
2020 AWS Cloud Engineer 양성과정

프로젝트 결과 보고서

빅 데이터 활용

하이브리드 인프라 구축

이름

이정근

이상준


전진수

이메일

mung0001@naver.com

wqe014@gmail.com


wjswlstn@yahoo.com

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	


개 정 이 력

개정 번호	개정 내용 요약	추가/수정 항목	개정 일자
1.0	최초 제정 승인	시작	2020.06.20
2.0	추가 수정 승인	수정	2020.06.25
3.0	최종 수정 승인	최종 수정	2020.06.31

문 서 규 칙


	Report	Version	Last Modified	④
	①	②	③	

- 작성은 Microsoft Word Office 365(v.1908)으로 작성되었으며, 저장시 PDF 형식으로 저장한다.
- 확인은 Adobe Arcrobat Reader로 사용한다.
- Report(①)에는 핵심 개정 내용을 표기한다.
- Verion (②)에는 문서의 개정 번호를 표기한다.
- Last Modified(③)에는 문서의 개정 일자를 표기한다.
- ④ 에는 프로젝트 명을 표기한다.


	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

목차

1. 프로젝트개요	7
1.1. 프로젝트 주제	7
1.2. 배경 및 목적	7
1.3. 일정	9
1.4. 인프라 전체 구성	10
1.5. 버전 및 장비 정보	11
2. 프로젝트 구현	14
2.1. 인프라 흐름 구성	14
1) On Premise 환경에서의 남는 자원 확인	14
2) Public 환경에서의 부족한 자원 지원	14
3) S3 동기화	15
4) EMR을 통한 데이터 분석	15
5) Lambda를 통한 데이터 정렬 및 RDS로 데이터 전달	16
6) 3 Tier 서비스 구현	16
2.2. Crawling 코드 작성	17
1) Crawling 대상 선정 기준	17
2) Crawling 코드 작성	17
2) Crawling 테스트 모듈화 구현	17
3) Crawling 코드 적용	19
2.3. RDS 및 S3 생성	19
1) RDS 생성	19
2) S3 생성	20
2.4. EMR 구현	22

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

1) Hadoop test jar 파일 생성	22
2) EMR 생성	22
3) EMR Hadoop jar 파일 적용	22
4) EMR 데이터 분석 및 저장	22
2.5. Lambda 구현	22
1) Lambda 생성	23
2) Lambda 데이터 전달 함수 작성	23
2.6. Web Page	24
1) 메인 페이지 구성	24
2) 트렌드 서비스 페이지 구성	24
3) 특가 상품 페이지 구성	25
2.7. On Premise 환경 구축 (OpenStack)	25
1) On Premise 테스트 환경 사양 선정	25
2) On Premise 테스트 서버 설정	25
3) Openstack template 테스트	26
2.8. Public 환경 구축 (AWS)	28
1) IAM 역할 생성 및 권한 부여	28
2) VPC, Subnet, Routing Table 설정.....	28
3) 보안그룹, ACL 설정	31
4) Public	34
5) Private	35
2.9. ETC	36
1) Auto Scaling	36
2) ELB 설정	37
3) CloudFront	39
3. 프로젝트 결과	41
3.1. 테스트	41
3.2. 비용산정 결과	43

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

3.3. 결과	44
1) 한계	44
2) 향후 발전방향	44
4. 참고문헌	45



	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

그림 목차


1) 그림 1 국내 데이터 및 분석 시장 전망	7
2) 그림 2 글로벌 데이터 증가량 추이	8
3) 그림 3 일정표	9
4) 그림 4 인프라 전체 구성도	10
5) 그림 5 On Premise 환경에서의 남은 자원	14
6) 그림 6 EMR을 통한 데이터 분석	15
7) 그림 7 3Tier 서비스 구현	16
8) 그림 8 트래픽 기준	18
9) 그림 9 Crawling test-1	18
10) 그림 10 Crawling test-2	18
11) 그림 11 Crawling test-3	18
12) 그림 12 Crawling test-4	18
13) 그림 13 Crawling test-result	19
14) 그림 14 RDS 생성	19
15) 그림 15 RDS 피라미터 그룹	20
16) 그림 16 S3 생성	20
17) 그림 17 Crawling 코드 업로드	20
18) 그림 18 Crawling 모듈화 코드 업로드	20
19) 그림 19 EMR 서비스 위한 Input, Output 생성	21
20) 그림 20 Hadoop tree	22
21) 그림 21 EMR Hadoop jar	22
22) 그림 22 EMR input test	23
23) 그림 23 EMR output test	23
24) 그림 24 Lambda	23
25) 그림 25 Project-Lambda-RDS	24
26) 그림 26 Web 메인 페이지	24
27) 그림 27 Web 트렌드	25
28) 그림 28 Web 특가상품	25
29) 그림 29 On Premise 테스트 서버	26
30) 그림 30 Openstack template	27
31) 그림 31 Openstack stack	27
32) 그림 32 VPC	28
33) 그림 33 Subnet	28
34) 그림 34 Public Routing table	29
35) 그림 35 Public Routing table subnet	29
36) 그림 36 Private Routing table	30
37) 그림 37 Private Routing table subnet	30
38) 그림 38 EMR Routing table subnet	30
39) 그림 39 ACL	31
40) 그림 40 EMR ACL	31
41) 그림 41 ex-elb-sg	32

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

42) 그림 42 public-sg	32
43) 그림 43 private-sg	33
44) 그림 44 RDS-sg	33
45) 그림 45 Crawling-sg	33
46) 그림 46 Worker	34
47) 그림 47 Apache2-000-default	34
48) 그림 48 Private	35
49) 그림 49 AMI	36
50) 그림 50 AS 시작 구성	36
51) 그림 51 As	36
52) 그림 52 AS 확인	36
53) 그림 53 ELB	37
54) 그림 54 NLB	37
55) 그림 55 ALB	37
56) 그림 56 CloudFront	38
57) 그림 57 CloutFront-web	38
58) 그림 58 정시 테스트1	39
59) 그림 59 정시 테스트2	40
60) 그림 60 정시 테스트3	40
61) 그림 61 불특정 시간 테스트1	41
62) 그림 62 불특정 시간 테스트2	41
63) 그림 63 불특정 시간 테스트3	42
64) 그림 64 S3 data	42
65) 그림 65 S3 Input	43
66) 그림 66 S3 Output	43
67) 그림 67 RDS	43
68) 그림 68 web Service	44

표 목차

1) 표 1 구현도구	11
2) 표 2 On Premise	12
3) 표 3 Public	12
4) 표 4 3 Tier web Instance	16
5) 표 5 3 Tier web Instance	16
6) 표 6 3 Tier RDS	17
7) 표 7 On Premise 서버	26
8) 표 8 AMI	26
9) 표 9 정시 테스트	40
10) 표 10 불특정 시간 테스트	42
11) 표 11 비용산정 표	

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

1. 프로젝트 개요

1.1. 프로젝트 주제

- 특정시간 이후의 프라이빗 클라우드 상의 남는 자원을 통한 데이터 수집
- 부족한 자원을 퍼블릭 클라우드를 통해 충족
- 퍼블릭 클라우드 상의 데이터 저장 및 분석 서비스 인프라 구축
- 데이터 크롤링 및 분석을 통한 특가, 트렌드 서비스 구현

1.2. 배경 및 목적

1) 프로젝트 주제 선정 이유 (내부적 요인)

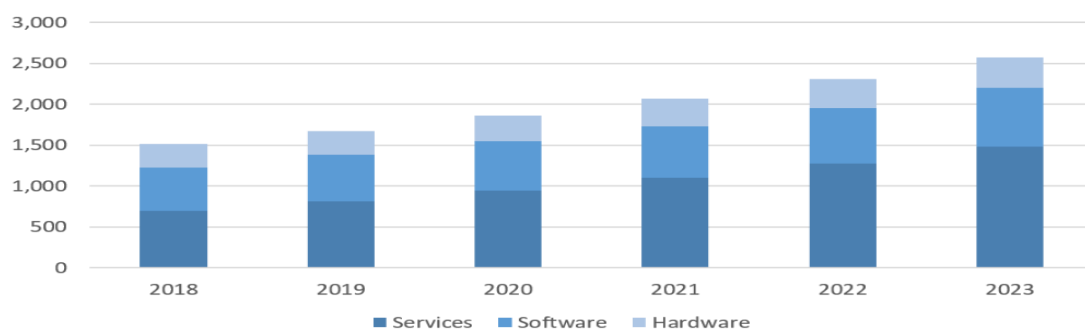
주제를 선정하는 과정에서 팀원들과 논의 중, 여러 아이디어가 나왔으며, 그 중 공통적으로 다뤄보고 싶었던 하이브리드 클라우드, 빅 데이터를 활용할 수 있는 주제로 선정

또한 학원에서의 강의내용들을 전체적으로 사용할 수 있는 프로젝트를 생각하고 있었고, 강의기간 중 학원생들끼리 개인적으로 진행했던 웹, 프로그래밍 등의 스터디에 대한 내용도 프로젝트에 함유하고 싶었기에, 이와 같은 프로젝트가 저희의 지금까지의 공부 및 노력들을 가장 잘 나타낼 수 있다 생각이 들었기에 프로젝트 주제로 선정하였음.

2) 프로젝트 주제 선정 이유 (외부적 요인)




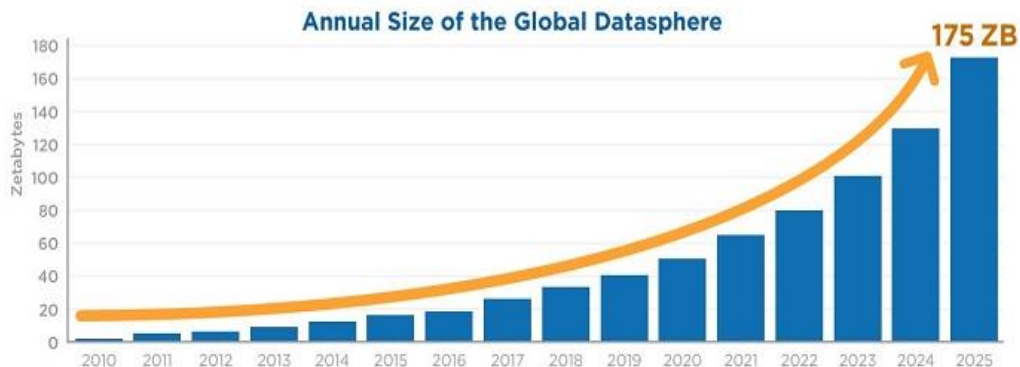
국내 빅데이터 및 분석 시장 전망 2019-2023년 [단위:십억]



<그림1>국내 데이터 및 분석 시장 전망

[출처]: IDC국내 데이터 시장현황과 전망.2019

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	




<그림2>글로벌 데이터 증가량 추이[출처]: Intel News
[출처]: Intel. 데이터 해석과 통계 처리 등 전문가 영역으로 확장

이미 빅 데이터는 의료, 금융, 부동산, 여행, 식품 등 다양한 분야에서 활용되어지고 있으며, IDC의 발표한 자료(그림 1)에 따르면 국내 빅 데이터시장은 2018년 1,500억원에 비해 2020년은 1,800억원 2023년은 2,500억원이 웃도는 것으로 예측되어지고 있다. 또한 인텔에서의 발표한 자료 (그림 2)에 2019년에 비해 2025년에는 175ZB까지 데이터 사용 및 수집량이 기하급수적으로 증가할 것이라 예측하였고, 이미 정부에서도 이전 2016년도부터 정부핵심사업으로 빅 데이터관련 사업을 항상 추진해왔고, 올해인 2020년에만 데이터 경제 활성화 사업에 730억원을 지원하였다.

하지만 대부분의 기업들은 빅데이터 도입에 선불리 도전할 수 없었던 데, 그 이유는 빅 데이터 도입에 따른 위험성 때문으로, 이 위험성의 핵심원인으로 마이크로 소프트의 고위임원 스노우플레이크 컴퓨팅의 CEO 밥 무글리아는 4가지를 주장하였는데, 우리는 이 문제점들을 퍼블릭을 통해 해결하는 것을 프로젝트 아이디어로 선정할 수 있었음.


