MPI based Cloud Computing in Environment Probing

Yucheng Guo, Chun Liu, Peng Wu, Xiayi Tang, Qingping Guo
Parallel Processing Lab
School of Computer Science and Technology
Wuhan University of Technology
Wuhan, China

Environment Probing Characteristics

- Massive Data: Sense Data, Aerial survey Data, Seismic Data…, Data CDs produced every year could reached the moon; Data Storage, Data Maintenance, Data Mining, Data Intensive Problem
- Science Computing Intensive Problem: Weather Forecast, Seismic Probing of Minerals, oil …
- Environment Probing Characteristics : Data Intensive as well as Science Computing Intensive Problems

Distributed Computing

- Parallel Computing: Vector Processor System, Data Driving System, Multi Transputer System, Multi processor system---Super Computers; Million Processors in One System
- Network Computing: Multi Computers Networked with a LAN, eg. NOW system in University of New York---MultiComputer Systems
- Grid Computing
- Cloud Computing

Parallel Computing vs Cloud Computing

- Parallel Computing focus on Scientific Computing, needs more strictly synchronization, solves computation intensive problems; Well-known standard: PVM, MPI(MPI 1, MPI 2, MPI 3 will be appeared soon)
- Cloud Computing focus on Data Intensive Processing, Distributed Data Base, Data Mining, using Asynchronous approach to achieve Data synchronization;

