



평가

사전평가

자가진단

필답평가

성취도평가

자가진단

수행평가

필답평가

포트폴리오

필답 평가

빅데이터 저장 및 분석 시스템 구축 기술 / 점수 : 96 점 (시험일 : 2020-03-25)

목록

Linux OS 이해



1. 리눅스 운영체제에 대한 설명 중 잘못된 것은?

정답 2 / 입력 답안 : 2

- 1) 리눅스는 시스템내 존재하는 모든 자원에 대해서 접근권한이나 소유체계가 엄격하므로 보안을 유지하기에 좋은 OS이다.
- 2) 리눅스 시스템은 여러 명이 사용할 수 없고 혼자서 사용할 수 있는 단일 시스템이다.
- 3) 리눅스의 관리자 계정은 root이다
- 4) 사용자가 입력한 명령어가 리눅스 내부에 전달되도록 셸을 이용하여 작업한다.



2. 리눅스의 설정 파일이 저장되는 폴더는 ?

정답 3 / 입력 답안 : 3

- 1) /usr
- 2) /root
- 3) /etc
- 4) /boot



3. hadoop계정의 홈디렉토리 위치는?

정답 3 / 입력 답안 : 3

- 1) /root/home
- 2) /home/home
- 3) /home/hadoop
- 4) /usr/hadoop



4. 현재 접속해 있는 디렉토리의 위치를 알고 싶은 경우 사용하는 명령어는?

정답 3 / 입력 답안 : 3

- 1) dir
- 2) ls
- 3) pwd
- 4) cd



5. 다음 중 리눅스 명령어에 대한 설명 중 잘못된 것은?

정답 3 / 입력 답안 : 3

- 1) 리눅스 명령어는 옵션과 함께 사용하며 여러 옵션과 함께 연결해서 사용하는 것이 가능하다.
- 2) 리눅스에서는 윈도우와 다르게 폴더 이동 명령을 주고 폴더를 이동했을 때 모든 경로가 보이지 않으므로 꼭 현재 위치한 폴더를 확인하고 작업하는 것이 좋다
- 3) hadoop계정을 이용하는 사용자도 리눅스 시스템의 모든 리소스를 접근하고 수정할 수 있다.
- 4) ls명령어의 옵션을 이용하면 숨김처리 되어 있는 파일이나 디렉토리도 확인할 수 있다.



6. 원격으로 접속하여 외부 호스트에 파일을 copy할때 사용하는 명령어는?

정답 1 / 입력 답안 : 1

- 1) scp
- 2) cp
- 3) mv
- 4) rm

○ 7. JVM에 의해 실행된 프로세스를 확인하고 싶을때 사용하는 명령어는?

정답 1 / 입력 답안: 1

- 1) jps
- 2) ls
- 3) js
- 4) ps

빅데이터 수집 시스템 개발

○ 1. 하둡의 완전 분산 모드를 구축하기 위해 설정해야 하는 작업이 아닌 것은?

정답 4 / 입력 답안: 4

- 1) VMWare에 머신을 생성한 후 centos를 설치한다.
- 2) 머신을 복제하여 네 대를 준비하고 방화벽을 해제한다.
- 3) hosts파일에 ip와 도메인을 등록한다.
- 4) hadoop내부에서 암호 통신을 하므로 암호 키를 외부의 인증기관으로 부터 발급받아 준비한다.

○ 2. 다음은 빅데이터를 설명하는 5v에 대한 내용이다. 잘못된 것은?

정답 1 / 입력 답안: 1

- 1) Volume: 다양한 사이즈의 데이터를 의미 하며 MB단위의 데이터 이상은 안된다.
- 2) Velocity: 다양한 데이터의 빠른 수집 및 저장, 처리, 분석을 의미 한다.
- 3) Variety: 다양한 형태의 데이터 및 다양한 디바이스에서의 데이터로 구성 된다.
- 4) Value: 가치 있는 데이터이여 하며 또한 가치 있는 데이터로 추출 해 낼 수 있어야 한다는 의미

○ 3. 다음 중 RDBMS로 부터 데이터를 수집하기 위한 hadoop 에코 시스템은?

정답 4 / 입력 답안: 4

- 1) R
- 2) flume
- 3) hive
- 4) sqoop

○ 4. 로그 데이터를 수집하기 위해 flume을 이용하려 한다. hadoop2머신에서 발생된 로그 데이터를 hadoop01의 HDFS에 전송하기 위해 정의해야 할 sink와 source의 타입은?

정답 3 / 입력 답안: 3

- 1) exec
- 2) spoolDir
- 3) avro
- 4) file_roll

○ 5. 다음은 flume을 구성하는 구성요소입니다. 옳지 않은 것은?

정답 4 / 입력 답안: 4

- 1) channel : 데이터를 쌓아놓는 일종의 큐
- 2) sink : 데이터가 저장될 목적지에 대한 정보
- 3) source : 저장할 원본 데이터의 타입
- 4) agent : 저장할 데이터의 종류

○ 6. 하둡 서버간 통신은 어떤 프로토콜을 사용 하는가?

