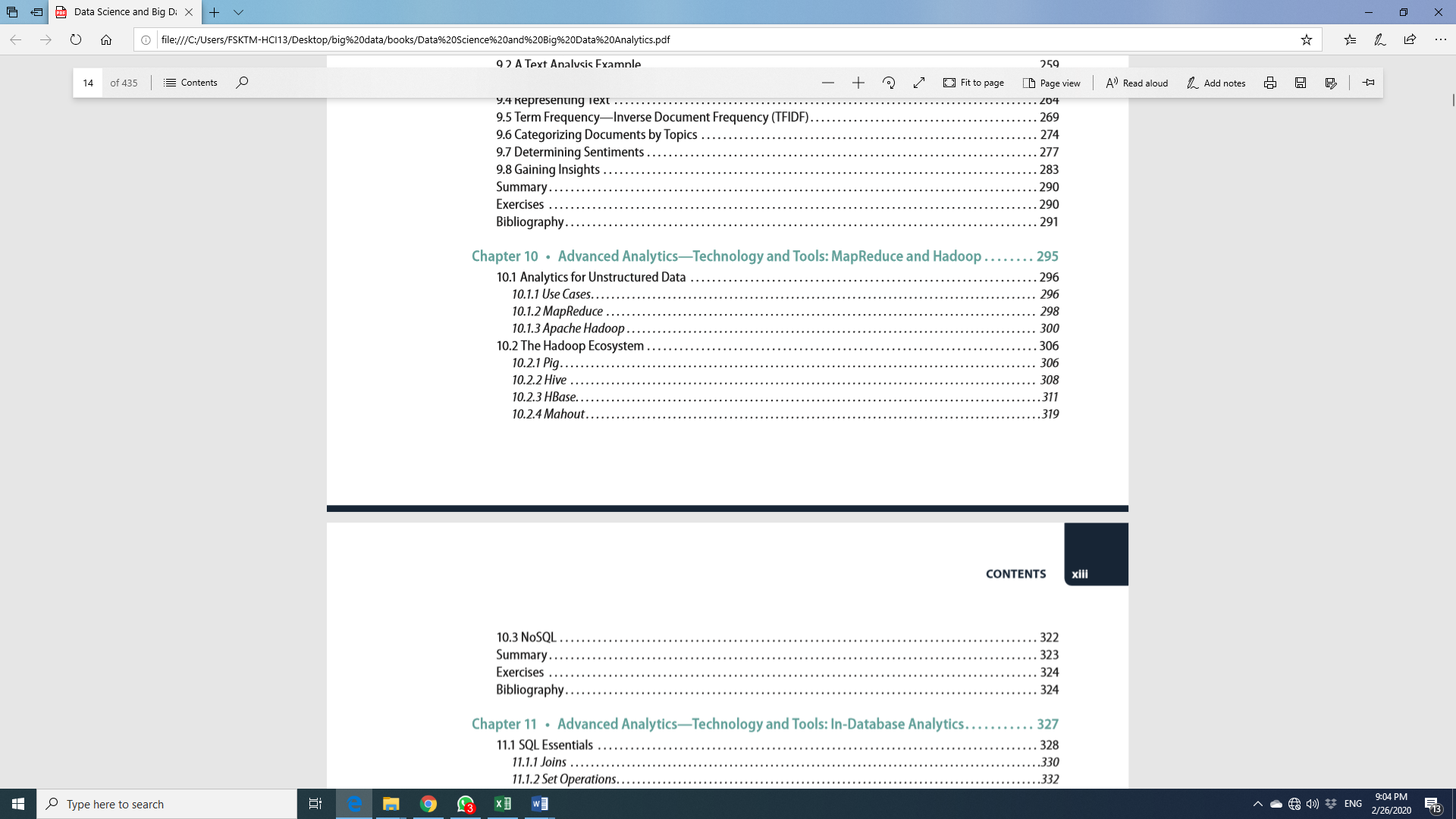
Assignment Week 3

1. Research and document one real world use case and actual implementations for Hadoop.
2. Compare and contrast Rhadoop, Hadoop, Pig, Hive, and HBase. List strengths and weaknesses of each tool set. Research and summarize three published use cases for each tool set.

Exercises 3 through 5 require some programming background and a working Hadoop environment. The text of the novel War and Peace can be downloaded from http://onlinebooks .library.upenn.edu/ and used as the dataset for these exercises. However, other datasets can easily be substituted. Document all processing steps applied to the data.

1. Use MapReduce in Hadoop to perform a word count on the speciﬁed dataset.
2. Use Pig to perform a word count on the speciﬁed dataset.
3. Use Hive to perform a word count on the speciﬁed dataset.



Ref: Data Science and Big Data Analytics, EMC Education Services.

Note:

* This assignment is a group assignment.
* All groups need to present the assignment in week 4 (12th March 2020). All of group members should be ready for Q&A about their assignment.
* Your submission should be documented in a word file <Assignment\_3\_GroupNumber.docx>. All the group members should submit in Spectrum individually.
* Please write the student numbers, student names, and group number in the first page.