How to study this tutorial

Igor Yakushin

`ivy2@uchicago.edu`

How to study this tutorial

- The main point of this tutorial is to show what useful tools are availabe in the Hadoop zoo, how to get started with various tools, how to use them on RCC clusters. Very often, getting started is the most difficult thing.
- The tutorial does not go deep into each tool. Once you know how to use those tools for simplest tasks, it would be much easier to understand the original documentation and go deeper on your own. Some tools, like Spark and Hive, are studied in more depth in Big Data Platform class
- In the first pass, study all the material in the exact order it is presented without skipping anything
- Do yourself all the examples presented in the tutorial

How to study this tutorial

- The slides and examples contain complete information but some things are easier to understand from the videos
- The videos do not have complete information, they do not repeat slides: they only show things that are easier to understand from videos.
- Therefore, do not skip slides and examples.
- After the first pass over the tutorial, you can use it as a reference and review topics in any order since it is organized in a modular way and it is easy to find what you might be looking for.
- If something is still not clear, ask me questions by e-mail or come see me during my office hours
- To get all the slides and examples, rather than downloading them one by one from canvas, do on midway or Hadoop cluster:

  ```
  git clone git@github.com:igory1999/Hadoop_video.git
  ```