China Big Data and HPC Initiatives Overview

Xuanhua Shi

Services Computing Technology and System Laboratory
Big Data Technology and System Laboratory
Cluster and Grid Computing Laboratory
Huazhong University of Science and Technology, Wuhan, China
xhshi@hust.edu.cn
Outline

- In-Memory Computing
- New Funding by MOST
  - HPC Initiatives (2016-2020)
  - Big Data Initiatives (2016-2020)
In-Memory Computing: Lifting the Burden of Big Data – Aberdeen Group

Figure 3: Satisfaction and Trust in Business Data

Table 1: More Data, More Speed, More Efficiency

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Use in-memory computing (n = 33)</th>
<th>Don't use (n = 163)</th>
<th>In-memory Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median amount of active business data</td>
<td>38 terabytes</td>
<td>18 terabytes</td>
<td>2.1 times more data</td>
</tr>
<tr>
<td>Median amount of data analyzed</td>
<td>14 terabytes (37% of all data)</td>
<td>4 terabytes (22% of all data)</td>
<td>3.5 times more data</td>
</tr>
<tr>
<td>Average response time for data analysis or query</td>
<td>42 seconds</td>
<td>75 minutes</td>
<td>107 times faster</td>
</tr>
<tr>
<td>Data volume processed per hour</td>
<td>1200 terabytes</td>
<td>3.2 terabytes</td>
<td>375 times more efficient</td>
</tr>
</tbody>
</table>

Source: Aberdeen Group, December 2011
Traditional systems exchange data pages between memory and disk when computing with big data: **Expensive IO cost**

- Speed of Disk: **ms**  \( \gg \)  Speed of Memory: **ns**
- In-Memory Computing: CPU reads data from memory and provides real-time data processing.
Downsides of DRAM Refresh

• **Energy consumption**: Each refresh consumes energy

• **Performance degradation**: DRAM bank unavailable while refreshed

• **QoS/predictability impact**: (Long) pause times during refresh

• **Refresh rate limits DRAM capacity scaling**
Emerging Non-volatile Memory (NVM) Technologies

Phase-change RAM (PCRAM)

Magnetic RAM (MRAM)

Resistive RAM (RRAM)

From left to right: Filament-based, Interface-based and PMC

Emerging Non-volatile Memory (NVM) Technologies
Storage Class Memory

- SCM is a new class of data storage and memory devices
- SCM blurs the distinction between
  - MEMORY (= fast, expensive, volatile) and
  - STORAGE (= slow, cheap, non-volatile)
- Characteristics of SCM
  - Solid state, no moving parts
  - Short access times (~ DRAM like, within an order-of-magnitude)
  - Low cost per bit (DISK like, within an order-of-magnitude)
  - Non-volatile (~ 10 years)
Reconstruction of Virtual Memory Architecture: Break the I/O Bottleneck

1980
- CPU
- RAM
- DISK
- TAPE
- Logic
- Memory
- Active Storage
- Archival

2008
- CPU
- RAM
- DISK
- TAPE
- FLASH SSD

2017
- CPU
- RAM
- SCM
- DISK
- TAPE

2018?
- CPU
- RAM
- SCM
- TAPE
New In-Memory Computing Architecture

The new in-memory computing architecture supports the shift from computing-centric to combination of computing and data.
Challenges of Hybrid Memory Systems

• How should SCM-based (main) memory be organized?

  - Partitioning
  - Should DRAM be a cache or main memory, or configurable?
  - What fraction? How many controllers?
Challenges of Hybrid Memory Systems

• Data allocation/movement (energy, performance, lifetime)
  – Who manages allocation/movement?
  – What are good control algorithms?
  – How do we prevent degradation of service due to wearout?

• Design of cache hierarchy, memory controllers, OS
  – Mitigate PCM shortcomings, exploit PCM advantages

• *Persistent* data can be *randomly* and *synchronously* addressed
  – Huge non-volatile address spaces, memory-mapped DB, persistent objects...
  – Should SCM be used like I/O or like memory or in a totally new way?
Challenges of Hybrid Memory Systems

• **Software Architecture**
  – Should one make SCM *visible* to applications software?
  – If visible, in which form?
    • New APIs, libraries, memory models, new I/O devices,...

