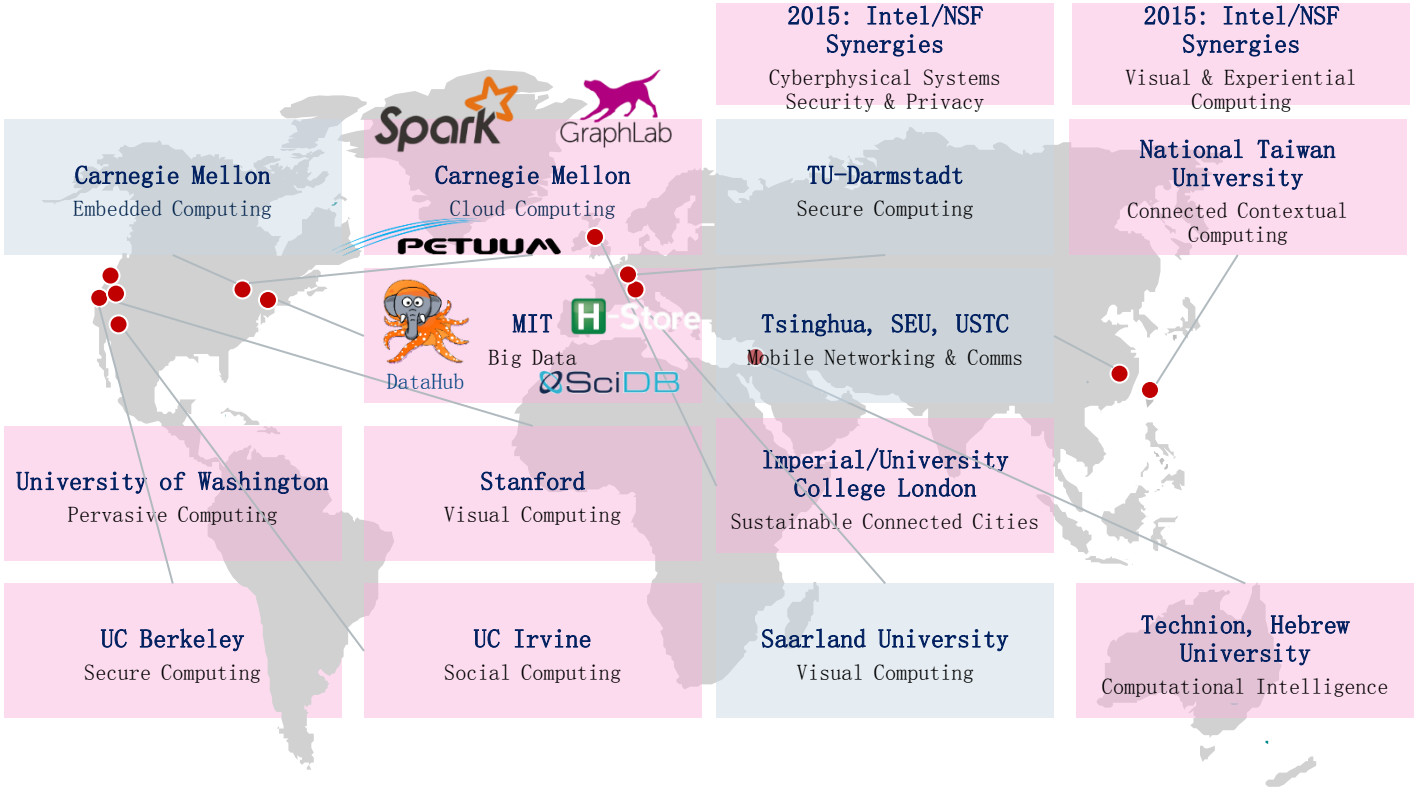
智能的未来是不是类脑

- BRAIN计划，Mind研究，记忆
- ANN需要不断的改进：普适性，反馈，时间因素，在线学习能力，…

需不需要新的类脑计算架构

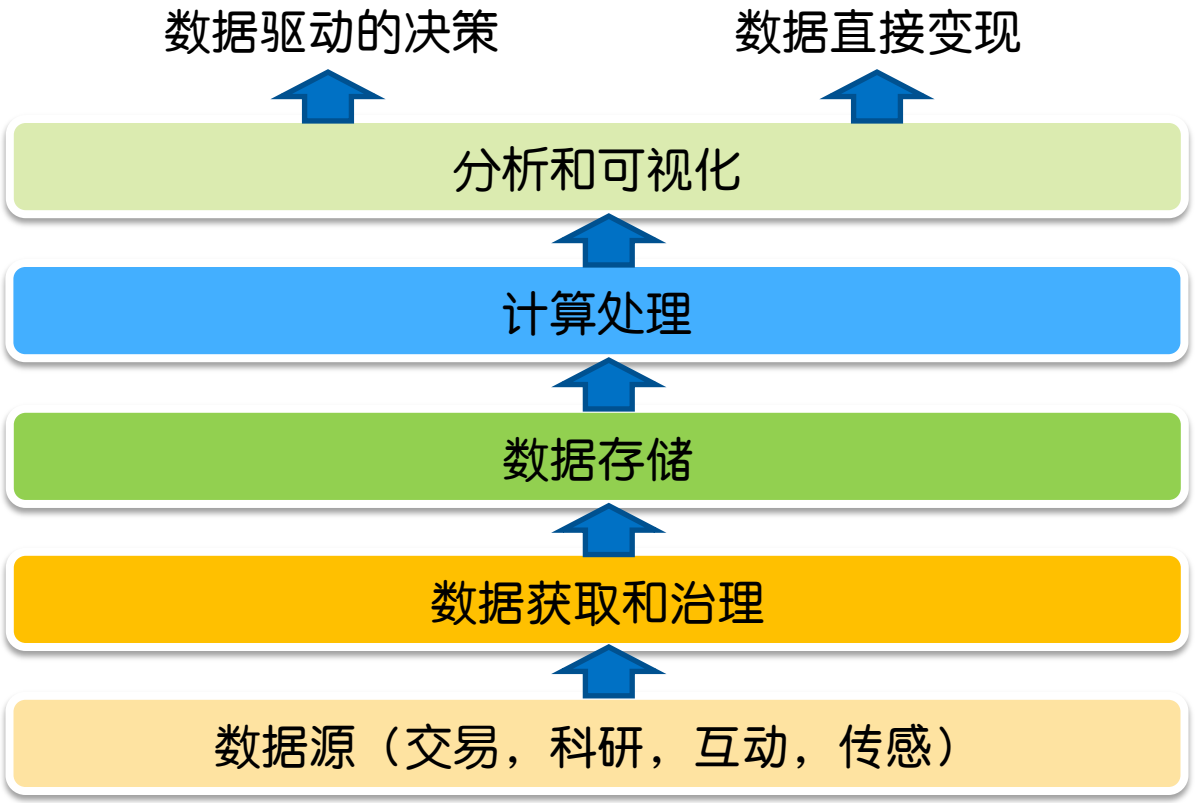
- 低功耗的识别、联想和推理能力
- ANN accelerator vs. Neuromorphic architecture (e.g. SNN-based)

英特尔大学联合研究


2015: Intel/NSF Synergies
Cyberphysical Systems
Security & Privacy
2015: Intel/NSF Synergies
Visual & Experiential Computing
Spark
GraphLab
Carnegie Mellon
Embedded Computing
Carnegie Mellon
Cloud Computing
PETUUM
TU-Darmstadt
Secure Computing
National Taiwan University
Connected Contextual Computing
DataHub
MIT
H-Store
Big Data
SciDB
Tsinghua, SEU, USTC
Mobile Networking & Comms
University of Washington
Pervasive Computing
Stanford
Visual Computing
Imperial/University College London
Sustainable Connected Cities
UC Berkeley
Secure Computing
UC Irvine
Social Computing
Saarland University
Visual Computing
Technion, Hebrew University
Computational Intelligence

敬谢聆听

生命周期



技术栈


应用和服务（包括数据服务）
可视化
数据分析
计算处理
数据存储和管理
硬件平台：计算、存储和互联
编程和管理工具
数据安全