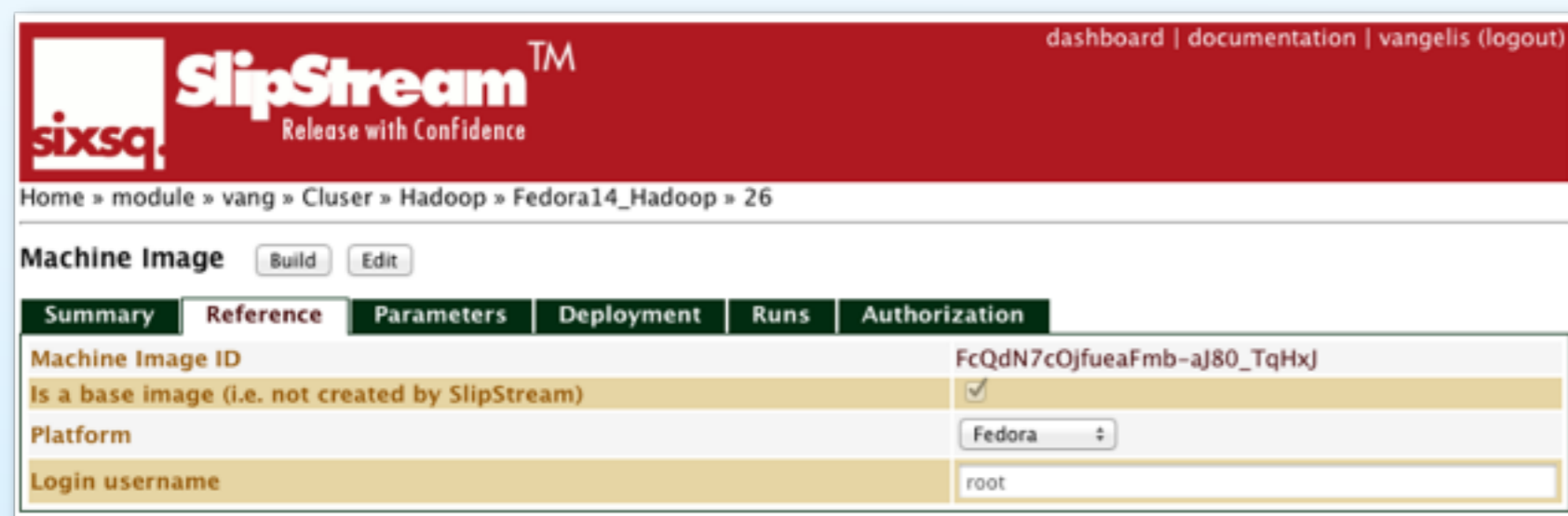


## MapReduce on StratusLab with SlipStream™

MapReduce is a patented software framework introduced by Google in 2004 to support distributed computing with large data sets on clusters of computers. Apache Hadoop is an open source implementation of MapReduce. StratusLab users can easily run MapReduce applications with Hadoop, using the available public services (reference IaaS cloud, Marketplace) and the power of the SlipStream™ platform.