• **Databases, Business Intelligence and Streams are first impacted**
  – Data-intensive HPC - predictable execution time of complex business analytics - streaming search
Challenges of Enabling and Exploiting NVM

• Enabling NVM and hybrid memory
  – How to tolerate errors?
  – How to enable secure operation?
  – How to tolerate performance and power shortcomings?
  – How to minimize cost?

• Exploiting emerging technologies
  – How to exploit non-volatility?
  – How to minimize energy consumption?
  – How to exploit NVM on chip?
Technology and System of In-Memory Computing for Big Data Processing

• Hybrid Memory Architecture for In-Memory Computing System
• System Software for In-Memory Computing System
• Parallel Processing for In-Memory Computing System
• Data Management for In-Memory Computing System
Technology and System of In-Memory Computing for Big Data Processing

• Project Overview
  – Total budget: 170M RMB
  – Period: January 2015 – December 2017
  – Participants
    • Inspur
    • Huawei
    • Sugon
    • Shanghai Jiaotong University
    • Huazhong University of Science and Technology
    • Chongqing University
    • National University of Defense Technology
    • Huadong Normal University
    • Jiangnan Computing Institute
Technology and System of In-Memory Computing for Big Data Processing

• Project Mission

  – Hybrid NVM-based high reliable, massive storage, and low power in-memory computing system, the capacity of NVM in each node should be in TB level, supporting zero bootup

  – System software and simulation platform for in-memory computing system

  – Parallel processing system for in-memory computing system

  – In-memory database for hybrid memory architecture to support decision making and other data management applications
System Architecture
Full System Simulator

Designed Based On:
MARSSX86+
NVMain

- More flexible (supports NVMMain, DRAMSim, HybridSim as main memory simulator)
- Simulate memory system precisely, easy to configure
Pros and Cons of DDPS

User Program

UDT

UDF

Distributed Data Process System

Stage1 → ... → StageN

Produce data objects with different lifetime

Platform with memory management

Apply Memory

Allocate

GC

Benefit from the high-level object-oriented language

The automatic memory management of platform effect the performance seriously。
Deca: Exploiting Raw Data of In-Memory Data Objects in Distributed Data-Parallel Systems

- Completely decomposing in-memory data objects to eliminate the references invocation and reduce frequent garbage collection in JVM
- A system to automatic converse the user codes, decompose data objects and manage in-memory raw data
- Speedup from 22.7x to 41.6x, compare with Apache Spark
Hadoop Engine: IO-intensive

1. I/O operations
2. I/O operations
3. I/O operations
4. I/O operations
5. I/O operations
6. I/O operations
7. I/O operations

Map Task

Local Disk

Hdfs

NetWork

Reduce Task

Final output

I/O wait

16/6/17
Mammoth: Memory-Centric MapReduce System

- A novel rule-based heuristic to prioritize memory allocation and revocation mechanism
- A multi-threaded execution engine, which realize global memory management
- Compatible with Hadoop
- Sources available at Github and ASF
- IEEE Computer Spotlight
Landscape of Disk-Based and In-Memory Data Management Systems (2014)
Outline

- In-Memory Computing
- New Funding by MOST
  - HPC Initiatives (2016-2020)
  - Big Data Initiatives (2016-2020)
HPC Initiatives (2016-2020)

• Two 100PFlops Supercomputer
  – One is located Wuxi, another is located in Guangzhou

• E-scale architecture

• E-scale processors

• High-speed network

• HPC software stack

• Co-design: Aircraft design and Weather forecasting

• Some typical applications
Big Data Initiatives (2016-2020)

- Big data infrastructure
  - New storage system
  - Data flow based data analyzed stack
  - Domain specific data management system

- Data-driven software

- Data analyze applications and Human-like intelligence
  - From data to knowledge
  - Large scale objects recognition
Thanks!