



2021 Woori Digital Academy
DAM(Digital Analytic Master) 과정
프로젝트 결과보고서
(FinancialApps ReviewAnalytics서비스 포함)



목 차



01

DAM 과정 소개
과정 내용

02

실습 및 과제 수행
서비스 구현 프로젝트 수행



DAM 과정 소개

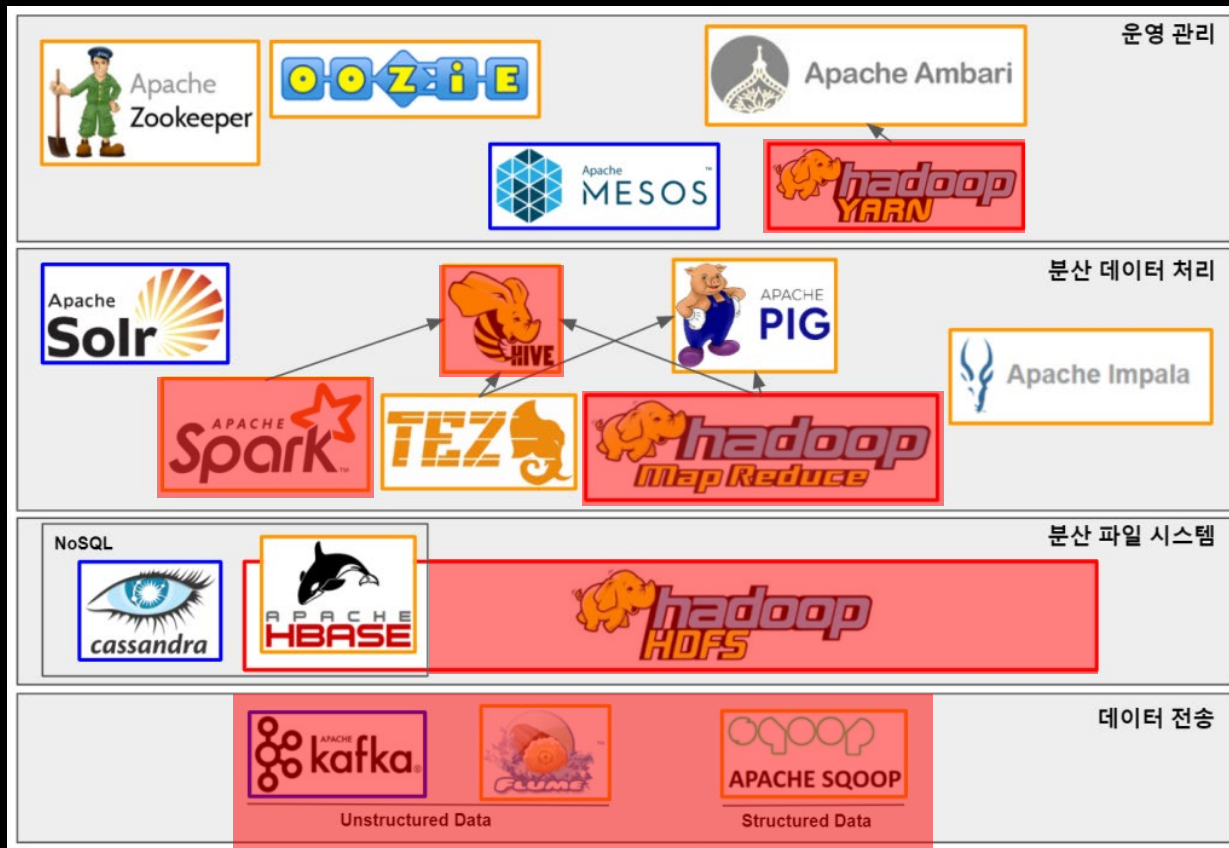
Digital Analytic Master

과정 내용



과정 내용(약식)

- ✓ Hadoop-빅데이터 분산처리
HDFS (파일시스템)
MapReduce(병렬처리)
Flume(로그수집)
Sqoop(ETL, 데이터 조회)
Hive(HDFS 데이터 조회)
- ✓ Spark-데이터 처리 엔진
Spark RDD(데이터 연산)
Spark SQL(데이터 조회)
Spark Streaming(실시간 처리)
Spark ML Lib(머신러닝 라이브러리)
- ✓ Kafka(실시간 분산 데이터 스트리밍)



