内存计算哪家强？

上海云人信息科技有限公司

个人介绍

- 吴朱华：上海云人信息科技有限公司的联合创始人兼CEO，国内资深的云计算和大数据专家，之前曾在IBM中国研究院参与过多款云计算产品的开发工作，同济本科，并曾在北京大学读过硕士。2010年底，他和另两位创始人组建了一支十多人的团队，在上海杨浦云基地办公。云人信息科技有限公司目前专注于大数据实时分析，尤其是互联网广告、运营商、证券金融和智能电网等有大数据实时分析需求的行业与企业。2011年中，发表业界最好的两本云计算书之一《云计算核心技术剖析》。在2013年以唯一云计算和大数据的代表初入选“2013年福布斯中国30位30岁以下的创业者”。

什么是内存计算？

- 内存技术就是把数据放在内存中吗？
- 计算机整体架构体系；
- Linux默认的Page Cache机制；
- 在很多场景下，把数据放在内存中得收益是有限的；

图 1.9　一个存储器层次模型的示例

TimesTen

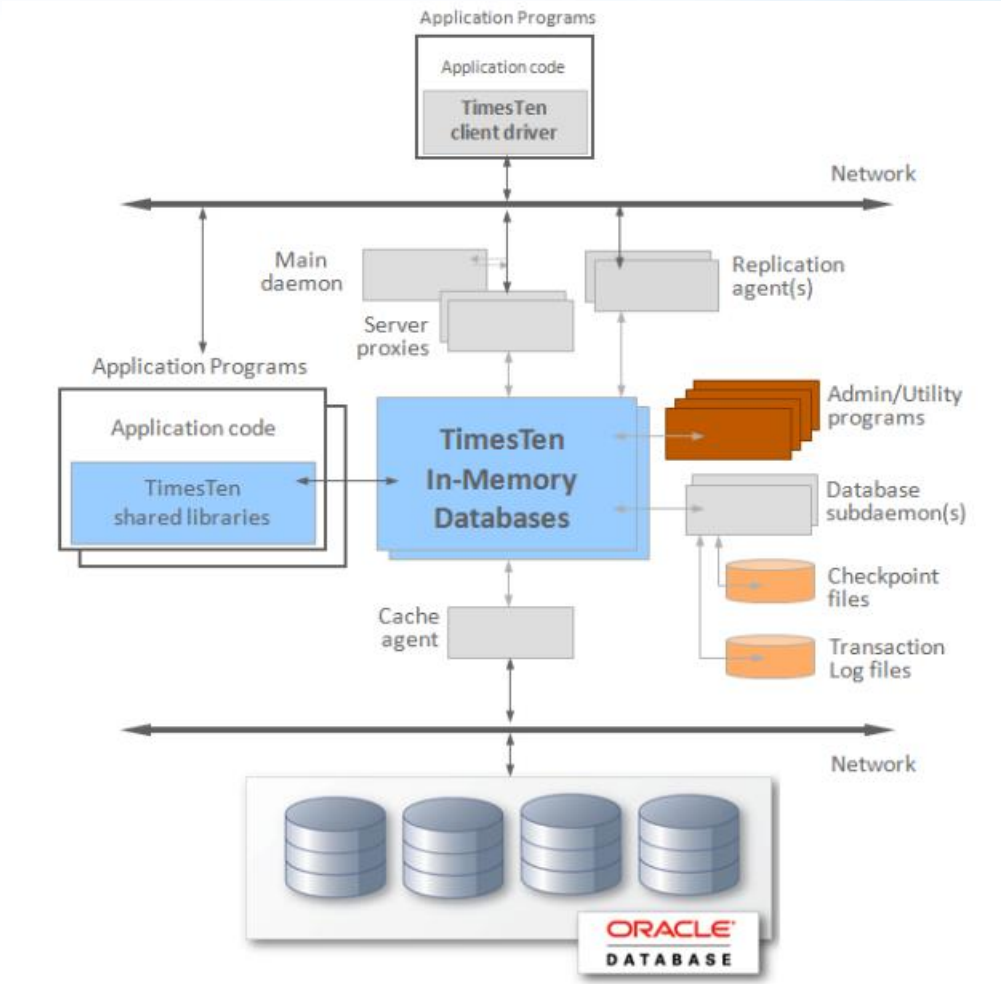
Oracle 内存数据库 TimesTen 是一个优化内存的关系数据库，提供了响应时间极短且吞吐量极高的应用程序，可满足各行业应用程序的需求。

