Deploying a Cassandra cluster on Amazon EC2

Clément LARDEUR

DataStax AMI

- AMI = Amazon Image Machine

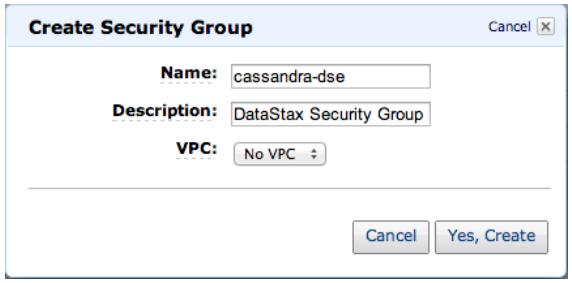
  https://aws.amazon.com/amis/datastax-auto-clustering-ami-2-2

- DataStax AMI does the following
  - Install the number of nodes specified
  - Uses RAID0 ephemeral disks instead of EBS
  - Sets the seed nodes cluster-wide
  - Install OpsCenter on the first node
  - Configure replication strategy using EC2Snitch

Create Security Group

- ➔ Security Group act as a Firewall
- ➔ Ports Needed

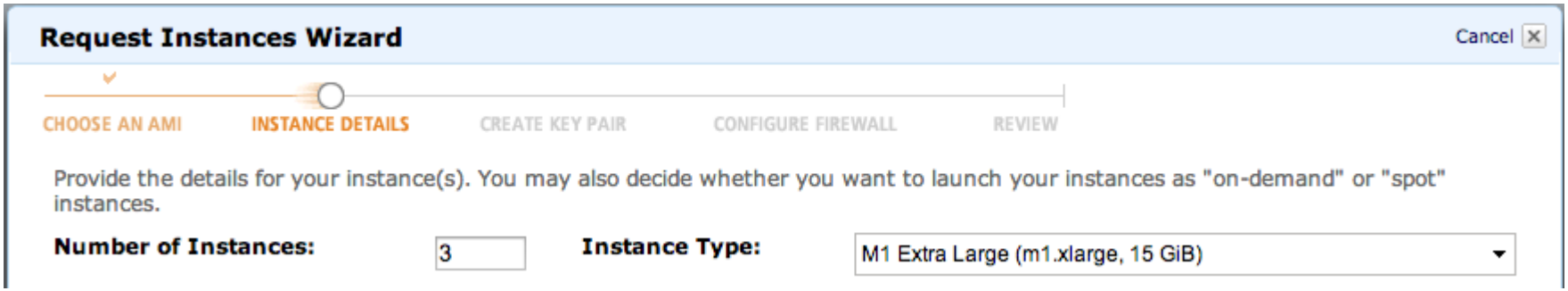
```
Public Facing:
    SSH:
        22: Default SSH port
    DataStax Enterprise Specific:
        8012: Hadoop Job Tracker client port
        8983: Portfolio Demo and Solr website port
        50030: Hadoop Job Tracker website port
        50060: Hadoop Task Tracker website port
    OpsCenter:
        8888: OpsCenter website port
Intranode:
    Cassandra:
        1024+: JMX reconnections
        7000: Cassandra intra-node port
        7199: Cassandra JMX monitoring port
        9160: Cassandra client port
    DataStax Enterprise Specific:
        9290: Hadoop thrift port
    OpsCenter:
        50031: OpsCenter job tracker proxy
        61620: OpsCenter intra-node monitoring ports
        61621: OpsCenter agent port
```



Port (Service)	Source	Action
1024 - 65535	sg-c86d0dbf (cassandra-dse)	Delete
7000	sg-c86d0dbf (cassandra-dse)	Delete
7199	sg-c86d0dbf (cassandra-dse)	Delete
9160	sg-c86d0dbf (cassandra-dse)	Delete
61620	sg-c86d0dbf (cassandra-dse)	Delete
61621	sg-c86d0dbf (cassandra-dse)	Delete
22 (SSH)	0.0.0.0/0	Delete
8888	0.0.0.0/0	Delete

Launch AMI

- Instance Details
  - Number of instances = Number of nodes
  - Instance Type :
    - Extra large for production
    - Large for development and light production
    - Small and Medium not supported !!



