Åke Edlund - “academic entrepreneur”

Uppsala, Rice, Berkeley, Technion

Cycore
EU-Supply
Ericsson Business Innovation

KTH-SICS
Cloud Innovation Center

Sony Ericsson
Parallel Systems

Startups
Academia
Industry

And advisor at severalnines
Project details

• Project time span: Jan-Dec, 2010
• ~3.5 FTE effort (*)
• Learning-by-doing, pilots and desk research
• Cloud recommendations to Nordic eScience community
• Dissemination and outreach, International collaboration
• Focus on moving non-HPC jobs to the cloud

(*) SE (1.5), NO (1), FIN (0.5), IS (0.3), DK (0.2)
Motivation - to release HPC resources to those who really need them

• Many single-node jobs being run on expensive HPC resources
• These single-node jobs could be run on less advanced systems - maybe cloud resources
• Possible cost advantage moving (some) of these to the cloud, without lowering the user experience?
The one year project (Jan-Dec 2010) outcome was a report describing:

1. state-of-the art of cloud computing;
2. cost of moving and running non-HPC jobs on a cloud computing environment;
3. how to do this in practice;
4. a list of identified risks/benefits on a short/long perspective.

The higher ambition with the project report is to gather decision information to help with the strategic long-term cloud plans for Europe.
We’re not alone: ECEE - Enabling Clouds for eScience
The open collaboration spot for cloud projects in Europe

NEON (coord.)
BalticCloud
NGS
SARA cloud
UCM
StratusLab
VENUS-C
GRNET cloud
SEECCI
CESGA
(more to join)

www.scientific-cloud.org
So, what’s the problem we’re trying to solve?
> 20% single node jobs, often more
Practice

• While there are many working solutions to this problem (to move single node jobs from HPC resources), the NEON project included the aspect of using today’s administrators to build and manage the (private) cloud test installations
Pilots

- NO: Cloud backed storage, eSysBio
- SE: private clouds (Multiple availability zones)
- FI: private clouds
- DK: DIY
- IS: public cloud ray tracing
The next deadline for quarterly research grant applications is December 17th, 2010

If you are awarded a Research Grant for free usage of AWS, the grant will be good for 1 year or until the usage credits have been fully utilized in the 1 year research grant time frame. Academic researchers who receive AWS grants may apply for future grants via the same application and review process with eligibility for a maximum of two grant awards in one calendar year. A number of academic researchers around the world have already chosen AWS to support their research including:

- The University of Oxford Malaria Atlas research on AWS
- The University of California, Berkeley RAD lab

Application Form

Apply for an AWS Research Grant here:

Name*
University*
Email (.edu or equivalent)*
Research project URL*

AWS services required for research:
- Amazon CloudFront
- Amazon Elastic Compute Cloud
- Amazon Elastic MapReduce
- Amazon Simple Queue Service
- Amazon Simple Storage Service
- Amazon SimpleDB

Description of research project, AWS solution, usage requirements (4000 character limit)*

* Indicates required field

Submit
Aug 12

Title: Running e-science environment for Systems Biology research (eSysBio) on AWS
This proposal briefly describes a pilot project for getting a part of the eSysbio (“e-science environment for Systems Biology research”, see esysbio.org) running on AWS.

-- Requested Amazon Resources Grant = USD 5,000
4 High CPU instances and corresponding storage, resulting in a AWS cost (according to online AWS calculator) on average USD 1,250 per month. 
We plan to run on these resources during 4 months of the project, resulting in a total of USD 5,000.

Aug 31

Dear Ake
On behalf of Amazon Web Services, we’re pleased to provide you a research grant in the form of AWS credits. Simply use the code below at the URL redemption site to automatically update your AWS account with the credits. Credits are set to expire 1 year from today. If credits are fully utilized and you continued using services, charges may apply to the credit card you have associated with the account. Also note that these credits may not be applied toward Reserved Instances fees.
We thank you for your grant inquiry and wish you all the best with your research work on AWS. If you publish your results and reference your use of AWS, we would be pleased if you shared that with us. Take care!
K.
Redemption site: http://aws.amazon.com/awscrédits
Code: PC1JXZDQ5KFS7FD
Private cloud, Sweden
NEON findings

Note: the findings listed here are from end of 2010.

