Managing a Hadoop cluster with Cloudera Manager

Mark Stetzer
@stetzer

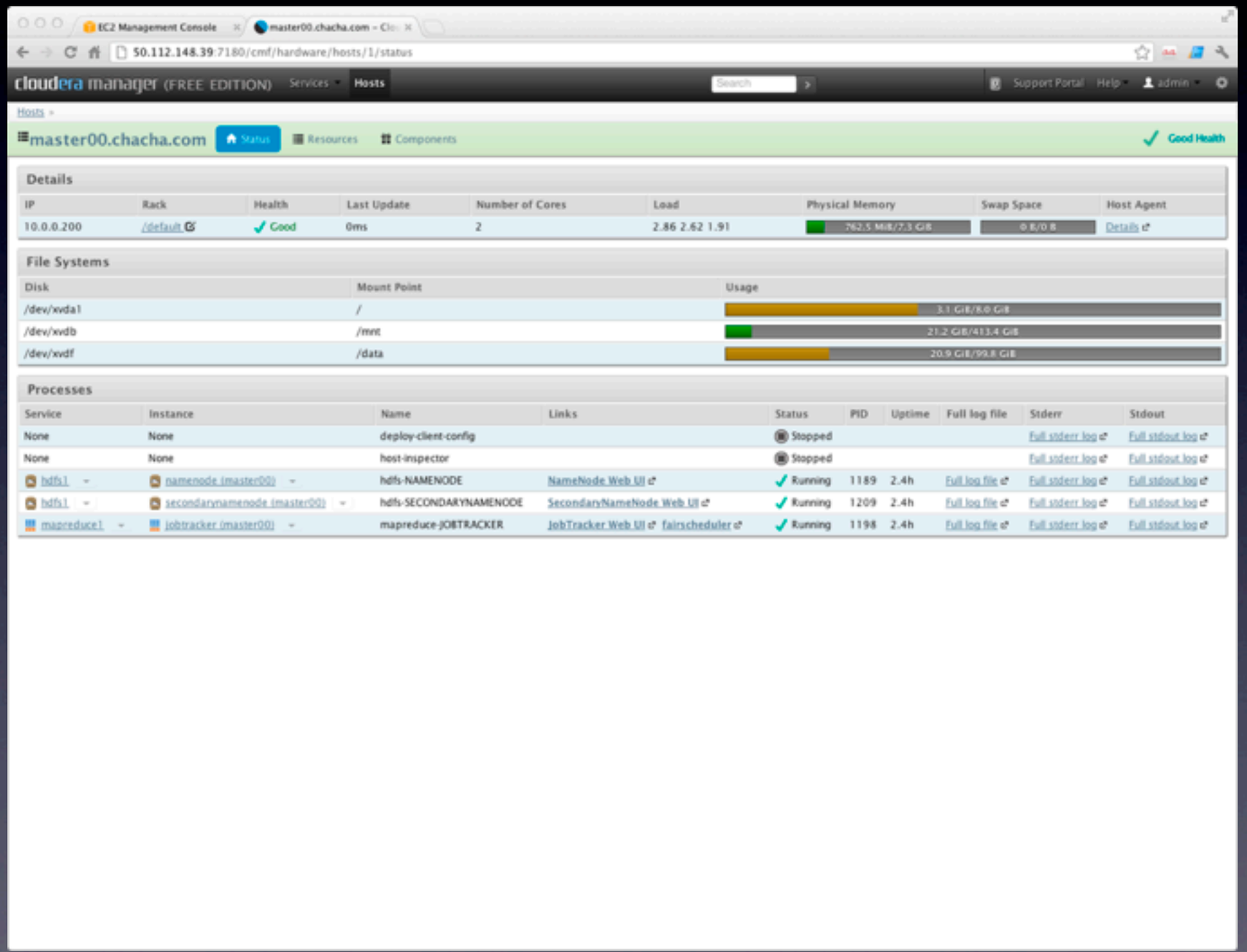
Why use CM?

