WHO IS SAP

SAP is the first global software company founded in 1972
- Market leader in enterprise SW (ERP, Planning, in-Memory Computing, ...)
- 67,000+ Employees Globally
Who is SAP? Mobile First, Cloud, Open Planform

Simplified User Experience

Live KPIs and Summaries, Connected to Precise Details in a Collaborative User Experience

Applications

Analytics

SAP HANA Platform

Suppress the Noise. Focus on the Key Factors.
SAP Business Applications and Analytics

- Sales Order Management
- Financial/Mgmt Accounting
- Business Intelligence
- Production Planning
- Talent Management
비즈니스에서의 빅 데이터 분석 문제

빅 데이터 예: 애플 앱 스토어
- 월평균 20억 건 다운로드 x 36개월 → 720억 (72 billion) 건의 다운로드 레코드
- 앱 다운로드 레코드당 100B → 7.2 TB (= 72 billion x 100B) 데이터 관리 필요

사람이 활용할 수 있는 정보로 가공하려면?
- 기본 연산
  - 선택 (filtering)
  - 분류 (grouping)
  - 집계 (aggregation)
- 7.2TB를 HDD에 저장하면, 읽는 데만 2시간 소요 (7,200GB / 1GB/s = 7,200s)
- 7.2TB를 DRAM에 저장해도, 읽는 데만 4분 소요 (7,200GB / 15GB/s / 2 channels = 240s)

http://www.scottlogic.com/blog/2014/03/20/app-store-analysis.html
Hardware Advances: Moore’s Law – Supercomputer

Moore’s Law: HW Power는 10년마다 100배씩 증가

2003년

슈퍼컴퓨터 (SGI)
12TB Memory + 1024 processors

2013년

범용 컴퓨터가 비싼 슈퍼컴퓨터 대체
24TB Memory + 960 processing cores
## Hardware Advances: Moore’s Law – DRAM

<table>
<thead>
<tr>
<th>Year</th>
<th>Price/GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>$5.50</td>
</tr>
<tr>
<td>2010</td>
<td>$12.37</td>
</tr>
<tr>
<td>2005</td>
<td>$189</td>
</tr>
<tr>
<td>2000</td>
<td>$1,107</td>
</tr>
<tr>
<td>1995</td>
<td>$30,875</td>
</tr>
<tr>
<td>1990</td>
<td>$103,880</td>
</tr>
<tr>
<td>1985</td>
<td>$859,375</td>
</tr>
<tr>
<td>1980</td>
<td>$6,328,125</td>
</tr>
</tbody>
</table>

Hardware Advances: Servers for In-Memory Computing

Blade Box (2007)

- 2 x 4-core CPU, 32GB DRAM
  (4GB DIMM x 4 x 2 memory channels/system)

HANA Server Box (2013)

- 4 x 10-core CPU, 2TB DRAM
  (32GB DIMM x 8 x 2 channels/cpu x 4 cpu)
Software Advances: Build for In-Memory Computing
Reduce Memory Access Stalls

- **In-Memory Computing**: It is all data-structures (not just tables)
- **Parallelism**: Take advantage of tens, hundreds of cores
- **Data Locality**: On-chip cache awareness

*Using Intel Ivy Bridge for approximate values. Actual numbers depend on specific hardware.*
In-Memory Computing – Columnar Data Structures

<table>
<thead>
<tr>
<th>Order</th>
<th>Store</th>
<th>...</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>456</td>
<td>France</td>
<td>...</td>
<td>1.99</td>
</tr>
<tr>
<td>457</td>
<td>Italy</td>
<td>...</td>
<td>4.99</td>
</tr>
<tr>
<td>458</td>
<td>Italy</td>
<td>...</td>
<td>9.99</td>
</tr>
<tr>
<td>459</td>
<td>Spain</td>
<td>...</td>
<td>0.99</td>
</tr>
</tbody>
</table>

SAP HANA: column order

Typical Database

```
SELECT Country, SUM(sales) FROM SalesOrders
WHERE Product = 'corn'
GROUP BY Country
```
SAP HANA: Dictionary Compression

**Column “Name“ (uncompressed)**
- Miller
- John
- Millman
- Zsuhwalski
- Baker
- Miller
- Jones
- Miller
- Johnson
- John
- Millman