첫 번째는 부실한 통합으로, 빅 데이터 도입 시 현재 인프라의 연동의 문제로, 빅 데이터는 단순히 매크로 프로그램이 아닌, 데이터 베이스와 인프라가 직접적으로 연관되어, 구축이 어려울 뿐 아니라, 오히려 현재 사용중인 인프라조차 망가질 수 있다는 점을 지적하였으며, 두 번째는 불분명한 목표로 단순히 빅데이터를 활용하자는 목표가 아닌, 뚜렷한 서비스를 제공하지 않는다는 점을 지적하였고, 세 번째는 기술 간극으로 부실한 통합과 연계된 문제로, 데이터 웨어하우스의 관한 내용으로, 빅 데이터도 단순히 쌓아두는 것이 아닌, 정리되며 재분석되어야 하는 데, 이를 해결할 용량이 시시각각 달라져 문제가 생길 수 있다는 점을 지적하였고, 마지막으로 네 번째는 기술 세대 차이는 데이터들은 항상 변하며 인프라들도 변화되고 그에 맞게 적용되어야 하지만, 사내의 환경에서는 독립적으로 만드는 데에는 한계가 있다는 것으로 우리는 이러한 문제점들을 클라우드 환경을 활용하여 독립적인 공간을 만듦으로써, 접근권한, 저장소, 데이터의 변동성 등의 문제를 해결할 수 있다고 생각했고, 이를 프로젝트 주제로 선정하였음.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

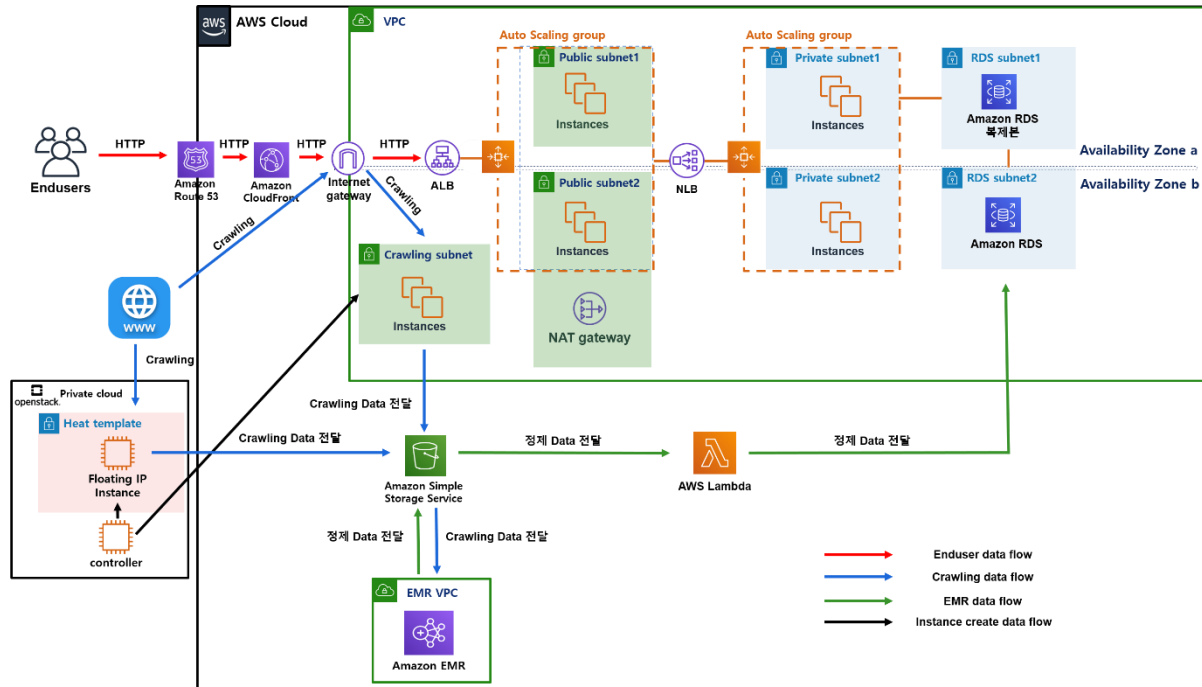
1.3. 일정

분류	단계	내용	선행작업	활용기술	소요시간
구상	1	아이디어 회의 및 설계			1 Day
	2	필요 시스템 분석			
	3	인프라 설계	하이브리드 환경		1 Day
설계	4	웹 페이지 환경 구성 설계	카테고리 기능 구성 게시판 기능 구성 게시물 기능 구성 트랜스 서비스 구성 가격 비교 서비스 구성	HTML, CSS, Js, Django	2 Day
	5	크롤링 및 하둡 설계	크롤링, 하둡 데이터 폴로어 설계		1 Day
구현	6	DB&Storage 설계	ERD 설계	S3, RDS, Mysql	1 Day
	7	크롤링 테스트 코드 작성	종소핑 크롤링 코드 작성 중고 쇼핑물 크롤링 코드 작성	Python	1 Day
	8	크롤링 이미지 생성 및 배포	크롤링 코드 CentOS 환경으로 배포 및 적용	CentOS, Cloud-utils, AMI	1 Day
	9	크롤링 수집 작업	크롤링 코드 테스트를 통한 자료 수집	Python, HTML, CSS	
	10	크롤링 처리 작업	크롤링 코드 자료 일관화	Python, HTML, CSS, Database	
	11	크롤링 오류 수정	크롤링 코드 오류 수정	Python, HTML, CSS, Database	1 Day
	12	크롤링 S3 연동	크롤링 코드 S3 저장 테스트	Python, HTML, CSS, S3	1 Day
	13	데이터베이스 양식 구축 작업	데이터 테이블 작성 및 적용	HADOOP, DB, 파일학장자	
	14	Bigdata 테스트 코드 작성	Hadoop 설치 및 테스트 코드 작성	CentOS, JAVA, HADOOP, DB	4 Day
	15	Bigdata 처리 작업	Hadoop 오프라인 테스트	JAVA, HADOOP, DB	1 Day
	16	Bigdata 오류 수정	Hadoop 테스트 오류 수정	JAVA, HADOOP, DB	
	17	Bigdata 저장소 연동	Hadoop 오프라인 저장소 연동	hadoop, Instance, EMR, DB	1 Day
	18	오픈스택 서비스 구축	오픈스택 서비스 구축 Rocky: Heat	CentOS7, Packstack	
	19	오픈스택 orchestration 구현	오픈스택레이션 서비스 테스트	Heat	1 Day
	20	EMR, 인스턴스 S3 연동	온프레미스 환경에서 크롤링 테스트 퍼블릭 환경에서의 크롤링 테스트 수집 된 자료 S3 저장 테스트	Instance, EMR, S3	1 Day
	21	S3 LAMBDA RDS 연동	S3 RDS 저장	S3, LAMBDA, RDS	2 Day
	22	RDS 위키복제본	RDS 복제본 생성	RDS	
	23	웹 페이지 구현	카테고리 기능 구현 DB table 생성 게시판 기능 구현 게시물 기능 구현 검색 기능 구현 트랜스 서비스 구현 가격 비교 서비스 구현	Django, Tomcat Java, security coding	3 Day
	24	웹 페이지 DB연동	웹 테스트용 DB 연동 웹 RDS 연동	RDS, DB, Django	1 Day
	25	테스트 및 오류수정	웹 서비스 기능 테스트	RDS, Apache2	
	26	웹 페이지 배포	퍼블릭 배포	EC2	
	27	web server	VPC 생성 서브넷 생성 리우팅 테이블 생성 인터넷 게이트웨이 연결 NAT 게이트웨이 연결 Apache 설치	VPC, EC2, EBS	1 Day
	28	서버 오토스케일링	로드 밸런서 생성 오토 스케일링 설정 CDN 설정 리우터 S3 등록	ELB, 오토스케일링, CDN, Route53	1 Day
	29	server 테스트	CloudWatch 모니터링 테스트	CloudWatch	
	30	app server	서브넷 생성 리우팅 테이블 생성 인터넷 게이트웨이 연결 로드 밸런서 설치 제각되어진 Web 등록 웹 홈페이지와 RDS 연동	VPC, EC2, EBS, RDS	1 Day
	31	app server 오토 스케일링	로드 밸런서 생성 오토 스케일링 설정	ALB	1 Day
	32	웹 서버 테스트	오토 스케일링 테스트		
	33	오류 수정	서비스 오류 수정		1 Day
시험/ 발표	34	최종 테스트	Log 및 CloudWatch 모니터링 테스트	CloudWatch	1 Day
	35	발표연습			

<그림3> 일정표


	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

1.4. 인프라 전체 구성










<그림4> 인프라 전체 구성도


- 인프라는 크게 On Premise 영역, Public 영역, Enduser 영역으로 나누어지며, 먼저 On Premise와 Public 영역에서 크롤링을 통해 데이터를 수집 후 이를 S3에 저장하고, 여기에서 하이브리드 환경을 활용하여, On premise의 자원의 고갈 및 자연재해로 인한 사용불가 시 이를 판단하며 부족한 할당량은 public에서 ec2를 사용해서 처리한다.
- S3로 모인 데이터는 시각적으로 나타낼 수 있도록 EMR을 통해 데이터를 가공하고, 분석한 데이터를 Lambda를 활용해 RDS에 저장한다.
- 외부의 Enduser들을 외부의 클라이언트들이 Route53 서비스를 통해 호스팅 된 도메인으로 접근하며 호스팅 된 도메인은 cloudFront 서비스를 이용하게 되며 CloudFront는 인터넷게이트웨이 이후 ELB (ALB)를 통해 private 대역의 RDS 및 was 대역의 데이터를 요청하고 제공하는 구조로 이루어져 있다.




	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

1.5. 버전 및 정보




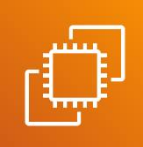

구현도구		
이미지	이름	역할
	Python2 Python3.6	Crawling, OpenStack, AWS CLI 계발 언어
	PIP PIP3	Crawling 계발 패키지
	Visual Studio Code	Crwaling 계발 도구
	Spring	Web 계발 도구
	MySQL 5.7	Web 계발 도구
	JAVA8	Web 계발 도구
	Apache	Web 구현 도구
	Tomcat8	Web 구현 도구


<표 1> 구현도구










	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

On Premise		
이미지	이름	역할
	Ubuntu 18.04	On Premise Crawling 운영체제
	CentOS 7	Openstack 설치 운영체제
	OpenStack	On Premise 환경


<표 2> On Premise

Public		
이미지	이름	역할
	Ubuntu 18.04	Public Crawling 운영체제
	Internet gateway	VPC 인터넷 연결
	NAT gateway	VPC 내 Subnet 통신
	Amazon EC2	Crawling image 및 Web Instances
	Amazon EC2 Auto Scaling	Elastic Web Service

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

			Amazon Lambda	S3 내 데이터 정제 및 RDS로의 데이터 이동
			Amazon CloudFront	Web 배포 Service
			Amazon Route 53	Web DNS Service
			Amazon VPC	Amazon network Service
			Elastic Load Balancing	Web Load Balancing Service
			Amazon EMR	데이터 분석 서비스
			Amazon RDS	Public DataBase (Mysql5.8)
			Amazon Simple Storage Service	Public Storage Service
			Amazon CloudWatch	Public Alarm Service

