Basics of Cloud Computing – Lecture 7

More AWS and Cloud-based Research at Mobile & Cloud Lab

Satish Srirama
Outline

• More Amazon Web Services
• How we are using cloud
• Cloud based Research @ Mobile & Cloud Lab
Cloud Providers and Services – we already discussed

- Amazon Web Services
  - Amazon EC2
  - Amazon S3
  - Amazon EBS
  - Amazon Elastic Load Balancing
  - Amazon Auto Scale
  - Amazon CloudWatch
- Eucalyptus
- OpenStack
- SciCloud
- Management providers
  - ElasticFox
  - RightScale
- PaaS
  - Google AppEngine
  - Windows Azure
MORE AWS
AWS we discuss

- AWS Management Console
- AWS Identity and Access Management
- AWS Elastic Beanstalk
- AWS CloudFormation
- Amazon Simple Workflow Service
- Amazon Elastic MapReduce
AWS Management Console

• Hope some of you have started using Amazon accounts
• You can manage your complete Amazon account with management console (Similar to Hybridfox)
  – AMI Management
  – Instance Management
  – Security Group Management
  – Elastic IP Management
  – Elastic Block Store
  – Key Pair management etc.
• Have different panes for different services
AWS Management Console - screenshot

https://console.aws.amazon.com/
AWS Identity and Access Management (IAM)

• How can an enterprise or group of people use a single credit card?

• Manage IAM users
  – Create new users and manage them
  – Create groups

• Manage permissions
  – Creating policies

• Manage credentials
  – Create and assign temporary security credentials
IAM policy

- Example policy giving access to complete EC2

```json
{
  "Statement": [
    {
      "Action": "ec2:*",
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "elasticloadbalancing:*",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "cloudwatch:*",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "autoscaling:*",
      "Resource": "*"
    }
  ]
}
```

http://aws.amazon.com/iam/
AWS Elastic Beanstalk

• Enables to easily deploy and manage applications in the AWS cloud
  – Simply upload a bundle of the applications built using .NET, PHP and Java technologies
• Automatically handles the deployment details of capacity provisioning, load balancing, auto-scaling, and application health monitoring
• **Something similar to PaaS**
• One retains full control over the AWS resources powering the application
  – You can access the underlying resources at any time
AWS Elastic Beanstalk

• AWS EB is built using familiar software stacks such as the Apache HTTP Server for PHP, IIS 7.5 for .NET, and Apache Tomcat for Java

• There is no additional charge for Elastic Beanstalk
  – Only the underlying AWS resources (e.g. Amazon EC2, Amazon S3) are charged

• Leverages AWS services such as Amazon EC2, S3, SNS, ELB, and Auto Scaling to deliver the same highly reliable, scalable, and cost-effective infrastructure

http://aws.amazon.com/elasticbeanstalk
AWS CloudFormation

- Provides an easy way to create and manage a collection of related AWS resources, provisioning and updating them in an orderly and predictable fashion
- It is based on templates model
  - Templates describe the AWS resources, the associated dependencies, and runtime parameters to run an app.
  - The templates describe stacks, which are set of software and hardware resources.
  - Something similar to CloudML and RightScale server templates
- Hides several details
  - How the AWS services need to be provisioned
  - Subtleties of how to make those dependencies work.
AWS CloudFormation

- Amazon provides several pre-built templates to start common apps as:
  - WordPress (blog)
  - LAMP stack
  - Gollum (wiki used by GitHub)
  - ...

- There is no additional charge for AWS CloudFormation. You pay for AWS resources (e.g. EC2 instances, Elastic Load Balancers, etc.)
Amazon Simple Workflow Service

• A workflow service for building scalable, resilient applications
• Reliably coordinates all of the processing steps within applications
  – such as business processes, sophisticated data analytics applications, or managing cloud infrastructure services
• Manages task execution dependencies, scheduling, and concurrency
• Provides simple API calls from code written in any language
• Capable to run on EC2 instances, or any of the customer’s machines located anywhere in the world
Amazon Simple Workflow Service

- Maintains application state
- Tracks workflow executions and logs their progress
- Holds and dispatches tasks
- Controls which tasks each of the application hosts will be assigned to execute

Amazon Elastic MapReduce

- Web interface and command-line tools for running Hadoop jobs on EC2
- Data stored in Amazon S3
- Monitors job and shuts machines after use

Running a job
- Upload job jar & input data to S3
- Create the cluster
- Create a Job Flow as steps
- Wait for the completion and examine the results

http://aws.amazon.com/elasticmapreduce/
Other interesting AWS

• Amazon Relational Database Service
  – Provides access to the capabilities of familiar database engines
  – MySQL, Oracle or Microsoft SQL Server

