

Item 161

git_comments:

git_commits:

1. **summary:** HBASE-1961 HBase EC2 scripts; updated readme; hardcode keypair to 'root'; copy ssh private key to master
message: HBASE-1961 HBase EC2 scripts; updated readme; hardcode keypair to 'root'; copy ssh private key to master git-svn-id: <https://svn.apache.org/repos/asf/hadoop/hbase/trunk@882242> 13f79535-47bb-0310-9956-ffa450edef68

github_issues:

github_issues_comments:

github_pulls:

github_pulls_comments:

github_pulls_reviews:

jira_issues:

1. **summary:** HBase EC2 scripts
description: Attached tarball is a clone of the Hadoop EC2 scripts, modified significantly to start up a HBase storage only cluster on top of HDFS backed by instance storage. Tested with the HBase 0.20 branch but should work with trunk also. Only the AMI create and launch scripts are tested. Will bring up a functioning HBase cluster. Do "create-hbase-image c1.xlarge" to create an x86_64 AMI, or "create-hbase-image c1.medium" to create an i386 AMI. Public Hadoop/HBase 0.20.1 AMIs are available: i386: ami-c644a7af x86_64: ami-f244a79b launch-hbase-cluster brings up the cluster: First, a small dedicated ZK quorum, specifiable in size, default of 3. Then, the DFS namenode (formatting on first boot) and one datanode and the HBase master. Then, a specifiable number of slaves, instances running DFS datanodes and HBase region servers. For example: {noformat} launch-hbase-cluster testcluster 100 5 {noformat} would bring up a cluster with 100 slaves supported by a 5 node ZK ensemble. We must colocate a datanode with the namenode because currently the master won't tolerate a brand new DFS with only namenode and no datanodes up yet. See HBASE-1960. By default the launch scripts provision ZooKeeper as c1.medium and the HBase master and region servers as c1.xlarge. The result is a HBase cluster supported by a ZooKeeper ensemble. ZK ensembles are not dynamic, but HBase clusters can be grown by simply starting up more slaves, just like Hadoop. hbase-ec2-init-remote.sh can be trivially edited to bring up a jobtracker on the master node and task trackers on the slaves.

jira_issues_comments:

1. launch-hbase-slaves can be run independently of launch-hbase-cluster to grow the cluster on demand. Each additional slave contributes one datanode and its instance local storage to DFS, and one region server.
2. Sweet Andrew. Should we add this to our contrib? Pity its hard to subclass bash scripts but it is what it is. Thats excellent that you can say how many RS instances and then also how many instances in zk cluster. You've posted amis already? Ones for hbase? Thats what these are: {code} i386: ami-c644a7af x86_64: ami-f244a79b {code} The scripts look great: {code} [stack@aa0-000-12 ec2]\$./bin/hbase-ec2 Usage: hbase-ec2 COMMAND where COMMAND is one of: list list all running HBase EC2 clusters launch-cluster <group> <num slaves> launch a cluster of HBase EC2 instances - launch-master then launch-slaves launch-master <group> launch or find a cluster master launch-slaves <group> <num slaves> launch the cluster slaves terminate-cluster <group> terminate all HBase EC2 instances delete-cluster <group> delete the group information for a terminated cluster login <group|instance id> login to the master node of the HBase EC2 cluster screen <group|instance id> start or attach 'screen' on the master node of the HBase EC2 cluster proxy <group|instance id> start a socks proxy on localhost:6666 (use w/foxyproxy) push <group> <file> scp a file to the master node of the HBase EC2 cluster <shell cmd>

