

grep - Assignment

The purpose of this assignment is to make you deep dive into practical aspects of using the **grep** command. Within the folder that contains this document, you will find a file called **cities.txt**, which contains over 500 different cities.

**Problems**

Provide a grep command (or concatenation of commands if needed) for the following queries (capitalization of letters matter):

1. List the cities that start with the letter “Z”
   * **Output**: Zaoshi, Zhukeng, Zhongxi, Zwettl, Zaragoza, Zheshart, Zhuji, Zadvorsko, Zhoujia.
2. List the cities that end with the word “gan”
   * **Output**: Tougan, Pedaringan, Banjar Sambangan, Salingogan
3. List the cities that start with “A”, “D”, or “E”
   * **Output:** there is 53.
4. List the cities that have “a” as their second letter, and “r” as their third letter, in any of its words.
   * **Output:** there is 22.
5. List the cities that don’t start with the letter “C”
   * **Output**: there is 477.
6. List the cities that have at least 14 characters (hint: use egrep)
   * **Output:** Saint-Laurent-du-Var, Charlottenberg, Atarodangwautu, Salon-de-Provence, Sendangkemulian, Nishi-Tokyo-shi, Karolino-Buhaz, Saint-Quentin-en-Yvelines, Challhuahuacho, Chervonopartyzans’k, Ambohitrolomahitsy.
7. Count how many cities start with the letter “B”
   * **Output:** there is 39.
8. Count how many times the letter “d” repeats
   * **Output:** there is 88.
9. List the cities that end with the same letter, twice (hint: use egrep)
   * **Output:** Izobil’nyy, Invercargill, Ingersoll, Villa Gesell, Ishii, Ingersoll
10. List the palindrome cities with five characters
    * **Output:** Anina, Navan, Noxon.
    * **Extra (though):** do it for any amount of characters.

final problem: In HDFS find two ways to copy a file from local to HDFS and find two ways to copy a file from HDFS to your local machine