



Siamak Faridani <faridani@gmail.com>

Hadoop: What I wish I knew before starting

CS 294-1 on Piazza <no-reply@piazza.com>

Thu, Apr 12, 2012 at 9:53 AM

To: faridani@berkeley.edu

Your classmate posted a new Note.

Hadoop: What I wish I knew before starting

Thought I'd share about the mishaps I've come across while trying to do this homework.

- To compile the jar file for hadoop in eclipse see @97
- If you want to use an xml parser so you get the entire page per map call see @98
- Call your JobConf's setNumReduceTasks **before** you create the Job object (for some reason I didn't and therefore – shocker – things didn't work).
- If you are using a combiner double check that it has the same output format as your mapper (I didn't do this by accident and it caused a very long debug process).
- The Job functions setOutputKeyClass and setOutputValueClass are for the *mapper's* output not the reducers, I don't know why and I wasn't able to figure it out from the API.

To everyone, if you found this note useful please share your own mishaps. Help your fellow classmates quickly get to the goal of the homework, which is to use hadoop to do data mining (not struggle to get hadoop to work). Click the edit button below and add to this list. :)

#hadoop #hw3 #icluster

Click here to view. Search or link to this question with @100.

Sign up for more classes at <http://piazza.com/berkeley>.

Thanks,
The Piazza Team

--

Contact us at team@piazza.com

You're receiving this email because faridani@berkeley.edu is enrolled in CS 294-1 at University of California, Berkeley (UC Berkeley). Sign in or un-enroll from this class.

If you already have a Piazza account under another email address, link faridani@berkeley.edu to that account here.