- <group|instance id> execute any command remotely on the master create-image create a HBase AMI {code} What you mean by a 'storage only cluster' above?
3. Yes I think we should drop this into our contrib/. "Storage only" indicates no mapreduce subsystem running. So, the cluster is for data storage and access only. But, it's trivial to start a jobtracker on the master and tasktrackers on the slaves with the addition of just two lines to hbase-ec2-init-remote.sh.
 4. Yes these are public AMIs for Hadoop+HBase 0.20.1: i386: ami-c644a7af x86_64: ami-f244a79b
 5. I think we should commit this to branch too, especially if public AMIs? Let me try them on our AMZ account just so someone else has exercised them... then lets commit to trunk and branch. We need to make the current wiki page obsolete too on commit. I can do that (the wiki page on ec2). Good stuff.
 6. There are a couple of minor fixups needed on commit, just fyi. I know what they are and can take care of them at that time. For example, hbase-ec2 needs a fixup to handle the additional parameter to launch-hbase-cluster.
 7. Here's a bit of feedback Andrew: + Edit the README. Make it about hbase rather than hadoop. Change the bin/hadoop-ec2 to be bin/hbase-ec2. + The usage is a little off when I type bin/hbase-ec2... for example, it says create-image when script is actually create-hbase-image, ditto for launch-master, etc. + I ran the create-hbase-image and it did this: {code}


```
stack@aron:~/checkouts/workspace/trunk/src/contrib/ec2$ ./bin/create-hbase-image INSTANCE_TYPE
is c1.xlarge. ARCH is x86_64. ./bin/create-hbase-image: line 32: ec2-describe-images: command not
found Starting a AMI with ID ami-f61dfd9f. ./bin/create-hbase-image: line 37: ec2-run-instances:
command not found Instance is . Polling server status (ec2-describe-instances ) ./bin/create-hbase-image:
line 45: ec2-describe-instances: command not found ./bin/create-hbase-image: line 45: ec2-describe-
instances: command not found ./bin/create-hbase-image: line 45: ec2-describe-instances: command not
found ./bin/create-hbase-image: line 45: ec2-describe-instances: command not found ./bin/create-hbase-
image: line 45: ec2-describe-instances: command not found ... {code} Does same if I add c1.xlarge as arg.
I did this: {code} stack@aron:~/checkouts/workspace/trunk/src/contrib/ec2$ ./bin/launch-hbase-cluster
stackcluster 3 3 Cluster name required! ... {code} Go easy.
```
 8. @stack: The former: You need both the EC2 API and AMI tools
 (<http://developer.amazonwebservices.com/connect/entry.jspa?externalID=351> ,
 <http://developer.amazonwebservices.com/connect/entry.jspa?externalID=368&categoryID=88>) installed
 and on the path. For Ubuntu, "apt-get install ec2-ami-tools". The latter: Must be something dumb I did.
 Let me check.
 9. The usage for hbase-ec2 is basically correct: {quote} if ["\$COMMAND" = "create-image"] ; then .
 "\$bin"/create-hbase-image \$* {quote} This is the same convention used by the Hadoop EC2 scripts. I did
 just add a 'launch-zookeeper' command that was missing.
 10. This is what you should see out of launch-hbase-cluster: {noformat} \$./launch-hbase-cluster cluster-0 4 3
 Creating/checking security groups Security group cluster-0-master exists, ok Security group cluster-0
 exists, ok Security group cluster-0-zookeeper exists, ok Starting ZooKeeper quorum ensemble. Starting
 an AMI with ID ami-c644a7af (arch i386) in group cluster-0-zookeeper Waiting for instance i-bf75efd7 to
 start: Started ZooKeeper instance i-bf75efd7 as ip-10-212-154-223.ec2.internal Public DNS
 name is ec2-174-129-186-94.compute-1.amazonaws.com. Starting an AMI with ID ami-c644a7af (arch
 i386) in group cluster-0-zookeeper Waiting for instance i-6d6af005 to start: Started ZooKeeper
 instance i-6d6af005 as ip-10-212-154-34.ec2.internal Public DNS name is ec2-67-202-48-84.compute-
 1.amazonaws.com. Starting an AMI with ID ami-c644a7af (arch i386) in group cluster-0-zookeeper
 Waiting for instance i-076af06f to start: Started ZooKeeper instance i-076af06f as ip-10-212-154-
 160.ec2.internal Public DNS name is ec2-174-129-153-78.compute-1.amazonaws.com. ZooKeeper
 quorum is ip-10-212-154-223.ec2.internal,ip-10-212-154-34.ec2.internal,ip-10-212-154-160.ec2.internal.
 Initializing the ZooKeeper quorum ensemble. ec2-174-129-186-94.compute-1.amazonaws.com hbase-
 ec2-init-zookeeper-remote.sh 100% 1201 1.2KB/s 00:00 starting zookeeper, logging to
 /mnt/hbase/logs/hbase-root-zookeeper-ip-10-212-154-223.out ec2-67-202-48-84.compute-
 1.amazonaws.com hbase-ec2-init-zookeeper-remote.sh 100% 1201 1.2KB/s 00:00 starting zookeeper,
 logging to /mnt/hbase/logs/hbase-root-zookeeper-ip-10-212-154-34.out ec2-174-129-153-78.compute-
 1.amazonaws.com hbase-ec2-init-zookeeper-remote.sh 100% 1201 1.2KB/s 00:00 starting zookeeper,
 logging to /mnt/hbase/logs/hbase-root-zookeeper-ip-10-212-154-160.out Testing for existing master in
 group: cluster-0 Starting master with AMI ami-f244a79b (arch x86_64) Waiting for instance i-bf6af0d7 to
 start..... Started as ip-10-245-101-219.ec2.internal Master is ec2-72-44-33-230.compute-
 1.amazonaws.com, ip is 72.44.33.230, zone is us-east-1d. Starting 4 AMI(s) with ID ami-f244a79b (arch
 x86_64) in group cluster-0 in zone us-east-1d i-3f6bf157 i-316bf159 i-336bf15b i-356bf15d {noformat}
 And then if you log on to the master a few minutes later: {noformat} \$ ssh -i id_rsa_root root@ec2-72-
 44-33-230.compute-1.amazonaws.com _| _|_) Fedora 8 _| (/ 64-bit _|_|_|_|_| Welcome to an EC2

