

분류번호 : 2001020903\_19v2

능력단위 명칭 : 빅데이터 수집시스템 개발

능력단위 정의 : 빅데이터 수집시스템 개발이란 활용목적에 적합한 데이터를 수집하기 위하여 빅데이터 수집시스템 구성, 내·외부 데이터 수집모듈 개발, 데이터 변환모듈 개발 및 수집 데이터 검증모듈을 개발하는 능력이다.

능 력 단 위 요 소	수 행 준 거
2001020903_19v2.1 빅데이터 수집시스템 설계하기	<ol style="list-style-type: none"><li>1.1 빅데이터 관련 기술 및 제품, 분석 방식을 조사하고 기존 시스템과의 연계를 고려하여 데이터의 유형에 따른 빅데이터 수집방식을 결정할 수 있다.</li><li>1.2 데이터 수집모델을 설계하기 위하여 데이터 요구사항 도출, 데이터 유형별 분류 및 특성파악을 할 수 있다.</li><li>1.3 데이터유형별 특성에 부합하는 데이터모델, 수집기능 및 제약조건을 설계할 수 있다.</li><li>1.4 완료된 데이터 수집모델 설계에 대한 결과의 타당성을 검증할 수 있다.</li></ol>
	<b>【지식】</b> <ul style="list-style-type: none"><li>• 관계형 데이터 모델링 방법</li><li>• 데이터웨어하우스 설계 방법</li><li>• NoSQL 활용 방법</li><li>• 데이터 수집 방법</li></ul>
	<b>【기술】</b> <ul style="list-style-type: none"><li>• 업무 도메인 분석 능력</li><li>• 데이터 요구사항 도출 능력</li><li>• 데이터 요구사항 분석 능력</li><li>• 데이터 요구사항 분류 능력</li><li>• 데이터 모델링 설계도구 활용 능력</li></ul>
	<b>【태도】</b> <ul style="list-style-type: none"><li>• 명확한 언어를 구사하려는 자세</li><li>• 정보를 적극적으로 수집하는 태도</li><li>• 사용자 요구사항의 적극적인 수용</li></ul>
2001020903_19v2.2 빅데이터 수집시스템 구성하기	<ol style="list-style-type: none"><li>2.1 수집된 데이터의 안정성과 신뢰성을 보증하는 수집시스템 구조를 파악할 수 있다.</li><li>2.2 일괄처리 데이터의 특성을 감안하여 빅데이터 수집시스템을 구성할 수 있다.</li><li>2.3 실시간 데이터의 특성을 감안하여 빅데이터 수집시스템을 구성할 수 있다.</li><li>2.4 주어진 테스트 절차에 의해 데이터 수집관리시스템 정상설치 여부를 검증할 수 있다.</li></ol>

<p>2001020903_19v2.2</p> <p>빅데이터 수집시스템 구성하기</p>	<p>【지식】</p> <ul style="list-style-type: none"> <li>• 관계형 데이터 모델링 방법</li> <li>• 데이터웨어하우스 설계 방법</li> <li>• NoSQL 활용 방법</li> <li>• 데이터 수집 방법</li> </ul>
	<p>【기술】</p> <ul style="list-style-type: none"> <li>• NoSQL 활용 능력</li> <li>• RDBMS 운영 관리 기술</li> <li>• 분산파일 시스템 기술</li> <li>• 웹서버 운영 관리 능력</li> <li>• 분산처리 프레임워크 기술</li> </ul>
	<p>【태도】</p> <ul style="list-style-type: none"> <li>• 분석적이고 논리적인 사고</li> <li>• 다양한 기술을 상호 비교하고 적절한 대안을 제시하는 자세</li> <li>• 최적의 안을 도출하려는 적극적인 태도</li> </ul>
<p>2001020903_19v2.3</p> <p>빅데이터 수집·변환 모듈 개발하기</p>	<p>3.1 수집대상 데이터의 유형별 접근방식에 따라 데이터를 수집하는 모듈을 작성할 수 있다.</p> <p>3.2 수집된 데이터를 대상으로 실제 시스템에서 사용하기 용이한 형식으로 변환하는 모듈을 구현할 수 있다.</p>
	<p>【지식】</p> <ul style="list-style-type: none"> <li>• 관계형데이터 수집·변환 관리 기법</li> <li>• 데이터웨어하우스 수집·변환 관리 기법</li> <li>• 데이터 수집·변환 기법</li> <li>• 웹서버 사용 방법</li> <li>• 빅데이터 플랫폼 유형과 작동 원리</li> </ul>
	<p>【기술】</p> <ul style="list-style-type: none"> <li>• 소프트웨어 표준 분석 기술</li> <li>• 하드웨어 규격 분석 기술</li> <li>• 이기종 하드웨어 및 소프트웨어 연동 기술</li> <li>• 유사 설치사례 분석 능력</li> <li>• 웹서버 관련 기술</li> <li>• NoSQL 활용 능력</li> <li>• 분산파일 시스템 기술</li> <li>• 분산처리 프레임워크 기술</li> </ul> <p>【태도】</p> <ul style="list-style-type: none"> <li>• 정보를 적극적으로 수집하는 태도</li> <li>• 문제해결을 위한 긍정적 자세</li> <li>• 주어진 시간 안에 과업완료를 위해 노력하는 책임감 있는 자세</li> </ul>

