|  |  |  |  |
| --- | --- | --- | --- |
| **평가일시** | 2018년 12월 07일 | | |
| **과정명** | 빅데이터분석(Hadoop과 R 실무) 실무자 양성과정 | | |
| **교과목** | 빅데이터 수집 | **훈련교사** | 임 동 조    (인) |
| **단원명** | 빅데이터 수집 | | |
| **세부내용** | 데이터 수집 계획 수립, 빅데이터 수집 시스템 구성하기, 내부 데이터  수집하기, 외부 데이터 수집하기, 데이터 변환하기, 수집 데이터 검증하기 | | |

|  |  |  |
| --- | --- | --- |
| **구분**  **(능력단위요소)** | 문 항 | 점 수 |
|  | [능력단위명]        : 빅데이터 수집  [능력단위요소명] : 수집 시스템 구축하기  5가지 능력단위요소로 평가합니다.  데이터 수집 계획 수립, 빅데이터 수집 시스템 구성하기, 내/외부 데이터 수집하기, 데  이터 변환하기, 수집 데이터 검증하기 | 100/  100  점 |
| 빅데이터 수집 시스  템 구성하기, 내/외부  데이터 수집하기 | 과제1  :  가상  머신  구축하기  및  리눅스  환경에서의  데이터  수집  (수행  준거  2-1  ~  2-3)  (2-1,  2-2)  각  2.5점,  (2-3)  5점  현재  원도우  환경이어서  원도우  환경위에  빅데이터  수집을  위한  리눅스  환경을  구축하려고  합니  다.  아래의  내용에  따라  가상  시스템을  구축해  주세요.  (가)  Linux  Ubuntu  64bit,  RAM  2GB,  HDD  20GB의  시스템을  환경의  가상  머신을  하나  만들  어  주세요.  (설치  완료  후,  확인을  위한  시스템  사양  부분을  확인  후,  화면  캡쳐를  해  주세요.)  (나)  scrapy  솔루션을  설치하고  구성해  주세요.  (설치시의  명령과  설치  확인  결과을  상세  내용을  작성해  주세요.)  (다)  설치  완료  후,  이에  대한  장애  발생  및  대응을  위한  스냅샵을  작성하고,  복구를  위한  가상  시스템  내보내기  메뉴를  이용하여  백업  파일을  하나  작성해  주세요.  (작업  절차  내용을  화면  캡쳐해  주세요.)  (수행준거  3-1,  3-2,  3-6)  15점  내부  DB  환경인  MYSQL에서  데이터를  가져올  수  있는지에  대한  실습입니다.  (가)  Python과  MySQL  DB의  Session을  연결을  수행해  주세요.  (나)  테이블을  확인  후,  수집하고자  하는  TABLE를  정한  후,  데이터를  수집을  수행해  주세요.  (다)  내부에서  수집한  데이터와  Python에서  수집한  데이터의  건수가  맞는지  확인해  주세요.  (수행준거  3-3,  3-7,  4-4)  15점  이번  수행  과제는  앞의  데이터  수집  실습  과제을  활용하여  수행해도  됩니다.  (가)  웹에서의  데이터를  실시간으로  수집하기  위해  cron  명령을  이용하여  python  프로그램을  10분  단위로  주기적으로  실행해  주세요. |

**평가방법: 포트폴리오  –  빅데이터  수집  과제  내용(2)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | (나)  외부  수집  데이터와  수집한  데이터의  건수  등의  데이터가  맞는지  확인해  주세요. | | | | |  |
| **성취수준** | **5수준**  **(90~100)** | | **4수준**  **(80~89)** | **3수준**  **(60~79)** | **2수준**  **(30~59)** | **1수준**  **(0~29)** | |
| **총점** | | | |  | | | |