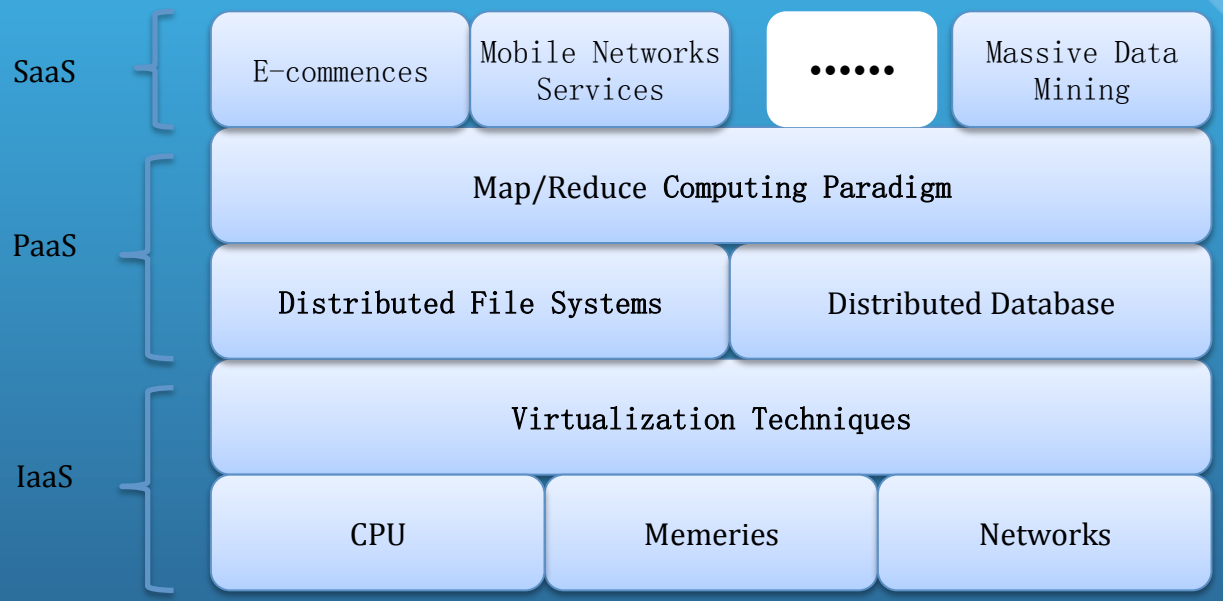
Well-Known Cloud Computing Projects

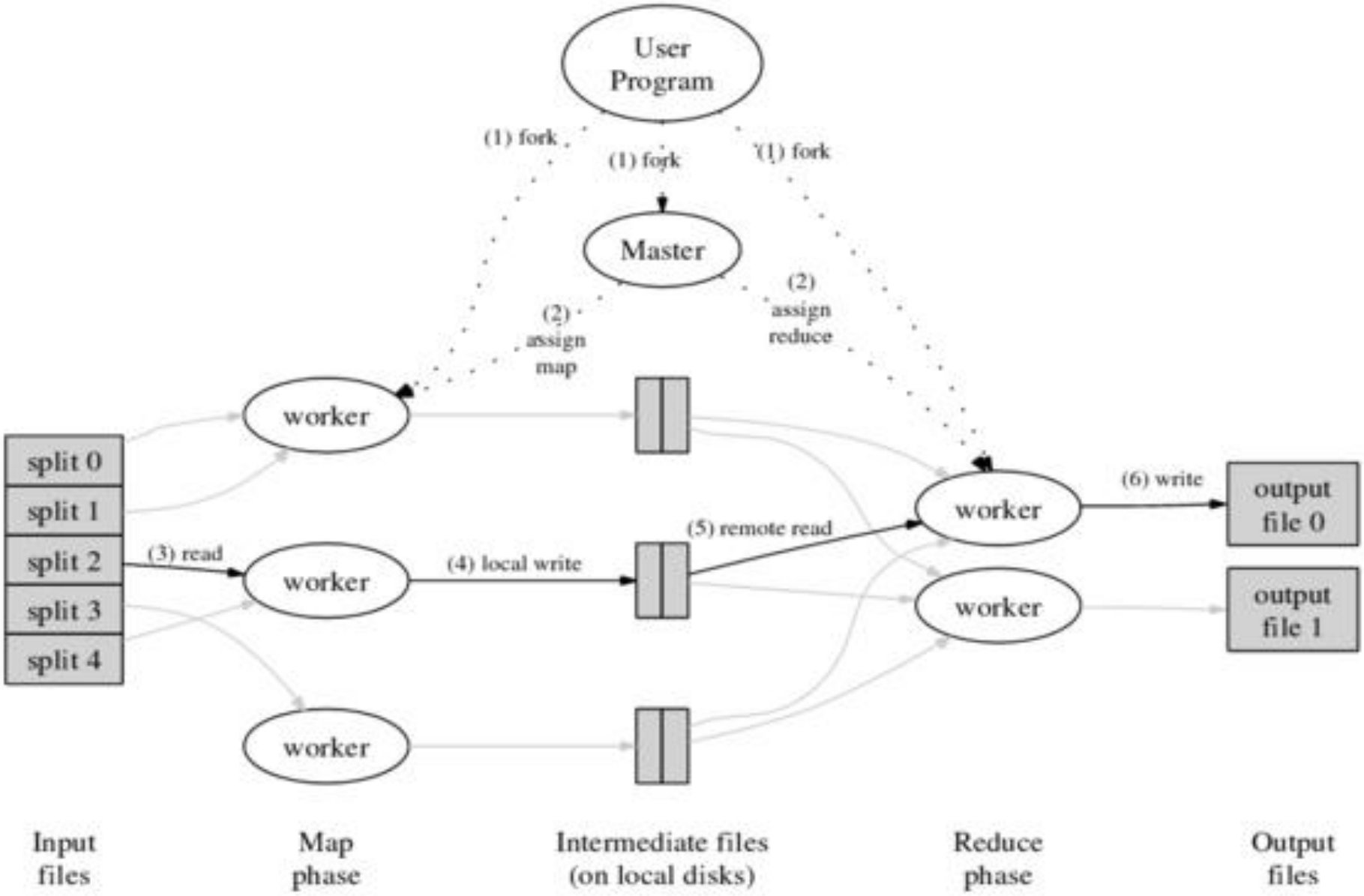
- 2004 GFS (Google File System) and MapReduce Paradigm, Google
- 2005 Hadoop Cloud Computing Platform, Apache
- 2006 EC2 Elastic Cloud Computing Platform, Amazon
- 2007 Blue Cloud, IBM
- 2008.05 Hydrazine Project, SUN Company
- 2008.07 Open Cirrus, the Cloud Computing Test Bed, HP、Intel and Yahoo
- 2009 Universe Computing System (UCS), Cisco
- 2009 Virtual Computing Environment Federation,（Cisco, EMC,Vmware)
- 2009.04, IBM/Google Cloud Computing Alliance, 14 Universities involved, Supported by the NSF, 5 million Dollars Project
- 2010年, Microsoft Azure, Microsoft Cloud Computing Service Platform

Well-known Cloud Computing Platforms

- IBM CloudBurst
- VCE Vblock
- HP CloudStart
- Oracle Exalogic
- Elastic Cloud
- 2009.01, E-commence Cloud Computing Center, Ali Software Company, Nanjing, China,
- 2009.07, China Chemical Enterprise Cloud Computing platform
- 2010 Dozens Research Project supported by NSFC
- 2011 Dozens more Research Project Reviewed by NSFC

General Cloud Computing Architecture




User Program
(1) fork
(1) fork
(1) fork
Master
(2) assign map
(2) assign reduce
worker
worker
worker
worker
worker
split 0
split 1
split 2
split 3
split 4
(3) read
(4) local write
(5) remote read
(6) write
output file 0
output file 1
Input files
Map phase
Intermediate files (on local disks)
Reduce phase
Output files

