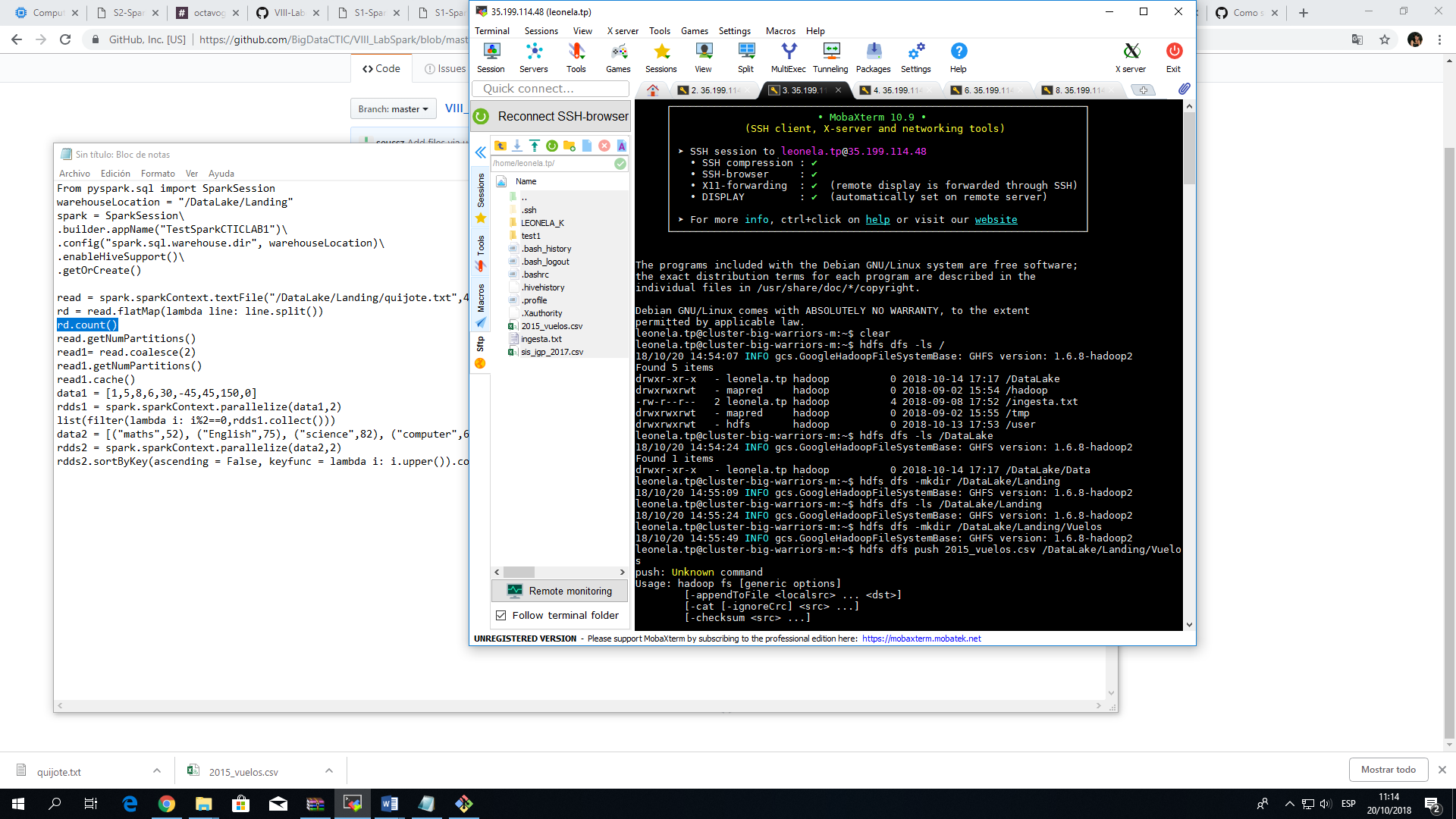