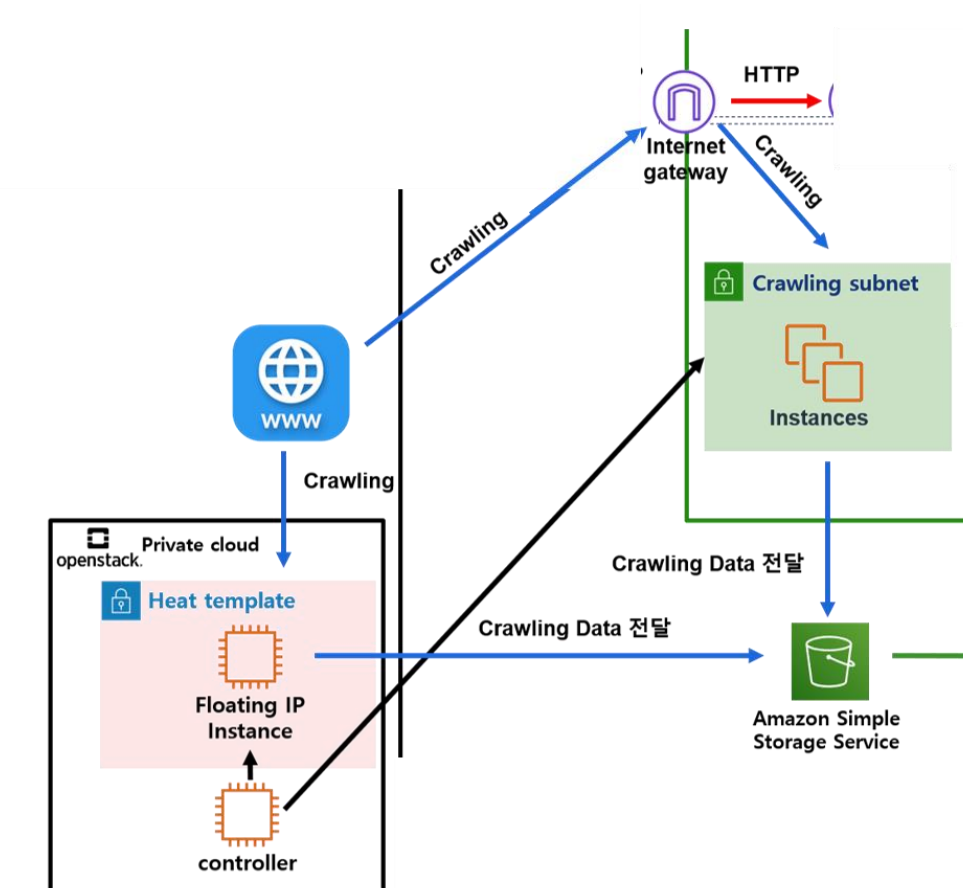
<표 3> Public

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

2. 프로젝트 구현


2.1. 인프라 흐름 구성

1) On Premise 환경에서 남는 자원 확인



<그림 5> On Premise 환경에서의 남는 자원

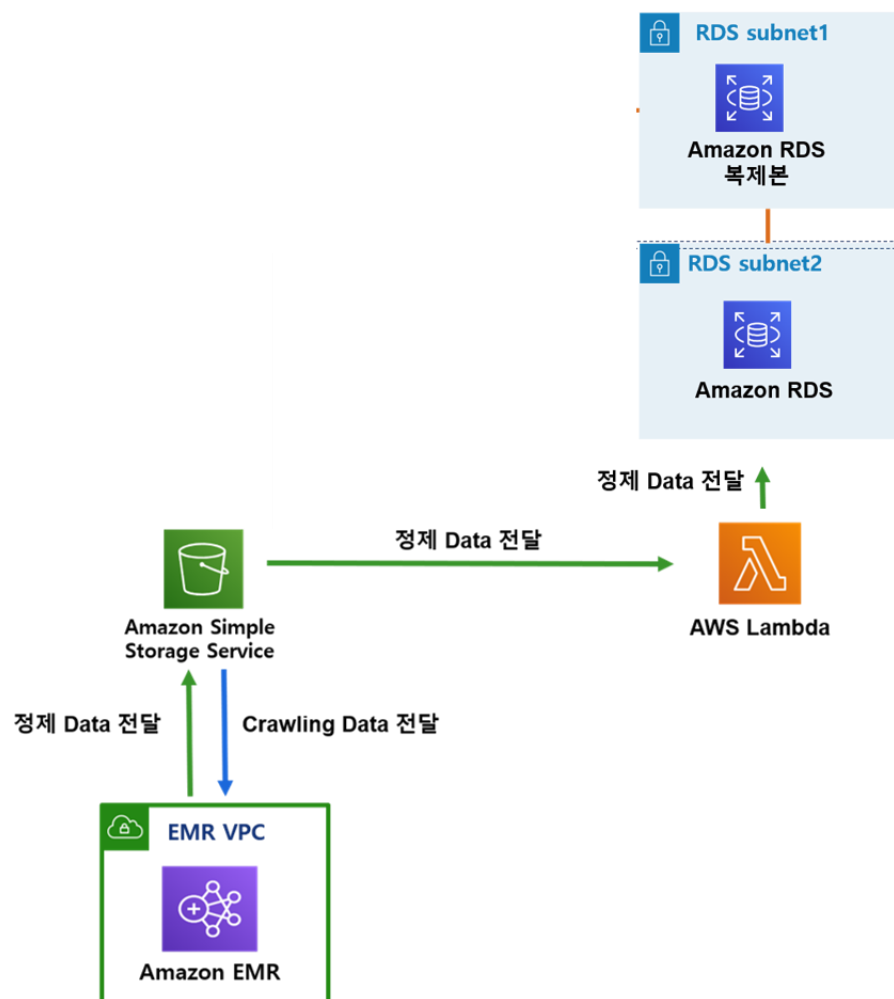
- 12시 이후가 되면 On Premise에서 매 시간 Compute노드들의 사양을 읽어, 남는 자원이 총 자원의 50%이상일 경우 비상자원 (남는 자원의 50)을 제외하고 템플릿을 통해 남는 자원의 최대크기의 인스턴스를 생성.
- 2) Public 환경에서의 부족한 자원 지원
 - 사용가능한 On Premise 자원의 크기 매일 다르기에, Controller 노드에서 부족한 자원만큼의 인스턴스 생성을 Public에 요청.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

3) S3 동기화


- 인프라상의 독립성을 위해 Crawling을 위한 인스턴스 및 EC2는 관리자 또한 접근할 수 없고, 자동적으로 업무를 수행 후 저장 및 종료.
- 이를 위해 Crawling의 대한 관리는 S3및 모듈화 된 코드를 통해 관리를 수행.

4) EMR을 통한 데이터 분석



<그림 6> EMR을 통한 데이터 분석

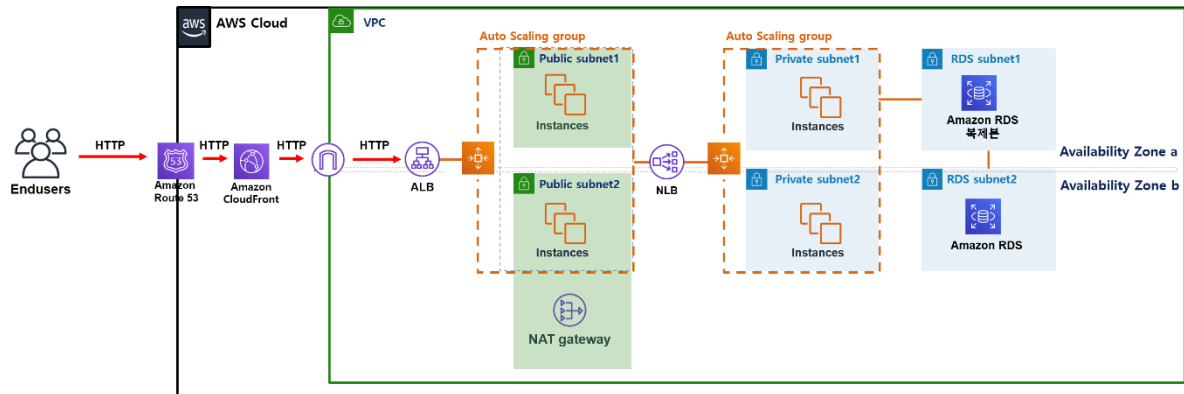
- S3상의 저장된 데이터는 EMR을 통해 데이터 분석을 수행.
- 분석된 데이터는 다시 S3에 저장.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

5) Lambda를 통한 데이터 정렬 및 RDS로 데이터 전달

- 분석된 데이터를 Lambda를 통해 RDS에 저장.

6) 3 Tier 서비스 구현



<그림 7> 3 Tier 서비스 구현

- 분석되어진 데이터를 RDS에 저장 후, 이를 시각적으로 나타낼 수 있는 3 Tier의 웹 서비스를 구현.


3 Tier Web Instance				
이름	유형	vCPUs	Memory(GB)	OS
Apache1-a	m1.micro	1	1	Ubuntu 18.04
Apache1-b	m1.micro	1	1	Ubuntu 18.04
Apache2-a	m1.micro	1	1	Ubuntu 18.04
Apache2-b	m1.micro	1	1	Ubuntu 18.04

<표 4> 3 Tier web Instance

- Web Instance는 각 서브넷당 최소 1개, 최대 2개로 As로 구성되어짐

3 Tier Was Instance				
이름	유형	vCPUs	Memory(GB)	OS
Tomcat1-a	m1.micro	1	1	Ubuntu 18.04
Tomcat2-a	m1.micro	1	1	Ubuntu 18.04
Tomcat1-b	m1.micro	1	1	Ubuntu 18.04
Tomcat2-a	m1.micro	1	1	Ubuntu 18.04

<표 5> 3 Tier was Instance

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

- Web Instance는 각 서버넷당 최소 1개, 최대 2개로 As로 구성되어짐.

3 Tier RDS			
이름	유형	엔진 버전	읽기 복제본 모드
root	db.t2.micro	MySQL 5.7.28	1

<표 6> 3 Tier RDS

- RDS는 MySQL 5.7버전을 사용하며, 읽기 복제본이 1 존재.

2.2. Crawling

1) Crawling 대상 선정 기준

- 일 평균 접속자가 50000 이상인 쇼핑몰 홈페이지
- 카테고리 태그 별 판매순위가 존재하는 쇼핑몰 홈페이지
- 상품의 이름, 이미지, 가격, URL이 존재하는 쇼핑몰 홈페이지
- Crawling 시 접속차단이 걸려있지 않은 홈페이지

2) Crawling 코드 작성

- Crawling 대상 선정 기준에 맞는 11st, Wemap, Gmarket 등의 대한 Crawling 코드를 작성

3) Crawling 테스트 코드 모듈화 구현

- On Premise 및 Public 환경의 업무 분담 및 스레싱을 위해서는 모듈화가 필요
- 외국의 크롤링 속도에 대한 자료를 찾아봤지만, 완벽한 결과가 나오지 않아, (Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019)을 참고하여 같은 환경에서 속도 테스트를 진행 후 결과를 통해 도출하였음.

트래픽 기준:

