Software provisioning on cloud resources with GridPilot

Frederik Orellana
Niels Bohr Institute
Copenhagen University
March 2009

1 Introduction

This note describes extensions of the software catalogue format introduced in [CATALOG]. The purpose of the extensions is to allow reusing the parsing logic of GridFactory when running jobs via GridPilot on Amazon's Elastic Compute Cloud [EC2], thus allowing provisioning software packages on EC2 instances in the TarPackage and ImagePackage formats as well as in a new, EC2 specific format, using the snapshot capabilities of Amazon's Elastic Blockstore service. The extensions amount to the introduction of two new instance package formats: “AMIPackage” and “EBSSnapshotPackage”.

2 AMIPackage

An AMIPackage is an instance package object describing an EC2 AMI. It has two mandatory blocks: “kb:manifest” and “kb:ami_id”, describing the location of the XML manifest and the AMI identifier respectively.

Notice that the parent MetaPackage of an AMIPackage, describes a virtual machine and as such should contain a “kb:VirtualMachine” block. This block should in principle contain the blocks <kb:service>ssh</kb:service> and <kb:privilege>administrator</kb:privilege>, but in practice all AMI instances are booted with a running SSH server and root login via an SSH key. GridPilot makes the assumption that this is the case and thus these two blocks are redundant.

3 EBSSnapshotPackage

An EBSSnaptshotPackage is an instance package object describing an Amazon Elastic Blockstore snapshot. The volume created from such a snapshot must have the same directory layout as an unpackacked TarPackage. An EBSSnaptshotPackage has a mandatory block “kb:snapshot_id” and optionally a block “kb:mountpoint”, describing the snapshot identifier and where the volume should be mounted respectively.

4 Catalogue format summary

For reference, here follows a listing of all valid catalogue elements, including the newly introduced ones (emphasized):

- **MetaPackage**
  - **rdf:about**
  - **kb:description**
  - **kb:instance**
  - **kb:registrykey**
  - **kb:command**
  - **kb:provides**
  - **kb:depends**
  - **kb:lastupdate**
  - **kb:tag**
  - **rdfs:label**
  - **kb:VirtualMachine**
    - **kb:os**
    - **kb: out_port**
    - **kb: in_port**
    - **kb:service**
    - **kb:privilege**
  - **kb:Requirements**
    - **kb:mem_kb**
    - **kb:out_port**
    - **kb:in_port**
    - **kb:privilege**

- **TarPackage**
    - **rdf:about**
    - **kb:url**
    - **kb:basesystem**
    - **kb:depends**
    - **kb:lastupdate**
    - rdfs:label
- **ImagePackage**
    - **rdf:about**
    - **kb:url**
    - **kb:basesystem**
    - **kb:depends**
    - **kb:lastupdate**
    - **kb:mountpoint**
    - rdfs:label
- ***AMIPackage***
    - **rdf:about**
    - **kb:url**
    - **kb:basesystem**
    - **kb:depends**
    - **kb:lastupdate**
    - **kb:ami_id**
    - **kb:manifest**
    - rdfs:label
- ***EBSSnapshotPackage***
    - **rdf:about**
    - **kb:url**
    - **kb:basesystem**
    - **kb:depends**
    - **kb:lastupdate**
    - **kb:snapshot_id**
    - **kb:mountpoint**
    - rdfs:label
- **BaseSystem**
    - **rdf:about**
    - **kb:description**
    - **kb:lastupdate**
    - rdfs:label
    - kb:url

The grayed-out elements are not used by GridPilot and their use is discouraged.

5 Bibliography

CATALOG, Frederik Orellana, “Software provisioning and virtualisation for grid services”, *in preparation*

GRIDFACTORY, Frederik Orellana, “GridFactory – a virtualised compute system for distributed clusters”, *in preparation*

GRIDPILOT, Frederik Orellana, “GridPilot – a graphical front-end to distributed compute systems”, *in preparation*

EC2, Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2/