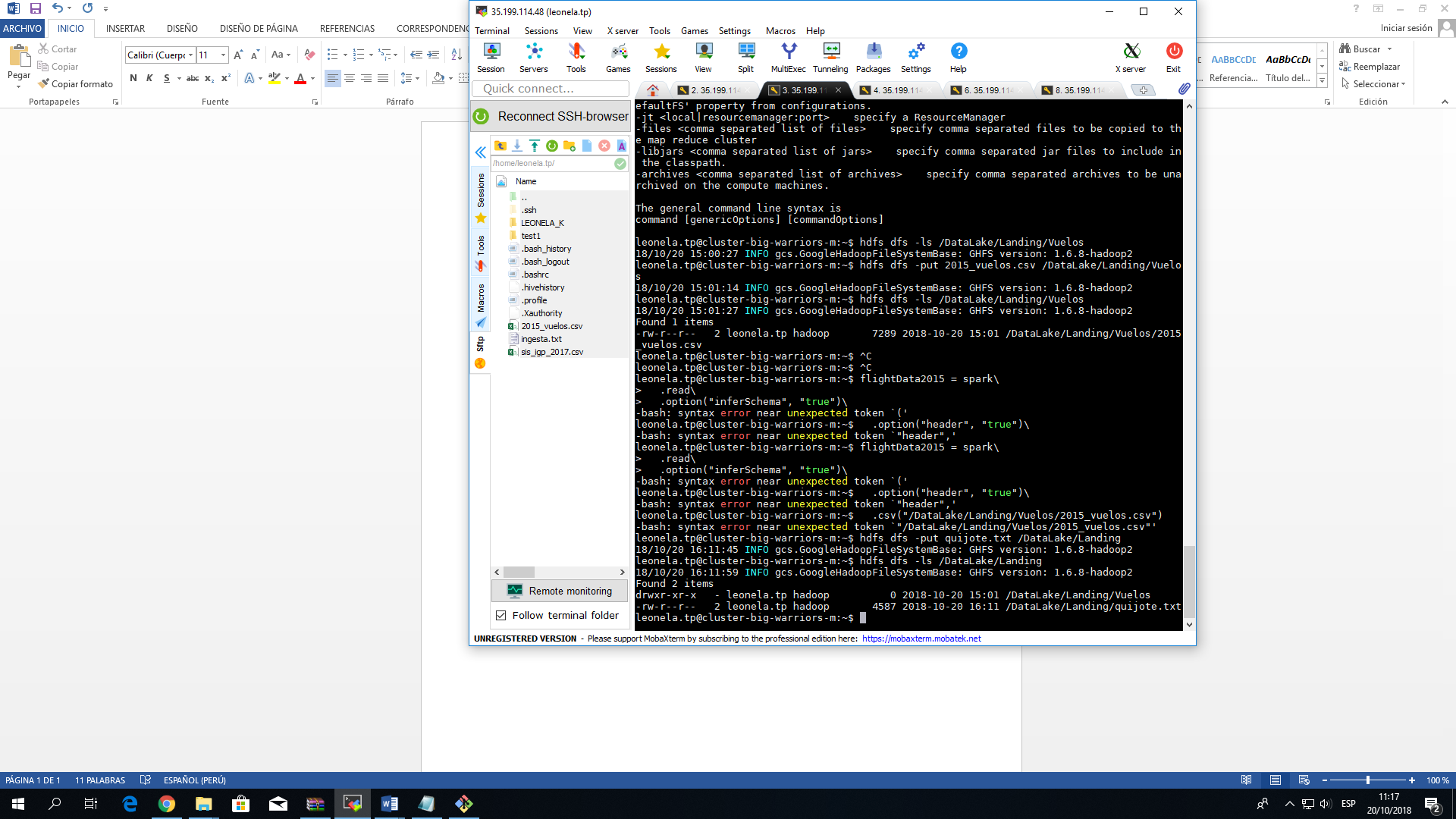
LABORATORIO 1 SPARK