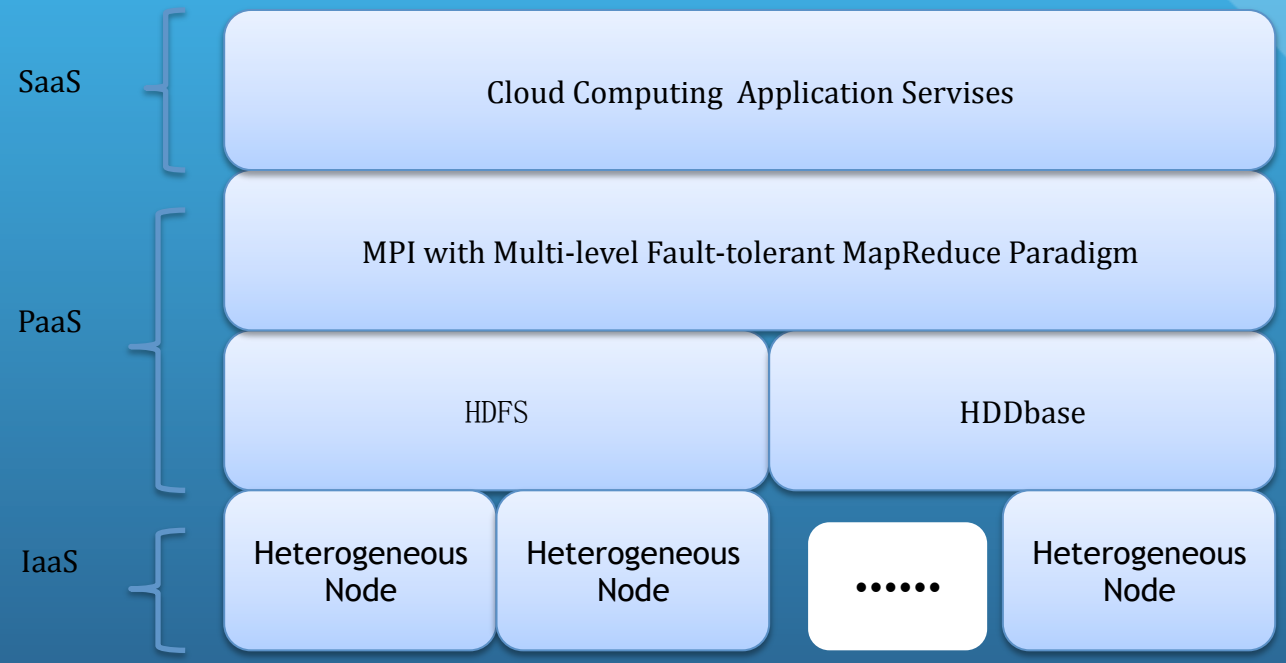
Two drawbacks in Distributed Computing Platforms

- MPI platform lacks massive data processing ability
- Current cloud computing platforms do not fully support intensive science computing

Data Processing intensive mixed with Scientific Computing intensive problems required a platform to support such kind of applications

Especially in Environment Probing Applications

MPI based Cloud Computing



- 云计算(Cloud Computing)是一种新近提出的计算模式，是分布式计算、并行计算和网格计算的发展。目前国内外诸多大型机构企业纷纷提出了各自的云计划。由于目前主要使用的云计算平台 均为国外公司所开发。其价格昂贵，后期的维护费用也十分惊人。而国内公司主要是在国外技术的基础上发展自己的私有云，没有独立的知识产权和核心技术。

- 另一方面，同时具有数据密集和计算密集两个特点的高性能海量数据计算，难以用现有的云计算技术解决。其原因在于：

- 目前主流云计算平台底层普遍采用虚拟化技术，其上所有软件和应用均运行在虚拟硬件之上，这种策略必然带来一定程度上的性能降低，有文献指出，其性能损耗可达20%左右。

- 现有云计算的任务发布和结果收集计算模型——MapReduce，其内部实现对中间数据采用先存储数据再读出转发处理的策略，当中间数据规模变大、个数增多时，这种模式必然产生大量的无用的磁盘I/O操作，使其效率不能满足高性能计算的应用需求。有文献指出，采用中间数据直接传递的实现方式，在一定条件下能将效率提高5倍之多。

- 因此，本项目拟自主研发一种高性能云计算平台。该平台不经过虚拟化，直接使用异构计算节点构建云平台底层；使用融合多层容错的MPI技术和多线程技术重写MapReduce编程模型，在计算中避免大量的无用的I/O操作，从而提高云计算的效率，以满足高性能云计算的需要。

- 本项目的关键技术和创新点归纳如下：
- 在节点异构环境下，不使用目前流行的虚拟化技术，而是利用MPI良好的异构环境开发能力，直接使用异构硬件架设云基础设施服务层，减少虚拟化对云底层硬件性能的影响，提高云平台效率，是本项目的一个重要创新点。
- 虽然MPI有进行高性能计算的优异能力，但容错容灾能力一直以来是MPI的一个重要缺陷。此缺陷限制了MPI在海量数据处理上的应用。因此，研究MPI容错容灾技术，实现多种不同层次的容错容灾，弥补MPI容错容灾能力的不足，是本项目的一个关键技术和创新点。
- 用MPI和多线程编程技术，对任务进行节点间和节点内的两级划分，节点间任务的粗划分，节点内进行任务的细化分。节点间实现MPI进程级并行，节点内实现线程级并行，充分挖掘异构节点计算能力，是本项目的关键技术之一。
- 使用多层次容错MPI优化和实现MapReduce编程模型，对中间结果进行处理，减少不必要的I/O操作，提高云计算速度和效率，也是本项目的重要创新和关键技术。