- Point & click management
- Central location to configure all nodes & services

Capabilities Overview

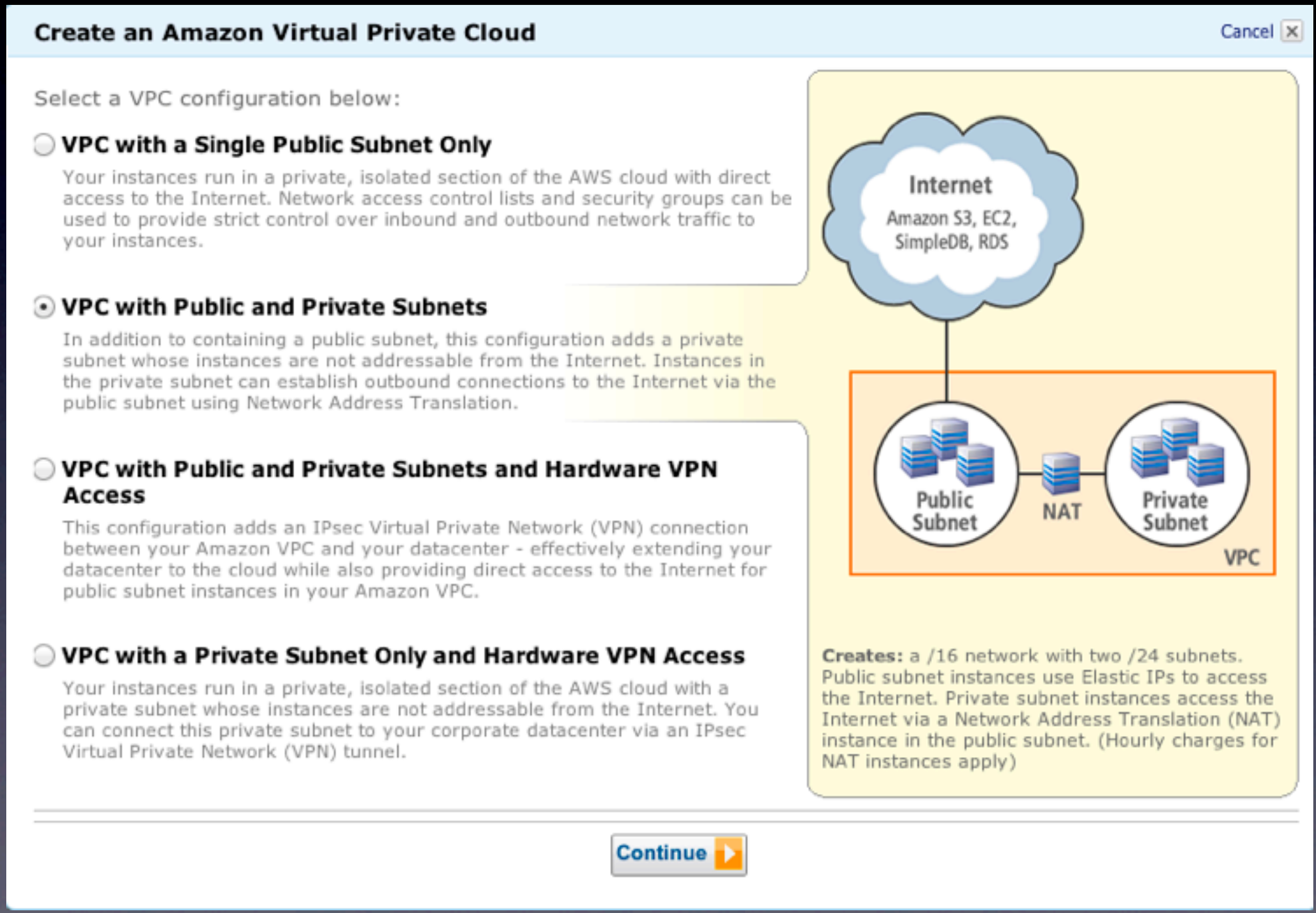
- Hosts dashboard showing load avg, disk usage, etc.
- Bundled config downloads
- CDH3 & CDH4 support
- HDFS HA (CDH4)
- Lucid, Precise, & Squeeze support
- API
- Gateway role (a.k.a. client box)
- ...and much more!

What does it look like?



First, some AWS setup...

Setting up a VPC



Configuring private IPs

- Set up DHCP to give instances same internal IP after restart

Assigning elastic IPs

- Instances in private subnet use NAT to get to Internet
- Instances in public subnet need elastic IPs

Mount EBS volume(s)

- mkfs.ext4 /dev/xvdf
- mkdir -m 000 /data
- echo "/dev/xvdf /data auto noatime 0 0" | sudo tee -a /etc/fstab
- mount /data

DNS

- Must configure DNS so hosts can address each other & reverse DNS works
- If instances can't identify themselves, Cloudera agents won't work correctly

Now take me through:

- Examining cluster health
- Adding a node to a cluster
- Configuring services on nodes
- Adding a new service

Questions?

Want to work with Hadoop at your day job?

http://about.chacha.com/about/careers/