Ingreso a mi DataLake

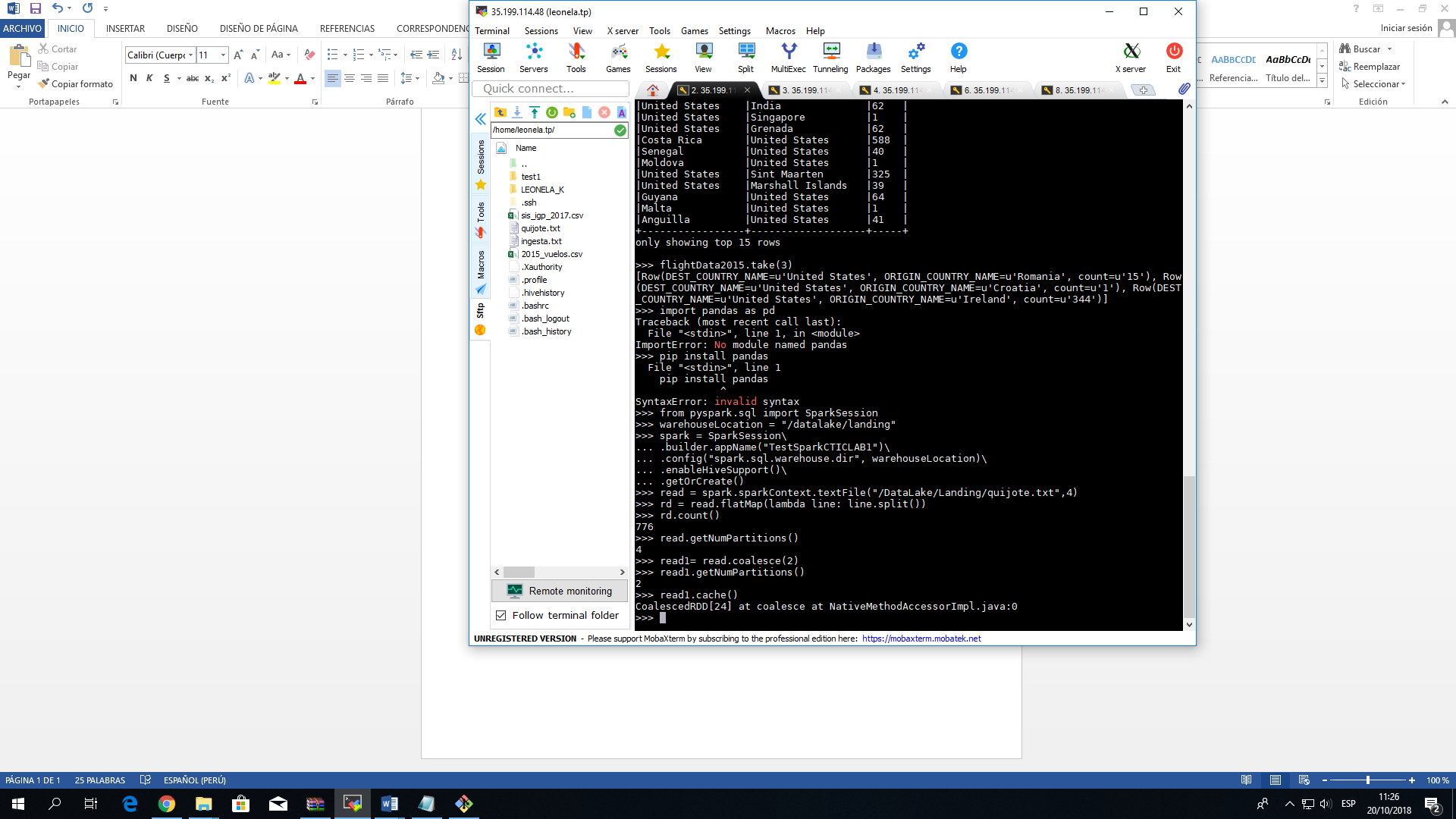
Creo mi carpeta Landing

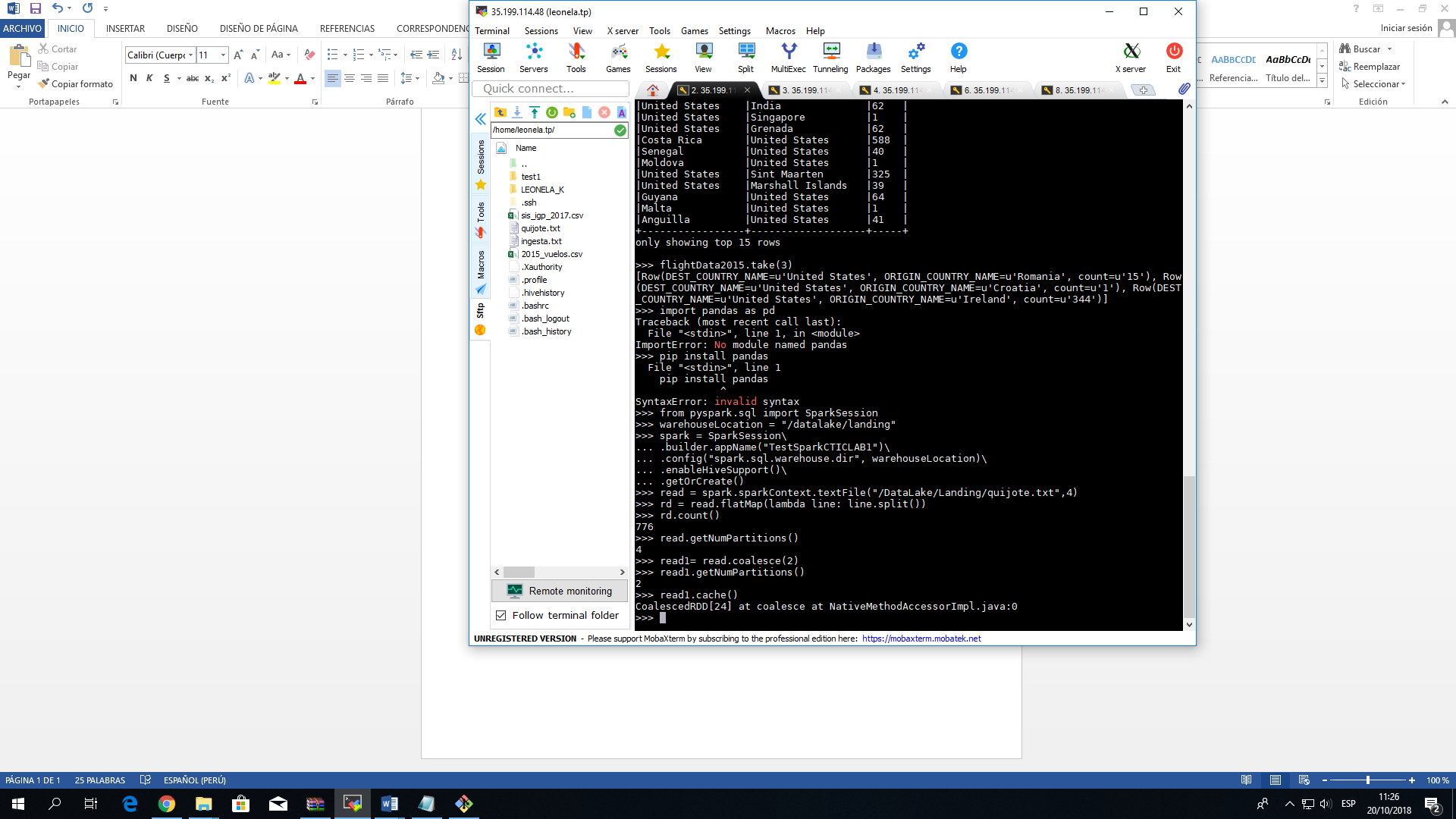


Subo mi archivo Quijote.txt a mi carpeta Landing del DataLake

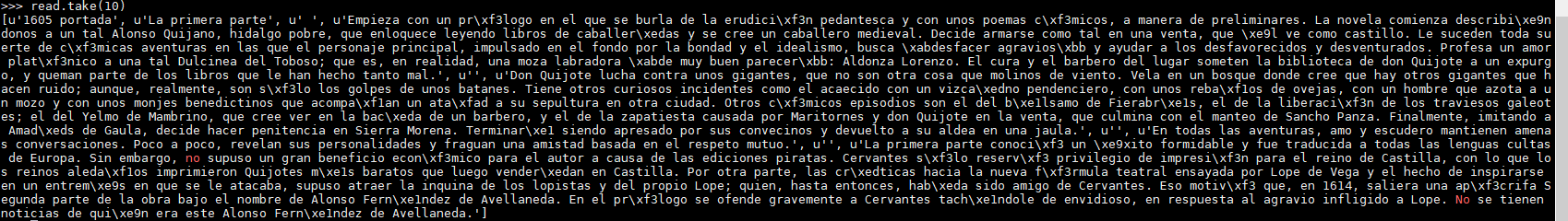


**Pregunta 1:** Crear un RDD a partir del archivo de la web http://www.daemcopiapo.cl/Biblioteca/Archivos/7\_6253.pdf tomar solo la primera pagina del primera capitulo, guardar el archivo como quijote.txt con 4 particiones





Verifico mi archivo



**Pregunta 2:** Encontrar la cantidad de palabras que existen



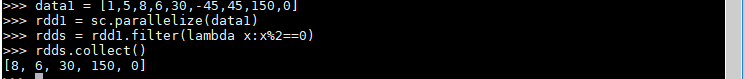
**Pregunta 3:**  Reducir el numero de particiones a 2



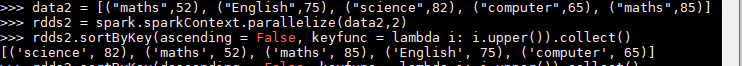
**Pregunta 4:**  Poner en cache mi RDD



**Pregunta 5:**  Generar el RDD de valores [1,5,8,6,30,-45,45,150,0] y quedarse con los elementos pares



**Pregunta 6:**  Generar el RDD de valores [("maths",52), ("English",75), ("science",82), ("computer",65), ("maths",85)] y ordenarlos de manera descendente teniendo los valores en mayúscula



**Pregunta 7:**  Commit a su repositorio