자연 시간


트래픽 없음

트래픽 많음

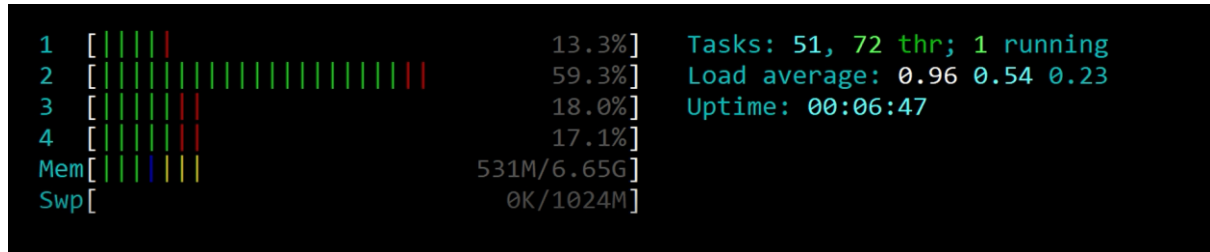
131_{ms}

151_{ms}

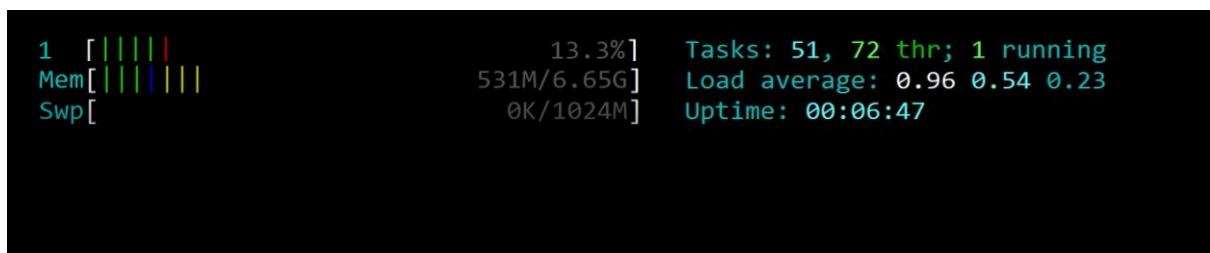
<그림 8> 트래픽 기준

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

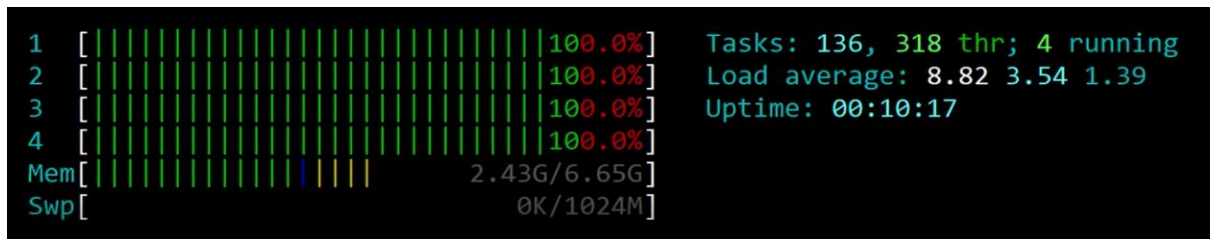
· Crawling Test : n cpu, n ram



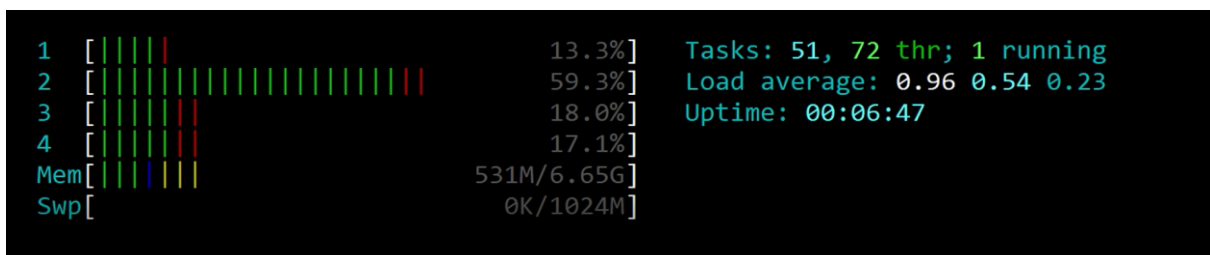
<그림9> Crawling test-1




<그림10> Crawling test-2



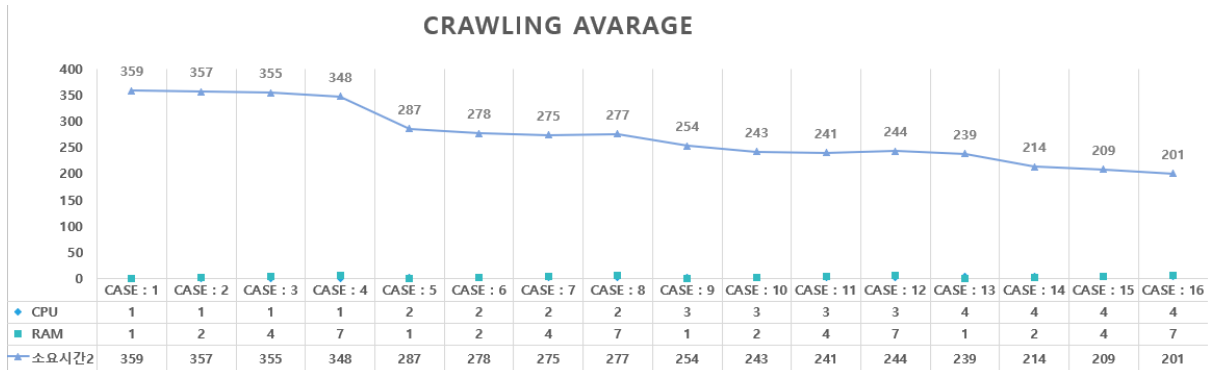
<그림11> Crawling test-3



<그림12> Crawling test-4

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

· 결과



<그림13> Crawling-test-result

- 그림 8의 트래픽 기준에서, 결과 코드 1개에 대한 최대 사용량은 CPU 1개, RAM 512M가 최대치 사용량이었으며, 이 때 걸린 시간은 평균 359초의 값이 도출되었음.

4) Crawling 코드 적용


- 분석 값을 토대로, 평균 45 -55분의 시간의 코드로 모듈화를 진행.
- 코드는 각 폴더 별 1시간 기준으로 모듈화 되어있으며, EC2 인스턴스 또한 t1.micro의 평균 40-50분 동안 작동이 확인.

2.3. RDS 및 S3 생성

1) RDS 생성

데이터베이스							그룹 리소스	수정	작업 ▼	S3에서
Q 데이터베이스 필터										
DB 식별자	역할 ▼	엔진 ▼	리전 및 AZ ▼	크기 ▼	상태 ▼					
<input checked="" type="radio"/> root	마스터	MySQL Community	ap-northeast-2c	db.t2.micro	사용 가능					
<input type="radio"/> root-rp	복제본	MySQL Community	ap-northeast-2a	db.t2.micro	사용 가능					

<그림 14> RDS 생성

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

RDS > 파라미터 그룹 > project-rds

project-rds

파라미터

Q char X

<input type="checkbox"/>	이름	값	허용된 값	수정 가능	소스	적용 유형	데이터 형식	타입
<input type="checkbox"/>	character_set_client	utf8	big5, dec8, cp850, hp8, koi8r, latin1, latin2, swe7, ascii, ujis, sjis, hebrew, tis620, euckr, koi8u, gb2312, greek, cp1250, gbk, latin5, armSCII8, utf8, cp866, keybcs2, macce, macroman, cp852, latin7, utf8mb4, cp1251, cp1256, cp1257, binary, geostd8, cp932, eucjms	true	user	dynamic	string	T
<input type="checkbox"/>	character-set-client-handshake	0, 1		true	engine-default	static	boolean	D
<input type="checkbox"/>	character_set_connection	utf8	big5, dec8, cp850, hp8, koi8r, latin1, latin2, swe7, ascii, ujis, sjis, hebrew, tis620, euckr, koi8u, gb2312, greek, cp1250, gbk, latin5, armSCII8, utf8, ucs2, cp866, keybcs2, macce, macroman, cp852, latin7, utf8mb4, cp1251, utf16, cp1256, cp1257, utf32, binary, geostd8, cp932, eucjms	true	user	dynamic	string	T

<그림 15> RDS 파라미터 그룹

- RDS를 자원 : db.t2.micro, 엔진 : mysql 5.8, VPC : Project-VPC, Subnet : RDS1-subnet, 보안그룹 : RDS-sg의 값으로 생성.
- 파라미터 값을 새로 생성하고, char 값을 utf-8번으로 수정.
- RDS에 파라미터 값을 적용.
- 다른 Az에 읽기 복제본을 생성.


2) S3 생성

+ 버킷 만들기 퍼블릭 액세스 설정 편집 비우기 삭제

1 버킷 1 리전

<input type="checkbox"/>	버킷 이름	액세스	리전	생성 날짜
<input type="checkbox"/>	project-datas	버킷 및 객체가 퍼블릭이 아님	아시아 태평양(서울)	6월 16, 2020 9:01:32 오전 GMT+0900

<그림 16> S3 생성

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

project-datas			
개요			
Q 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.			
업로드 + 폴더 만들기 다운로드 작업 아시아 태평양(서울) 보기 1 대상 6			
이름	마지막 수정	크기	스토리지 클래스
__init__.py	6월 18, 2020 7:17:57 오후 GMT+0900	0 B	스탠다드
crawl_11st.py	6월 22, 2020 3:14:43 오후 GMT+0900	3.6 KB	스탠다드
crawl_Auction.py	6월 22, 2020 3:14:43 오후 GMT+0900	3.4 KB	스탠다드
crawl_Gmaket.py	6월 22, 2020 3:14:43 오후 GMT+0900	3.6 KB	스탠다드
crawl_g9.py	6월 22, 2020 3:14:43 오후 GMT+0900	3.4 KB	스탠다드
crawl_wemap.py	6월 22, 2020 3:14:43 오후 GMT+0900	3.4 KB	스탠다드

<그림 17> Crawling 코드 업로드


업로드 + 폴더 만들기 다운로드 작업 아시아 태평양(서울)			
_20	--	--	--
_3	--	--	--
_4	--	--	--
_5	--	--	--
_6	--	--	--
_7	--	--	--
_8	--	--	--
_9	--	--	--
include	--	--	--
10H.py	6월 19, 2020 12:50:59 오후 GMT+0900	534.0 B	스탠다드
11H.py	6월 19, 2020 12:50:59 오후 GMT+0900	534.0 B	스탠다드
12H.py	6월 19, 2020 12:50:59 오후 GMT+0900	534.0 B	스탠다드
13H.py	6월 19, 2020 12:50:59 오후 GMT+0900	534.0 B	스탠다드
14H.py	6월 19, 2020 12:51:01 오후 GMT+0900	534.0 B	스탠다드
15H.py	6월 19, 2020 12:51:00 오후 GMT+0900	534.0 B	스탠다드

<그림 18> Crawling 모듈화 코드 업로드

code	--	--
crawl_ver4	--	--
data	--	--
input	--	--
log	--	--
output	--	--
example01.jar	6월 16, 2020 4:03:00 오후 GMT+0900	4.3 KB
testfile.json	6월 16, 2020 3:47:55 오후 GMT+0900	226.0 B

<그림 19> EMR 서비스를 위한 Input, Ouput 생성

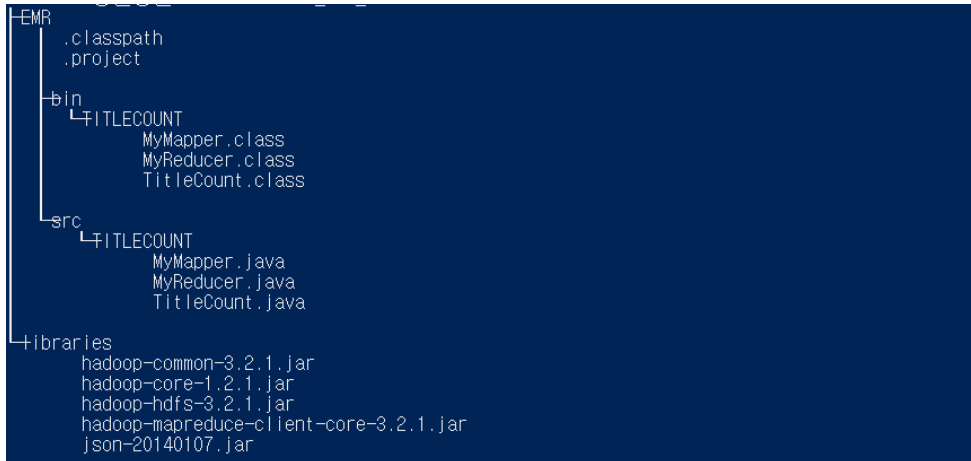
- S3는 Public이 아닌 endpoint를 통해서만 접속이 가능하게 설정.
- On Premise 상의 Instance와 Public 상의 ec2는 code의 디렉터리에서 파일을 가져와 Crawling 작업 후 data의 디렉터리에 해당날짜로 생성된 파일안에 데이터를 저장하도록 설정.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

- 이 후 데이터는 EMR 서비스를 통해 가공되어질 수 있게, 미리 폴더를 생성하였음.

2.4. EMR 구현

1) Hadoop test jar 파일 생성



<그림 20> Hadoop tree


- Wordcount jar 파일을 프로젝트에 맞게 수정을 진행하였음.

2) EMR 생성

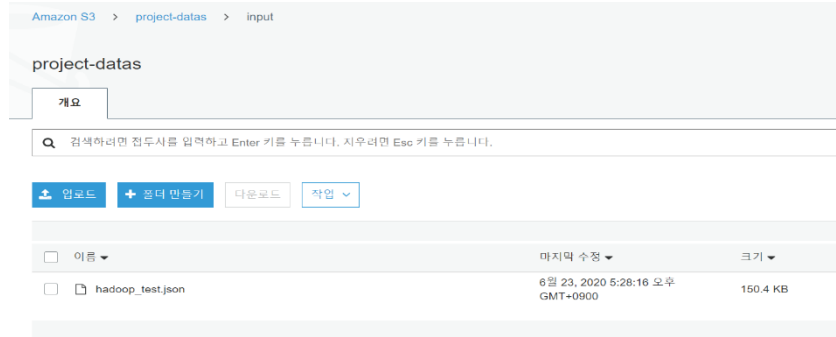
- 버전 : emr-5.30.0 / Hadoop 2.8.5
- VPC : EMR VPC
- 마스터 : r3.xlarge 1, 코어 : r3.xlarge 1
- 보안그룹 : EMR-sg

3) EMR Hadoop jar 파일 적용

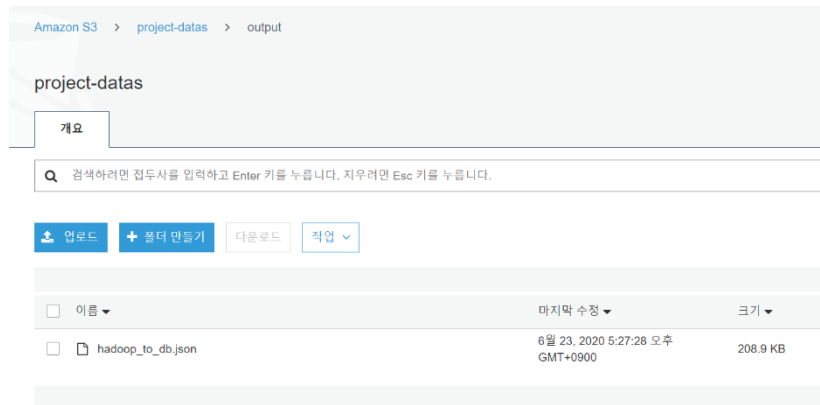
<그림 21> EMR Hadoop jar

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

4) EMR 데이터 분석 및 저장



<사진 22> EMR input test

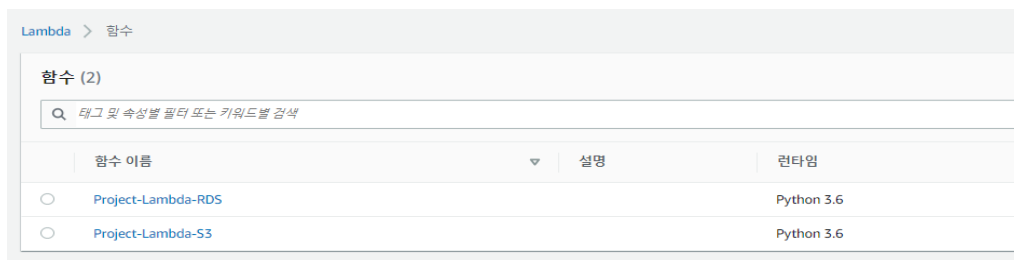


<사진 23> EMR ouput test

- S3의 Input 디렉토리 데이터가 EMR을 통해 Output 폴더에 저장.