MAPREDUCE
(Processing using
different languages)



HIVE & DRILL
(Analytical
SQL-on-Hadoop)



MAHOUT &
SPARK MLlib
(Machine learning)



PIG
(Scripting)



HBASE
(NoSQL Database)



ZOOKEEPER
& AMBARI
(Management
& Coordination)



SPARK (In-Memory,
Data Flow Engine)



KAFKA & STORM
(Streaming)



Spark
MLlib

SOLR & LUCENE
(Searching
& Indexing)



OOZIE
(Scheduling)



Resource
Management

YARN

Storage



Flume



Unstructured/
Semi-structured Data

Sqoop



Structured Data



실습 및 과제 수행

과정 내용을 활용한 데이터 서비스 구현
Financial Apps Review Analytics



프로젝트 목표

금융 앱 서비스 강·약점 분석

- ✓ 국내,외 금융사 앱 리뷰의 긍/부정 분류를 통한 강/약점 분석
- ✓ 오픈소스 중심으로 저비용 조건하에
기 구축된 그룹 내 서비스를 오마주하여 개발
- ✓ Hadoop의 HDFS/MapReduce를 통한
Word Count 처리 및 Rank

Demand 요약 및 Notification/Report

- ✓ 앱 리뷰 내 언급 횟수를 기반으로
사용자 요구를 카테고리 분류하고 보고서를 제공하는 서비스
- ✓ 멀티플랫폼(Android,iOS,Windows,ChromeOS,Linux) Notification

프로젝트 주제 선정 배경

핀테크 앱, 금융뱅킹앱 추월...토스·카뱅 앱 접속건
수 1·2위

한국 소비자들은 지난해 금융 앱을 약 410억회 방문했고, 일주일 평균 은행 월렛 앱은 4.6회, 핀테크 월렛 앱에는 11.7회 방문한 것으로 나타났다.

뱅크 월렛 앱은 1위가 KB스타뱅킹, 2위 NH스마트뱅킹, 3위 신한 쏠, 4위 우리은행 원터치개인뱅킹, 5위에는 신한페이판이 올랐다.

애플, 아마존에 이어 구글 등 빅테크 기업이 금융 분야로의 진출에 속도를 내고 있다.

23일 외신 등에 따르면 구글은 지난 18일(현지시간) 은행 계좌 개설과 간편송금, 생활비 관리 등의 기능이 추가된 '구글 페이'를 출시했다.

