

분류번호 : 2001020905_19v2

능력단위 명칭 : 빅데이터 처리시스템 개발

능력단위 정의 : 빅데이터 처리시스템 개발이란 저장된 데이터를 처리목적에 따라 크기, 종류, 저장 구조를 고려하여 처리 및 가공하기 위한 분산처리, 실시간처리, 이벤트처리 모듈을 개발하는 능력이다.

능 력 단 위 요 소	수 행 준 거
2001020905_19v2.1 빅데이터 처리시스템 설계하기	<ol style="list-style-type: none">1.1 빅데이터 관련 기술, 제품, 분석을 고려하여 처리데이터의 크기, 종류, 저장구조에 따른 빅데이터 처리방식을 결정할 수 있다.1.2 데이터 처리모델을 설계하기 위하여 데이터 요구사항을 도출하고 데이터 유형별로 분류하여 특성을 파악할 수 있다.1.3 데이터의 크기, 종류, 저장 구조 특성에 부합하는 데이터 처리모델 및 제약조건을 설계할 수 있다.1.4 완료된 데이터 처리모델 설계에 대한 결과의 타당성을 검증할 수 있다.
	【지식】 <ul style="list-style-type: none">• 데이터 저장 방식 설계 기법• 데이터 처리 모듈 구성 방식• 분산 데이터 저장 구조• 분산 데이터 처리 기법• 빅데이터 처리시스템 검증 기법
	【기술】 <ul style="list-style-type: none">• 분산파일 시스템 기술• 대규모 데이터 병렬 처리 기술• 업무 도메인 분석 능력• 빅데이터 요구사항 분석 능력• 빅데이터 처리모델 설계도구 활용 능력
2001020905_19v2.2 빅데이터 처리시스템 구성하기	【태도】 <ul style="list-style-type: none">• 빅데이터 처리 모델 및 서비스에 따라 최적의 처리 방식 선정 의지• 다양한 기술을 상호 비교하고 적절한 대안을 제시하는 자세• 상호간 의견 조율 및 최적의 안을 도출하려는 태도• 다양한 요구사항을 긍정적으로 수용하려는 자세
	<ol style="list-style-type: none">2.1 빅데이터 처리시스템설계서에 따라 빅데이터 처리 하드웨어와 소프트웨어를 설치하고 처리시스템 개발환경설정을 할 수 있다.2.2 빅데이터 처리시스템을 효율적으로 사용하기 위해 사용자 혹은 작업 종류에 따라 시스템 자원을 조정할 수 있다.2.3 주어진 테스트 절차에 의해 데이터 처리시스템 설치완료 여부를 확인할 수 있다. 【지식】 <ul style="list-style-type: none">• 분산 데이터 저장 구조

2001020905_19v2.2 빅데이터 처리시스템 구성하기	<ul style="list-style-type: none"> • 분산 데이터 처리 기법 • 데이터웨어하우스 데이터 처리 기법 • 데이터 처리 기법 • 빅데이터 플랫폼 유형과 작동 원리
	【기술】 <ul style="list-style-type: none"> • 분산파일 시스템 기술 • 대규모 데이터 병렬 처리 기술 • 하드웨어 및 소프트웨어 구축 능력 • 분산 병렬 컴퓨팅 솔루션 구축 능력
	【태도】 <ul style="list-style-type: none"> • 설치 매뉴얼 및 절차 준수 • 정보를 적극적으로 수집하는 태도 • 문제해결을 위한 긍정적 자세 • 주어진 시간 안에 과업완료를 위해 노력하는 책임감 있는 자세
2001020905_19v2.3 분산처리 수행모듈 개발하기	3.1 분산처리 모델에 따라 대규모 데이터를 병렬 처리하는 분산처리 수행모듈을 작성할 수 있다. 3.2 분산처리를 수행하기 위하여, 스크립트를 작성하고 등록할 수 있다. 3.3 주어진 테스트 절차에 의해서 데이터 분산처리 완료 및 성공 여부를 테스트하고, 실패 시 문제점을 파악하고 처리하는 모듈을 수정할 수 있다.
	【지식】 <ul style="list-style-type: none"> • 분산 데이터 저장 구조 • 분산 데이터 처리 기법 • 데이터 저장 처리 기법 • 정보시스템 구축 방법론 • 정보시스템 개발 생명주기
	【기술】 <ul style="list-style-type: none"> • 분산파일 시스템 기술 • 대규모 데이터 병렬 처리 기술 • 대용량 데이터 질의 처리 기술 • 분산 병렬처리 프로그래밍 기술 • 스크립트 언어 프로그래밍 기술 • 프레임워크 활용 개발 능력
2001020905_19v2.4 실시간 수행모듈 개발하기	【태도】 <ul style="list-style-type: none"> • 시스템에 대한 분석적 사고 • 문제해결을 위한 긍정적 자세 • 소스 코드를 심도 있게 분석하고 개선하려는 자세 • 적극적으로 의사소통하려는 태도
	4.1 실시간처리를 위하여 스키마, 컬럼, 조건, 처리 방식을 결정하고 실시간 수행모듈을 작성할 수 있다.