2.5. Lambda 구현

1) Lambda 생성



<그림 24> Lambda

- Project-Lambda-S3는 crawling을 통해 추출된 데이터의 형식을 EMR이 읽을 수 있도록 S3 속 Input 디렉토리에 저장.
- Project-Lambda-RDS는 EMR을 통해 정제된 데이터와 Crawling 통해 추출한 데이터를 RDS에 저장.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

2) Lambda 데이터 전달 .

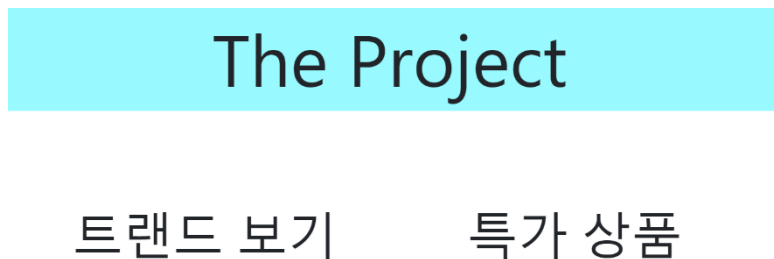
```
mysql> select * from Wordcount limit 7\G
***** 1. row *****
wid: 1
word: #16cm이상
count: 1
***** 2. row *****
wid: 2
word: #2020리뉴얼
count: 1
***** 3. row *****
wid: 3
word: #5년AS
count: 1
***** 4. row *****
wid: 4
word: #가벼움
count: 1
***** 5. row *****
wid: 5
word: #갓성비
count: 1
***** 6. row *****
wid: 6
word: #괴물흡입력
count: 1
***** 7. row *****
wid: 7
word: #국민스킨
count: 1
7 rows in set (0.00 sec)
```

<사진 25> Project-Lambda-RDS

- Project-Lambda-RDS 를 통해 데이터가 RDS 에 저장되었음을 확인할 수 있음.


2.6. Web Page

1) 메인 페이지



<사진 26> Web 메인 페이지

- 메인 페이지에서는 시각적으로 구현된 트렌드 및 특가 상품에 이어주는 역할을 수행.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

음.

2) On Premise 테스트 서버 설치

운영체제	CPU	RAM	IP	Hostname	역할
CentOS7	6	8	192.168.0.75	Controller	Controller
CentOS7	4	8	192.168.0.76	Compute	Compute, Network

<표 7> On Premise 서버

```

+-----+-----+
| ID | Name |
+-----+-----+
| 02a8e27a8c1e4a6f8b92c5347bc7628f | aodh |
| 0df12a68efe24b6cb18ffb87a447c563 | nova |
| 0e703b28145a4b81b7f3fb15139d6e50 | heat-cfn |
| 26ffffce219134de08be05774f2eed0b9 | heat-admin |
| 290bbb9b66ae4f8dbf7dd466e0635ce4 | placement |
| 617404347ccb4a01bf149bfc50763e7 | heat |
| 7990ee104fc1465490559066ad73f7df | gnocchi |
| 984015bfd50a4e7e99cf3640a6712bef | swift |
| b2d10fa42dc645dd9e00716613daced | neutron |
| ba5960d4458f465a824e72c8ab6ea322 | admin |
| bb2da23eb18b466280eb7358fd8e5ee3 | ceilometer |
| c6aaef55f2c8487481b8bc46781104dd | glance |
| fa6bec49bf2e4f9f9c1f16396f02ed22 | cinder |
+-----+-----+

[root@controller ~]# openstack host list
+-----+-----+-----+
| Host Name | Service | Zone |
+-----+-----+-----+
| controller | conductor | internal |
| controller | scheduler | internal |
| controller | consoleauth | internal |
| compute | compute | nova |
+-----+-----+-----+


[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host | Project | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total) | 4 | 8102 | 256 |
| compute | (used_now) | 2 | 4307 | 0 |
| compute | (used_max) | 2 | 3795 | 10 |
| compute | ab70628506f0441ab7857fe67e73088a | 2 | 3795 | 10 |
+-----+-----+-----+-----+-----+

[root@controller ~]# openstack stack list
+-----+-----+-----+-----+-----+
| ID | Creation Time | Updated Time | Stack Name | Project | Stack Status |
+-----+-----+-----+-----+-----+
| 2f9d6f08-5fdd-4bca-9ac6-1374034f8aff | crawling | 2020-06-19T01:52:54Z | None | ab70628506f0441ab7857fe67e73088a | CREATE_COMPLET
E
+-----+-----+-----+-----+-----+

[root@controller ~]#

```

<그림 29> On Premise 테스트 서버

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

3) Openstack template 테스트

```

root@controller:~
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+
| Host   | Project | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+
| compute | (total) | 4   | 8102      | 256      |
| compute | (used_now) | 0   | 512       | 0         |
| compute | (used_max) | 0   | 0         | 0         |
+-----+-----+-----+-----+

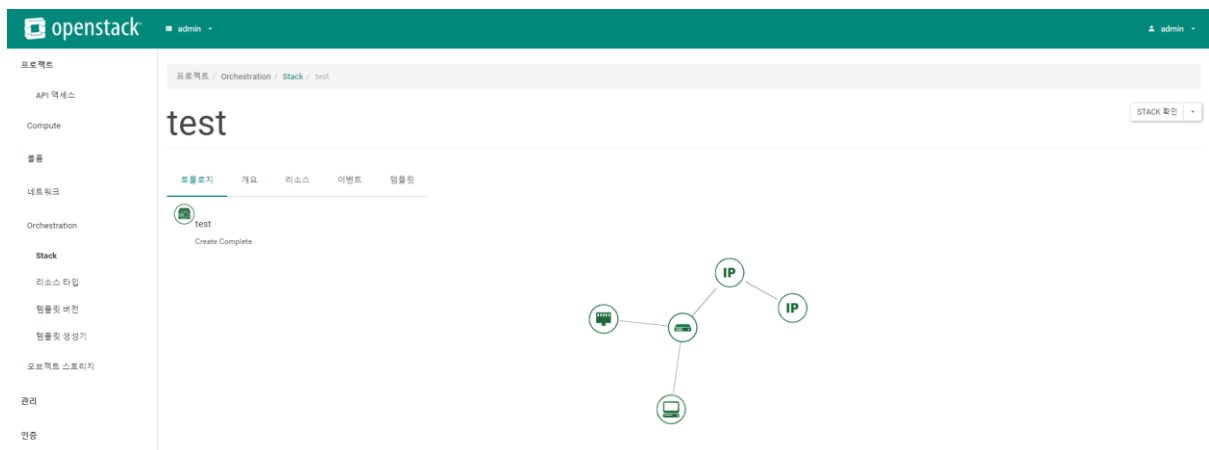
[root@controller ~]# heat stack-create test -f ~/templates/Heat.yaml
WARNING (shell) "heat stack-create" is deprecated, please use "openstack stack create" instead
WARNING (shell) "heat stack-list" is deprecated, please use "openstack stack list" instead
+-----+-----+-----+-----+
| id          | stack_name | stack_status | creation_time | update_time | project |
+-----+-----+-----+-----+
| 723c02e1-3f1e-415a-b9da-5fa337ab3303 | test       | CREATE_IN_PROGRESS | 2020-06-22T03:07:58Z | None        | ab70628506f0441ab7857fe67e73088a |
+-----+-----+-----+-----+

[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+
| Host   | Project | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+
| compute | (total) | 4   | 8102      | 256      |
| compute | (used_now) | 2   | 4307      | 0         |
| compute | (used_max) | 2   | 3795      | 10        |
| compute | ab70628506f0441ab7857fe67e73088a | 2   | 3795      | 10        |
+-----+-----+-----+-----+

[root@controller ~]#


```

<그림 30> Openstack template



<그림 31> Openstack stack

- On Premise 속 오케스트레이션 서비스를 위해 Heat 를 설치 및 테스트를 완료하였으며, S3 동기화 시켜, 명령을 할당받아 자동적으로 코드를 받아와서, 값을 S3 에 저장하고 자동적으로 종료& 삭제되게 구현.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

- 보안그룹은 :80 번만 열려있으며, 키 페어가 존재하지 않아 접근이 불가능하게 설정.

2.8. Public 환경 구축 (AWS)

1) IAM 역할 생성 및 권한 부여

사용자명	그룹	권한	역할
Jklee	Admin	AdministratorAccess	관리자
Crawling	Crawling_IAM	AmazonS3FullAccess	Crawling data 저장을 위한S3 권한
Onpremise	On-Premises-IAM	AWSOpsWorksRegisterCLI_OnPremises AWSOpsWorksRegisterCLI_EC2	CLI를 통해퍼블릭 EC2생성을 위한 역할

<표 8> IAM

2) VPC, Subnet, Routing Table 설정

<input type="checkbox"/>	Name	VPC ID	상태	IPv4 CIDR	IPv6 CIDR
<input type="checkbox"/>	Project-VPC	vpc-0b3f667b25d5f8b51	available	10.0.0.0/16	-
<input type="checkbox"/>	EMR-VPC	vpc-0c259d49eca20f5c0	available	172.31.0.0/16	-


<그림 32> VPC

- 10.0.0.0/16 : enduser가 접속하는 환경으로 이중화 된 가용영역 설정 및 악의적인 사용을 방지하기 위해 public/ private 대역으로 나누어 설정.
- 172.31.0.0/16 : EMR용 vpc로 public 대역으로 구성.

<input type="checkbox"/>	Name	서브넷 ID	상태	VPC	IPv4 CIDR	사용 가능한 IPv	IPv6 CIDR	가용 영역
<input checked="" type="checkbox"/>	EMR-a	subnet-0bc3d04b0b2ae806	available	vpc-0c259d49eca20f5c0 EMR-VPC	172.31.0.0/20	4091	-	ap-northeast-2a
<input type="checkbox"/>	EMR-b	subnet-03be05adcbd300cdd	available	vpc-0c259d49eca20f5c0 EMR-VPC	172.31.16.0/20	4091	-	ap-northeast-2b
<input type="checkbox"/>	EMR-c	subnet-0228a398909d253ef	available	vpc-0c259d49eca20f5c0 EMR-VPC	172.31.32.0/20	4091	-	ap-northeast-2c
<input type="checkbox"/>	Project-Crawling	subnet-08c3d5f827972b684	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.0.0/24	250	-	ap-northeast-2a
<input type="checkbox"/>	Project-Private1	subnet-01b98c9ff94ba182	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.11.0/24	251	-	ap-northeast-2a
<input type="checkbox"/>	Project-Private2	subnet-035575517cfd0f5	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.12.0/24	251	-	ap-northeast-2c
<input type="checkbox"/>	Project-Public1	subnet-0a64a4b48f26e1bb9	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.1.0/24	248	-	ap-northeast-2a
<input type="checkbox"/>	Project-Public2	subnet-0738537b7000290a0	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.2.0/24	251	-	ap-northeast-2c
<input type="checkbox"/>	Project-RDS1	subnet-0714523034c6ea81a	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.21.0/24	251	-	ap-northeast-2a
<input type="checkbox"/>	Project-RDS2	subnet-00c26f3b049e4daac	available	vpc-0b3f667b25d5f8b51 Project-VPC	10.0.22.0/24	250	-	ap-northeast-2c

<그림 33> Subnet

- 10.0.0.0/24 대역은 크롤링을 하기 위한 Subnet.
- 10.0.1.0/24 및 10.0.2.0/24는 NAT Gateway 및 Eip 사용으로 private subnet의 원격접속을 위해 사용.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

- 10.0.11.0/24, 10.0.12.0/24 was 대역의 tomcat 서비스를 사용.
- 10.0.21.0/24, 10.0.22.0/24 RDS 대역의 mysql 서비스를 사용.
- EMR 서비스를 사용하기 위한 Public Subnet으로 설정 및 10.0.11.0/24, 10.0.12.0/24 was 대역의 tomcat 서비스를 사용..