Available cloud services have improved rapidly, and many new are appearing as we speak. Still the findings highlight the need of hands-on testing with existing personnel and resources.
Private cloud technology is **not mature enough** yet to provide a transparent user experience.
Private cloud technology is **not mature enough** yet to provide a transparent user experience.

**Successful counter examples near by:**

- SARA (NL) and their Virtual Private HPC Cluster.
- CSC (FIN), using OpenNebula for external bioinformatics users.
Public cloud technology is mature enough but lacks certain features that will be necessary to include cloud resources in a transparent manner in a national infrastructure -- (e.g. quota management).

Example: eSysBio
Public clouds are **competitive in the low end** for non-HPC jobs (low memory, low number of cores) on price.

**NEON findings (3/6)**
NEON findings (4+5+6/6)

A significant fraction (ca. 20%) of the jobs running on the current supercomputer infrastructure are potentially suitable for cloud-like technology.
This holds in particular for single-threaded or single-node jobs with small/medium memory requirements.

There is a backlog of “real” supercomputer jobs that suffers from the non-HPC jobs on the supercomputer infrastructure.

Available storage capacity is not accessible in a user-friendly way; most storage clouds are only accessible via programmable interfaces.
<table>
<thead>
<tr>
<th>Gap analysis - Private Clouds</th>
<th>Gap analysis - Public Clouds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heavy reliance on system and network administrator expertise</strong> for end users</td>
<td><strong>More reliance on system administrator expertise for end users</strong> - same downside as for private cloud - though network knowledge might be less of an issue.</td>
</tr>
<tr>
<td><strong>Separate management software</strong> - either you pay for an Enterprise version, or you pay for separate management software/services</td>
<td><strong>Separate management software</strong> - this depends on the particular cloud solution and the level of management required. Obviously, public cloud offering offer some level of management if only for billing purposes. Public clouds develop themselves rapidly in this area, but not in a compatible way.</td>
</tr>
<tr>
<td><strong>Identity management integration limited to users uploading their own certificates</strong></td>
<td><strong>Identity management integration limited to users uploading their own certificates</strong></td>
</tr>
<tr>
<td><strong>Quota management hard or non-existent</strong></td>
<td><strong>Quota management hard or non-existent</strong> - this is obviously not in the interest of public cloud providers. Why limit usage if that’s what they sell?</td>
</tr>
<tr>
<td><strong>Rapid update cycle of core infrastructure software</strong></td>
<td></td>
</tr>
<tr>
<td><strong>No other middleware offerings out of the box</strong> - services like databases, queues etc. must be deployed and maintained by the providing site.</td>
<td></td>
</tr>
<tr>
<td><strong>No mature storage offering</strong> - storage offerings are either no part of private clouds, or perform sub-par</td>
<td></td>
</tr>
</tbody>
</table>
Key findings

• Private cloud technology not mature enough
  - Private clouds require above average control of the network topology
  - Private clouds need “current” hardware.
  - Private cloud stacks are not as complete as public cloud offerings

• Public cloud mature enough

• Public cloud pricing competitive for low end non-HPC jobs - which is >>20% of current DC jobs.
What about cost? DCI perspective

Guide to where costs go in the data center.

Amortized Cost Component Sub-Components

~ 45% Servers CPU, memory, storage systems
~ 25% Infrastructure Power distribution and cooling
~ 15% Power draw Electrical utility costs
~ 15% Network Links, transit, equipment

The general rule-of-thumb, based on today’s cost and pricing picture, is to use clouds if the total need of compute time is below 12 months, and extra storage need below 6 months.
What about cost?
- The picture

“The Economics of the Cloud” (Nov, 2010)

Example from Norway. BTW, it was really hard to get real numbers from the datacenters...

