Large Scale Sky Computing Applications with Nimbus

Pierre Riteau
Université de Rennes 1, IRISA
INRIA Rennes – Bretagne Atlantique
Rennes, France
Pierre.Riteau@irisa.fr
INTRODUCTION TO SKY COMPUTING
IaaS clouds

- On demand/elastic model
- Pay as you go
- Access to virtual machines with administrator privileges
  - Portable execution stack
- Commercial providers
  - Amazon EC2 => “infinite” resource pool (e.g. 10K cores)
- Scientific clouds => limited number of resources
  - Science Clouds
  - FutureGrid Nimbus clouds
Sky Computing

- Federation of multiple IaaS clouds
- Creates large scale infrastructures
- Allows to run software requiring large computational power
Sky Computing Benefits

• Single networking context
  – All-to-all connectivity

• Single security context
  – Trust between all entities

• Equivalent to local cluster
  – Compatible with legacy code
LARGE-SCALE SKY COMPUTING EXPERIMENTS
Sky Computing Toolkit

• Nimbus
  – Resource management
  – Contextualization (Context Broker)

• ViNe
  – All-to-all connectivity

• Hadoop
  – Task distribution
  – Fault tolerance
  – Resource dynamicity
Context Broker

• Service to configure a complete cluster with different roles

• Supports clusters distributed on multiple clouds (e.g. Nimbus and Amazon EC2)

• VMs contact the context broker to
  – Learn their role
  – Learn about other VMs in the cluster

• Ex. : Hadoop master + Hadoop slaves
  • Hadoop slaves configured to contact the master
  • Hadoop master configured to know the slaves
<?xml version="1.0" encoding="UTF-8"?>
<cluster xmlns="http://www.globus.org/2008/06/workspace/metadata/logistics">
  <workspace>
    <name>hadoop-master</name>
    <image>fc8-i386-nimbus-blast-cluster-004</image>
    <quantity>1</quantity>
    <nic wantlogin="true">public</nic>
    <ctx>
      <provides>
        <role>hadoop_master</role>
        <role>hadoop_slave</role>
      </provides>
      <requires>
        <role name="hadoop_slave" hostname="true" pubkey="true"/>
        <role name="hadoop_master" hostname="true" pubkey="true"/>
      </requires>
    </ctx>
  </workspace>
  <workspace>
    <name>hadoop-slaves</name>
    <image>fc8-i386-nimbus-blast-cluster-004</image>
    <quantity>16</quantity>
    <nic wantlogin="true">public</nic>
    <ctx>
      <provides>
        <role>hadoop_slave</role>
      </provides>
      <requires>
        <role name="hadoop_master" hostname="true" pubkey="true"/>
      </requires>
    </ctx>
  </workspace>
</cluster>
ViNe

- Project of the University of Florida (M. Tsugawa et al.)
- High performance virtual network
- All-to-all connectivity
Hadoop

- Open-source MapReduce implementation
- Heavy industrial use (Yahoo, Facebook…)
- Efficient framework for distribution of tasks
- Built-in fault-tolerance
- Distributed file system (HDFS)
Sky Computing Architecture

- Distributed Application
- MapReduce App
- Hadoop
- ViNe
- IaaS Software
- IaaS Software
Grid’5000 Overview

• Distributed over 9 sites in France
• ~1500 nodes, ~5500 CPUs
• Study of large scale parallel/distributed systems
• Features
  – Highly reconfigurable
    • Environment deployment over bare hardware
    • Can deploy many different Linux distributions
    • Even other OS such as FreeBSD
  – Controlable
  – Monitorable (metrics access)
• Experiments on all layers
  – network, OS, middleware, applications
Grid’5000 Node Distribution
FutureGrid: a Grid Testbed

- NSF-funded experimental testbed
- ~5000 cores
- 6 sites connected by a private network
Resources used in Sky Computing Experiments

- 3 FutureGrid sites (US) with Nimbus installations
  - UCSD (San Diego)
  - UF (Florida)
  - UC (Chicago)
- Grid’5000 sites (France)
  - Lille (contains a white-listed gateway to FutureGrid)
  - Rennes, Sophia, Nancy, etc.
- Grid’5000 is fully isolated from the Internet
  - One machine white-listed to access FutureGrid
  - ViNe queue VR (Virtual Router) for other sites
ViNe Deployment Topology

All-to-all connectivity!
Experiment scenario

• Hadoop sky computing virtual cluster already running in FutureGrid (SD, UF, UC)
• Launch BLAST MapReduce job
• Start VMs on Grid’5000 resources
  – With contextualization to join the existing cluster
• Automatically extend the Hadoop cluster
  – Number of nodes increases
    • TaskTracker nodes (Map/Reduce tasks execution)
    • DataNode nodes (HDFS storage)
  – Hadoop starts distributing tasks in Grid’5000
  – Job completes faster!
Job progress with cluster extension

Job progression

- Number of completed tasks
- Number of nodes

- Red: Progress with extension
- Green: Progress without extension
- Blue: Number of nodes

Time in seconds:
- 0
- 240
- 480
- 720
- 960
- 1200
- 1440
- 1680
- 1920
- 2160
- 2400
- 2640
- 2880
- 3120
Scalable Virtual Cluster Creation (1/3)

- Standard Nimbus propagation: scp

Diagram:
- Nimbus Repository
- VM1
- VM2
- VM3
- VM4
- VMM A
- VMM B
Scalable Virtual Cluster Creation (2/3)

- Pipelined Nimbus propagation: Kastafior/TakTuk
Scalable Virtual Cluster Creation (3/3)

- Leverage Xen Copy-on-Write (CoW) capabilities

![Diagram showing the process of scalable virtual cluster creation using CoW capabilities.](Image)
Propagation Performance

![Chart showing the propagation performance for different number of VMs to instantiate. The y-axis represents the instantiation time in seconds, and the x-axis represents the number of VMs. The chart includes bars for SCP, TakTuk, and QCOW.](image-url)
CONCLUSION
Conclusion

• Sky Computing to create large scale distributed infrastructures

• Our approach relies on
  – Nimbus for resource management, contextualization and fast cluster instantiation
  – ViNe for all-to-all connectivity
  – Hadoop for dynamic cluster extension

• Provides both infrastructure and application elasticity
Ongoing & Future Works

- Elastic MapReduce implementation leveraging Sky Computing infrastructures (presented at CCA ‘11)
- Migration support in Nimbus
  - Leverage spot instances in Nimbus
Acknowledgments

• Tim Freeman, John Bresnahan, Kate Keahey, David LaBissoniere (Argonne/University of Chicago)
• Maurício Tsugawa, Andréa Matsunaga, José Fortes (University of Florida)
• Thierry Priol, Christine Morin (INRIA)
THANK YOU!

QUESTIONS?