TimesTen (TimesTen) 通过改变数据在运行时驻留位置的假设来提供实时性能。通过在内存中管理数据，并相应地优化数据结构和访问算法，数据库操作能够以最大效率执行，从而大大提高响应速度和吞吐量。TimesTen 是一个可嵌入到应用程序中的数据库，通过消除了进程间通信和网络开销，进一步提高数据库操作的性能。

TimesTen的特色

- 它有一个完整关系数据库的引擎；
- 在数据结构方面，为内存做了很多优化；
- 插入，事务和查询性能优异；
- 可以和应用服务器直连，无需IPC的成本；

TimesTen的架构图



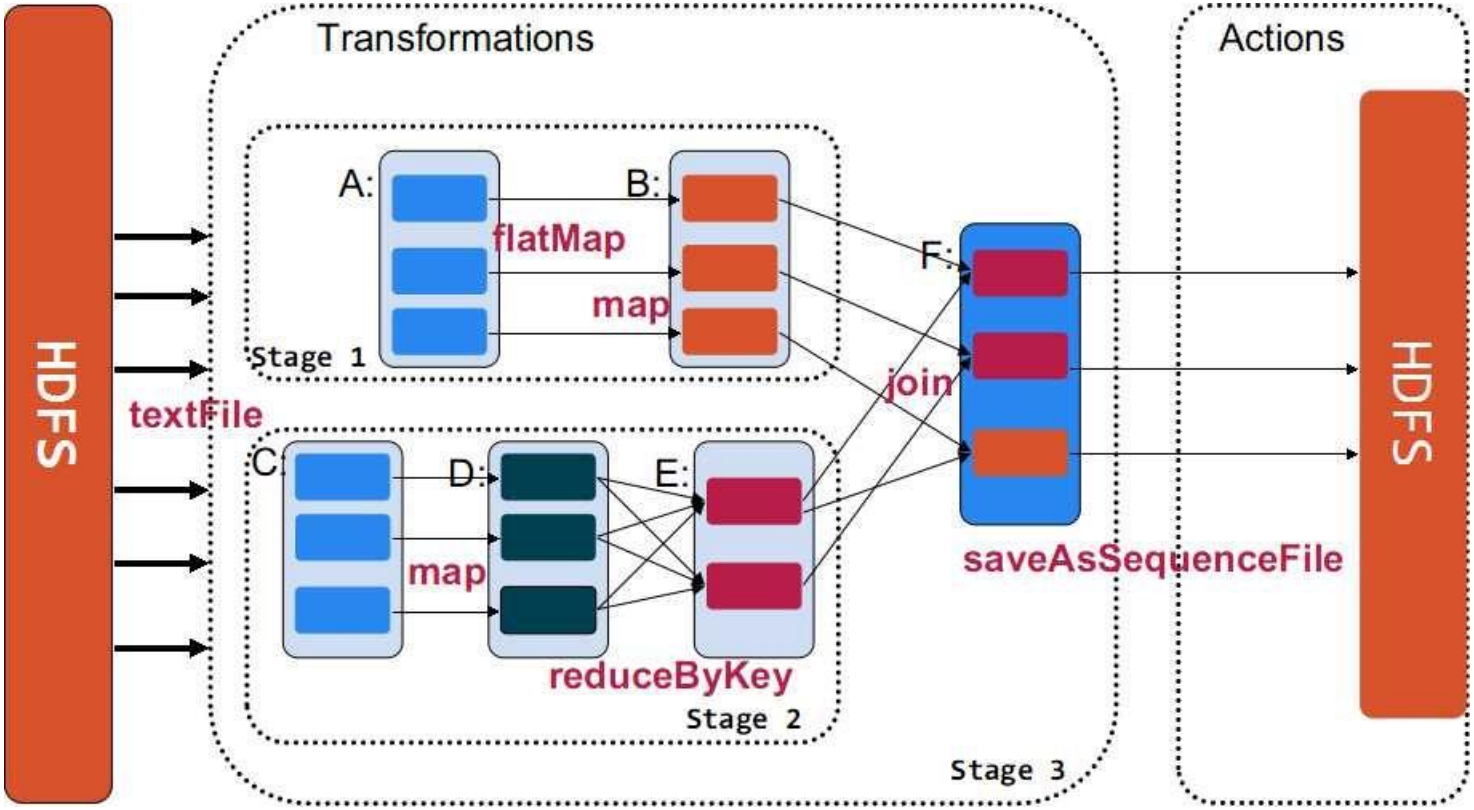