**Column “Name“ (dictionary compressed)**

- Value-ID array:
  - One element for each row in column
  - Value IDs:
    - 4
    - 1
    - 5
    - N
    - 0
    - 4
    - 3
    - 4
    - 2
    - 1
    - 5

- Dictionary:
  - 0: Baker
  - 1: John
  - 2: Johnson
  - 3: Jones
  - 4: Miller
  - 5: Millman

- Value ID, implicitly given by sequence in which values are stored

- Sorted
  - Value
  - Zsuhwalski
## SAP HANA: Multi-Core Parallelization

### Diagram

![Diagram showing multi-core parallelization](image)

### Table

<table>
<thead>
<tr>
<th>Col A</th>
<th>Col B</th>
<th>Col C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000032</td>
<td>4545</td>
<td>2500</td>
</tr>
<tr>
<td>67867868</td>
<td>76</td>
<td>21</td>
</tr>
<tr>
<td>2345</td>
<td>6347264</td>
<td>78675</td>
</tr>
<tr>
<td>89988657</td>
<td>435</td>
<td>3432423</td>
</tr>
<tr>
<td>234123</td>
<td>3434</td>
<td>89089</td>
</tr>
<tr>
<td>21</td>
<td>1252</td>
<td>562356</td>
</tr>
<tr>
<td>2342343</td>
<td>342455</td>
<td></td>
</tr>
<tr>
<td>78787</td>
<td>3333333</td>
<td></td>
</tr>
<tr>
<td>9999993</td>
<td>8789</td>
<td></td>
</tr>
<tr>
<td>13427777</td>
<td>4523523</td>
<td></td>
</tr>
<tr>
<td>23423</td>
<td>6767312</td>
<td></td>
</tr>
<tr>
<td>123123123</td>
<td>789976</td>
<td></td>
</tr>
<tr>
<td>1212</td>
<td>20002</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>2346098</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>454544711</td>
<td>78787</td>
<td>3665364</td>
</tr>
</tbody>
</table>

**Processed by:**

- Core 1
- Core 2
- Core 3
- Core 4
HANA Development Process

Engine Developer

Workstation

Submit a change

Central Code

HANA

Make & Test Farm

~2,000 builds a day
~85,000 test suite executions a day (> 350,000 regression test cases are grouped into ~900 test suites)
~1,000 servers with 40 cores and 1TB memory

Check out

Merge a change

Submit a change

Merge a change
Development Tools

- Git, Gerrit, Jenkins, BugZilla, Jira, …
- Static Code Analysis: Coverity, ConQAT, …
- Performance Profiling
- Code Coverage
Performance Profiling
Code Coverage

Git Branch: orange
Head Commit: 8db2c29542b3190b464177900d1f4716e24e12cc
Make ID / Test Run: 1560595
Test: Regression_NewDB_COV_part1..12
Date of Coverage: 2014-06-11
Current View: $OWN/sys/src/ptime/query/plan_executor/dml/qe_parallel_util.cc

```
104     :         case PROJECT_FETCH:
105          :             {
106     742654 :                 config_threads = config.prefetchThread;
107     742638 :                 break;
108          :             }
109          :               }  //FIXME? need to decide policy
110          :         case PROC_INSERT_TREX_FETCH:
111          :             {
112     844158 :                 config_threads = max_thread;
113     844158 :                 break;
114          :             }
115          :               }  //FIXME : hash build is currently not parallelized.
116          :         case HASHJOIN_FETCH:
117          :             {
118     24321147 :                 config_threads = config.hashBuildThread;
119     24329125 :                 break;
120          :             }
121          :               }  //FIXME : hash build is currently not parallelized.
122          :         case SHARED_SUBTREE_FETCH:
123          :             {
124     1589096 :                 config_threads = config.subtreeFetchThread;
125     1589083 :                 break;
126          :             }
127          :         default:
128          :             {
129     738712942 :                 if(config_threads < 1) return 1;
130     738712942 :                 else if(config_threads < max_thread) return config_threads;
131     684177859 :                 else return max_thread;
132          :             }
```
One Global HANA Platform Team
Working Globally

HANA Workshop in Seoul, 2012