



DEEP  
LEARNING  
INSTITUTE



DLI Accelerated Data Science Teaching Kit

# Lecture 10.5 - How to Try Hadoop



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

# What can you use Hadoop for?

- As a “swiss knife”
- Works for many types of analyses/tasks (but not all of them)
  - More recent technologies (e.g., **Spark**) may be faster for machine learning tasks
- What if you want to write less code?
  - There are tools to make it easier to write MapReduce program (**Pig**), or to query results (**Hive**)

# How to try Hadoop?

- Hadoop can run on a single machine (e.g., your laptop)
  - Takes < 30min from setup to running
- Or a “home-grown” cluster
  - Research groups often connect retired computers as a small cluster
- Amazon Web Service (AWS), Microsoft Azure
  - You only pay for what you use, e.g., compute time, storage



DEEP  
LEARNING  
INSTITUTE



DLI Accelerated Data Science Teaching Kit

# Thank You