(수행  준거  2-1  ~  2-3)  (2-1,  2-2)  각  2.5점,  (2-3)  5점

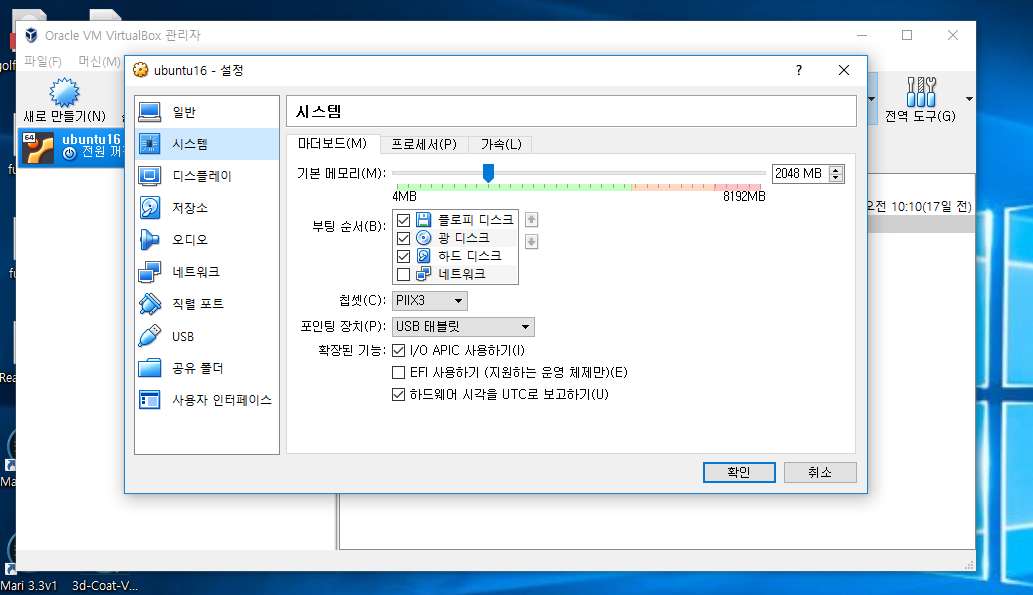
현재  원도우  환경이어서  원도우  환경위에  빅데이터  수집을  위한  리눅스  환경을  구축하려고  합니

다.  아래의  내용에  따라  가상  시스템을  구축해  주세요.

(가)  Linux  Ubuntu  64bit,  RAM  2GB,  HDD  20GB의  시스템을  환경의  가상  머신을  하나  만들

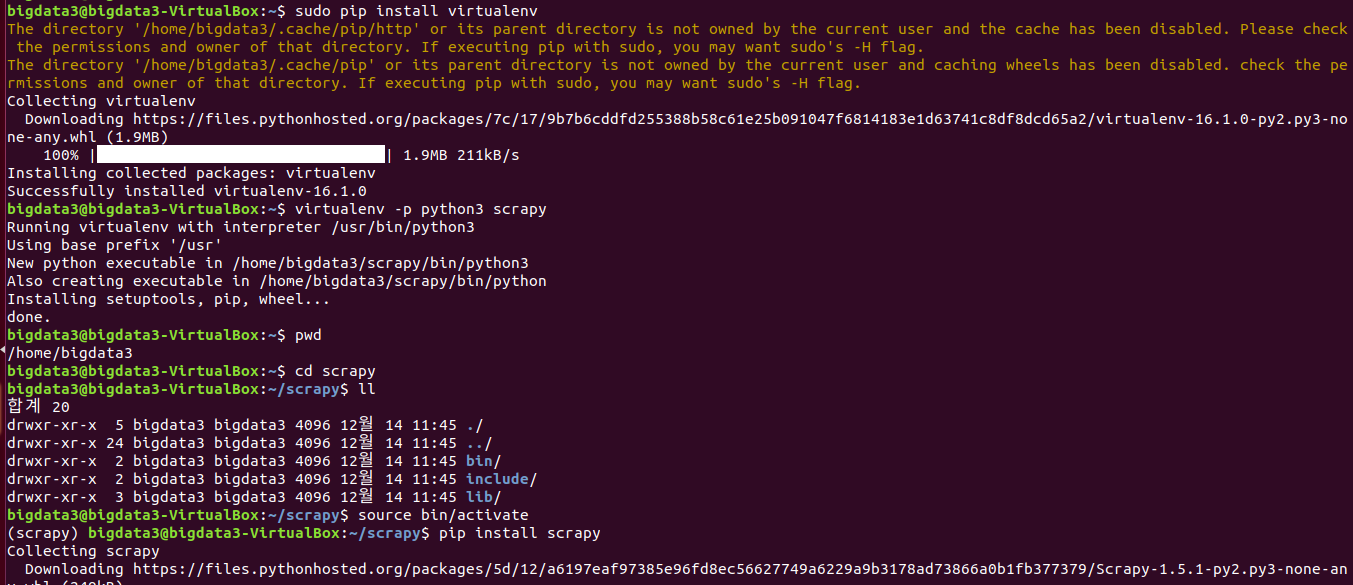
어  주세요.

(설치  완료  후,  확인을  위한  시스템  사양  부분을  확인  후,  화면  캡쳐를  해  주세요.)



(나)  scrapy  솔루션을  설치하고  구성해  주세요.