2001020905_19v2.4 실시간 수행모듈 개발하기	<p>4.2 실시간처리 중 데이터 정합성 유지를 위한 로깅, 검증, 예외처리 공통모듈을 작성할 수 있다.</p> <p>4.3 주어진 테스트 절차에 의해서 데이터 실시간처리 완료 및 성공 여부를 테스트하고, 실패 시 문제점을 파악하고 처리하는 모듈을 수정할 수 있다.</p>
	<p>【지식】</p> <ul style="list-style-type: none"> • 실시간 처리 대상 선정 방식 • 실시간 트랜잭션 데이터 처리 기법 • 인 메모리 컴퓨팅 기법 • 정보시스템 구축 방법론 • 정보시스템 개발 생명주기
	<p>【기술】</p> <ul style="list-style-type: none"> • 실시간 원천 데이터 특성 분석 능력 • 분산파일 시스템 기술 • 인 메모리 데이터 관리 기술 • NoSQL 활용 능력 • 로드 밸런싱(Load Balancing) 기술 • 프레임워크 활용 개발 능력
2001020905_19v2.5 이벤트처리 수행모듈 개발하기	<p>【태도】</p> <ul style="list-style-type: none"> • 시스템에 대한 분석적 사고 • 문제해결을 위한 긍정적 자세 • 소스 코드를 심도 있게 분석하고 개선하려는 자세 • 프로세스 흐름에 대해 주의 깊게 관찰하는 태도
	<p>5.1 스트림데이터의 특정 패턴에 대한 쿼리를 작성할 수 있다.</p> <p>5.2 이벤트 감지와 처리를 하기 위한 규칙을 정의하고, 규칙에 따라 이벤트처리를 수행할 수 있는 모듈을 작성할 수 있다.</p> <p>5.3 주어진 테스트 절차에 의해서 데이터 이벤트처리 완료 및 성공 여부를 테스트하고, 실패 시 문제점을 파악하고 처리하는 모듈을 수정할 수 있다.</p>
	<p>【지식】</p> <ul style="list-style-type: none"> • 이벤트 감지 및 처리 기법 • 스트림 데이터 처리 기법 • 인 메모리 컴퓨팅 기법 • 정보시스템 구축 방법론 • 정보시스템 개발 생명주기
	<p>【기술】</p> <ul style="list-style-type: none"> • 이벤트 패턴화 정의 능력 • 분산파일 시스템 기술 • 인 메모리 데이터 관리 기술 • 복합 이벤트 처리(CEP:Complex Event Processing) 기술 • 연속 질의 처리(Continuous Query Processing) 기술

2001020905_19v2.5 이벤트처리 수행모듈 개발하기	• 프레임워크 활용 개발 능력
	【태도】 <ul style="list-style-type: none"> • 시스템에 대한 분석적 사고 • 문제해결을 위한 긍정적 자세 • 소스 코드를 심도 있게 분석하고 개선하려는 자세 • 프로세스 흐름에 대한 주의 깊게 관찰하는 태도