2001020903_19v2.3 빅데이터 수집·변환 모	<ul style="list-style-type: none"> <li>• 새로운 하드웨어 및 소프트웨어를 적극적으로 수용하고 배우려는 자세</li> </ul>
2001020903_19v2.4 수집 데이터 검증모듈 개발하기	<p>4.1 구현된 수집시스템에서 생성한 데이터의 유효성여부를 검증할 수 있는 규칙설정 기능을 개발할 수 있다.</p> <p>4.2 구현된 수집시스템에서 생성한 데이터의 유효성여부를 검증하는 기능을 개발할 수 있다.</p> <p>4.3 개발된 검증모듈을 실행하여 테스트 데이터의 유효성여부를 검증하고, 모듈의 문제점이 발생할 경우 검증모듈을 개선할 수 있다.</p>
	<p>【지식】</p> <ul style="list-style-type: none"> <li>• 데이터 유효성 검증 방법</li> <li>• 정보시스템 구축 방법론</li> <li>• 정보시스템 개발 생명주기</li> </ul>
	<p>【기술】</p> <ul style="list-style-type: none"> <li>• 웹서버 운영 관리 능력</li> <li>• NoSQL 활용 능력</li> <li>• 분산파일 시스템 관련 기술</li> <li>• 분산처리 프레임워크 기술</li> </ul>
	<p>【태도】</p> <ul style="list-style-type: none"> <li>• 시스템에 대한 분석적 사고</li> <li>• 현황분석을 위한 수리적 사고</li> <li>• 이슈 및 위험을 판단하기 위한 논리적 사고</li> <li>• 문제해결을 위한 긍정적 자세</li> </ul>

## □ 적용범위 및 작업상황

### 고려사항

- 이 능력단위는 활용목적에 적합한 데이터를 수집하기 위하여 빅데이터 수집시스템 구성, 내·외부 데이터 수집모듈 개발, 데이터 변환모듈 개발 및 수집 데이터 검증모듈을 개발하는 업무에 적용한다.
- 정형데이터란 관계형 데이터베이스와 같이 구조화 되어 있는 데이터를 말한다.
- 준정형데이터란 비구조화 되어 있으나 정형화된 메타데이터를 가지고 있는 데이터를 말한다.
- 비정형데이터란 비구조화 되어 있으며 정형화된 메타데이터를 가지고 있지 않은 데이터를 말한다.
- 일괄처리 데이터란 배치(Batch)형태의 정해진 절차에 따라 처리하는 시스템에서 처리하는 데이터를 의미하고, 보통 비실시간(Non-Real Time) 데이터를 말한다.
- 실시간 데이터란 처리대상 데이터가 발생과 동시에 최단시간 안에 데이터를 수집, 처리해서 반영하는 시스템에서 처리하는 데이터를 의미한다.
- 데이터의 유형을 고려한다는 의미는 데이터의 형태가 정형, 준정형, 비정형인 지를 구분하는 것을 포함하여 센서 데이터, RFID 데이터, 소셜 네트워크 데이터, 금융 및 이동통신 데이터 등 발생하는 데이터의 분야별 특성을 고려해야 한다는 것이다.
- 대용량 실시간 서비스 데이터 특성을 감안한 빅데이터 저장방식은 하둡 기반의 빅데이터 배치 저장 방식을 배제한, 아파치 스파크(Apache Spark), 아파치 스톰(Apache Storm) 등 실시간 스트리밍 데이터 저장 프레임워크에 대한 고려를 말한다.
- 이 능력단위에서 필요로 하는 기술 중 분산파일 시스템은 HDFS(Hadoop File System)을 말하며 NoSQL 기술은 MongoDB, Apache Hbase, Apache Cassandra 등을 말한다.