Spark

现在Apache Spark可以说是最火的开源大数据项目，就连EMC旗下专门做大数据Pivotal也开始抛弃其自研十几年GreenPlum技术，转而投入到Spark技术开发当中，并且从整个业界而言，Spark火的程度也只有IaaS界的OpenStack能相提并论。

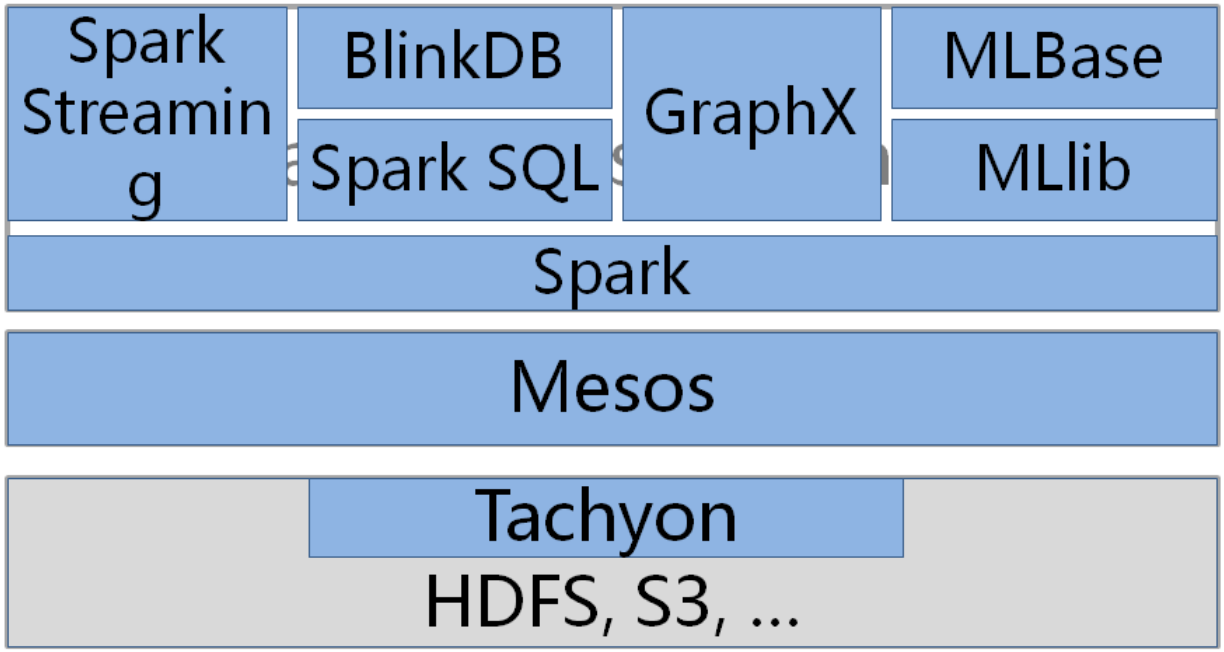
Spark核心机制

- RDD(Resilient Distributed Datasets)；
- 灵活并且有多种类型的算子；
- 常见数据结构Key-Value（Sequence File）；
- 一些个人见解；