핀테크, 인터넷은행의
시장점유 증가

시중은행의 비대면 서비스 확대

빅테크 기업의 금융 진출

우리금융의
MAU 증대를 위한
비대면 서비스 활성화 전략 필요



구축 Preview

개발 Workflow

개발환경

시스템 구성도

Data Flow

비고. 그룹 내 관련 서비스



개발 Workflow





개발환경

Server Instance



Managing Source Codes



git



IDEs



colab

KMS, Wiki, Managing Project



Languages



PIMS



Database



Data Sources

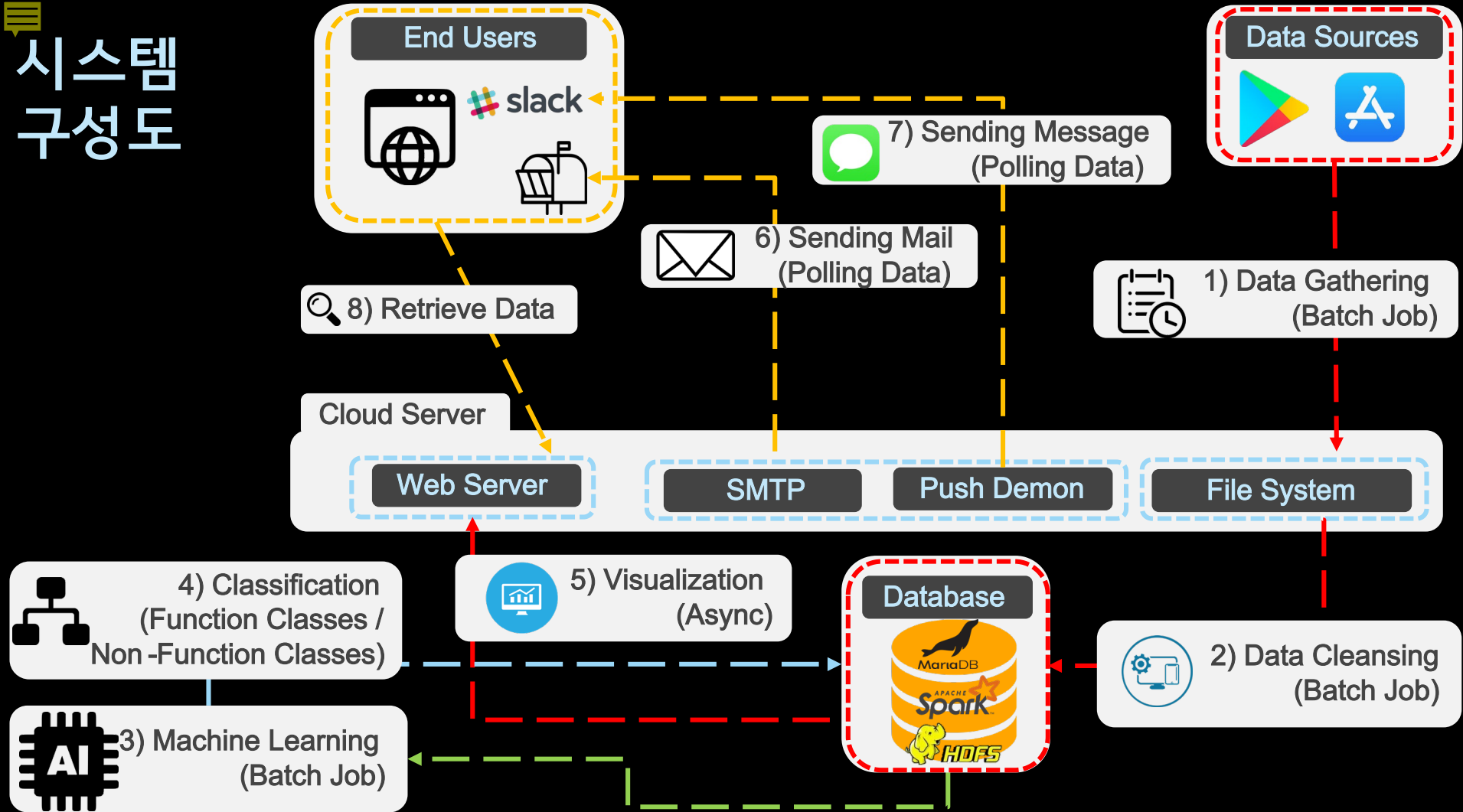


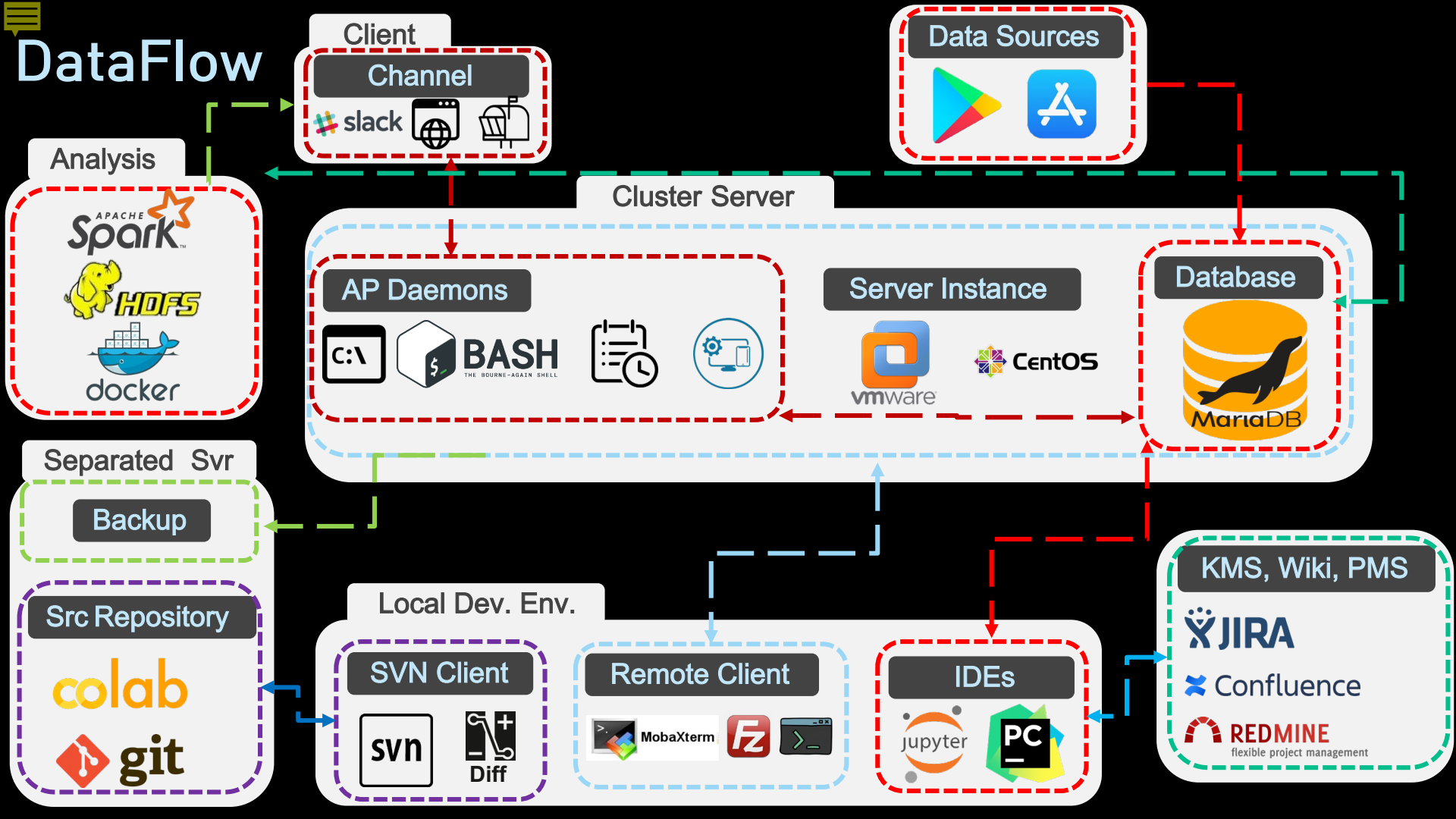
Channel





시스템 구성도





(비고)그룹 내 관련서비스(앱리뷰모니터링)