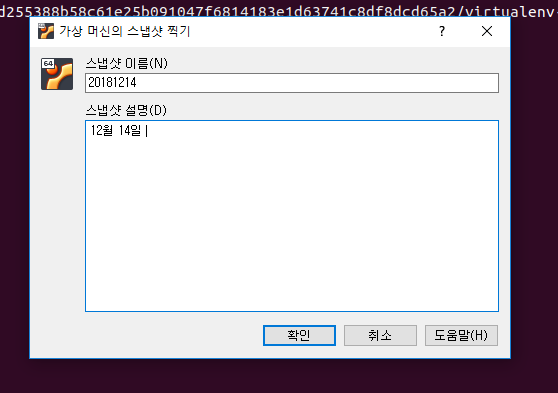
(설치시의  명령과  설치  확인  결과을  상세  내용을  작성해  주세요.)

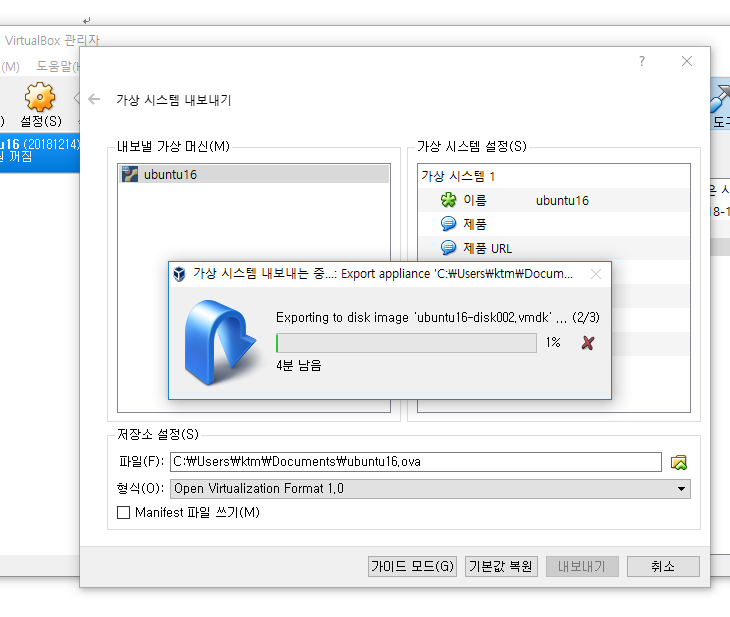


(다)  설치  완료  후,  이에  대한  장애  발생  및  대응을  위한  스냅샵을  작성하고,  복구를  위한  가상

시스템  내보내기  메뉴를  이용하여  백업  파일을  하나  작성해  주세요.

(작업  절차  내용을  화면  캡쳐해  주세요.)





(수행준거  3-1,  3-2,  3-6)  15점

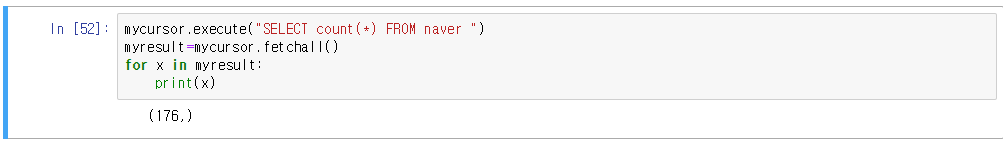
내부  DB  환경인  MYSQL에서  데이터를  가져올  수  있는지에  대한  실습입니다.

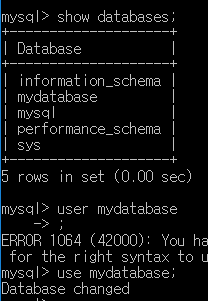
1. Python과  MySQL  DB의  Session을  연결을  수행해  주세요.

(나)  테이블을  확인  후,  수집하고자  하는  TABLE를  정한  후,  데이터를  수집을  수행해  주세요.

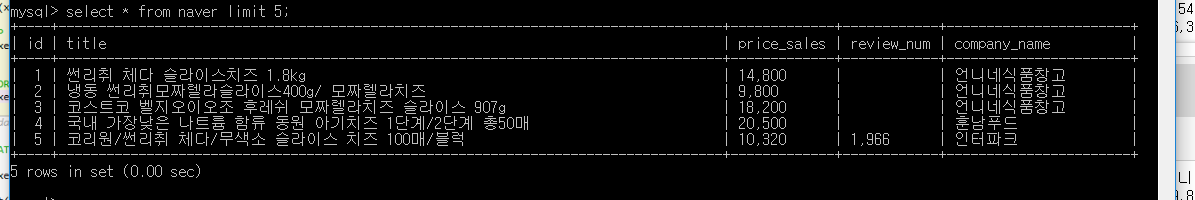
(다)  내부에서  수집한  데이터와  Python에서  수집한  데이터의  건수가  맞는지  확인해  주세요.











나머지는 첨부파일 참조 부탁 드립니다!!!