Public Image :-> Base --[see /etc/ec2/release-notes]-- [root@ip-10-245-101-219 ~]# jps -l 1358
 org.apache.hadoop.hdfs.server.namenode.NameNode 1567 org.apache.hadoop.hbase.master.HMaster
 1820 sun.tools.jps.Jps 1434 org.apache.hadoop.hdfs.server.datanode.DataNode [root@ip-10-245-101-219
 ~]# hbase shell HBase Shell; enter 'help<RETURN>' for list of supported commands. Version: 0.20.1,
 r822817, Wed Oct 7 11:55:42 PDT 2009 hbase(main):001:0> status 'simple' 4 live servers ip-10-242-133-
 139.ec2.internal:60020 1258079538311 requests=0, regions=2, usedHeap=26, maxHeap=987 ip-10-242-
 97-203.ec2.internal:60020 1258079557660 requests=0, regions=0, usedHeap=37, maxHeap=987 ip-10-
 245-101-187.ec2.internal:60020 1258079561915 requests=0, regions=0, usedHeap=37, maxHeap=987 ip-
 10-245-111-47.ec2.internal:60020 1258079556528 requests=0, regions=0, usedHeap=37, maxHeap=987
 0 dead servers {noformat}

11. Tested terminate-hbase-cluster now also.

12. **body:** More testing. Good to be dropped in now.

label: test

13. I'm +1 on adding ec2 contrib if only so I can help work on the stuff as I find issues.

14. This latest tarball still has usage that does not mention hbase: {code}

```
stack@aron:~/checkouts/workspace/trunk/src/contrib/ec2$ ./bin/hbase-ec2 Usage: hbase-ec2
COMMAND where COMMAND is one of: list list all running HBase EC2 clusters launch-cluster
<name> <slaves> <zoos> launch a HBase cluster launch-zookeeper <name> <zoos> launch the
zookeeper quorum launch-master <name> launch or find a cluster master launch-slaves <name> <slaves>
launch the cluster slaves terminate-cluster <name> terminate all HBase EC2 instances delete-cluster
<name> clean up after a terminated cluster login <name|instance id> login to the master node screen
<name|instance id> start or attach 'screen' on the master proxy <name|instance id> start a socks proxy on
localhost:6666 push <name> <file> scp a file to the master node <shell cmd> <group|instance id>
execute a command on the master create-image create a HBase AMIde} {code} .. it should be 'launch-
hbase-cluster' instead of 'launch-cluster'? etc. I did the apt-get install of ami tools. Got this error when I
tried below command: {code} $ ./bin/launch-hbase-cluster x 3 3 Creating/checking security groups
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 34: ec2-
describe-group: command not found Creating group x-master
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 37: ec2-
add-group: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 38: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 39: ec2-
authorize: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 51: ec2-describe-group: command not found Creating group x
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 54: ec2-
add-group: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 55: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 56: ec2-
authorize: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 65: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 66: ec2-
authorize: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 73: ec2-describe-group: command not found Creating group x-zookeeper
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 76: ec2-
add-group: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 77: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 78: ec2-
authorize: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 80: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 81: ec2-
authorize: command not found /home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-
cluster-secgroups: line 82: ec2-authorize: command not found
/home/stack/checkouts/workspace/trunk/src/contrib/ec2/bin/init-hbase-cluster-secgroups: line 83: ec2-
authorize: command not found {code}
```

15. You need both the api and ami tools. Thought you had one but not the other. You need both. Click through both links in the readme and the comment above on this issue. Maybe on ubuntu 'apt-get install ec2-api-tools' will pull down the rest. The help is correct. It is for that script. See the if-else switch inside. It is not help for the *-hbase-* worker scripts. What we need is a wiki page on this. Will put one up.