Public Routing Table rtb-09e4384201096d964 3개의 서브넷 예 vpc-0b3f667b25d5f8b51 ... 672			
라우팅 테이블: rtb-09e4384201096d964			
요약	라우팅	서브넷 연결	Edge Associations
라우팅 편집			
보기 모든 라우팅 ▼			
대상	대상	상태	전파됨
10.0.0.0/16	local	active	아니요
0.0.0.0/0	igw-0266ef61638111509	active	아니요

<그림 34> Public Routing table

요약	라우팅	서브넷 연결	Edge Associations	라우팅 전파
서브넷 연결 편집				
서브넷 ID	IPv4 CIDR	IPv6 CIDR		
subnet-0a64a4b48f26e1b...	10.0.1.0/24	-		
subnet-08c3d5f827972b6...	10.0.0.0/24	-		
subnet-0738537b7000290...	10.0.2.0/24	-		

<그림 35> Public Routing table subnet

- Public Routing table은 트래픽이 외부로 바로 나갈 수 있어야 하기 때문에 인터넷 게이트웨이를 따로 설정하였으며 10.0.0.0/16에서 public subnet만 구성.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

Name

라우팅 테이블 ID

명시적으로 다음과 연결

Edge associations

기본

VPC ID

소유자

Private Routing Table

rtb-057320230b1bdb6f9

4개의 서브넷

-

아니요

vpc-0b3f667b25d5f8b51 | ...

672130635098

라우팅 테이블: rtb-057320230b1bdb6f9

요약

라우팅

서브넷 연결

Edge Associations

라우팅 전파

태그

라우팅 편집

보기

모든 라우팅

대상	대상	상태	전파됨
10.0.0.0/16	local	active	아니요
0.0.0.0/0	nat-0425d6f9c5d999221	blackhole	아니요

<그림 36> Private Routing table

요약

라우팅

서브넷 연결

Edge Associations

라우팅 전파

태그

서브넷 연결 편집

서브넷 ID	IPv4 CIDR	IPv6 CIDR
subnet-01b98c9ffb94ba18...	10.0.11.0/24	-
subnet-00c26f3b049e4da...	10.0.22.0/24	-
subnet-035575517cffad0f...	10.0.12.0/24	-
subnet-0714523034c6ea8...	10.0.21.0/24	-

<그림 37> Private Routing table subnet

- Private Routing table은 트래픽이 외부로 나갈 수 없도록 해야하기 때문에 NAT gateway를 설정을 진행, 또한 public subnet의 원격접속을 통해 필요 sw 설치 및 설정을 해야 하기 때문에 10.0.0.0/16에서 private subnet만 구성.

EMR Routing Table

rtb-08c9103daf379d5a8

-

-

예

vpc-0c259d49

라우팅 테이블: rtb-08c9103daf379d5a8

요약

라우팅

서브넷 연결

Edge Associations

라우팅 전파

태그


라우팅 편집

보기

모든 라우팅

대상	대상	상태
172.31.0.0/16	local	active
0.0.0.0/0	igw-0d0dc3c01673e156a	active

<그림 38> EMR Routing table

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	


- EMR Routing table은 기본적인 Public으로 설정.

3) 보안그룹, ACL 설정

 acl-0e5b091589a9... 7개의 서브넷 예 vpc-0b3f667b25d5f8b51 Project-VPC					
보기 모든 규칙 ▼					
규칙 #	유형	프로토콜	포트 범위	소스	허용/거부
100	SSH (22)	TCP (6)	22	121.140.73.126/32	ALLOW
101	HTTP (80)	TCP (6)	80	0.0.0.0/0	ALLOW
102	사용자 지정 TCP 규칙	TCP (6)	8009	0.0.0.0/0	ALLOW
103	MySQL/Aurora (3306)	TCP (6)	3306	0.0.0.0/0	ALLOW
*	모두 트래픽	모두	모두	0.0.0.0/0	DENY


<그림 39> ACL

- 관리자만 22번 포트 사용하며 Enduser는 3tier를 통해 RDS까지 설정을 해야 하므로 관련된 포트 개방하였고, 이후 다른 포트에 대해 접속을 방지해야 하므로 마지막 규칙에는 모든 트래픽 거부로 설정.

 acl-070689b403ba... 3개의 서브넷 예 vpc-0c259d49eca20f5c0 EMR-VPC 672130635098					
네트워크 ACL: acl-070689b403ba09eb4					
세부 정보 인바운드 규칙 아웃바운드 규칙 서브넷 연결 태그					
인바운드 규칙 편집					
보기 모든 규칙 ▼					
규칙 #	유형	프로토콜	포트 범위	소스	허용/거부
100	모두 트래픽	모두	모두	0.0.0.0/0	ALLOW
*	모두 트래픽	모두	모두	0.0.0.0/0	DENY

<그림 40> EMR ACL

- EMR을 위한 기본 vpc로 퍼블릭 환경으로 작업을 해야하기 때문에 기본적인 ACL으로 설정.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

Name	보안 그룹 ID	보안 그룹 이름	VPC ID	설명	소유자	인바운드 규칙
-	sg-007dff247859bf5cc	ex-elb-sg	vpc-0b3f667b25d5f8b51	ex-elb-sg	672130635098	1 권한 항목

sg-007dff247859bf5cc - ex-elb-sg

세부 정보 | **인바운드 규칙** | 아웃바운드 규칙 | 태그

인바운드 규칙			
유형	프로토콜	포트 범위	소스
HTTP	TCP	80	0.0.0.0/0

<그림 41> ex-elb-sg

- EX-ELB-SG :Enduser기준 Internet gateway에서 가장 먼저 진입하는 ELB로써 ALB는 보안그룹을 설정해야 하며, Inbound는 80포트 source는 전체로 설정.


sg-05d528d597c9861f6 - Public-sg

세부 정보 | **인바운드 규칙** | 아웃바운드 규칙 | 태그

인바운드 규칙				
유형	프로토콜	포트 범위	소스	설명 - 선택 사항
HTTP	TCP	80	sg-007dff247859bf5cc (ex-elb-sg)	-
SSH	TCP	22	121.140.73.126/32	-

<그림 42> public-sg

- Public-SG [web] : 위의 EX-elb를 통해 진입이 가능하지만 기본적으로public subnet으로 Enduser가 진입이 가능하지만 트래픽의 분산을 위해 ex-elb-sg를 기준으로 설정을 하였습니다. port 22는 관리자만 진입이 가능하도록 관리자의 ip로 설정.
- IN-ELB-SG : web 내부에서 동작하는 ELB로써 NLB를 사용하게 되며, NLB의 주체는 ip로써 대상그룹으로만 확인을 하게되므로 보안그룹을 따로 설정하지 않음.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

인바운드 규칙					인바운드 규칙 편집
유형	프로토콜	포트 범위	소스	설명 - 선택 사항	
SSH	TCP	22	0.0.0.0/0	-	
사용자 지정 TCP	TCP	8009	10.0.11.0/24	-	
사용자 지정 TCP	TCP	8009	10.0.12.0/24	-	
사용자 지정 TCP	TCP	8009	sg-05d528d597c9861f6 (Public-sg)	-	

<그림 43> Private-sg

- Private-SG [was] : 개념은 in-elb-sg[nlb]에서 들어오는 트래픽에 대해서만 Security Group 설정을 하면 되지만, web 대역[public]의 ip가 유지됩니다. 그렇기 위해서는 was 대역[private] 입장에서는 web ip를 허용해야합니다. 즉, IN-ELB-SG의 대상 그룹 추가 및 private-sg의 대상을 추가 설정.
- 연동된 포트는 8009로 설정.


sg-091c112464eb77990 - rds-launch-wizard					
세부 정보	인바운드 규칙	아웃바운드 규칙	태그		
인바운드 규칙					인바운드 규칙 편집
유형	프로토콜	포트 범위	소스	설명 - 선택 사항	
MYSQL/Aurora	TCP	3306	sg-0467dae49f4d0229f (Private-sg)	-	

<그림 44> RDS-sg

- DB-SG : Enduser에게 제공하는 곳으로 db에 접근은 private-sg만 가능하도록 설정하였습니다. 연동된 포트는 3306으로 설정.

<input checked="" type="checkbox"/>	-	sg-0fd636b2814c6e2bc	Crawling-sg	vpc-0b3f667b25d5f8b51	Crawling-sg	672130635098	1 권한 항목
sg-0fd636b2814c6e2bc - Crawling-sg							
세부 정보	인바운드 규칙	아웃바운드 규칙	태그				
인바운드 규칙							인바운드 규칙 편집
유형	프로토콜	포트 범위	소스	설명 - 선택 사항			
HTTP	TCP	80	0.0.0.0/0	-			

<그림 45> Crawling-sg

 주)SL정보 INFORMATION	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

- Crawl-SG : 퍼블릭 대역으로 자동화 된 설정을 바탕으로 동작하므로 port 80만 허용.

4) Public 설정

- 방화벽의 80번 포트를 개방.

```
workers.tomcat_home=/usr/share/tomcat8
workers.java_home=/usr/lib/jvm/java-8-openjdk-amd64

worker.list=tomcat8
worker.tomcat8.port = 8009
worker.tomcat8.host = NLB-bfa411d460737c28.elb.ap-northeast-2.amazonaws.com
worker.tomcat8.type = ajp13
worker.tomcat8.lbfactor = 1
```


<그림 46> Worker

- Tomcat과의 연동을 위해 libapche2-mod-jk 설치후 worker 파일을 생성 후 jk.conf파일에 읽도록 적용.

```
ServerAdmin webmaster@localhost
DocumentRoot /var/lib/tomcat8/webapps/ROOT
SetEnvIF Request_URI "/*.html" no-jk
JkMount /*.jsp tomcat8
DocumentRoot /var/www/html/
```

<그림 47> apache2-000-default

- html이면 Apache에서, jsp이면 tomcat에서 파일을 처리하도록 000-default 파일을 수정.
- Apache를 자동시작을 활성화.
- Public-AMI 이미지 파일을 생성.

 주)SL정보 INFORMATION	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

5) Private 설정

- 방화벽의 8009포트를 개방.

```

    <SSLHostConfig>
      <Certificate certificateKeyFile="conf/localhost-rsa-key.pem"
        certificateFile="conf/localhost-rsa-cert.pem"
        certificateChainFile="conf/localhost-rsa-chain.pem"
        type="RSA" />
    </SSLHostConfig>
  </Connector>
-->

<!-- Define an AJP 1.3 Connector on port 8009 -->
<Connector port="8009" protocol="AJP/1.3" redirectPort="8443" />


<!-- An Engine represents the entry point (within Catalina) that processes
every request. The Engine implementation for Tomcat stand alone
analyzes the HTTP headers included with the request, and passes them
on to the appropriate Host (virtual host).
Documentation at /docs/config/engine.html -->

<!-- You should set jvmRoute to support load-balancing via AJP ie :
<Engine name="Catalina" defaultHost="localhost" jvmRoute="jvm1">
-->
<Engine name="Catalina" defaultHost="localhost">
"/etc/tomcat8/server.xml" 167L, 7511C 117,0-1 72%

```

<그림 48> Private 설정

- 연동을 위해 libapche2-mod-jk를 설치 후, server.xml 수정를을 수정하여 8009포트를 시작하게 설정.
- 웹 페이지를 설정한 경로에 수정.
- Tomcat의 자동시작을 활성화.
- Private-AMI 이미지 파일을 생성.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

2.9. ETC

1) Auto Scaling

<input type="checkbox"/>	Name	AMI 이름	AMI ID
<input type="checkbox"/>	Public-AMI	Apache-AMI	ami-0741526e3a09a873a
<input type="checkbox"/>	Private_AMI	Tomcat-AMI	ami-0b97bd7770f022478
<input type="checkbox"/>		Crawling	ami-0c303dbc88db13a61

<그림 49> AMI

필터:

<input type="checkbox"/>	이름	AMI ID	인스턴스 유형	스팟 가격	생성 시간
<input type="checkbox"/>	Public-as	ami-0741526e...	t2.micro		2020년 6월 23일 오후 4시 30분 ...
<input type="checkbox"/>	Private-as	ami-0b97bd77...	t2.micro		2020년 6월 23일 오후 2시 39분 ...

<그림 50> AS 시작 구성

- Public-as 시작 구성 : AMI : Public-AMI, 자원 : t2.micro, 보안그룹 : Public-sg
- Private-as 시작 구성: AMI : Private-AMI, 자원 : t2.micro, 보안그룹 : Private-sg









필터:

✕


<input type="checkbox"/>	이름	시작 구성 / 템플릿	인스턴스	목표 용량	최소	최대	가용 영역
<input type="checkbox"/>	Private-as	Private-as	2 ⓘ	1	1	2	ap-northeast-2a, ap-northea...
<input type="checkbox"/>	Public-as	Public-as	2 ⓘ	1	1	2	ap-northeast-2a, ap-northea...