• NoSQL databases
  – Simple DB
  – DynamoDB
CLOUD BASED RESEARCH @ MOBILE & CLOUD LAB
Scientific Computing on the Cloud

• Public clouds provide very convenient access to computing resources
  – On-demand and in real-time
  – As long as you can afford them

• High performance computing (HPC) on cloud
  – Virtualization and communication latencies are major hindrances [Srirama et al, SPJ 2011; Batrashev et al, HPCS 2011]
    • Things have improved significantly over the years
  – Research at scale
    • Cost-to-value of experiments
Adapting Computing Problems to Cloud

- Reducing the algorithms to cloud computing frameworks like MapReduce [Srirama et al, FGCS 2012]
- Designed a classification on how the algorithms can be adapted to MR
  - Algorithm → single MapReduce job
    • Monte Carlo, RSA breaking
  - Algorithm → $n$ MapReduce jobs
    • CLARA (Clustering), Matrix Multiplication
  - Each iteration in algorithm → single MapReduce job
    • PAM (Clustering)
  - Each iteration in algorithm → $n$ MapReduce jobs
    • Conjugate Gradient
- Applicable especially for Hadoop MapReduce
Issues with Hadoop MapReduce

• It is designed and suitable for:
  – Data processing tasks
  – Embarrassingly parallel tasks

• Has serious issues with iterative algorithms
  – Long “start up“ and “clean up“ times ~17 seconds
  – No way to keep important data in memory between MapReduce job executions
  – At each iteration, all data is read again from HDFS and written back there at the end
  – Results in a significant overhead in every iteration
Alternative Approaches

• Restructuring algorithms into non-iterative versions
  – CLARA instead of PAM [Jakovits & Srirama, Nordiccloud 2013]

• Alternative MapReduce implementations that are designed to handle iterative algorithms [Jakovits and Srirama, HPCS 2014]
  – E.g. Twister, HaLoop, Spark

• Alternative distributed computing models
  – Bulk Synchronous Parallel model [Valiant, 1990] [Jakovits et al, HPCS 2013]
  – Building a fault-tolerant BSP framework (NEWT) [Kromonov et al, HPCS 2014]
Communication pattern of Cluster vs Cloud

- Cloud has huge troubles with communication/transmission latencies
  - Virtualization technology is the culprit
- Performance Comparison of virtual machines and Linux containers (e.g. Docker)

[Srirama et al, SPJ 2011]
Remodeling Enterprise Applications for the Cloud

• Remodeling workflow based applications for the cloud
  – To reduce communication latencies among the components
  – Intuition: Reduce inter-node communication and to increase the intra-node communication

• Auto-scale them based on optimization model and CloudML

[Image of workflow diagram]

[Srirama and Viil, HPCC 2014]
Migrating Scientific Workflows to the Cloud

- Workflow can be represented as weighted directed acyclic graph (DAG)
- Partitioning the workflow into groups with graph partitioning techniques [Srirama and Viil, HPCC 2014]
  - Such that the sum of the weights of the edges connecting to vertices in different groups is minimized
  - Utilized Metis’ multilevel k-way partitioning
- Scheduling the workflows with tools like Pegasus
  - Considered peer-to-peer file manager (Mule) for Pegasus
Report: Mobile cloud to grow beyond $11 billion in 2018

Written by CooperEgg // July 12, 2012 // No Comment // Cloud Performance

The proliferation of smartphones, tablets and other mobile devices is contributing to change in the private sector, as businesses continue to leverage these gadgets in an attempt to enhance efficiency and potentially gain a competitive advantage. According to a new report by Global Industry Analysts, the evolution of mobility is also changing the cloud computing landscape, pushing the mobile cloud market to generate more than $11 billion in revenue by 2018.
Mobile Applications

• One can do interesting things on mobiles directly
  – Today’s mobiles are far more capable
  – Location-based services (LBSs), mobile social networking, mobile commerce, context-aware services etc.

• It is also possible to make the mobile a service provider
  – Mobile web service provisioning [Srirama et al, ICIW 2006; Srirama and Paniagua, MS 2013]
  – Challenges in security, scalability, discovery and middleware are studied [Srirama, PhD 2008]
  – Mobile Social Network in Proximity [Chang et al, ICSOC 2012; PMC 2014]
However, we still have not achieved

• Longer battery life
  – Battery lasts only for 1-2 hours for continuous computing

• Same quality of experience as on desktops
  – Weaker CPU and memory
  – Storage capacity

• Still it is a good idea to take the support of external resources for building resource intensive mobile applications
Mobile Cloud Applications

• Bring the cloud infrastructure to the proximity of the mobile user
• Mobile has significant advantage by going cloud-aware
  – Increased data storage capacity
  – Availability of unlimited processing power
  – PC-like functionality for mobile applications
  – Extended battery life (energy efficiency)
Mobile Cloud – Our interpretation

• We do not see Mobile Cloud to be just a scenario where mobile is taking the help of a much powerful machine!!!