16. Committed scripts into src/contrib/ec2 on trunk. Two changes from the tarball on this issue: 1) Set file descriptor limit for root to 32768. See <http://wiki.apache.org/hadoop/Hbase/Troubleshooting#A6> 2) Set epoll user descriptor limit to 32768. See <http://pero.blogs.aprilmayjune.org/2009/01/22/hadoop-and-linux-kernel-2627-epoll-limits/>
17. **body:** Playing with scripts more: + Has hardcoded JAVA_VERSION in hbase-ec2-env.sh. Is that intentional? + If I do the following, it doesn't work: {code} \$./bin/hbase-ec2 list -h Required option '-K, -private-key KEY' missing (-h for usage) No running clusters. {code} The -h is not passed to ec2-describe-instances. I see that the hadoop scripts have the same issue. So it seems like we require people to fill in keys into the hbase-ec2-env.sh. Is that right? If so, lets add stuff to the readme?
label: documentation
18. Yeah, sorry if I'm being a bit thick here Andrew but seems like I need to have defines for EC2_PRIVATE_KEY and for EC2_CERT in my environment to use these scripts? Is that right?
19. bq. Has hardcoded JAVA_VERSION in hbase-ec2-env.sh. Is that intentional? Yes. That's used when building the AMI only. Also you can see the location of HBase and JVM packages is hardcoded, and is currently a bucket of mine in S3. We should host HBase tarballs somewhere on ASF systems instead. I put up JVM packages into S3 because Sun's java download URLs are not stable (or at least have not been in the past). bq. The -h is not passed to ec2-describe-instances Argument handling by the scripts is basically as-is from parent. bq. So it seems like we require people to fill in keys into the hbase-ec2-env.sh The requirement is: * Fill in AWS_ACCOUNT_ID * Fill in AWS_ACCESS_KEY_ID * Fill in AWS_SECRET_ACCESS_KEY * Fill in KEY_NAME * Make sure a file named id_rsa_\${KEY_NAME} exists in EC2_KEYDIR * Define EC2_PRIVATE_KEY defined in the environment. This is usually done when setting up the API tools. Probably we should add a line for that in hbase-ec2-env.sh to call that out. The user must also put the private key into EC2_KEYDIR as pk*.pem and their cert in there as cert*.pem. All should go into the readme and up on the wiki. I'll add this to the readme now.
20. **body:** Tightened up configuration a bit and documented configuration required for hbase-ec2-env.sh in README as r880937.
label: documentation
21. @apurtell: I'm having a bit of trouble making these stuff work. I've edited hbase-ec2-env.sh adding in my vitals but it seems like i need to export EC2_PRIVATE_KEY, etc., otherwise the ec2 programs don't see their content. For example, I seem to have to do this to get the private key and cert into place so the amz programs will pick them up: {code} Index: bin/list-hbase-clusters
===== --- bin/list-hbase-clusters (revision 881125) +++ bin/list-hbase-clusters (working copy) @@ -23,7 +23,7 @@ .
"\$bin"/hbase-ec2-env.sh # Finding HBase clusters -CLUSTERS=`ec2-describe-instances | awk
"'RESERVATION' == \$1 && \$4 ~ /-master\$/, 'INSTANCE' == \$1' | tr '\n' '\t' | grep running | cut -f4 | rev
| cut -d '-' -f2- | rev` +CLUSTERS=`ec2-describe-instances -K \$EC2_PRIVATE_KEY | awk
"'RESERVATION' == \$1 && \$4 ~ /-master\$/, 'INSTANCE' == \$1' | tr '\n' '\t' | grep running | cut -f4 | rev
| cut -d '-' -f2- | rev` [-z "\$CLUSTERS"] && echo "No running clusters." && exit 0 {code} (See how I added the -K above). ... or if I export the keys, it works as follows: {code} # Your AWS private key file -- must begin with 'pk' and end with '.pem' -EC2_PRIVATE_KEY= +export
EC2_PRIVATE_KEY=/Users/stack/.ec2/... {code} Maybe its because I'm on a macintosh? I see that the hadoop scripts don't export variables nor pass with -K nor -C. Let me try over on linux.
22. I think people generally set up their environment per the directions on configuring the ami and api tools. That's what I did, using those tools on their own for some time before taking on this issue. Better to add -K and -C parameters to the commands then rely on the environment as you demonstrate. Glad to have you finding these things. Thanks for sticking with it.
23. Rather than add -K and -C I think need to do export of the variables in hbase-ec2-env.sh. Whats odd is that it would seem the hadoop scripts should have the same issue. Let me play some more.
24. @stack: I already committed modifications which put -K and -C into a variable \$TOOL_OPTS which is passed on the command line to every tool. I'm sure the Hadoop scripts have this same issue as all of this was copied from them.
25. Andrew, thats amazing that you have the ascent to java license inline... and how it installs everything, java, hadoop, hbase, and... ganglia included. I got this: {code} Unable to read instance meta-data for product-codes Creating bundle manifest... ec2-bundle-vol complete. ERROR: Error talking to S3: Server.AccessDenied(403): Only the bucket owner can access this property Done Client.InvalidManifest: HTTP 403 (Forbidden) response for URL http://s3.amazonaws.com:80/hbase-images/hbase-0.20.1-x86_64.manifest.xml: check your S3 ACLs are correct. Terminate with: ec2-terminate-instances i-971a79ff {code} I should have changed this, S3_BUCKET, in hbase-ec2-env it seems so its a bucket I