### 자료 및 관련 서류

- 빅데이터 수집 계획서
- 빅데이터 수집 관리시스템 설치 매뉴얼
- 빅데이터 수집 관리시스템 설계도
- 빅데이터 수집모델 설계도
- 빅데이터 현황집계 자료
- 빅데이터 수집 관리시스템 운영지침
- 빅데이터 수집모듈 검증결과서

### 장비 및 도구

- 컴퓨터, 프린터, 문서작성프로그램
- 빅데이터 수집 도구
- 통계 및 수치 집계 소프트웨어
- 데이터 모델링 도구

### 재료

- 해당없음



## □ 평가지침

### 평가방법

- 평가자는 능력단위 빅데이터 수집시스템 개발의 수행준거에 제시되어 있는 내용을 평가하기 위해 이론과 실기를 나누어 평가하거나 종합적인 결과물의 평가 등 다양한 평가 방법을 사용할 수 있다.
- 피평가자의 과정평가 및 결과평가 방법

평 가 방 법	평 가 유 형	
	과 정 평 가	결 과 평 가
A.포트폴리오		
B.문제해결 시나리오	V	
C.서술형시험	V	V
D.논술형시험		V
E.사례연구	V	
F.평가자 질문		V
G.평가자 체크리스트	V	V
H.피평가자 체크리스트		
I.일지/저널		
J.역할연기		
K.구두발표		
L.작업장평가		
M.기타		

### 평가지 고려사항

- 수행준거에 제시되어 있는 내용을 성공적으로 수행할 수 있는지를 평가해야 한다.
- 평가자는 다음 사항을 평가해야 한다.
  - 빅데이터 관련 기술 및 제품, 분석 방식을 조사 여부
  - 기존 시스템과의 연계를 고려한 빅데이터 수집방식 결정 능력
  - 데이터 요구사항 도출, 데이터 유형별 분류 및 특성파악 능력
  - 데이터모델, 수집기능 및 제약조건 설계 능력
  - 완료된 데이터 수집모델 설계 결과의 타당성 검증 능력
  - 데이터의 안정성과 신뢰성을 보증하는 수집시스템 구조 파악 능력
  - 일괄처리 데이터의 특성을 감안한 빅데이터 수집시스템 구성 능력
  - 실시간 데이터의 특성을 감안한 빅데이터 수집시스템 구성 능력
  - 테스트 절차에 따른 데이터 수집관리시스템 정상설치 여부 검증 능력
  - 데이터의 유형별 접근방식별 데이터 수집 모듈 작성 능력
  - 실제 시스템에서 사용하기 용이한 형식으로 변환하는 모듈 구현 능력
  - 생성한 데이터의 유효성여부를 검증할 수 있는 규칙설정 기능 개발 능력
  - 데이터의 유효성여부 검증 기능 개발 능력
  - 테스트 데이터 유효성여부 검증 및 검증모듈 개선 능력

## □ 직업기초능력

순번	직업기초능력	
	주요영역	하위영역
1	의사소통능력	경청 능력, 기초외국어 능력, 문서이해 능력, 문서작성 능력, 의사표현 능력
2	수리능력	기초연산 능력, 기초통계 능력, 도표분석 능력, 도표작성 능력
3	문제해결능력	문제처리 능력, 사고력
4	정보능력	정보처리 능력, 컴퓨터활용 능력
5	기술능력	기술선택 능력, 기술이해 능력, 기술적용 능력

## □ 개발·개선 이력

구 분		내 용
직무명칭(능력단위명)		빅데이터플랫폼구축(빅데이터 수집시스템 개발)
개발·개선연도	현재	2019
버전번호		v2
개발·개선기관	현재	정보기술·사업관리 인적자원개발위원회(한국소프트웨어산업협회)
향후 보완 연도(예정)		2020