Spark核心机制图


HDFS
textFile
Transformations
Stage 1
A:
B:
flatMap
map
Stage 2
C:
D:
map
E:
reduceByKey
F:
join
Stage 3
saveAsSequenceFile
Actions
HDFS

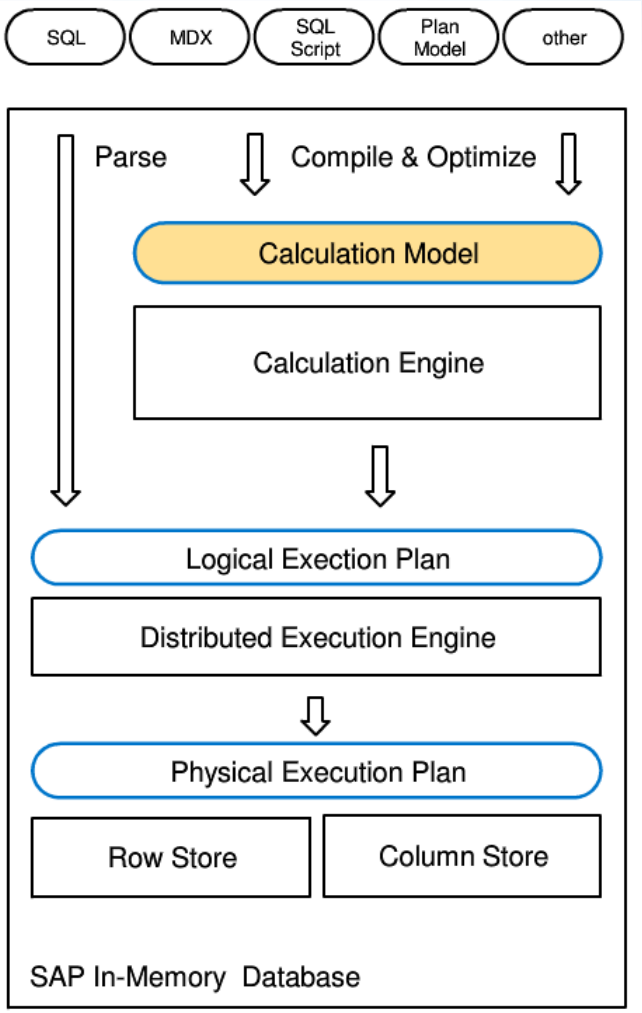
Spark生态系统



HANA

- 从2010年开始的内存计算技术产品，在开源界最火的可能，在商用界，最有代表性的莫过于SAP的旗舰级内存技术平台HANA。

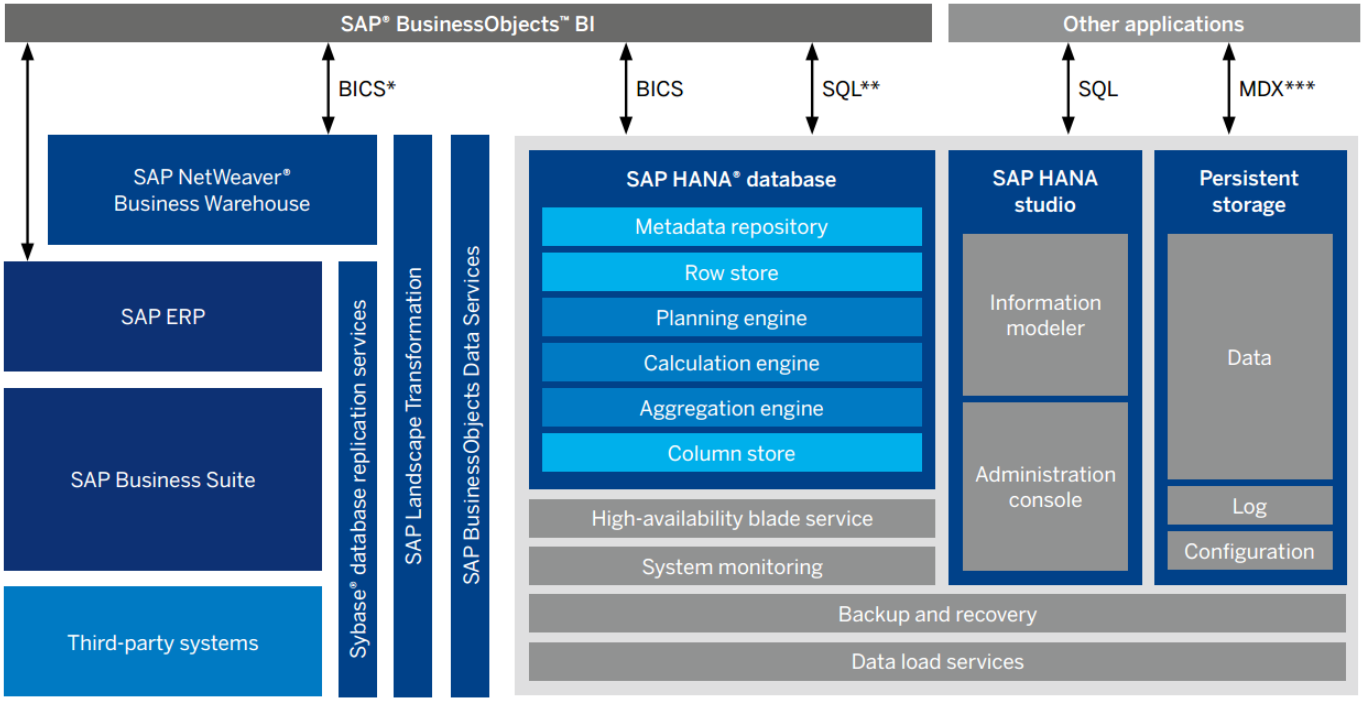
HANA计算引擎图


SQL
MDX
SQL Script
Plan Model
other
Parse
Compile & Optimize
Calculation Model
Calculation Engine
Logical Exection Plan
Distributed Execution Engine
Physical Execution Plan
Row Store
Column Store
SAP In-Memory Database

HANA的核心机制

- 充分利用INTEL最新的CPU特性，并且设计Cache友好的数据结构；
- 有公共L语言，并通过LLVM来进行动态编译；
- 使用SSD做快照，对性能的压榨不遗余力；
- 同时支持行列存；