have access to.. Thats no prob. I tried running `$./bin/hbase-ec2 launch-cluster stackcluster 3 3...` and all went well till zk nodes came up: {code} Starting ZooKeeper quorum ensemble. Starting an AMI with ID ami-c644a7af (arch i386) in group stackcluster-zookeeper Waiting for instance i-6b1c7f03 to start: Started ZooKeeper instance i-6b1c7f03 as ip-10-245-59-97.ec2.internal Public DNS name is ec2-72-44-33-220.compute-1.amazonaws.com. Starting an AMI with ID ami-c644a7af (arch i386) in group stackcluster-zookeeper Waiting for instance i-471c7f2f to start: Started ZooKeeper instance i-471c7f2f as ip-10-245-58-191.ec2.internal Public DNS name is ec2-67-202-46-119.compute-1.amazonaws.com. Starting an AMI with ID ami-c644a7af (arch i386) in group stackcluster-zookeeper Waiting for instance i-c51c7fad to start: Started ZooKeeper instance i-c51c7fad as ip-10-244-206-65.ec2.internal Public DNS name is ec2-174-129-119-249.compute-1.amazonaws.com. ZooKeeper quorum is ip-10-245-59-97.ec2.internal,ip-10-245-58-191.ec2.internal,ip-10-244-206-65.ec2.internal. Initializing the ZooKeeper quorum ensemble. ec2-72-44-33-220.compute-1.amazonaws.com lost connection ec2-67-202-46-119.compute-1.amazonaws.com lost connection ec2-174-129-119-249.compute-1.amazonaws.com lost connection ... {code} They seem up in the console but the above seems to have stopped the script going on to start the regionservers? I tried it twice and got same lost connection both times. Terminate cluster is sweet the way it asks you if you want to shut down all.

