

Deploying HPC platforms using ProActive



The main steps

1. **Swarm.xml** to allow container-to-container network using **overlay network**, natively supported by docker engine **swarm** mode (a swarm is a cluster of Docker engines/nodes offering services: add/remove nodes,..). An overlay network requires a key-value store (here **consul** supported by docker) to holds information about the network state.
2. Deploy your HPC platform via **docker containers**, for a portable deployment. To start/control docker containers on (remote) hosts, we use the docker-machine command relying on ssh.
 - a. **HDFS.xml** if you just need a dedicated file system adapted to big data
 - b. **Spark.xml** for a big data processing platform
 - c. **HDFS.xml** and **Spark.xml** if you need both
3. Test your platform
 - a. **submit_spark_PI.xml** to only test your Spark platform (compute PI)
 - b. **submit_write_read_persons_HDFS.xml** to test both your HDFS and Spark platforms (write and read objects in Spark from/to the HDFS)

Swarm wkw params

instance_name	try.activeeon.com	base name of the docker containers (consul, swarm)
consulIP	localhost	IP of the targeted consul machine
swarmMasterAgentIP	localhost	IP of the targeted swarm master machine
swarmAgentIPsFile	undefined	file name of the targeted swarm agent machines IPs
userSSH	cperUser	username used by docker-machine to ssh
subnet	25.25.25.0/24	subnet for container networking
networkName	my-net	network name
dashboard_host_name	try.activeeon.com	hostname of the consul web portal
dashboard_port	8500	port of the consul web portal

HDFS wkw params

instance_name	hdfsContainer	base name of the HDFS docker containers
dashboard_port	6000	port of the HDFS web portal
namenode_IP_address	127.0.0.1	IP of the targeted namenode machine
optional_datanode_IP_addresses_file	undefined	file name of the targeted datanode machines IPs (if required)
datanode_starting_port	50010	Port range start of the datanodes
fs_name	25.25.25.2	name of the default file system
network	my-net	network name
hdfs_portal_host_name	try.activeeon.com	hostname of the HDFS web portal

Spark wkw params



Activeeon
SCALE BEYOND LIMITS

instance_name	sparkContainer	base name of the Spark docker containers
dashboard_port	5000	port of the Spark web portal
network	my-net	network name
master_IP_address	127.0.0.1	IP of the targeted master machine
optional_worker_IP_addresses_file	undefined	file name of the targeted worker machines IPs (if required)
optional_hdfs_IP_address	127.0.0.1	IP of the HDFS namenode machine (if required for a HDFS/Spark install)
spark_portal_host_name	try.activeeon.com	hostname of the Spark web portal

submit_spark_pi wkw params



Activeeon
SCALE BEYOND LIMITS

spark_master_url	spark://sparkContainerMaster:7077
network	my-net

spark master url (displayed on the spark web portal)

network name



spark_master_url	<input type="text" value="spark://sparkContainerMaster:7077"/>
hdfs_directory	<input type="text" value="hdfs://hdfsContainerNamenode:9000/persons"/>
network	<input type="text" value="my-net"/>

spark master url (displayed on the spark web portal)

hdfs directory for IO operations

network name