PLAYSTORE, APPSTORE ▾

은행APP

평가별점

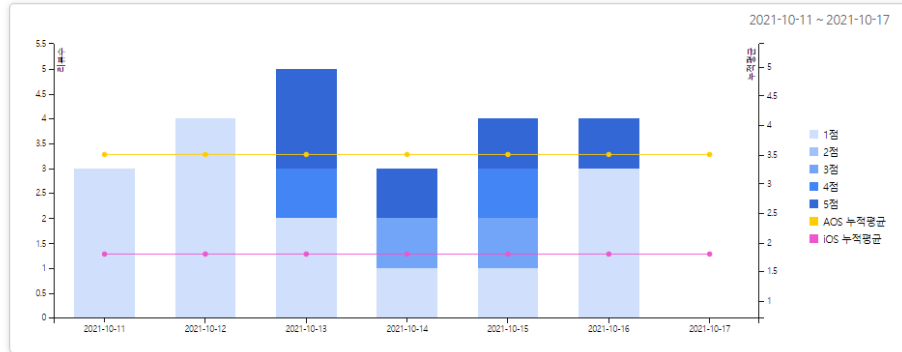
준색 입력

검색

데이터소스	은행APP	별점	리뷰제목	리뷰요약	전체 PLAYSTORE APPSTORE	전체 우리의 WON 댁팅	전체 신한 SOL	전체 KB 스타뱅킹	전체 하나원큐	전체 카카오뱅크	전체 케이뱅크	담당직원명	
PLAYSTORE	KB 스타뱅킹	★★★★☆		이틀전부터 로그인한 실패도 안되고, 재설치해도 그러네요. 은행앱이 이러면 어떻게 중요아	2021-10-18 00:00:00	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	프로프로그램	2021-10-18 00:00:00
PLAYSTORE	i-ONE Bank	★★★★★		모바일인증서로 로그인 되지 않습니다 v3 스타뱅킹 둘다 전화 저장관한 접근권한 허 했어도 재부팅 > 다시 KB스타뱅킹을 설치를 부탁합니다. 그리고 나서 KB모바일인증서가 없어졌을 경우 발급 진행해주시고, 모바일인증서가 있다면 [전체매뉴>인증센터>모바일 인증서>인증방법 등록변경>2면비밀번호]모르는 경우를 통해 KB모바일인증서 발급 부탁드립니다. 보안프로그램인 v3 mobile plus도 재설치 부탁드립니다. 안나오면 조치로 해결이 안될 경우 고객센터(1588-9999)로 전화주시면 보다 정확한 상담 도와드리겠습니다. 감사합니다.>	2021-10-18 00:00:00	안녕하세요	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	정말 귀찮	2021-10-18 00:00:00
PLAYSTORE	KB 스타뱅킹	★★★★☆		그동안 무슨일이 있었길래 앱이.... 너무 이상하겠네요. 리뷰를 남기지조차 않는데	2021-10-18 00:00:00	안녕하세요	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	이해 주셔서 감사드립니다	2021-10-18 00:00:00
PLAYSTORE	KB 스타뱅킹	★★★★☆		진짜 국민은행 두번이나 안온다 계좌발달번호 찾으려니 타행인증서 등록 하려니 로그인	2021-10-18 00:00:00	안녕하세요. ***고객님, 고객님의 비대면 금융 업무 불편하셨을 것을 대단히 진심	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00
PLAYSTORE	NH스마트뱅킹	★★★★☆		2시간전에는 어플이 들어가졌는데 지금은 들어가지질 않네요..	2021-10-18 00:00:00	안녕하세요. ***고객님, 먼저 NH스마트뱅킹 이용에 불편을 드려 죄송함	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00
PLAYSTORE	NH스마트뱅킹	★★★★☆		왜 네트웍스 오류가 뜨는지 ㅋㅋ	2021-10-18 00:00:00	안녕하세요. 고객님의, 먼저 NH스마트뱅킹 이용에 불편을 드려 죄송합니다.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00
PLAYSTORE	NH스마트뱅킹	★★★★☆		지금 접이안되요 왜그런지	2021-10-18 00:00:00	안녕하세요. ***고객님, 먼저 NH스마트뱅킹 이용에 불편을 드려 죄송함	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00
PLAYSTORE	NH스마트뱅킹	★★★★☆		뭐사야된다고요 예러다지 못시자나요? 머하는거임	2021-10-18 00:00:00	안녕하세요. 고객님의, 먼저 NH스마트뱅킹 이용에 불편을 드려 죄송합니다.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00
PLAYSTORE	NH스마트뱅킹	★★★★☆		어, 답답하네 진짜 예러코드 계속고요? 오류리겟다커요? 어플다시 깔아도 노답이네요? 이정도	2021-10-18 00:00:00	안녕하세요. 고객님의, 먼저 NH스마트뱅킹 이용에 불편을 드려 죄송합니다.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2021-10-18 00:00:00

◀ 1 2 3 4 5 ... 86 ▶

우리WON뱅킹 별점 트렌드차트



타행 앱 별점 차이



2. Review Text Contents

CONFIDENTIAL



(비고)그룹 내 관련서비스(앱리뷰모니터링)

3. Internal Email Push

모든 항목 읽지 않음

기준: 날짜 ▾ 새로운 항목 ↓

◀ 오늘

BigData App Review
[등록알림] 우리WON뱅킹 APP 리뷰(★★★★★)
오전 1:02

◀ 어제

BigData App Review
[등록알림] 우리WON뱅킹 APP 리뷰(★★★★★)
(월) 오후 8:02

2021-10-18 (월) 오후 8:02

BigData App Review <BigDataAppReview@woorifg.com>
[등록알림] 우리WON뱅킹 APP 리뷰(★★★★★)

CONFIDENTIAL

PLAYSTORE
2021-10-18
★★★★★

우리은행 '우리 WON 뱅킹' 앱(APP)에 대한 PlayStore(AppStore)에 등록된 평가내용 공유합니다.(★★★★★)
잘쓰다가 갑자기 공동인증이 사라졌다고 다시만들라고합니다 불안해서 못쓰겠네요



※ 본 메일은 '우리 WON 뱅킹' 앱(APP) 리뷰 뿐만 아니라, 고객의 단순 불만이 포함되어 있습니다.