26. Is the 'lost connection' what hbase-1982 is about?
27. No, HBASE-1982 is about handling what happens if you ask for 100 slaves (assuming your EC2 account is set up for that many concurrent instances), etc. "Lost connection" is a complaint from SSH. Check your networking. Firewall in the way on your end I wonder. Or any errors seen while setting up the security groups? Add "-v" to SSH_OPTS in hbase-ec2-env. Additional "-v" switches increment the debug level, up to "-v -v -v".
28. Drop the "-q" from SSH_OPTS before adding "-v"s.
29. Hi Andy, I ran into the same problem to stack had -- "lost connection" when try to start ZK. Then, I found that using KEY_NAME other than "root" fails to connect to the ZK because the user with the KEY_NAME doesn't exist on the ZK server. For example, I tried KEY_NAME=hbase, and I couldn't login to the ZK server by "ssh -i ~/.ec2/id_rsa_hbase hbase@ec2- ... I created a key pair named "root" and I got my cluster up and running. Thanks, Tatsuya
30. **body:** Nice find Tatsuya. I'll hardcode the key name to root. In retrospect \$KEY_NAME is confusing as a variable because it cannot currently be changed.
label: code-design
31. Nice Tatsuya.
32. **body:** Feedback up on hbase-user@ from Naresh Rapolu: {quote} Your scripts are working fine. We restarted everything and tested, and they are working fine. A few issues though : - While starting, launch-hbase-cluster gives the following error. error: "fs.epoll.max_user_instance" is an unknown key. It occurs during starting zookeeper instances. - We needed MapReduce along with HBase. The note on the JIRA page that you only need to add only two lines in hbase-ec2-env.sh is insufficient. The following changes need to be made. 1. hbase-ec2-env.sh should write mapred.job.tracker property into hadoop-site.xml (Also shouldnt you be having core-site.xml and hdfs-site.xml as it is hadoop-0.20.1 ??? Infact because of this , there are warning messages all over the place when you are using hdfs through command line). 2. HADOOP_CLASSPATH in hadoop/conf/hadoop-env.sh needs to be changed in the underlying AMI, to include hbase, zookeeper jars and conf directory. Probably you can modify the public AMI, and recreate the bundle as the paths to these are known apriori. 3. For other users, the following three lines should be added in hbase-ec2-env.sh For master: "\$HADOOP_HOME"/bin/hadoop-daemon.sh start jobtracker "\$HADOOP_HOME"/bin/hadoop-daemon.sh start tasktracker For slave: "\$HADOOP_HOME"/bin/hadoop-daemon.sh start tasktracker. {quote} Incorporate these suggestions. bq. error: "fs.epoll.max_user_instance" is an unknown key This is a bit of future proofing. That's not a known sysctl key until kernel 2.6.27, at which point oddly low epoll user descriptor limits go into effect. See <http://pero.blogs.aprilmayjune.org/2009/01/22/hadoop-and-linux-kernel-2627-epoll-limits/>. At some point there may be a 2.6.27 based AKI. I could /dev/null the message but then some other more serious potential problem with sysctl would be hidden. bq. Also shouldnt you be having core-site.xml and hdfs-site.xml as it is hadoop-0.20.1 Yes. What I did for this initial work is adapt the Hadoop EC2 scripts, which target 0.19.
label: code-design
33. Committed to branch now also. Cancelling patch. Will keep this issue open for tracking related activities.
34. Thanks very much for providing these scripts! I've pulled the latest from svn (as of 2009-12-05) and when I run "bin/hbase-ec2 launch-cluster sethbase 5 5" I can see this in the output: [Deprecated] Xalan: org.apache.xml.res.XMLErrorResources_en_US Starting ZooKeeper quorum ensemble.

/Users/sethladd/Code/hbase/trunk/src/contrib/ec2/bin/launch-hbase-zookeeper: line 70: seq: command not found
 ZooKeeper quorum is . Initializing the ZooKeeper quorum ensemble. Testing for existing master in group: sethbase [Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US [Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US Starting master with AMI (arch x86_64)
 Required parameter 'AMI' missing (-h for usage) Waiting for instance to start.[Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US .[Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US Notice the "line 70: seq: command not found" Also notice the "Required parameter 'AMI' missing I am running this on Mac OS X 10.6. I have also just downloaded the EC2 AMI and API tools (as of 2009-12-05) Also, it appears like nothing is starting (though I don't know how long it usually takes) but I just see a ".[Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US" repeated over and over. Any ideas or tips? Thanks again for the scripts and the help!