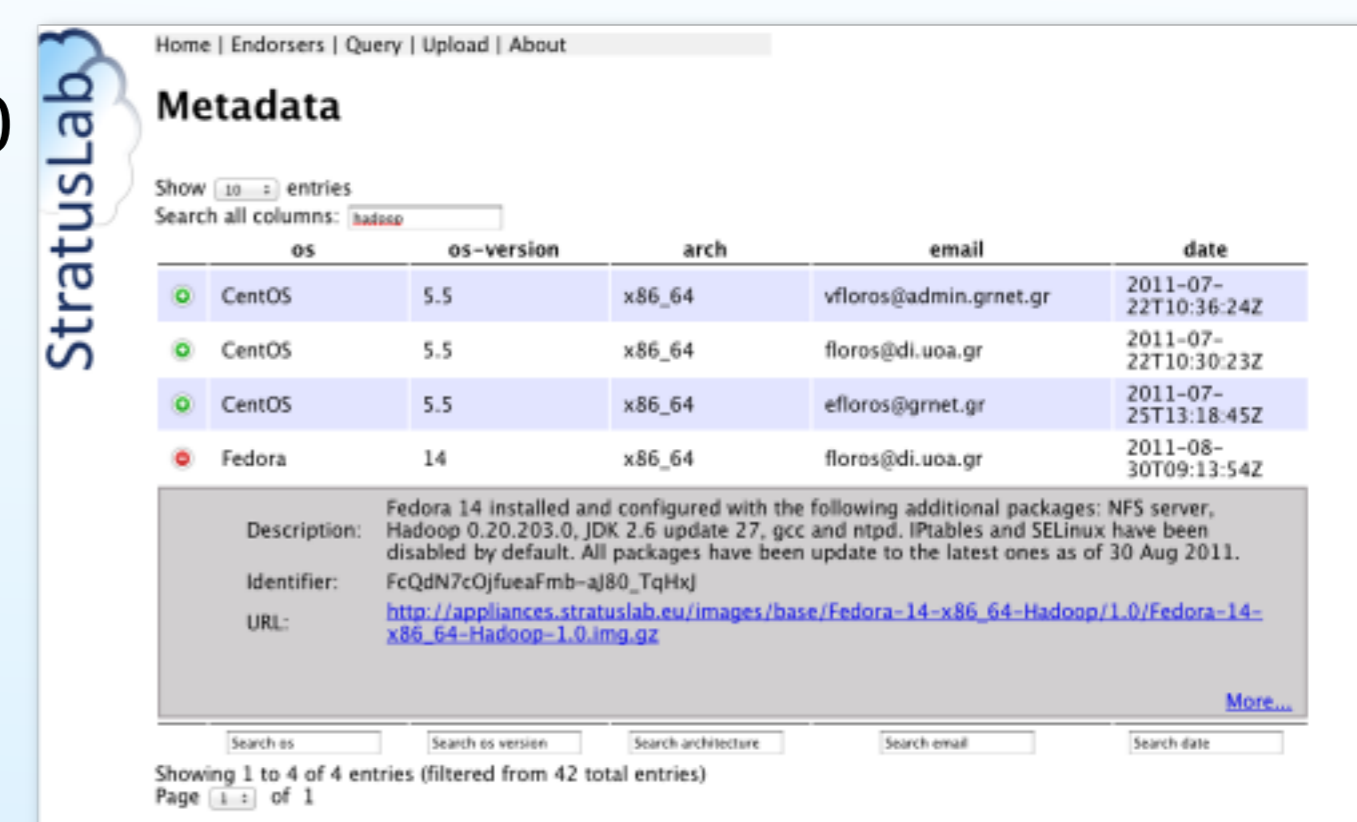
### 1. Define machine images

- Reference Hadoop appliance from Cluster module.
- Develop necessary logic that prepares environment for passwordless SSH logins and invokes application



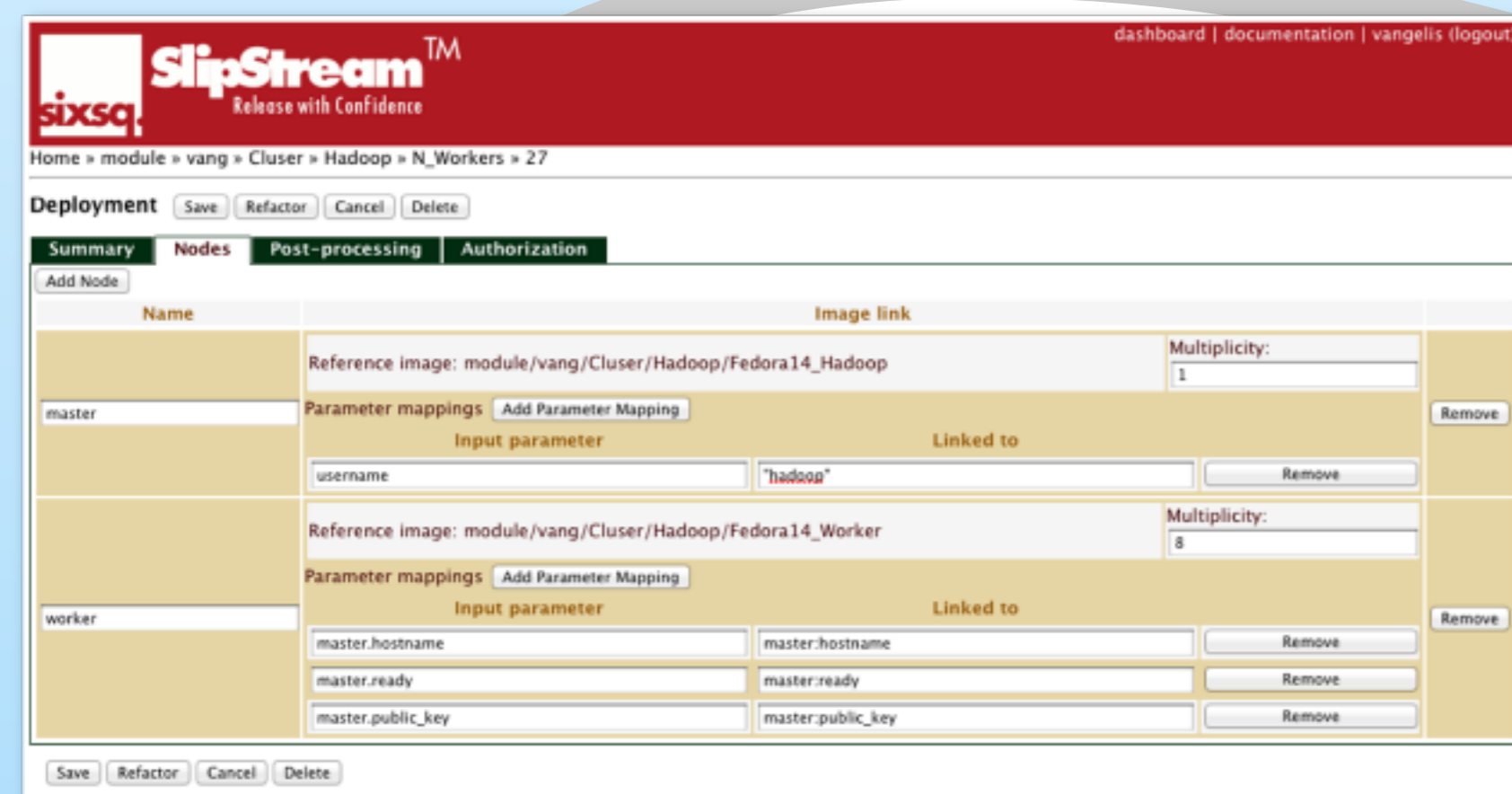
### Hadoop appliance available from StratusLab Marketplace