Picture: NOK/node/year versus utility level
What about energy consumption?
What’s your PUE (Power Usage Effectiveness)?

PUE = Total Facility Power / IT Equipment Power

Typical = 2+  Exceptional = 1.5  Google and others = 1.2
Industry: IT energy usage expected to rise 15% annually

Joyce Dickerson, Director, Sustainable IT, Stanford University

We're far from the PUE that many public clouds manage to achieve

Goal: Keep energy usage flat, while meeting increasing demand on IT infrastructure
## Risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doing nothing</td>
<td><strong>Users will, and may disrupt current practices and administration</strong></td>
</tr>
<tr>
<td>Data sensitivity</td>
<td>Local data storage, only computing in cloud. Safe harbor policy when dealing with US based cloud providers.</td>
</tr>
<tr>
<td>Economic lock-in private clouds</td>
<td>Wait and watch, via e.g. ECEE</td>
</tr>
<tr>
<td>Economic lock in public clouds (data)</td>
<td>Volume deals and tenders. Keep data local when possible.</td>
</tr>
<tr>
<td>Economic DoS</td>
<td>Negotiate volume tresholds; active monitoring tools that monitor overall usage and cost structure.</td>
</tr>
</tbody>
</table>
## Risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doing nothing</td>
<td>Users will, and may disrupt current practices and administration</td>
</tr>
<tr>
<td>Data sensitivity</td>
<td>Local data storage, only computing in cloud. Safe harbor policy when dealing with US based cloud providers.</td>
</tr>
<tr>
<td>Economic lock-in private clouds</td>
<td>Wait and watch, via e.g. ECEE</td>
</tr>
<tr>
<td>Economic lock in public clouds (data)</td>
<td>Volume deals and tenders. Keep data local when possible.</td>
</tr>
<tr>
<td>Economic DoS</td>
<td>Negotiate volume tresholds; active monitoring tools that monitor overall usage and cost structure.</td>
</tr>
</tbody>
</table>

Goal with public cloud service to coordinate this (see next slides)
Now, 2011

- **Norway and Sweden**: Sigma and SNIC starting up a public cloud service for their constituencies in a joint project
- **Finland**: CSC and national efforts building Cloud sized DC
- **All**: experimenting on private clouds and local solutions. Eucalyptus - OpenNebula - OpenStack
- Many: participating in international (mostly EU) projects.
- Nordic collaboration: no explicit plans, opportunity to do a shared public cloud project (see next slide)
- International collaboration: ECEE ([www.scientific-cloud.org](http://www.scientific-cloud.org))
NEON 2.0

A Nordic Cross Country Cloud Collaboration

Opportunity to establish a Nordic public cloud service:

- Sharing resources
- Sharing data
- Sharing users
- Sharing support

Cost, organizational and operational benefits through larger tenders and shared support (*)

(*)Same applies to smaller settings but with smaller benefits

Starting with Sweden and Norway
So, now you’re putting all your single node jobs into the public cloud?

Well, step-by-step, learning by doing...

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Now that we have fully restored functionality to all affected services, we would like to share more details with our customers about the events that occurred with the Amazon Elastic Compute Cloud ("EC2") last week, our efforts to restore the services, and what we are doing to prevent this sort of issue from happening again. We are very aware that many of our customers were significantly impacted by this event, and as with any significant service issue, our intention is to share the details of what happened and how we will improve the service for our customers.

The issues affecting EC2 customers last week primarily involved a subset of the Amazon Elastic Block Store ("EBS") volumes in a single Availability Zone within the US East Region that became unable to serve read and write operations. In this document, we refer to these as "stuck" volumes. This caused instances trying to use these affected volumes to also get "stuck" and therefore always attempt to read or write to them. In order to restore these volumes and stabilize the EBS cluster in that Availability Zone, we disabled all control APIs (e.g. Create Volume, Attach Volume, Detach Volume, and Create Snapshot) for EBS in the affected Availability Zone for much of the duration of the event.

