**Data Engineering Week2 Homework**

**17기 신예진**

**RDBMS와 Hadoop의 특징, 공통점과 차이점**

**RDBMS**

* Relational DataBase Management System의 약자로 관계형 데이터베이스를 생성하고 관리할 수 있는 소프트웨어이다.
* RDBMS 내의 모든 데이터는 2차원 테이블 형태로 표현되며, 테이블은 다른 테이블과 관계를 맺고 있다.
* 관계를 나타내기 위해 foreign key를 사용하며, 이 foreign key를 이용하여 join이 가능하다.

**Hadoop**

* 분산 환경에서 빅데이터를 저장하고 처리할 수 있는 자바 기반의 오픈소스 프레임워크이다.
* 여러 대의 컴퓨터 클러스터에 대규모 데이터셋을 분산 처리할 수 있게 해주는 프레임워크이다.
* hdfs와 mapreduce를 시작으로 하둡생태계가 확장, 발전되었다.

**공통점**

* 데이터를 저장하는 용도로 사용되며, 정형 데이터를 처리할 수 있다는 공통점이 있다.

**차이점**

|  |  |  |
| --- | --- | --- |
|  | RDBMS | Hadoop |
| 데이터 다양성 | 구조화된 데이터 저장 | 여러가지 형태의 데이터 저장(정형, 반정형, 비정형) |
| 데이터 저장고 | 평균 데이터 양을 저장 | 훨씬 데이터 양을 저장 |
| 속도 | 읽기가 빠름 | 읽기 및 쓰기가 빠름 |
| 확장성 | 수직확장성 | 수평확장성 |
| 하드웨어 | 고급 서버 사용 | 상용 하드웨어 사용 |
| 처리량 | 처리량이 높음 | 처리량이 낮음 |

출처

<https://khj93.tistory.com/entry/Database-RDBMS와-NOSQL-차이점>

<https://ko.strephonsays.com/rdbms-and-hadoop-8502>

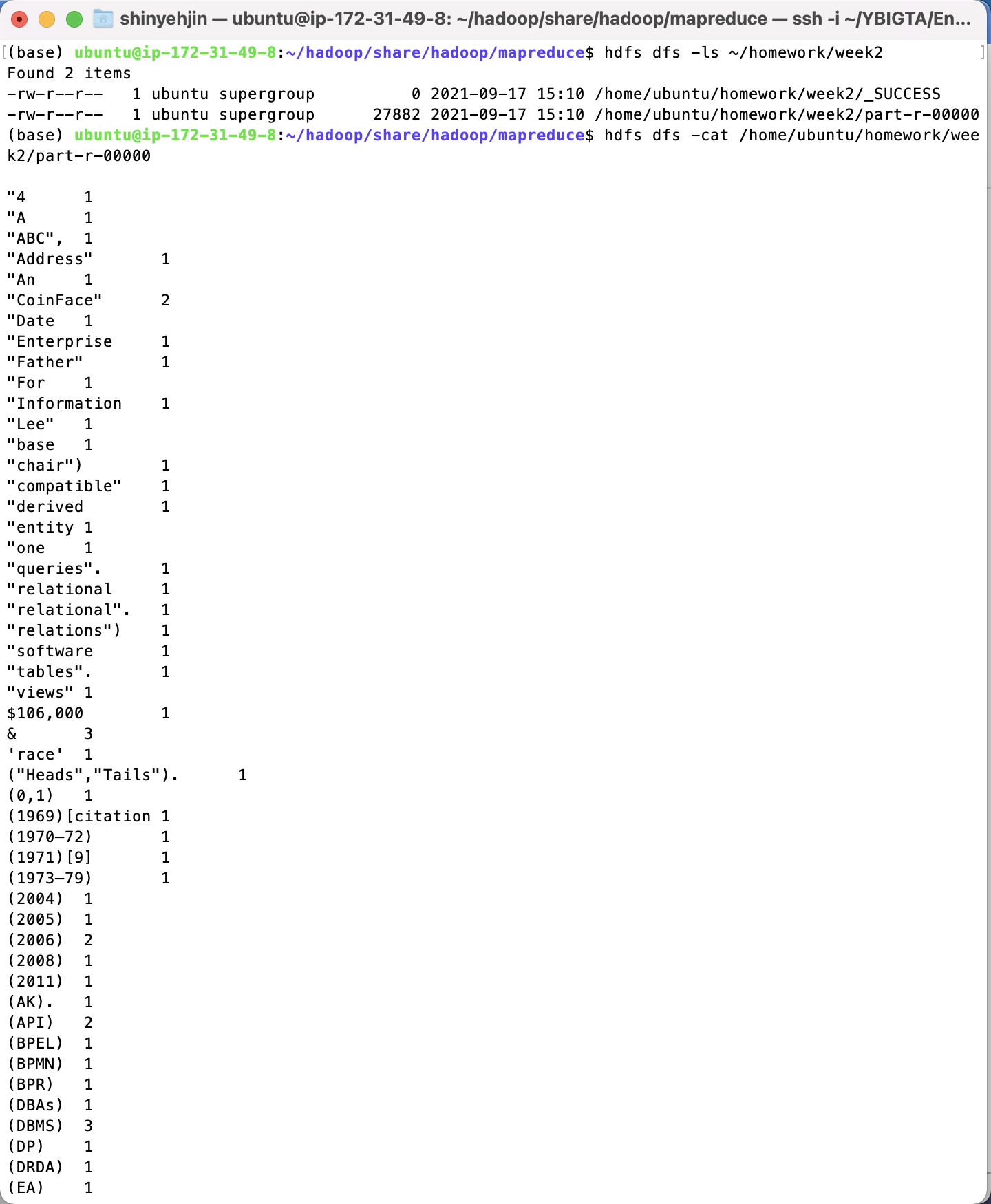
Wordcounter 실행

**wordcounter**

테이블이(가) 표시된 사진

자동 생성된 설명

**결과파일 스크린샷 (폴더명은 제맘대로 했어요..ㅎ)**



**input 파일 일부 (wiki 아무거나 3개 긁어서 사용했습니다)**

텍스트, 신문, 스크린샷이(가) 표시된 사진

자동 생성된 설명