- Fedora 14 base OS
- Hadoop 0.20.203.0
- JDK 1.6 runtime



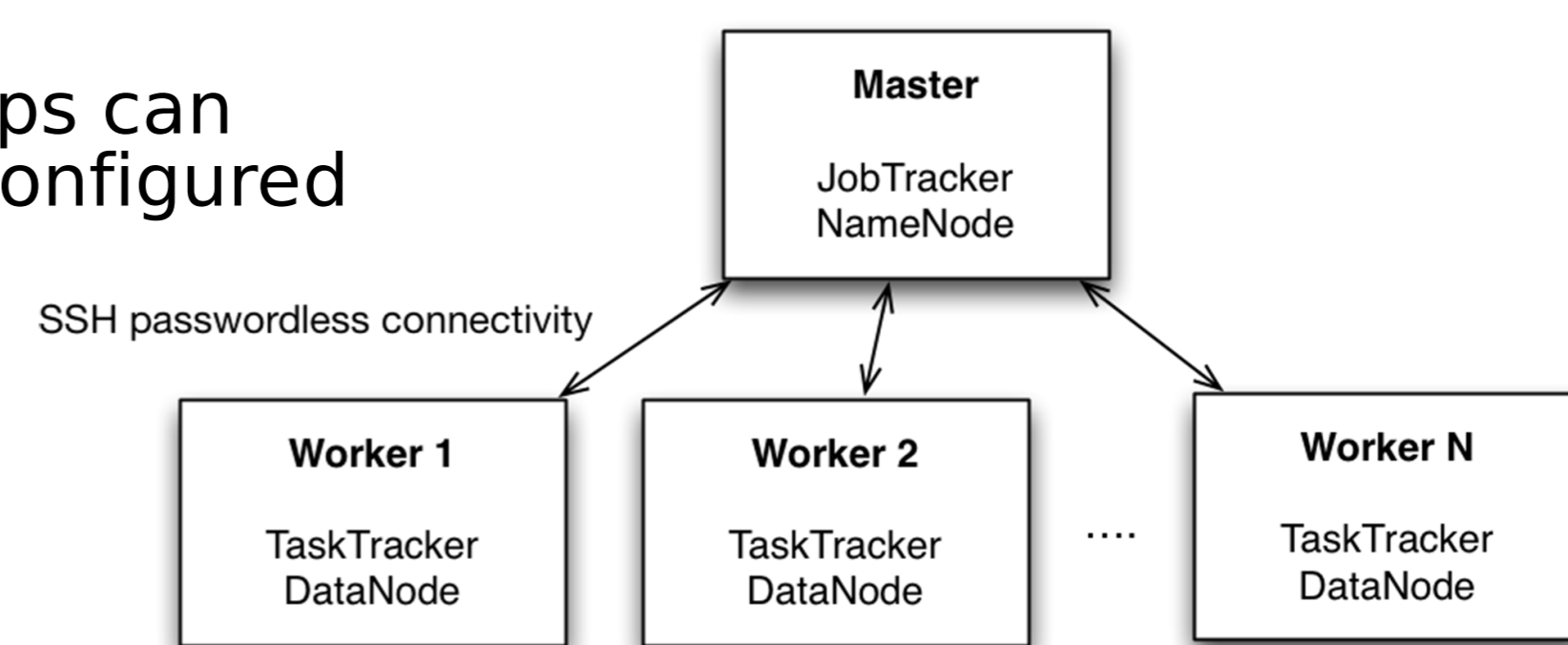
### 2. Define deployment

- Select instance type per node (master, worker)
- Set worker cardinality
- Define master/worker dependencies



