Cluster-Based Computing

Junghoo Cho

cho@cs.ucla.edu

Cluster-Based Computing - Junghoo Cho - cho@cs.ucla.edu

Computing on Clusters

- Many corporations manage *data centers* with a large numbe clusters
  - > 10,000s machines in one data center
  - Commodity linux boxes
- Handle a large amount of user traffic and data

Challenges in Cluster Computing

- Q: What are challenges in operating machines at this scale?
- Hardware failures
  - Power and heat issues
  - Main source of failures: power supply, hard drive, network
- Difficulty of ensuring consistency among nodes

Unavoidable Failures

- Q: Assuming 99.9% uptime (9 hour downtime per year), how
  machines are down at any point with 10,000 machines?
- A nightmare for system administrator
  - Need to automate most of maintenance tasks, including initial depl
    replacement machine synchronization, etc.
- Very important to have software infrastructure to
  - Manage failed nodes
  - Monitor loads on individual machines
  - Schedule and distribute tasks and data to nodes

Example: Kubernetes

- Automatic deployment and management of containerized applications
- Progressive rollout of application changes
- Automatic scaling and load balancing of apps based on CPU
- Automatic restart of failed, unresponsive nodes

Remarks on Cluster Computing

- **DO NOT ASSUME ANYTHING!!!**
  - Explicitly define failure scenarios and their likelihood
  - Failure WILL happen. plan ahead for it
  - Make sure your code is covered for likely scenarios
  - Choose simplicity over generality
- Minimize state sharing among nodes
  - Decide who wins in case of conflict
- Minimize network bandwidth usage

Starting A New Web Site

- Q: You want to start a new site, called http://cs144.com. How do it?
  1. Buy the domain name cs144.com
     - GoDaddy.com, register.com, … (~$10/year)
  2. Get a "web server" with a public IP and update DNS to the IP

Provisioning Web Server

- Q: How can we provision a web server?
  1. Set up a physical machine
     1. Buy a machine (≈$1,000/PC)
     2. Buy an internet connection from an ISP (≈$100/month)
     3. Install OS and necessary software
  2. "Rent" a machine from a cloud hosting companies
     - Amazon Web Service, Google Cloud Platform, Windows Azure, …

Example: Physical Machine

- Our class server

Three Types of Cloud Service

- Q: If we "rent" from a cloud hosting company, exactly what d

1. Infrastructure as a service (IaaS)
2. Platform as a service (PaaS)
3. Software as a service (SaaS)

Infrastructure as a Service (IaaS)

- Rent a "virtual machine" and run your own virtual machine ir
  - e.g., Amazon Elastic Compute Cloud, Microsoft Azure Virtual Machir Compute Engine, …
- No hardware to manage, we manage all software including (

Platform as a Service (PaaS)

- Rent computing "platform" on which we program our app
  - Storage, database, middleware, …, via programmable APIs
- No need to manage underlying software stack, just write the
  - Provides service quality guarantee
    - "99% queries will finish in 100ms"
  - Scalability is built-in as part of service guarantee
  - "They solve our problems for money"
  - Issues of "vendor lock-in"

Software as a Service (SaaS)

- Rent fully working "off-the-shelf" software over internet
  - Google G Suite, Office 365, Salesforce.com, …
- No hardware or software to manage, just use the app

Amazon Web Services

- Amazon EC2 (Elastic Compute Cloud, virtual machine)
- Amazon ECS (Elastic Container Service)
- Amazon S3 (Simple Storage Service, distributed filesystem)
- Amazon Aurora (Relational Database Service)
- Amazon DynamoDB (NoSQL datastore)
- Amazon ElastiCache (in-memory object caching)
- AWS Lambda (event-driven functional programming API)
- Amazon Elastic Load Balancing
- Amazon CloudFront (content distribution network)
- …

What We Learned

- Challenges in cluster computing
  - Unavoidable machine failure
  - Distribution of states and tasks
- Map-reduce programming pattern
- Cluster software infrastructure
  - Kubernetes
- Cloud service provider
  - IaaS, PaaS, SaaS
  - Amazon Web Service, Microsoft Azure, Google Cloud Platform