HANA产品全貌

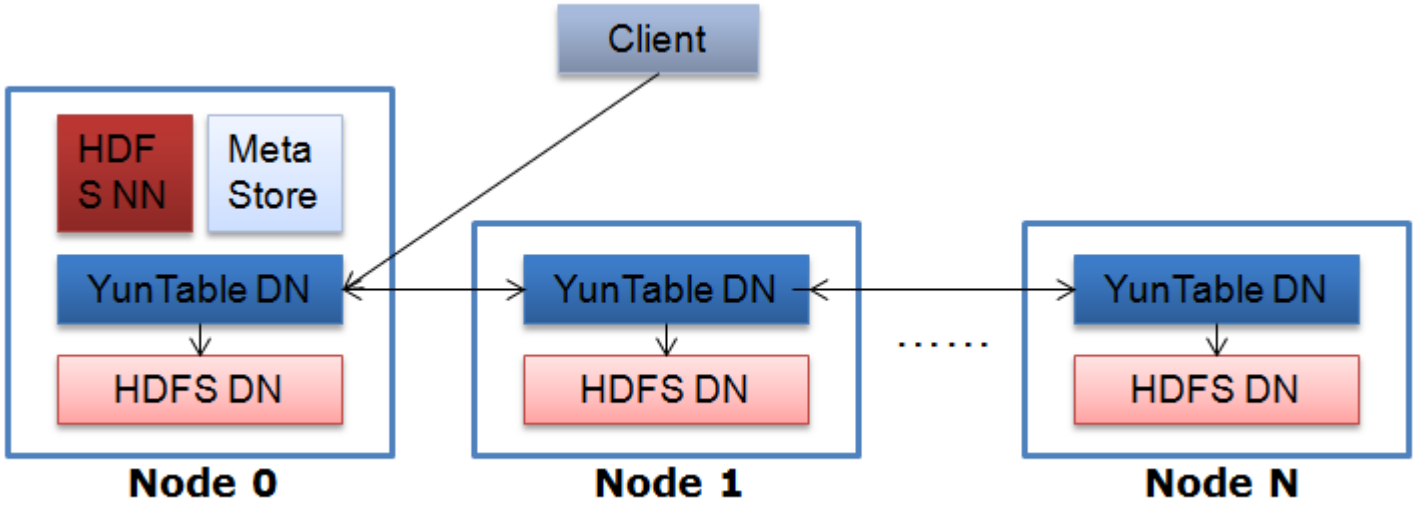


*BICS = business intelligence consumer services, **SQL = structured query language, ***MDX = multidimensional expression

YunTable

- YunTable MPP内存计算数据库，采用MPP（大规模并行处理），列存2.0，动态数据分发，In-Memory Computing（内存计算）等多项创新技术，现在版本是4.0版本，发布日期为2014年11月底；
- 在内存计算设计方面，我们比较接近HANA做法，使用SIMD指令集，以及基于LLVM的动态编译技术来进行提速，即使数据不缓存在内存中。

系统架构


Client
HDF
S NN
Meta
Store
YunTable DN
HDFS DN
YunTable DN
HDFS DN
.....
YunTable DN
HDFS DN
Node 0
Node 1
Node N

核心特性

- 支持MPP，自动线性动态扩展至数百台集群；
- 提供全面的SQL支持，并提供多平台的SQL驱动；
- 在大数据情况下，对数据进行秒级的实时分析，包括复杂查询，已经多个大表之间的Join；
- 数据保存在HDFS上面，保证数据可靠性；
- 采用通用的x86硬件，降低使用成本；

