# **StratusLab** Enhancing Grid Infrastructures with Virtualization and Cloud Technologies

# **MapReduce on StratusLab with SlipStream<sup>TM</sup>**

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 laaS cloud, Marketplace) and the power of the SlipStream<sup>TM</sup> platform.

### **1. Define machine images**

 Reference Hadoop appliance from Cluster module.

• Develop necessary logic that prepares environment for passwordless SSH logins and invokes application

Sixse.	dashboard   documentation   vangelis (logout)
Home » module » vang » Cluser » Hadoop » Fedora14_Hadoop » 26	
Machine Image Build Edit	
Summary Reference Parameters Deployment Runs	Authorization
Machine Image ID	FcQdN7cOjfueaFmb-aJ80_TqHxJ
Is a base image (i.e. not created by SlipStream)	₫
Platform	Fedora 🗘
Login username	root

### Hadoop appliance available from StratusLab Marketplace

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

Show	10 == entries				
Searc	h all columns: 📊 os	os-version	arch	email	date
۰	CentOS	5.5	x86_64	vfloros@admin.grnet.gr	2011-07- 22T10:36:24Z
۰	CentOS	5.5	x86_64	floros@di.uoa.gr	2011-07- 22T10:30:23Z
۰	CentOS	5.5	x86_64	efloros@gmet.gr	2011-07- 25T13:18:45Z
۰	Fedora	14	x86_64	floros@di.uoa.gr	2011-08- 30T09:13:54Z
			DK 2.6 update 27, gcd	following additional package and ntpd. IPtables and SELing update to the latest ones as	ux have been
	Description: Identifier: URL:	disabled by default. A FcQdN7cOjfueaFmb-a http://appliances.stra x86_64-Hadoop-1.0.	aj80_TqHxJ ituslab.eu/images/bas	se/Fedora-14-x86_64-Hadoo	-

### **2. Define deployment**

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

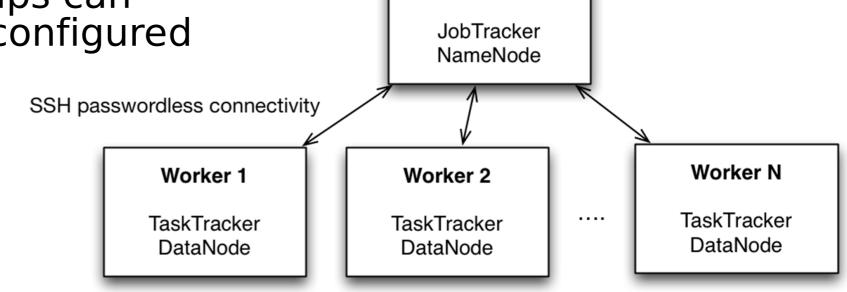
	<b>Stream</b> ease with Confidence							
	lluser » Hadoop » N_Workers » 27							
mmary Nodes	Post-processing Authorization							
Name		Image link						
	Reference image: module/vang/Cluser/Had	oop/Fedora14_Hadoop	Multiplicity:					
ster	Parameter mappings Add Parameter Mapping							
	Input parameter	Linked to	Remove					
	Reference image: module/vang/Cluser/Had		Multiplicity:					
rker	Parameter mappings Add Parameter Mapping Input parameter	Linked to		Remove				
	master.hostname	master:hostname	Remove					
	master.ready	master:ready	Remove					
	master.public_key	master:public_key	Remove					

## **3.** Run deployment

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

### **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



Master

# **Application monitoring**

Hadoop web console available from master node

master Hadoop Map/Reduce Administration

sixsq.		rean se with Confide 7-4136-a395		03a		dash	iboard   docu	imentation   1	vangelis (logo
lun									
Summary	Runtime Pa	rameters							
Module		vang/Cluse	r/Hadoop/N_	Workers/28					
Category		Deployment	t						
Description									
User		vangelis							
Start		2011-09-1	4 14:22:02.5	32 UTC					
End									
Status		Running							
UUID		0a33dfd9-d	1747-4136-a	395-aa8321	f9603a				
Results		Disable aut	o refresh						
Machines									
orchestrator	worker.1	worker.2	worker.3	worker.4	worker.5	worker.6	worker.7	worker.8	master.1
Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)	Running (Inactive)

Started: Wed Sep 14 11:08:09 UTC 2011
Version: 0.20.203.0, r1099333
Compiled: Wed May 4 07:57:50 PDT 2011 by com
Identifier: 201109141108

### Cluster Summary (Heap Size is 15.12 MB/966.69 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Graylisted Nodes	Excluded Nodes
0	0	1	<u>9</u>	0	0	0	0	18	18	4.00	Q	Q	Q
Schedu	ling In	formation	ı										
Queue Na	me State	Scheduling	) Informa	ation									
default	runnir	g N/A											
none													
Comple	eted Jo	bs											
Comple	eted Jo	Priority	User			Map Ma Total Co		educe %	Reduce Total	Reduces Completed	Job Sche Informati		Diagnostic
Jobid	9141108 0	Priority		warrie C	00.00%		mpleted C					ion	
Jobid	9141108 0	Priority		word 1	00.00%	Total Co	mpleted C	complete	Total		Informati	ion	Info

Website: http://stratuslab.eu Support: support@stratuslab.eu Contact: contact@stratuslab.eu Copyright © 2011, Members of the StratusLab collaboration

Centre National de la Recherche Scientifique France

earch.org Universidad Complutense Greek Research and de Madrid Technology Network S. A. Spain Greece

grnet



StratusLab is co-funded by the European Community's Seventh Framework Programme (Capacities) Grant Agreement INFSO-RI-261552.