정답 3 / 입력 답안: 3

- 1) TCP/IP
- 2) UDP
- 3) SSH
- 4) HTTP

✖ 7. 다음 하둡 설정 파일의 설명이 틀린것을 고르시오

정답 2 / 입력 답안: 1

- 1) hadoop-env.sh: 하둡이 실행되기 위해 필요한 환경 설정을 정의하는 파일로 자바의 설치 경로를 정의
- 2) core-site.xml: 맵리듀스에서 공통적으로 사용될 환경 정보 설정

- 3) hdfs-site.xml: HDFS 하둡 관리 시스템에 대한 환경 정보 설정
- 4) mapred-site.xml: 맵리듀스에서 사용할 환경 정보 설정

8. 클라이언트의 요청을 받고 데이터 노드에 저장된 블록의 정보를 갖고 있는 HDFS의 마스터는?

정답 3 / 입력 답안 : 3

- 1) HDFS관리시스템
- 2) 잡트래커
- 3) 네임노드
- 4) 보조네임노드

빅데이터 저장 시스템 개발

1. 하둡의 기본 블록사이즈는?

정답 2 / 입력 답안 : 2

- 1) 64KB
- 2) 64MB
- 3) 32MB
- 4) 32KB

2. HDFS 에 파일을 저장하는 명령어 는?

정답 1 / 입력 답안 : 1

- 1) put
- 2) cat
- 3) jar
- 4) cp

3. HDFS에 대한 설명 중 틀린 것은?

정답 1 / 입력 답안 : 1

- 1) 네임노드는 모든 HDFS를 관리하는 마스터 이며 클라이언트의 요청을 받아 데이터를 자신의 저장구조에 저장한다.
- 2) 데이터노드는 네임노드에게 지속적으로 heart beat를 보내 통신한다.
- 3) 데이터노드에는 데이터를 블록단위로 나누어 저장하며 복제본도 같이 저장한다.
- 4) 노드가 죽거나 이상 상황이 발생되면 문제가 생긴 노드에 저장된 블록을 다른 데이터 노드에 복제한다.

4. 맵퍼의 출력데이터가 리듀서에 전달되기 전에 출력데이터를 줄이기 위해 사용하는 것은?

정답 1 / 입력 답안 : 1

- 1) combiner
- 2) partitioner
- 3) reducer
- 4) comparator

5. document기반으로 json의 데이터 포맷을 사용하는 nosql은

정답 1 / 입력 답안 : 1

- 1) mongodb
- 2) oracle
- 3) ms-sql
- 4) mysql

6. 하둡의 잡을 관리하는 데몬은?

정답 1 / 입력 답안 : 1

- 1) 잡트래커
- 2) 네임노드
- 3) 보조네임노드
- 4) 파티셔너

7. 맵테스크의 출력결과를 어떤 리듀서로 보낼지 결정하는 것은?

정답 3 / 입력 답안 : 3

- 1) 맵퍼
- 2) 리듀서
- 3) 파티셔너
- 4) 네임노드

빅데이터 분석 시스템 개발

- ☐ 1. R에서 다양한 통계 패키지를 설치 하고 나서 사용하기 위한 명령어는 무엇인지 고르시오?

정답 2 / 입력 답안 : 2

- 1) `install("")`
- 2) `install.packages("")`
- 3) `install.pack()`
- 4) `install.library()`

- ☐ 2. 데이터 분석 시 사용되는 형식의 설명 중 맞는 것을 고르시오?

정답 3 / 입력 답안 : 3

- 1) `vector`: 1차원, 타입이 다른 복수 데이터 처리
- 2) `matrix`: 2차원, 타입이 다른 복수 데이터 처리
- 3) `data frame`: 2차원, 타입이 같은 복수 데이터 처리
- 4) `list`: 1차원, 타입이 다른 복수데이터 처리

- ☐ 3. 설치된 외부 라이브러리를 사용하는 함수는 무엇인지 고르시오

정답 3 / 입력 답안 : 3

- 1) `insert`
- 2) `import`
- 3) `library`
- 4) `package`

- ☐ 4. csv 파일을 읽어 들일때 특정 컬럼에 문자열을 `Factor`로 인식 하지 않게 하기 위한 키워드는 무엇인지 고르시오?

정답 1 / 입력 답안 : 1

- 1) `stringAsFactors = F`
- 2) `stringAsFactors = T`
- 3) `stringNotFactors = T`
- 4) `stringNotFactors = F`

- ☐ 5. 4행 3열의 `data frame`이 있다. 이중 2번 열만 빼고 데이터를 출력 하시 위한 명령어중 틀린 것을 고르시오

정답 3 / 입력 답안 : 3

- 1) `[-,c(2)]`
- 2) `[,c(1,3)]`
- 3) `[2,]`
- 4) `[, -2]`

- ☐ 6. 2차원의 동일한 사이즈의 데이터를 관리하며 열 별로 다른 타입의 데이터를 저장할 수 있는 구조는?

정답 2 / 입력 답안 : 2

- 1) `list`
- 2) `dataframe`
- 3) `matrix`
- 4) `vector`