核心技术

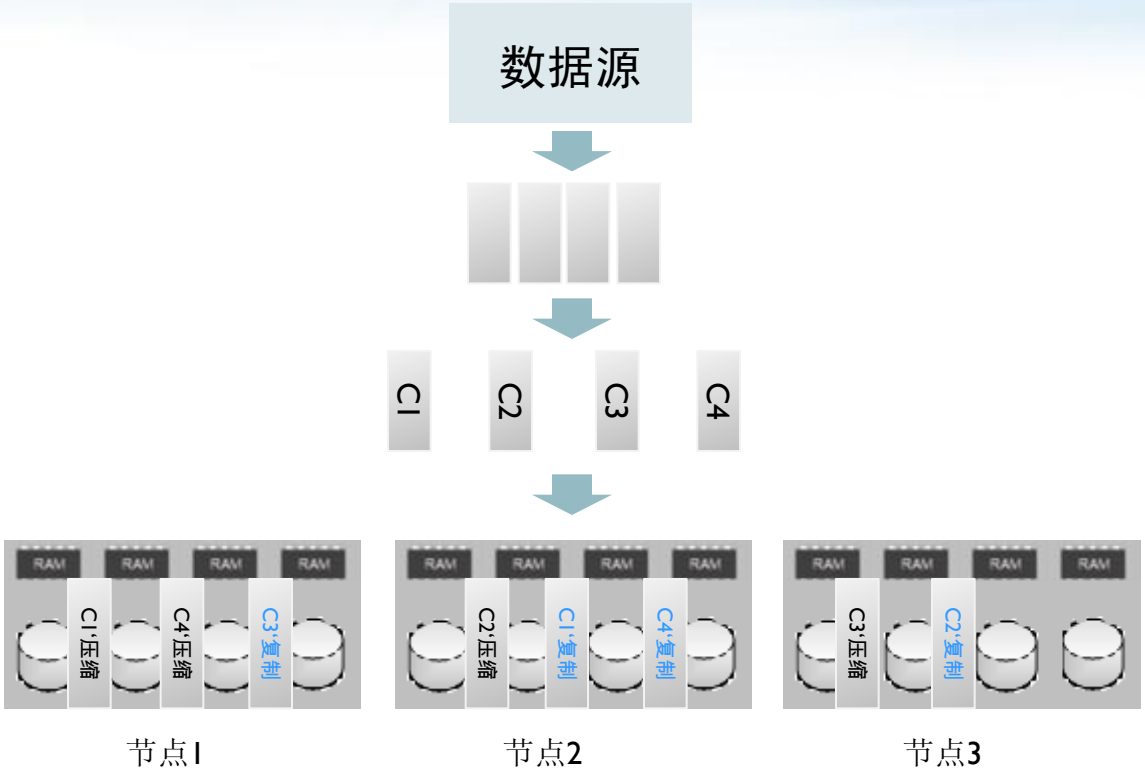
MPP

列存2.0

动态数据分发

In-Memory Computing

MPP（大规模并行处理）



并行：数据复制分布存储在不同的节点上并行处理

高可用性：任何节点宕机将不影响数据完整和业务连续性

列存 2.0

数据源原始结构

赵	25	男
钱	25	男
孙	24	男
李	30	男
周	31	女

映射到存储

- 基本功能
  - 减少I/O
  - 高效的数据压缩
- 高级功能
  - 快速数据过滤
  - 字典Encoding
  - 数据自动排序

行式的数据组织

列式的数据组织

存储地址

数据动态分发

1. **BROADCAST**：让较小的表能够完整地分发到每个DataNode上，这样能快速地完成两表之间的Join，并且尽可能地降低网络I/O；

2. **REHASH**：让两个大表能够一起动态再分配（Hash机制），使的每个DataNode能获取到其所需两张大表的部分数据，从而完成Join；

3. **LOCALIZE**：在导入数据的时候，预先完成Broadcast和Hash来进一步提升性能；

4. **PIPLINE**：只要收到Join数据的一部分，就开始执行Join，而不是等待所有数据；

内存计算

1. **Vector Processing**：在执行SQL指令的时候，使用最新的INTEL SSE4.1 和 SSE4.2 指令集来加快指令执行；

2. **Runtime Code Generation**：能在执行查询前，动态编译查询指令，从而避免传统非常耗时的Switch-Case，并且提高Cache命中率；

3. **SSD/Disk Support**：数据不一定需要限定在内存中，数据在SSD和硬盘中依旧可以享受内存计算的好处；

查询示例1 TPC-H Q3

```
select
  l_orderkey, sum(l_extendedprice*(1-l_discount)) as revenue, o_orderdate, o_shippriority
from
  customer c join orders o
    on c.c_mktsegment = 'FURNITURE' and c.c_custkey = o.o_custkey
  join lineitem l
    on l.l_orderkey = o.o_orderkey
where
  o_orderdate < '1995-03-06' and l_shipdate > '1995-03-06'
group by l_orderkey, o_orderdate, o_shippriority
order by revenue desc, o_orderdate
limit 10;
```

在4台中低配服务器上面，做了2张大表（分别6亿，1.5亿）和1张中表（1.5千万）的三表Join，查询耗时仅为18.7秒！！！！！！！

PeopleYun

YunTable路线图

1. 5.0 版本，将于15年3月发布，主要提供文本分析工具和数据挖掘功能，以及OLAP展现工具，使YunTable成为业界第一流的内存计算平台；

总结

- 个人认为内存计算这个概念很类似云计算，很难用学术语言去界定；
- 多个内存技术都在很多方面有了突破，新的基于SIMD和LLVM的计算引擎，新的基于内存的数据结构，以及多种新的计算模型；
- 用户在选择内存计算时候，可以根据自己的需求来选择；

Q & A

Thank You !