메일발송문의 : 80700-4437



구축 서비스 소개

WebService / Push / Batch / Etc



Report(평점 변화 추이, 긍/부정 Words)
<http://kaydash.iptime.org>

Report(평점 변화 추이, 긍/부정 Words)

<http://kaydash.iptime.org>

네 긍정 KEYWORD-WORD CLOUD



Updated yesterday at 11:59 PM

부정 KEYWORD-WORD CLOUD



Updated yesterday at 11:59 PM

앱리뷰 분석 시스템

종합분석

III. 비교분석

리뷰분석

④ 실시간 리뷰분석

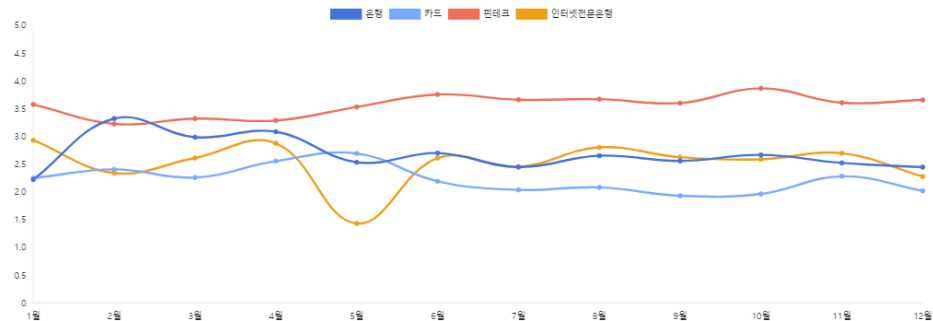
종합 분석

4대 은행(4), 카드(5), 핀테크(2), 인터넷전문은행(2) 종합분석

리뷰 평점 추이

부야별

2021년도



Updated yesterday at 11:59 PM

카드 Web Service

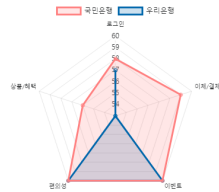
Differential Analytics

우리은행 vs 국민은행
우리카드 vs 현대카드

금융사별 분석

조회 대상 : 국민은행 (분야: 은행)

키워드별 평점 분석



리뷰분석 건수 : 국민은행 (104 건) / 우리은행 (36 건)

키워드 감성 상세

키워드	최근 평가
이벤트	이벤트도다른쪽보다많.. 오후때문에회원가입이..
이체/결제	입출금카톡알림정제식.. 단축이제한번에다건이..
로그인	안녕하세요갑자기저문.. 앱트스만큼알고있고행..
기능/편의성	이벤트도다른쪽보다많.. 업데이트후에로그인이..

키워드 약점 상세

키워드	최근 평가
이벤트	회원가입에어려웁니까.. 전파너무작중나네요최..
이체/결제	방금에이불통장에입금.. 보안이슈가심각합니다..
로그인	국민은행을따날때가끔.. 앱데이트후에로그인이..
기능/편의성	국민은행을따날때가끔.. 앱데이트후에로그인이..

키워드 감성 KEYWORD-WORD CLOUD



Updated yesterday at 11:59 PM

키워드 약점 KEYWORD-WORD CLOUD

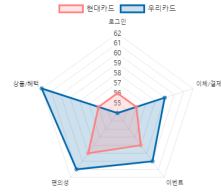


Updated yesterday at 11:59 PM

금융사별 분석

조회 대상 : 현대카드 (분야: 카드사)

키워드별 평점 분석



리뷰분석 건수 : 현대카드 (103 건) / 우리카드 (102 건)

키워드 감성 상세

키워드	최근 평가
이벤트	혜택이 다양하고 좋아요.. 좋은알인정보알아서줄..
이체/결제	스마일페이랑네이버카.. 카드사용이 편리하긴한..
로그인	앱회원가입이 쉽다비.. 인터페이스편하고세현..
기능/편의성	편리하게 사용하고있습.. 서비스최고..

