Challenges – March 10, 2017 – Madrid, Spain

- Overview

    - Build a CM-managed CDH cluster and secure it

- Place your work in the `challenges/labs` folder

    - All text files require Markdown (`.md`) extension and formatting
    - All screenshots must be in PNG format
    - You will create your own files for the challenges

- You can consult with each other and research online

    - Submit only your own work!

- Update your GitHub repo often -- don't wait until the end!
- If you break your cluster, or your cluster breaks you:

    - Tell an instructor (`mfernest` or `rsiwicki`)
    - Review the work you have pushed to GitHub
    - Create a new Issue to describe what you think happened

Challenge Setup

- Create the Issue `Challenges Setup`
- Make sure you have both `mfernest` and `rsiwicki` as Collaborators
- Assign the Issue to yourself and label it `started`
- In the file challenges/labs/0_setup.md:

    - List the cloud provider you are using (AWS, GCE, Azure, other)
    - List the nodes you are using by IP address and name
    - List the Linux release you are using
    - Demonstrate the disk capacity available on each node is >= 30 GB
    - List the command and output for `yum repolist enabled`

- Add the following Linux accounts to all nodes

    - User neymar with a UID of 2010
    - User ronaldo with a UID of 2016
    - Create the group `barca` and add `ronaldo` to it
    - Create the group `merengues` and add `neymar` to it

- List the /etc/passwd entries for neymar and ronaldo

    - Not the entire file!

- List the /etc/group entries for barca and merengues

    - Not the entire file!

- Push these updates to your GitHub repo
- Label your Issue `submitted`
- Assign the Issue to both instructors

Challenge 1: Install a MySQL server

- Create the Issue `Install MySQL` or `Install MariaDB` as appropriate
- Assign the Issue to yourself and label it `started`
- Install a MySQL 5.6 or MariaDB 5.5 server, as appropriate, on the first node listed in `0_setup.md`

    - Use the appropriate YUM repository to install the package.
    - Copy the repo you're using to `challenges/labs/1_my-database-server.repo.md`

- On all cluster nodes

    - Install the appropriate DB client package and JDBC connector jar

- Start the database service
- Create the following databases

    - `scm`
    - `rman`
    - `hive`
    - `oozie`
    - `hue`
    - `sentry`

- Record the following in `challenges/labs/1_db-server.md`

    - The hostname of your db server node
    - The command and output for display your database server's version
    - The command and output for listing your created databases

- Push this work to your GitHub repo
- Label the Issue `submitted` and assign it to both instructors

Challenge 2: Install Cloudera Manager 5.9.x

- Create the Issue `Install CM`
- Assign yourself to the Issue and label it `started`
- Install Cloudera Manager on the second node listed in `0_setup.md`
- Configure the CM repo to install the latest release

    - List the command and output for `ls /etc/yum.repos.d` in `challenges/labs/2_cm.md`
    - Copy the `cloudera-manager.repo` file to `challenges/labs/2_cloudera-manager.repo.md`

- Configure Cloudera Manager

    - Use the `scm_prepare_database.sh` script to write your `db.properties` file

        - List the full command line in `2_cm.md`

- Start the Cloudera Manager server. Then in `challenges/labs/2_db.properties.md`:

    - Add the complete first line from your server log
    - Add the complete line that contains the phrase "Started Jetty server"
    - Add the full contents of your `db.properties` file

- Push these changes to your GitHub repo and label the Issue 'submitted`
- Assign the issue to both instructors

Challenge 3 – Install CDH 5.9.x

- Create the Issue `Install CDH`
- Assign the issue to yourself and label it `started`
- Install the latest CDH release; deploy Coreset services only

  - Name your cluster using your GitHub handle

- Create user directories in HDFS for `neymar` and `ronaldo`
- Add the following to 3_cm.md:

  - The command and output for `hdfs dfs -ls /user`
  - The output from the CM API call `../api/v14/hosts`

- Login to Hue to install the Hive sample data

  - Capture a Hue screen that lists the Hive tables to `challenges/labs/3_hue_hive.png`

- Push this work to your GitHub repo and label the Issue `submitted`
- Assign the issue to both instructors

Challenge 4 – HDFS Testing

- Create the Issue Test HDFS
- Assign the issue to yourself and label it started
- As user neymar, use teragen to generate a 65,536,000-record dataset into eight files

    - Set the block size for this file to 16 MB
    - Name the target directory tgen640
    - Use the time command to capture job duration

- Put the following in the file challenges/labs/4_teragen.md

    - The full teragen command and job output
    - The result of the time command
    - The command and output of hdfs dfs -ls /user/neymar/tgen640
    - The command and output to show how many blocks are stored under this directory

- Push this work to your GitHub repo and label the Issue submitted
- Assign the issue to both instructors

Challenge 5 – Kerberize the cluster

- Create the Issue `Kerberize cluster`
- Assign the issue to yourself and label it `started`
- Install an MIT KDC on the same node as the CM server

  - Name your realm after your GitHub handle
  - Use ES as a suffix
  - For example: `RSIWICKI.ES`

- Create Kerberos principals for `neymar`, `ronaldo`, and `cloudera-scm`

  - Grant `cloudera-scm` the privileges needed to create principals and generate keytabs

- Enable Kerberos for the cluster
- Run the `terasort` program as `neymar` using the output target `/user/neymar/tsort640m`

  - Copy the command and full output to `challenges/labs/5_terasort.md`

- Run the Hadoop `pi` program as the user `ronaldo`

  - Copy the command and full output to `challenges/labs/5_pi.md`

- Copy the **text files** stored in `/var/kerberos/krb5kdc/` to your repo:

  - Add the prefix `5_` and the suffix `.md` to the original file name
  - Example: `5_kdc.conf.md`

- Push this work to your GitHub repo and label the Issue `submitted`
- Assign the issue to both instructors

Challenge 6 – Upgrade CM and CDH

- Create the Issue `Update Cluster`
- Update Cloudera Manager to the latest available release

  - Capture the command used to determine the API version available, along with the output.
  - Store them in the file `6_cm_latest.png`

- Upgrade CDH to the latest available release

  - Capture the CDH version and the services that are running
  - Store it in the file `6_cdh_latest.png`

- Assign the issue to both instructors
- Push all work to your GitHub repo

Once you finish, or when time is called:

- Commit any outstanding changes from your repo to GitHub
- Notify mfernest@cloudera.com that you have stopped pushing to your repo

  - Continue working, if you like, after sending this note

- Please fill out [this survey form (https://goo.gl/forms/pmHeHx03zRu3cnlc2)](https://goo.gl/forms/pmHeHx03zRu3cnlc2)
- Anything else you'd like to express about the class?

  - Please add your comments to `labs/7_feedback_final.md` -- don't forget to commit them!

- Now take it easy. You've worked very hard all week. Safe travels!