For two periods during the first day of the issue, the degraded EBS cluster affected the EBS APIs and caused high error rates and latencies for EBS data to EBS data across the entire US East Region. As with any complicated operational issue, this one was caused by a number of factors, and it caused interacting with one another and therefore gives us many opportunities to protect the service against any similar event reoccurring.

Overview of EBS System

It is helpful to understand the EBS architecture so that we can better explain the event. EBS is a distributed, replicated block data store that is optimized for consistency and low latency read and write access from EC2 instances. There are two main components of the EBS service:

(i) a set of EBS clusters (each of which runs entirely inside of an Availability Zone) that store user data and serve requests to EC2 instances; and
(ii) a set of control plane services that are used to coordinate user requests and propagate them to the EBS clusters running in each of the Availability Zones in the Region.

An EBS cluster is comprised of a set of EBS nodes. These nodes store replicas of EBS's volume data and serve read and write requests to EC2 instances. EBS volume data is replicated across multiple EBS nodes for durability and availability. Each EBS node employs a peer-to-peer based, fast failover strategy that aggressively provisions new replicas if one of the copies ever fails or becomes unavailable. The nodes in an EBS cluster are connected to each other via two networks: the primary network is a high bandwidth network used for normal operation for all inter-node communication with other EBS nodes, and the secondary network, used for replication, is a lower bandwidth network used as a back-up network to allow EBS nodes to reliably communicate with other nodes in the EBS cluster and provide overflow capacity for data replication. This network is not designed to handle all traffic from the primary network but rather provide highly-reliable connectivity between EBS nodes across the EBS clusters.

When a node loses connectivity to a node to which it is replicating data to, it assumes the other node failed. To preserve durability, it must find a new node to which it can replicate its data (this is called re-mirroring). As part of the re-mirroring process, the EBS node searches its EBS cluster for another node with an available server space, establishes connectivity to that new node, and begins replicating data.

In a normally functioning cluster, finding a location for the new replica occurs in a single operation. However, if a node that has copies of the data hold onto the data until they can confirm that another node has taken ownership of that portion of the data. This provides an additional level of protection against customer data loss. Also, when data is a "primary" volume, it becomes "stuck" to that data is blocked until the system has identified a new primary (or writable) replica. This is required for consistency of EBS volume data under all potential failure modes. From the perspective of an EC2 instance trying to do I/O on a volume while this is happening, the volume will appear "stuck".

In addition to the EBS clusters, there is a set of control plane services that accepts user requests and propagates them to the appropriate EBS cluster. There is one set of EBS control plane services per EC2 Region, but the control plane itself is highly distributed across the Availability Zones to provide availability and fault tolerance. These control plane services also act as the authority to the EBS clusters when they elect primary replicas for each volume in the cluster (for consistency, there must only be a single primary replica for each volume at any time).
Summary

• Practical testing in 5 countries.

• **Private clouds still a challenge to many administrators** - i.e. admins need to be educated/ convinced more on clouds AND private clouds need to be even more user friendly. Getting better already.

• Public clouds is an alternative many eScience researchers are trying now, we want to be a **HUB** for our users. 2011 focus for NEON project.

• We’re just in the beginning - so far very IaaS focussed. **Expecting more interest higher up in the stack - SaaS for eScience.**
Coordinator

Data storage dev, and technical lead norway

Thanks! Questions?

Åke Edlund
edlund@pdc.kth.se

Jukka Kommeri
jukka.antero.kommeri@cern.ch

Maarten Koopmans
maarten@vrijheid.net

Miika Tuisku
miika.tuisku@hip.fi

Zeeshan Ali Shah
zashah@pdc.kth.se

Pekka Lehtovuori
Pekka.Lehtovuori@csc.fi

Ilja Livenson
livenson@kth.se

Klaus Marius Hansen
klausmh@diku.dk

Frederik Orellana
orellana@nbi.dk

Helmut Neukirchen
helmut@hi.is

Ebba Hvannberg
ebba@hi.is

Coordinator