35. For the record, I replaced line 70 of launch-hbase-zookeeper with this line: for inst in {1..\$NO_INSTANCES} ; do and that seemed to do the trick. The above might make it more cross platform, assuming that bash is in use. Still getting the "Required parameter 'AMI' missing (-h for usage)" error but I'll look at that next.
36. The config needs S3_BUCKET to be "iridiant-bundles" for now, if you want to use one of the public images I put up. Or, use create-hbase-image to make yourself a private AMI. In the latter case set S3_BUCKET to your private bucket. I committed an update to hbase-ec2-env.sh and also your patch to launch-hbase-zookeeper. Thanks for the feedback Seth!
37. **body:** Thanks Andrew, that allowed my zookeeper instances to start! I received a 503 when (apparently) trying to start the hbase nodes. But I'll try again. org.codehaus.xfire.fault.XFireFault: Server returned error code = 503 for URI : https://ec2.amazonaws.com. Check server logs for details at org.codehaus.xfire.fault.XFireFault.createFault(XFireFault.java:89) at org.codehaus.xfire.client.Invocation.invoke(Invocation.java:83) at org.codehaus.xfire.client.Invocation.invoke(Invocation.java:114) at org.codehaus.xfire.client.Client.invoke(Client.java:336) at org.codehaus.xfire.client.XFireProxy.handleRequest(XFireProxy.java:77) at org.codehaus.xfire.client.XFireProxy.invoke(XFireProxy.java:57) at \$Proxy12.describeInstances(Unknown Source) at com.amazon.aes.webservices.client.Jec2.describeInstances(Jec2.java:1390) at com.amazon.aes.webservices.client.Jec2.describeInstances(Jec2.java:1354) at com.amazon.aes.webservices.client.cmd.DescribeInstances.invokeOnline(DescribeInstances.java:49) at com.amazon.aes.webservices.client.cmd.BaseCmd.invoke(BaseCmd.java:719) at com.amazon.aes.webservices.client.cmd.DescribeInstances.main(DescribeInstances.java:58) Caused by: org.codehaus.xfire.XFireRuntimeException: Server returned error code = 503 for URI : https://ec2.amazonaws.com. Check server logs for details at org.codehaus.xfire.transport.http.HttpChannel.sendViaClient(HttpChannel.java:130) at org.codehaus.xfire.transport.http.HttpChannel.send(HttpChannel.java:48) at org.codehaus.xfire.handler.OutMessageSender.invoke(OutMessageSender.java:26) at org.codehaus.xfire.handler.HandlerPipeline.invoke(HandlerPipeline.java:131) at org.codehaus.xfire.client.Invocation.invoke(Invocation.java:79) ... 10 more.[Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US Started ZooKeeper instance i-cd5014a5 as ip-10-212-149-139.ec2.internal [Deprecated]
 Xalan: org.apache.xml.res.XMLErrorResources_en_US Public DNS name is ec2-75-101-218-238.compute-1.amazonaws.com. ZooKeeper quorum is ip-10-245-63-79.ec2.internal,ip-10-212-151-194.ec2.internal,ip-10-212-149-139.ec2.internal. Initializing the ZooKeeper quorum ensemble. ec2-72-44-53-128.compute-1.amazonaws.com lost connection ec2-174-129-77-191.compute-1.amazonaws.com lost connection ec2-75-101-218-238.compute-1.amazonaws.com lost connection
label: code-design
38. Also, there's a bit of inconsistency. ~/Code/hbase/trunk/src/contrib/ec2 : bin/hbase-ec2 Usage: hbase-ec2 COMMAND where COMMAND is one of: list list all running HBase EC2 clusters launch-cluster <name> <slaves> <zoos> launch a HBase cluster But the README says: 4) ./bin/hbase-ec2 launch-cluster <name> <nr-zoos> <nr-slaves>, e.g ./bin/hbase-ec2 launch-cluster testcluster 3 3 Notice that they disagree with order of command line arguments. Reading the code, it looks like the correct order is nr-slaves and then nr-zoos
39. The 503 errors are internal to EC2 and there is little we can do about them except perhaps roll back and try again. I think perhaps you and I were testing around the same time as I got these this afternoon as well for a while. "Lost connection" means SSH did not succeed, probably because it is not fully/properly

configured yet: you do not have a keypair named "root" and/or have not configured the path to the ssh private key file in hbase-ec2-env.sh.

40. Thanks again Andrew. You are correct, the 503 were transient. I tried it later and it worked. My SSH problem was because of incorrect permissions on id_rsa_root. I would suggest adding the following note in the README, under: " Make sure the private part of your AWS SSH keypair exists in the same directory as EC2_PRIVATE_KEY with the name id_rsa_root." add: "Also, ensure that the permissions on id_rsa_root are 600 (ONLY owner readable/writable)"
41. I think I've discovered another issue. In launch-hbase-master, line 91 reads: scp \$SSH_OPTS \$PRIVATE_KEY_PATH "root@\$MASTER_EC2_HOST:/root/.ssh/id_rsa" I think \$PRIVATE_KEY_PATH should actually be \$EC2_ROOT_SSH_KEY I'll give it a shot.
42. Thanks for all the feedback Seth. You are correct on both points and I have committed your suggestions.
43. **body:** Quick spelling tweak. Currently, in the README, it says: "Also, insure that" it should be "Also, ensure that"
- label:** documentation
44. Closed per HBASE-2543.