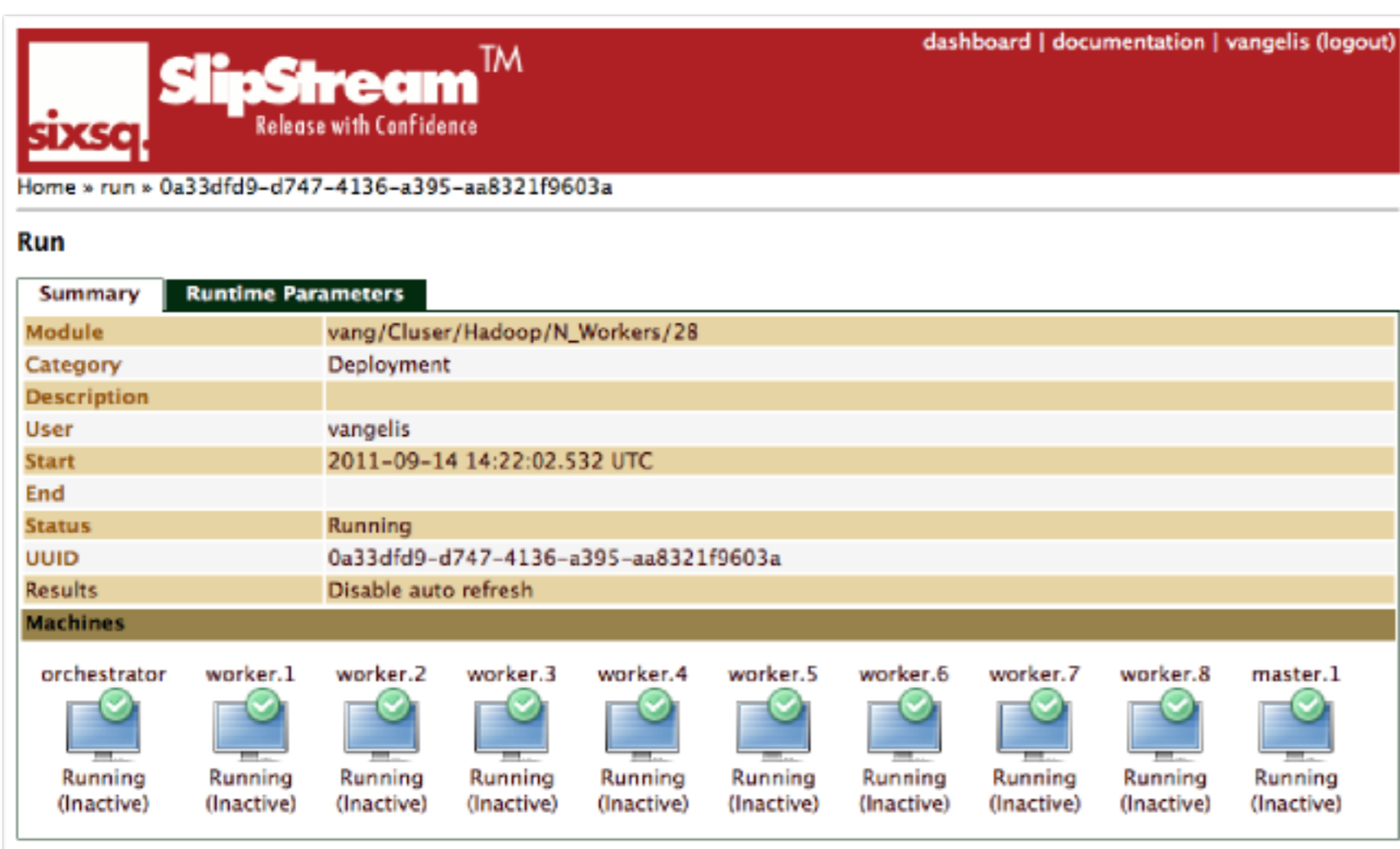
### Pre-configured Hadoop Cluster Architecture

- 1 Master node - Running JobTracker and NameNode service
- N workers acting as DataNodes and TaskTrackers
- Other setups can easily be configured



### 3. Run deployment

- Monitor deployment progress
- Hadoop web console available from master node
- Application output available for off-line retrieval upon termination



### Application monitoring

- Hadoop web console available from master node