• We do not see cloud as just a pool of virtual machines

• Mobile Cloud based system should take advantage of some of the key intrinsic characteristics of cloud efficiently
  – Elasticity & AutoScaling
  – Utility computing models
  – Parallelization (e.g., using MapReduce)
Mobile Cloud Binding Models

[Flores et al, MoMM 2011]

Task Delegation

[Flores and Srirama, MCS 2013]

Code Offloading

[Flores & Srirama, JSS 2014]
MCM – enables

- Interoperability between different Cloud Services (IaaS, SaaS, PaaS) and Providers (Amazon, Eucalyptus, etc)
- Provides an abstraction layer on top of API
- Composition of different Cloud Services
- Asynchronous communication between the device and MCM [Warren et al, IEEE PC 2014]
- Means to parallelize the tasks and take advantage of Cloud’s intrinsic characteristics
MCM applications

• **CroudSTag** [Srirama et al, MobiWIS 2011]
  – Social group formation with people identified in Pictures/Videos

• **Zompopo** [Srirama et al, NGMAST 2011]
  – Intelligent calendar, by mining accelerometer sensor data

• **Bakabs** [Paniagua et al, iiWAS-2011]
  – Managing the Cloud resources from mobile

• **Sensor data analysis**
  – Human activity recognition
  – Context aware gaming
  – MapReduce based sensor data analysis [Paniagua et al, MobiWIS 2012]

• **SPiCa: A Social Private Cloud Computing Application Framework** [Chang et al, MUM 2014]
Code Offloading - Major Components

• Major research challenges
  – What, when, where and how to offload?
• Mobile
  – Code profiler
  – System profilers
  – Decision engine
• Cloud based surrogate platform

[Flores and Srirama, MCS 2013]
Challenges and technical problems

• Inaccurate code profiling
  – Code has non-deterministic behaviour during runtime
    • Based on factors such as input, type of device, execution environment, CPU, memory etc.
  – Some code cannot be profiled (e.g. REST)

• Integration complexity
  – Dynamic behaviour vs Static annotations
    • E.g. Static annotations cause unnecessary offloading

• Dynamic configuration of the system

• Offloading scalability and offloading as a service
  – Surrogate should have similar execution environment
  – Should also consider about resource availability of Cloud

[Flores et al, IEEE Communications Mag 2015]
Practical adaptability of offloading

Applications that can benefit became limited with increase in device capacities
Internet of Things (IoT)

- “The Internet of Things allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any path/network and Any service.” [European Research Cluster on IoT]
- More connected devices than people
- Cisco believes the market size will be $19 trillion by 2025
IoT - Scenarios

• Environment Protection
• Smart Home

[Kip Compton]  
[Perera et al, TETT 2014]
Internet of Things – Challenges

How to provide energy efficient services?

How do we communicate automatically?

How to interact with ‘things’ directly?

[Chang et al, ICWS 2015]

[Chang et al, SCC 2015; Liyanage et al, MS 2015]

3/22/2016 Satish Srirama
Cloud-based IoT

Remote Cloud-based processing

Connectivity nodes & Embedded processing

Sensing and smart devices
Further research focus in IoT

• We have established IoT and Smart Solutions Lab with Telia company support

• Interesting topics
  – Discovery of IoT devices
  – Study of IoT based middlewares
  – Study of available IoT platforms
    • Amazon IoT
    • Open IoT
IoT Data Processing on Cloud

• Enormous amounts of unstructured data
  – In Zetabytes ($10^{21}$ bytes) by 2020 [TelecomEngine]
  – Has to be properly stored, analysed and interpreted and presented

• Big data acquisition and analytics
  – Is MapReduce sufficient?
    • MapReduce is not good for iterative algorithms [Srirama et al, FGCS 2012]
  – IoT mostly deals with streaming data
    • Message queues such as Apache Kafka can be used to buffer and feed the data into stream processing systems such as Apache Storm
    • Apache Spark streaming

• How to ensure QoS aspects such as security of data?
  – Anonymization and Expiry of data?
    • Especially for the personal data
WE ALWAYS WELCOME NEW IDEAS!

email: srirama@ut.ee
This week in lab

• Work with Google AppEngine
Next Week

• Summarize what we have learnt
• How to prepare for the examination
References

- Check Amazon videos and webinars at http://aws.amazon.com/resources/webinars/
References - continued