키워드 약점 상세

키워드	최근 평가
이벤트	쓰레기같은 리뷰발방안.. 눈치없이자꾸놀리는헛..
이체/결제	발리다시입데이트해봐.. 앱카드결제역종업식제..
로그인	1결제후며칠지니아결.. 지문등록하고휴대폰인..
기능/편의성	앱카드결제역종업식제.. 1결제후며칠지니아결..

키워드 감성 KEYWORD-WORD CLOUD



Updated yesterday at 11:59 PM

키워드 약점 KEYWORD-WORD CLOUD



Updated yesterday at 11:59 PM

Web Service

Text Analytics

앱리뷰 분석 시스템

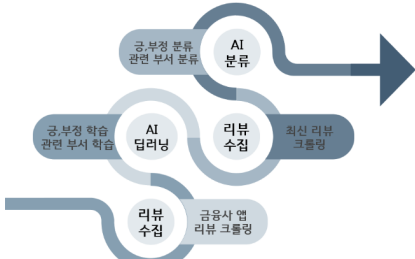
총괄조회
④ 종합분석
상세조회
③ 비교분석
② 리뷰분석
① 실시간 리뷰분석

리뷰 분석기

리뷰 입력

보안이 믿음직스러워요 조회

[리뷰 분석 프로세스]



```
graph LR; A[리뷰 수집] --> B[금융사 앱 리뷰 크롤링]; B --> C[리뷰 수집]; C --> D[AI 딥러닝]; D --> E[금, 부정 학습 관련 부서 학습]; E --> F[AI 분류]; F --> G[금, 부정 분류 관련 부서 분류]; G --> H[최신 리뷰 크롤링]; H --> I[리뷰 수집]; I --> J[리뷰 수집];
```



리뷰 분석

리뷰 내용

보안이 믿음직스러워요

금/부정 분류 결과

입력하신 리뷰는 **긍정** 리뷰입니다.

부서별 업무 연관도

입력하신 리뷰는 **디지털전략부** 와 연관이 높습니다

앱리뷰 분석 시스템

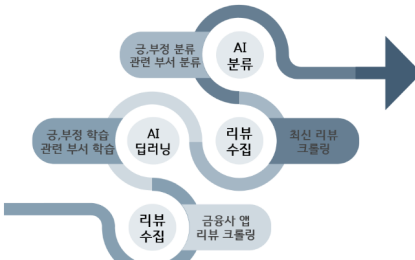
총괄조회
④ 종합분석
상세조회
③ 비교분석
② 리뷰분석
① 실시간 리뷰분석

리뷰 분석기

리뷰 입력

성능이 좋아요 조회

[리뷰 분석 프로세스]



```
graph LR; A[리뷰 수집] --> B[금융사 앱 리뷰 크롤링]; B --> C[리뷰 수집]; C --> D[AI 딥러닝]; D --> E[금, 부정 학습 관련 부서 학습]; E --> F[AI 분류]; F --> G[금, 부정 분류 관련 부서 분류]; G --> H[최신 리뷰 크롤링]; H --> I[리뷰 수집]; I --> J[리뷰 수집];
```



리뷰 분석

리뷰 내용

성능이 좋아요

금/부정 분류 결과

입력하신 리뷰는 **긍정** 리뷰입니다.