Launch AMI

➔ Instance Options

Basic Options

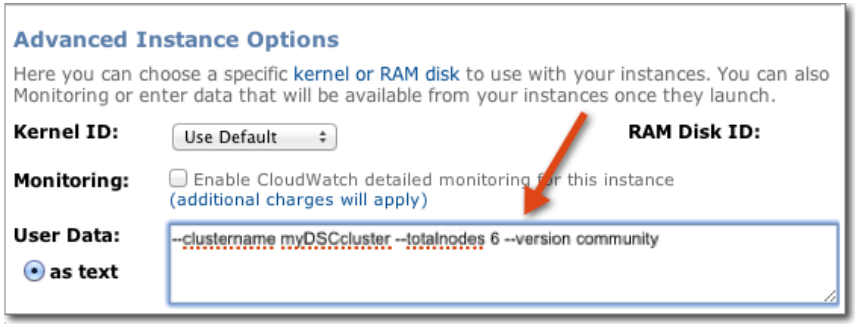
```
--clustername <name>
    The name of the Cassandra cluster
    REQUIRED

--totalnodes <#>
    Cluster size
    REQUIRED

--version [ community | enterprise ]
    Installs either DataStax Enterprise or
    DataStax Community Edition
    REQUIRED
```

DataStax Enterprise Specific

```
--username <user>
    The username provided during DSE registration
    --password is REQUIRED for a DSE installation

--password <pass>
    The password provided during DSE registration
    --username is REQUIRED for a DSE installation

--analyticsnodes <#>
    Number of analytics nodes that run with Hadoop
    Default: 0

--searchnodes <#>
    Number of search nodes that run with Solr
    Default: 0
```

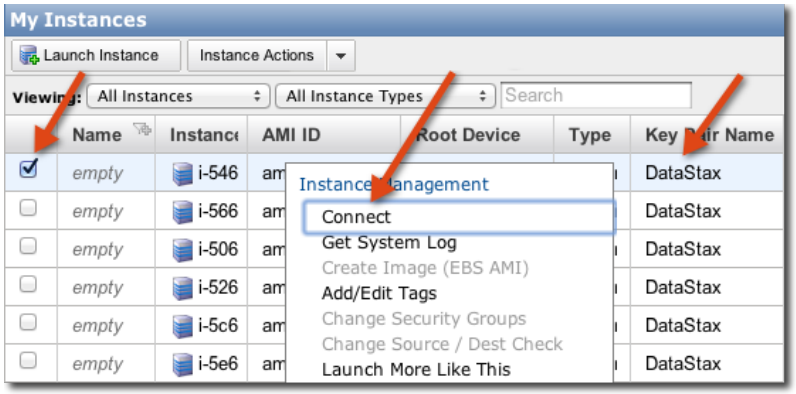
Advanced Options

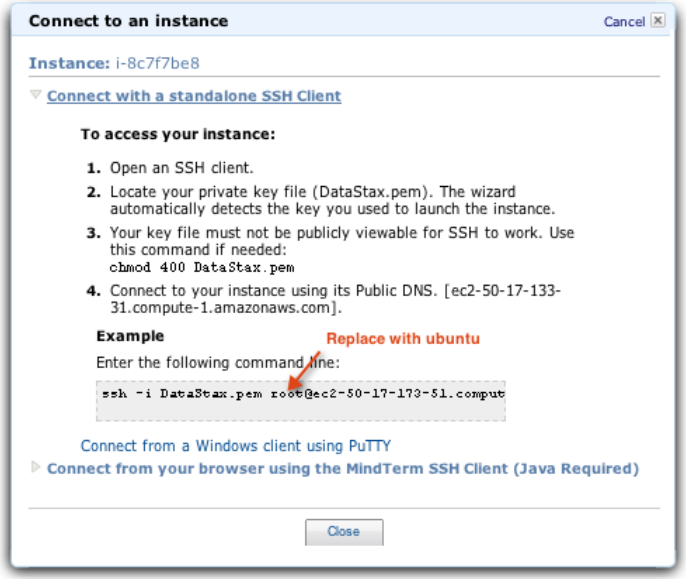
```
--release <release_version>
    Allows for the installation of a previous DSE
version
    Example: 1.0.2-1
    Default: Ignored

--opscenter no
    Disables the installation of OpsCenter on the
cluster
    Default: yes
```



Connecting to instances

- Using SSH Client
  - use your private key file (ex: DataStax.pem)
  - change the connection user from `root` to `ubuntu`





Configure OpsCenter Agents

➔ OpsCenter URL

http://<public-dns-of-first-instance>:8888/

➔ Install OpsCenter agents

- click on 'Fix'
  - username = `ubuntu`
  - credentials = the entire content of your private key (.pem)

⇨ The Dashboard shows the agent is connected

Expanding your cluster

➔ Adding node by reusing the DataStax AMI

http://www.datastax.com/documentation/cassandra/1.2/webhelp/index.html#cassandra/install/../../cassandra/install/expandAMI.html

1. stop cassandra
2. remove the data directories
3. configure `cluster_name` and `seeds` in file `cassandra.yaml`
4. start cassandra

➔ Extends the DataStax AMI

- adding `--seeds` option parameter

Demo