Introduction to Compute Cloud

Tao Zou

CS 5220
Applications of Parallel Computers
About me

• 3\textsuperscript{rd} year PhD student in Computer Science
  – Databases/Cloud Computing/Distributed Systems
• Have been burning thousands of dollars in
  – Amazon Web Services \((\text{Amazon Pays})\)
  – Microsoft Azure \((\text{Microsoft Pays})\)
  – Rackspace Cloud \((\text{My advisor Pays})\)
• Who is paying for the cost?
Where is the Cloud?
Where is the Cloud?

- Datacenter
Where is the Cloud?
What is a Cloud?

• Service, rather than a product

• Cloud Models:
  
  **Infrastructure as a Service**
  - Provides raw computing resources.
  
  Do anything
  
  Amazon EC2
  
  Rackspace

  **Platform as a Service**
  - Provides programming languages and tools.
  
  Program anything
  
  Microsoft Azure
  
  Google AppEngine

  **Software as a Service**
  - Provides applications.

  Use anything
  
  Salesforce.com
  
  Google Docs

- Hardware focus, higher flexibility
- Application focus, less flexibility,
Amazon Web Service: Elastic Compute Cloud (EC2)

Compute
Amazon Elastic Compute Cloud (EC2)
Amazon Elastic MapReduce
Auto Scaling

Content Delivery
Amazon CloudFront

Database
Amazon SimpleDB
Amazon Relational Database Service (RDS)
Amazon ElastiCache

Deployment & Management
AWS Elastic Beanstalk
AWS CloudFormation

E-Commerce
Amazon Fulfillment Web Service (FWS)

Industry-specific Clouds
AWS GovCloud (US)

Messaging
Amazon Simple Queue Service (SQS)
Amazon Simple Notification Service (SNS)
Amazon Simple Email Service (SES)

Monitoring
Amazon CloudWatch

Networking
Amazon Route 53
Amazon Virtual Private Cloud (VPC)
Elastic Load Balancing
AWS Direct Connect

Payments & Billing
Amazon Flexible Payments Service (FPS)
Amazon DevPay

Storage
Amazon Simple Storage Service (S3)
Amazon Elastic Block Store (EBS)
AWS Import/Export

Support
AWS Premium Support

Web Traffic
Alexa Web Information Service
Alexa Top Sites

Workforce
Amazon Mechanical Turk
Amazon Elastic Compute Cloud (EC2)
Amazon Elastic Compute Cloud (EC2)
Amazon Elastic Compute Cloud (EC2)
Amazon Elastic Compute Cloud (EC2)

Give me 10 instances
Amazon Elastic Compute Cloud (EC2)

Give me 10 instances
Amazon Elastic Compute Cloud (EC2)

Give me 10 instances
Run Computation $\rightarrow$ takes X hours
Shutdown all my instances
Amazon Elastic Compute Cloud (EC2)

Give me 10 instances
Run Computation $\rightarrow$ takes X hours
Shutdown all my instances
Amazon Elastic Compute Cloud (EC2)

Give me 10 instances
Run Computation → takes X hours
Shutdown all my instances
EC2 Pricing Model: Pay As You Go

- Pay only for what you use
  - Machine hours (10 \cdot [X])
  - Type of instances

<table>
<thead>
<tr>
<th>Region: US East (Virginia)</th>
<th>Linux/UNIX Usage</th>
<th>Windows Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard On-Demand Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small (Default)</td>
<td>$0.085 per hour</td>
<td>$0.12 per hour</td>
</tr>
<tr>
<td>Large</td>
<td>$0.34 per hour</td>
<td>$0.48 per hour</td>
</tr>
<tr>
<td>Extra Large</td>
<td>$0.68 per hour</td>
<td>$0.96 per hour</td>
</tr>
<tr>
<td>Micro On-Demand Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micro</td>
<td>$0.02 per hour</td>
<td>$0.03 per hour</td>
</tr>
<tr>
<td>Hi-Memory On-Demand Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra Large</td>
<td>$0.50 per hour</td>
<td>$0.62 per hour</td>
</tr>
<tr>
<td>Double Extra Large</td>
<td>$1.00 per hour</td>
<td>$1.24 per hour</td>
</tr>
<tr>
<td>Quadruple Extra Large</td>
<td>$2.00 per hour</td>
<td>$2.48 per hour</td>
</tr>
<tr>
<td>Hi-CPU On-Demand Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>$0.17 per hour</td>
<td>$0.29 per hour</td>
</tr>
<tr>
<td>Extra Large</td>
<td>$0.68 per hour</td>
<td>$1.16 per hour</td>
</tr>
<tr>
<td>Cluster Compute Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadruple Extra Large</td>
<td>$1.60 per hour</td>
<td>$1.98 per hour</td>
</tr>
<tr>
<td>Cluster GPU Instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadruple Extra Large</td>
<td>$2.10 per hour</td>
<td>$2.60 per hour</td>
</tr>
</tbody>
</table>

- Large instances \(\rightarrow\) total cost (0.34 \cdot 10 \cdot [X])
EC2 Pricing Model: Instance Types

- **Large Instance**
  - 7.5 GB memory
  - 4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each)
  - 850 GB instance storage
  - I/O Performance: High

- **Cluster Compute Quadruple Extra Large Instance**
  - 23 GB of memory
  - 33.5 EC2 Compute Units (2 x Intel Xeon X5570, quad-core “Nehalem” architecture)
  - 1690 GB of instance storage
  - 64-bit platform
  - I/O Performance: Very High (10 Gigabit Ethernet)

- **EC2 Compute Units = CPU capacity of a 1.0-1.2 GHz 2007 Xeon processor**
Why Cloud?

• Time-to-solution
  – 1000 cores * 1 hour = 1 core * 1000 hours

• Scalability: Can easily get 1000 cores
  – For $170 per hour

• Instance Availability
  – No queuing time
How to Use EC2
AWS Management Console

- OR using command line
EC2: Start Instance

- Load OS Image
  - Similar to VMware, VirtualBox
  - But to all instances in parallel

- Takes < 1 minutes
EC2: Start Instance

- `ec2-run-instances AMI [-n INSTANCE_COUNT]`
EC2: Login using SSH

- ssh -i KeyPair root@ec2-107-20-54-150.compute-1.amazonaws.com

```
Welcome to an EC2 Public Image

Fedora 8
64-bit

Welcome to an EC2 Public Image

[see /etc/ec2/release-notes]--

[root@ip-10-203-26-198 ~]#
```

- mpirun -hostfile ~/hosts -n 100 fishSim.exec
EC2: Save Image

• ec2-bundle-instance

• Takes ~10 minutes (except Cluster/GPU Compute Instance)

• Typical Procedure
  – Start one instance from a default image
  – Install MPI, OpenMP,...
  – Save Image
  – Start 100 instances from the saved image
EC2: Terminate Instances

- `ec2-terminate-instances INSTANCEID [INSTANCEID ...]`
How to get it for free

• AWS in Education
  – http://aws.amazon.com/education/

• For teaching:
  – CS 5300 - The Architecture of Large-Scale Information Systems

• For research:
  – Quarterly research grant applications
    • Next deadline: February 10, 2012
    • Renewable: maximum of two grant awards per academic year