□ 적용범위 및 작업상황

고려사항

- 이 능력단위는 저장된 데이터를 처리목적에 따라 크기, 종류, 저장 구조를 고려하여 처리 및 가공하기 위한 분산처리, 실시간처리, 이벤트처리 모듈을 개발하는 업무에 적용한다.
- 요건 문서화는 유관부서와 협의를 위한 요청일, 시스템명, 처리요구사항, 부서명, 담당자명 등을 포함한다.
- 빅데이터 처리시스템설계서란 데이터 처리 기능을 구현하는 빅데이터 플랫폼 기반의 하드웨어, 소프트웨어 구성안을 작성한 문서이다.
- 스크립트는 반복적인 처리 프로세싱을 위해서 개발자들이 손쉽게 프로그램을 다룰 수 있도록 지원해 주는 도구를 포함한다.
- 인 메모리 컴퓨팅은 데이터 적재 및 처리를 메모리상에서 수행하는 아키텍처로 높은 성능이 요구되는 실시간, 이벤트 처리 업무에 적용한다.
- 실시간 처리 방식이란, 정의된 시간 안에 Insert, Update, Delete, Select 작업을 수행하여 결과를 전달하는 프로세싱을 말한다.
- 연속 질의 처리(Continuous Query Processing)란 SQL 기반의 언어를 사용하여 일정 시간 동안 지속적인 쿼리를 유지하는 데이터처리를 말한다.
- CEP(Complex Event Processing)는 여러 이벤트 소스로 부터 발생한 이벤트를 대상으로 실시간으로 의미 있는 데이터를 추출하여 대응되는 처리를 하는 것을 말한다.

자료 및 관련 서류

- 분석 대상 변수 리스트
- 빅데이터 처리시스템설계서
- 소프트웨어 설치 및 운영 매뉴얼
- 시스템 아키텍처(하드웨어, 소프트웨어 아키텍처)
- 요구사항 명세서
- 빅데이터 저장 구조 정의서
- 데이터 모델 정의서
- 분산 병렬 프로그래밍 기술 자료

장비 및 도구

- 컴퓨터, 프린터, 문서작성프로그램
- 프로그램 개발도구
- 이벤트 처리 도구
- 빅데이터 처리 성능 측정 도구

재료

- 해당없음

□ 평가지침

평가방법

- 평가자는 능력단위 빅데이터 처리시스템 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.
 - 처리데이터의 크기, 종류, 저장구조에 따른 빅데이터 처리방식 결정 능력
 - 데이터 요구사항 도출, 데이터 유형별 분류 및 특성 파악 능력
 - 데이터의 크기, 종류, 저장 구조 특성에 부합하는 데이터 처리모델 및 제약조건을 설계 능력
 - 완료된 데이터 처리모델 설계에 대한 결과의 타당성을 검증 능력
 - 빅데이터 처리 하드웨어와 소프트웨어 설치 및 처리시스템 개발환경설정 능력
 - 사용자 혹은 작업 종류에 따른 시스템 자원 조정 능력
 - 테스트 절차에 따른 데이터 처리시스템 설치완료 여부 확인 능력
 - 대규모 데이터를 병렬 처리하는 분산처리 수행모듈 작성 능력
 - 분산처리를 수행하기 위한 스크립트 작성 및 등록 여부
 - 데이터 분산처리 완료 및 성공 여부를 테스트 및 처리 모듈 수정 능력
 - 스키마, 컬럼, 조건, 처리 방식 결정 및 실시간 수행모듈 작성 능력
 - 데이터 정합성 유지를 위한 로깅, 검증, 예외처리 공통모듈 작성 능력
 - 데이터 실시간처리 완료, 성공 여부 테스트 및 처리 모듈 수정 여부
 - 특정 패턴에 대한 쿼리문 정의 및 조건문 작성 능력

- 이벤트 감지와 처리를 위한 규칙 정의 및 규칙에 따른 이벤트처리 수행 모듈 작성 능력
- 데이터 이벤트처리 완료 및 성공 여부 테스트 및 모듈 수정 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 처리시스템 개발)
개발·개선연도	현재	2019
버전번호		v2
개발·개선기관	현재	정보기술·사업관리 인적자원개발위원회(한국소프트웨어산업협회)
향후 보완 연도(예정)		2020