<그림 51> As

- Public-as 최소 1, 최대 2 as 보안그룹 Public-sg
- Private-as 최소 1, 최대 2 as 보안그룹 Private-sg

<input type="checkbox"/>		i-020436af866a01161	 terminated	-	Project-Key	
<input type="checkbox"/>		i-02c3cc947eebd0408	 terminated	-	Project-Key	
<input type="checkbox"/>	Public2	i-04c8f5d768aa2707b	 terminated	-	Project-Key	
<input type="checkbox"/>	Public1	i-0983016265f486b80	 running	-	Project-Key	10.0.1.213
<input type="checkbox"/>		i-0789cd6ac35a21778	 terminated	-	Project-Key	
<input type="checkbox"/>	Private2	i-0931565a91afb0c32	 terminated	-	Project-Key	
<input type="checkbox"/>		i-082bab34fd35cd8d9	 terminated	-	Project-Key	
<input type="checkbox"/>	Private1	i-0aa9f11b29988cf56	 running	-	Project-Key	10.0.11.133

<그림 52> As 확인

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

2) ELB 설정

Q 태그 및 속성별 필터 또는 키워드별 검색					
<input type="checkbox"/> 이름	DNS 이름	상태	VPC ID	가용 영역	유형
<input type="checkbox"/> ex-elb	ex-elb-1640004473.ap-northeast-2.elb.amazonaws.com	active	vpc-0b3f667b25d5f8b51	ap-northeast-2c, ap-northeast-2a	application
<input type="checkbox"/> Inn-elb	Inn-elb-ab00e25cdd1c3787.elb.ap-northeast-2.amazonaws.com	active	vpc-0b3f667b25d5f8b51	ap-northeast-2a, ap-northeast-2c	network

<그림 53> ELB

In-elb 8009 TCP ip Inn-elb vpc-0b3f667b25d5f8b51

대상 그룹: In-elb

설명 대상 상태 검사 모니터링 태그

로드 밸런서는 등록 프로세스가 완료되고 대상이 초기 상태 검사를 통과하자마자 새로 등록된 대상에 대한 라우팅 요청을 시작합니다. 대상에 대해

[편집](#)

등록된 대상

IP 주소	포트	가용 영역	상태	설명
10.0.12.184	8009	ap-northeast-2c	healthy	이 대
10.0.11.95	8009	ap-northeast-2a	healthy	이 대

가용 영역

가용 영역	대상 개수
ap-northeast-2a	1
ap-northeast-2c	1

<그림 54> NLB

- NLB 설정 : as Private Subnet 그룹의 인스턴스 TCP 8009
- NLB 대상그룹 : as Private Subnet 그룹의 인스턴스 TCP 8009

ex-elb 80 HTTP instance ex-elb vpc-0b3f667b25d5f8b51

대상 그룹: ex-elb

설명 대상 상태 검사 모니터링 태그

로드 밸런서는 등록 프로세스가 완료되고 대상이 초기 상태 검사를 통과하자마자 새로 등록된 대상에 대한 라우팅 요청을 시작합니다. 경우 대상을 등록 취소할 수 있습니다.

[편집](#)

등록된 대상


인스턴스 ID	이름	포트	가용 영역	상태	설명
i-0983016265f486b80	Public1	80	ap-northeast-2a	healthy	이 대
i-020436af866a01161	Public2	80	ap-northeast-2c	healthy	이 대

가용 영역

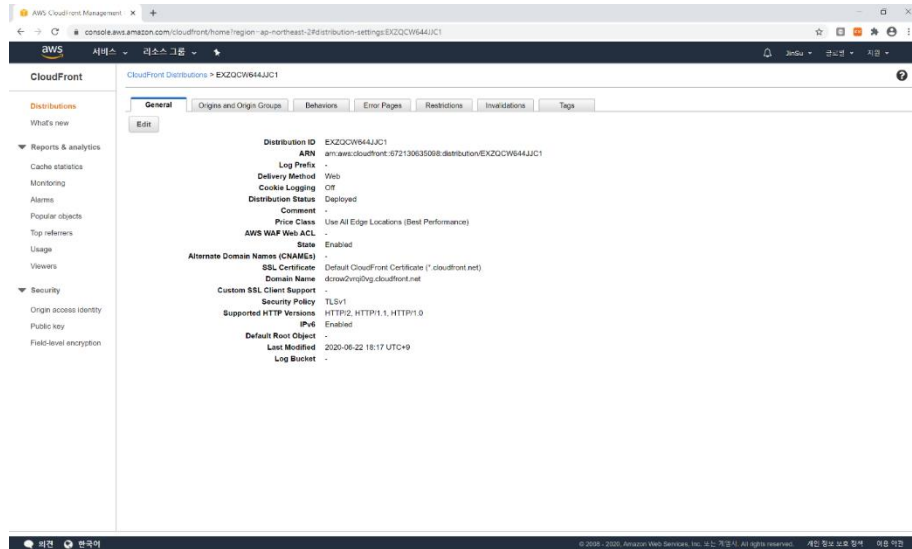
가용 영역	대상 개수
ap-northeast-2a	1
ap-northeast-2c	1

<그림 55> ALB

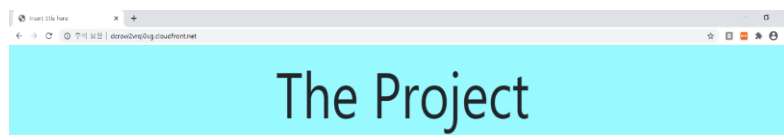
- ALB 설정 : as Private Subnet 그룹의 인스턴스 TCP 8009
- ALB 보안 그룹 : ex-elb-sg
- ALB 대상그룹 : as Private Subnet 그룹의 인스턴스 TCP 8009

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

3) CloudFront




<그림 56> CloudFront



트랜드 보기 특가 상품

<그림 57> CloudFront

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

3. 프로젝트 결과

3.8. 테스트


- 테스트 기준
 - 현 학원의 Compute Node (CPU 4, RAM 8G) 기준으로 진행.
 - 테스트 시간 00:00 ~ 05:00 : Crawling 을 통한 데이터 수집.
 - 트리거 작동 조건 : 00:00 ~ 04:00 Compute Node 상의 50% 이상의 자원이 남아 있을 때 비상시 자원 (50%)를 제외하고 사용.
 - 만약 04:00 시가 되어도 남은 자원이 없으면, Public 100% 사용.
 - On Premise 의 인스턴스 및 EC2 는 보안그룹이 오직 80 포트만 열려있고, 접근 가능한 키 페어를 가지고 있지 않음. (접근 불가 .
 - 05:00 : EMR 을 통한 데이터 분석 후, Lambda 를 통한 RDS 로의 데이터 이동.
 - 웹 서비스는 상시 운영.
- 정시 테스트

```

root@controller:~
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total) | 4   | 8102      | 256      |
| compute | (used_now) | 0   | 512       | 0         |
| compute | (used_max) | 0   | 0          | 0         |
+-----+-----+-----+-----+-----+
[root@controller ~]# date
2020. 06. 22. (월) 23:59:57 KST
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total) | 4   | 8102      | 256      |
| compute | (used_now) | 2   | 4307      | 0         |
| compute | (used_max) | 2   | 3795      | 10        |
| compute | ab70628506f0441ab7857fe67e73088a | 2   | 3795      | 10        |
+-----+-----+-----+-----+-----+
[root@controller ~]# date
2020. 06. 23. (화) 00:01:07 KST

```

<그림 58> 정시 테스트 1

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

New EC2 Experience
Tell us what you think

인스턴스 시작 연결 작업

EC2 대시보드 New
이벤트 New
태그
보고서
제한
▼ 인스턴스
인스턴스
인스턴스 유형
시작 템플릿
스팟 요청
Savings Plans
예약 인스턴스
전용 호스트 New

태그 및 속성별 필터 또는 키워드별 검색

Name	인스턴스 ID	인스턴스 상태	IPv4 퍼블릭 IP	키 이름	프라이빗 IP 주소
<input type="checkbox"/>	i-00e4b1fc6b3ed947b	running	13.125.197.17		10.0.0.172
<input type="checkbox"/>	i-01647a9a1a21c17f0	running	52.79.205.222		10.0.0.124
<input type="checkbox"/>	i-028b864bef3cb380	running	13.209.83.138		10.0.0.137
<input type="checkbox"/>	i-042e855444e15f6a5	running	3.34.192.226		10.0.0.92
<input type="checkbox"/>	i-0652dad34a404c2a	running	3.34.135.114		10.0.0.31
<input type="checkbox"/>	i-065d5b414b03ebca0	terminated	-	Project-Key	
<input type="checkbox"/>	i-06b646d2b71ca7dce	running	52.78.156.207		10.0.0.16
<input type="checkbox"/>	i-0b33b932d36544e05	running	13.124.224.104		10.0.0.126
<input type="checkbox"/>	i-0bbcb8901edf85403	running	13.124.103.145		10.0.0.8
<input type="checkbox"/>	i-0d1ba4ceb3ac755933	running	52.79.249.157		10.0.0.161
<input type="checkbox"/>	i-0eac162cde0c5c92d	running	13.124.240.164		10.0.0.250

<그림 59> 정시 테스트 2

New EC2 Experience
Tell us what you think

인스턴스 시작 연결 작업

EC2 대시보드 New
이벤트 New
태그
보고서
제한
▼ 인스턴스
인스턴스
인스턴스 유형
시작 템플릿
스팟 요청
Savings Plans
예약 인스턴스
전용 호스트 New


태그 및 속성별 필터 또는 키워드별 검색

Name	인스턴스 ID	인스턴스 상태	IPv4 퍼블릭 IP	키 이름	프라이빗 IP 주소
<input type="checkbox"/>	i-00e4b1fc6b3ed947b	terminated	-		
<input type="checkbox"/>	i-01647a9a1a21c17f0	terminated	-		
<input type="checkbox"/>	i-028b864bef3cb380	terminated	-		
<input type="checkbox"/>	i-042e855444e15f6a5	terminated	-		
<input type="checkbox"/>	i-0652dad34a404c2a	terminated	-		
<input type="checkbox"/>	i-065d5b414b03ebca0	terminated	-	Project-Key	
<input type="checkbox"/>	i-06b646d2b71ca7dce	terminated	-		
<input type="checkbox"/>	i-0b33b932d36544e05	terminated	-		
<input type="checkbox"/>	i-0bbcb8901edf85403	terminated	-		
<input type="checkbox"/>	i-0d1ba4ceb3ac755933	terminated	-		
<input type="checkbox"/>	i-0eac162cde0c5c92d	terminated	-		

<그림 60> 정시 테스트 3

작동시간	On Premise 상의 사용가능 CPU	On Premise 상의 사용가능 RAM (G)	Public 상의 지원자원	생성되는 EC2 수
2020.06.23.00:00	4 * 0.5 = 2	8 * 0.5 = 4	2/ 4G	2.micro * 10

<표 9> 정시 테스트

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

· 불특정시간 테스트

```

root@controller:~
[root@controller ~]# date
2020. 06. 24. (수) 23:59:33 KST
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 3    | 4819      | 0         |
| compute | (used_max)                            | 3    | 4307      | 11        |
| compute | ab70628506f0441ab7857fe67e73088a    | 3    | 4307      | 11        |
+-----+-----+-----+-----+-----+
[root@controller ~]# date
2020. 06. 25. (목) 00:30:33 KST
You have new mail in /var/spool/mail/root
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 3    | 4819      | 0         |
| compute | (used_max)                            | 3    | 4307      | 11        |
| compute | ab70628506f0441ab7857fe67e73088a    | 3    | 4307      | 11        |
+-----+-----+-----+-----+-----+
[root@controller ~]# date
2020. 06. 25. (목) 00:59:06 KST
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 3    | 4819      | 0         |
| compute | (used_max)                            | 3    | 4307      | 11        |
| compute | ab70628506f0441ab7857fe67e73088a    | 3    | 4307      | 11        |
+-----+-----+-----+-----+-----+
[root@controller ~]# date
2020. 06. 25. (목) 01:17:04 KST
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 3    | 4819      | 0         |
| compute | (used_max)                            | 3    | 4307      | 11        |
| compute | ab70628506f0441ab7857fe67e73088a    | 3    | 4307      | 11        |
+-----+-----+-----+-----+-----+

```


<그림 61> 불특정 시간 테스트 1

```

[root@controller ~]# date
2020. 06. 25. (목) 01:59:42 KST
You have new mail in /var/spool/mail/root
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 1    | 1024      | 0         |
| compute | (used_max)                            | 1    | 512       | 1         |
| compute | ab70628506f0441ab7857fe67e73088a    | 1    | 512       | 1         |
+-----+-----+-----+-----+-----+
[root@controller ~]# openstack host show compute
+-----+-----+-----+-----+-----+
| Host   | Project                                | CPU | Memory MB | Disk GB |
+-----+-----+-----+-----+-----+
| compute | (total)                               | 4    | 8102      | 256      |
| compute | (used_now)                            | 2    | 4563      | 0         |
| compute | (used_max)                            | 2    | 4051      | 11        |
| compute | ab70628506f0441ab7857fe67e73088a    | 2    | 4051      | 11        |
+-----+-----+-----+-----+-----+

```

<그림 62> 불특정 시간 테스트 2

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

작동시간	On Premise 상의 사용가능 CPU	On Premise 상의 사용가능 RAM (G)	Public 상의 지원자원	생성되는 EC2 수
2020.06.25.02:00	3 * 0.5 = 1	7 * 0.5 = 3.5	3/ 4.5G	2.micro * 10

<표 10> 불특정 시간 테스트

New EC2 Experience
Tell us what you think

인스턴스 시작 연전 작업

EC2 대시보드 New

이벤트 New

태그

보고서

제한

▼ 인스턴스

인스턴스

인스턴스 유형

시작 템플릿

스팟 요청

Savings Plans

예약 인스턴스

전용 호스트 New

용량 예약

▼ 이미지

AMI

변들 작업

태그 및 속성별 필터 또는 키워드별 검색

Name	인스턴스 ID	인스턴스 상태	IPv4 퍼블릭 IP	키 이름	프라이빗 IP 주소
	i-0005c1c692816bb47	terminated	-		
	i-00d62835857e1e149	terminated	-		
	i-0143b122bbce6f62d8	terminated	-		
	i-01b777032b3a6b8a3	terminated	-		Project-Key
	i-01c2b7415d7e79dd9	terminated	-		
	i-0289db6c27d2c3e45	terminated	-		
	i-02f0ddafa3607401	terminated	-		
	i-05f562d6ff6f14fa3	terminated	-		
	i-07116386af8b998d2	terminated	-		
	i-07b83bd65286f6cfc	terminated	-		
	i-09366671183f3f57e	terminated	-		
	i-0a23e7fd899dd24d1	terminated	-		
	i-0b736ba49b8871c40	terminated	-		
	i-0ca824778685ba39c	terminated	-		
	i-0cf0e555b379983b	terminated	-		Project-Key
	i-0d0f0bed8b225cb20	terminated	-		

<그림 63> 불특정 시간 테스트 3

S3, Lambda, EMR, RDS test

project-datas

개요

검색: 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 자우려면 Esc 키를 누릅니다.

업로드 + 폴더 만들기 다운로드 작업


아시아 태평양(서울) 2

보기 1 대상 9

이름	마지막 수정	크기	스토리지 클래스
20200616	--	--	--
20200618	--	--	--
20200619	--	--	--
20200620	--	--	--
20200621	--	--	--
20200622	--	--	--
20200623	--	--	--
20200624	--	--	--
20200625	--	--	--

보기 1 대상 9

<그림 64> S3 data

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

Amazon S3 > project-datas > data > 20200625 > 11st

project-datas

개요

🔍 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.

업로드


+ 폴더 만들기

다운로드

작업 ▼

<input type="checkbox"/> 이름 ▼
<input type="checkbox"/> 가구.json
<input type="checkbox"/> 가전.json
<input type="checkbox"/> 남성의류.json
<input type="checkbox"/> 뷰티.json
<input type="checkbox"/> 스포츠,레저,자동차,취미.json
<input type="checkbox"/> 식품.json
<input type="checkbox"/> 여성의류.json
<input type="checkbox"/> 출산,유아.json
<input type="checkbox"/> 패션잡화.json

<그림 65> S3 Input




	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	


Amazon S3 > project-datas > output

project-datas


개요

🔍 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.

 업로드
  폴더 만들기
 다운로드
  작업 ▼

<input type="checkbox"/>	이름 ▼
<input type="checkbox"/>	 dataset_hadoop.json

<그림 66> S3 Output

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

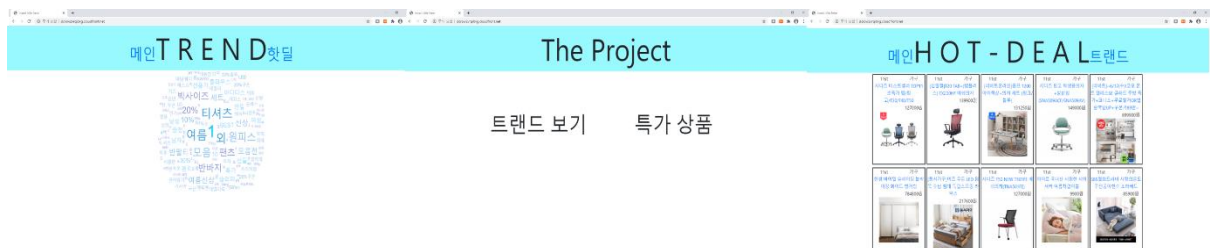
```

root@ip-10-0-1-218: ~
6194 | 추가 2인 | 1 | | http://image.gmarket.co.kr/service_image/2020/06/16/2020
6195 | 추가 극한 | 1 | | | 55200 | http:/
6196 | 추가 인경 | 1 | | /item.gmarket.co.kr/Item?goodscode=169542407&ver=637280161127101316
6197 | 추가 인종 | 1 | |
6198 | 추가 동은 | 2 | | | 1791 | 2 | 8 | 링크통 베이비 섹터세제 2200ml 6개
6199 | 후드 | 4 | | | http://image.gmarket.co.kr/service_image/2020/06/16/2020
6200 | 후드집업 | 1 | | | 30900 | http:/
6201 | 후레쉬 | 1 | | 0616155712433140_0_0.jpg
6202 | 후아 | 1 | | /item.gmarket.co.kr/Item?goodscode=1527100464&ver=637280161127101316
6203 | 후회할 | 2 | | | 1792 | 2 | 8 | 머리가 마시는 아인슈타인 컨디션 200mlx24입
6204 | 후지 | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/16/2020
6205 | 후채오리 | 1 | | | 34930 | http:/
6206 | 후라후프 | 1 | | 0616155712433140_0_0.jpg
6207 | 후라후프 | 2 | | | /item.gmarket.co.kr/Item?goodscode=1756347589&ver=637280161127101316
6208 | 후드지 | 1 | | | 1793 | 2 | 8 | 30%증복+10%카드 오가닉담 내외/우주복 BEST
6209 | 후이열 | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/16/2020
6210 | 후션 | 2 | | | 0616155712433140_0_0.jpg
6211 | 후러 | 12 | | | /item.gmarket.co.kr/Item?goodscode=1808607588&ver=637280161127101316
6212 | 후라X아티스트플라보 | 2 | | | 1794 | 2 | 8 | 리베로 기저귀 밴드형/편타형 4팩 + 사은품
6213 | 후라신상 | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/16/2020
6214 | 후대 | 3 | | | 0616155712433140_0_0.jpg
6215 | 후대용 | 13 | | | /item.gmarket.co.kr/Item?goodscode=747669469&ver=637280161127101316
6216 | 후대용고개기 | 1 | | | 1795 | 2 | 9 | 14K Gold-Pin 귀걸이 외 유물리 모음
6217 | 후대용산통기 | 2 | | | http://image.gmarket.co.kr/service_image/2020/06/12/2020
6218 | 후대용가방 | 1 | | | /item.gmarket.co.kr/Item?goodscode=767588368&ver=637280158481721017
6219 | 후단 | 1 | | | 1796 | 2 | 9 | [결산세일 30%증복] 제이에스티나X이유
6220 | 후단 | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/12/2020
6221 | 후마플로 | 1 | | | /item.gmarket.co.kr/Item?goodscode=1223881788&ver=637280158481721017
6222 | 후수제까지 | 1 | | | 1797 | 2 | 9 | [나폴] 남녀공용 선글라스
6223 | 후해리~ | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/12/2020
6224 | 후냉방 | 1 | | | 0612103258992023_0_0.jpg
6225 | 후단 | 1 | | | /item.gmarket.co.kr/Item?goodscode=644502406&ver=637280158481721017
6226 | 후노개 | 1 | | | 1798 | 2 | 9 | [30%증복] 로즈몽 상반기 결산세일
6227 | 후노개집성복 | 1 | | | http://image.gmarket.co.kr/service_image/2020/06/12/2020
6228 | 후말리아 | 2 | | |
6229 | 후비스캐스자물 | 1 | | |
6230 | 후아플문산 | 2 | | |
6231 | 후즈플로 | 1 | | |
6232 | 후 | 4 | | |
6233 | 후장버드 | 2 | | |
6234 | 후문 | 1 | | |
6235 | 후색 | 1 | | |
6236 | 후색 | 5 | | |
6237 | 후스터 | 1 | | |
6238 | 후색 | 1 | | |
6238 rows in set (0.01 sec)

mysql>

```

<그림 67> RDS



<그림 68> web Service


3.9. 비용산정 결과

<표 11> 비용산정 표

3.10. 결과

1) 한계

- Crawling 은 인터넷 속도 즉 트래픽의 영향을 가장 크게 받지만, 이를 계산하는 방법에 문제가 있어, 동일 환경에서 테스트 후, 통계를 작성하여 진행.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	


- 사내의 On Premise 의 평균자원의 대한량에 대한 정보가 없어, 현재 학원의 컴퓨터들을 기준으로 진행하였음.
- 하이브리드 클라우드를 통해, Private 의 기존 문제점들을 최소화할 수 있으나, 권한부분에서 완전히 독립적인 인프라로는 만드는 것을 불가능하였음.

2) 향후 발전방향

- Public 을 사용할 경우, 해외 사이트에 대한 Crawling 의 대해 높은 효율을 나타내었음.
- 단순히 데이터뿐만이 아닌, 하이브리드 클라우드를 이용하여 여러 서비스를 독립시킬 때, 각 장점에 맞춰 커스텀마이징이 가능하다는 것을 알게 되었고, 현재 네이버, KT, 뉴타코닉 등의 기업이 이와 같은 인프라를 운용 및 준비 중에 있는 것을 아알게 되었음.

3) 최종요약

- On Premise 환경에서의 남는 자원과 Public 에서의 자원 할당은 충분히 성공적으로 이루어 졌으며, 데이터의 독립 및 인프라의 독립 또한 충분히 가능하게 이루어졌음.
- On Premise 의 환경에 대한 기준을 확립할 방법이 없어, 이 때문에 자원과 비용을 계산하는 데에 기 문제가 다수 발생하였음, 현재의 프로젝트는 강의 시 사용한 컴퓨터를 기준으로 진행하였음.
- 해외 사이트를 크롤링해서 데이터를 수집하거나 해외 사이트에 서비스를 개시할 경우 국내에서의 운영보다 확실히 좋은 성과를 기대할 수 있었음.

	Report	Version	Last Modified	Dada Bank Project
	Final Report	3.0	2020.06.31	

4. 참고문헌

- ‘의료 빅데이터’... 4차 산업혁명 시대의 핵심 자원으로 주목 BIOTIME 2020.05 나지영 기자
- 新정치 패러다임...선거 승리하려면 ‘빅데이터’ 읽어라 2020.04 신승훈 기자
- 글래드 호텔X삼성전자X삼성카드, 고객 빅데이터 분석해 최상의 서비스 제공 아시아기자협회 2020.05.07 이주형 기자
- 스포츠 승리팀 예측에 AI·빅데이터 활용 2018.07.05
- 우리금융 AI, 빅 데이터로 ‘초개인화’ 마케팅 매일경제 2020.06.01 최승진 기자
- 향후 카드사들은 마이데이터 사업 등을 통해 빅데이터 전문회사로 성장할 것” INSIDE 2020.05 - 장명현 여신금융연구소 연구원.
- 바다·치킨·카페·맛집...빅데이터로 본 제주여행 ‘유유자적’ 파이낸셜 2020.05.22 최승훈 기자
- SK텔레콤, 빅데이터 마케팅 서비스 ‘T-Deal’ 개시 2020.03 박미영 기자
- 도 빅데이터 협의회 “플랫폼 개발, 내달까지 차질 없이 추진” 2020.06.01 경남신문 김희진 기자
- [김호준의 중소기업탐구] 중소기업의 ‘빅데이터’ 활용은 가능한가 2020.03 김호준
- 중소 10곳중 5곳 “빅데이터 꼭 필요한데 구할방법 없어요” 2019.08.22 윤진호 기자
- CIO 실패 가능성85% 빅데이터 프로젝트의 문제와 해법 2019.05.21 Andy Patrizlo