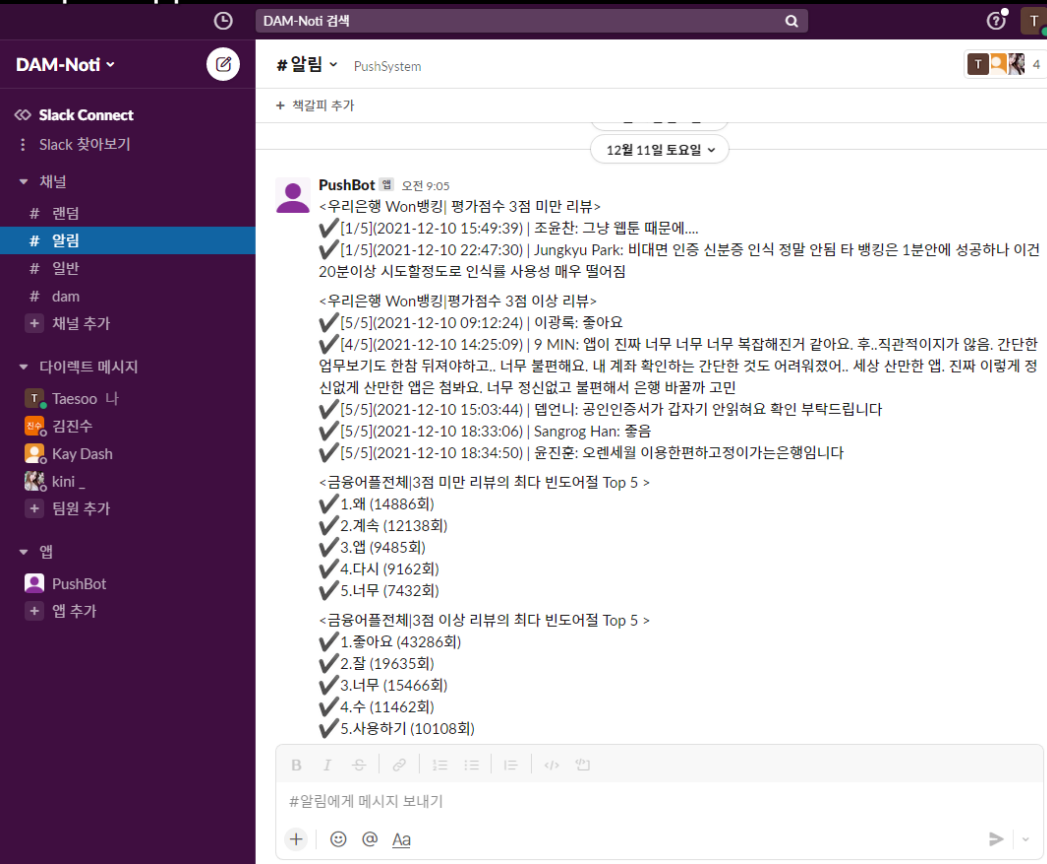
부서별 업무 연관도

입력하신 리뷰는 **디지털전략부** 와 연관이 높습니다

Push Service(Web / Mobile)

App Push(Slack, daily AM 09:00 ~ 09:05 Push)

https://app.slack.com/



DAM-Noti 검색

#알림 PushSystem

+ 채널피 추가

12월 11일 토요일

PushBot 오전 9:03

<우리은행 Won뱅크| 평가점수 3점 미만 리뷰>

- ✓[1/5](2021-12-10 15:49:39) | 조운찬: 그냥 웹툰 때문에...
- ✓[1/5](2021-12-10 22:47:30) | Jungkyu Park: 비대면 인증 신분증 인식 정말 안됨 타 뱅킹은 1분안에 성공하나 이런 20분이상 시도할정도로 인식을 사용성 매우 떨어짐

<우리은행 Won뱅크| 평가점수 3점 이상 리뷰>

- ✓[5/5](2021-12-10 09:12:24) | 이광록: 좋아요
- ✓[4/5](2021-12-10 14:25:09) | 9 MIN: 앱이 진짜 너무 너무 너무 복잡해진거 같아요. 후.. 직관적이지가 않음. 간단한 업무보기도 한참 되려야하고.. 너무 불편해요. 내 계좌 확인하는 간단한 것도 어려워졌어.. 세상 산만한 앱. 진짜 이렇게 정신없게 산만한 앱은 첨봐요. 너무 정신없고 불편해서 은행 바꿀까 고민
- ✓[5/5](2021-12-10 15:03:44) | 텀연나: 공인인증서가 갑자기 안읽혀요 확인 부탁드립니다
- ✓[5/5](2021-12-10 18:33:06) | Sangrog Han: 좋음
- ✓[5/5](2021-12-10 18:34:50) | 윤진준: 오펜세월 이용한편하고정기가는은행입니다

<금융어플전체3점 미만 리뷰의 최다 빈도어절 Top 5 >

- ✓1.왜 (14886회)
- ✓2.계속 (12138회)
- ✓3.앱 (9485회)
- ✓4.다시 (9162회)
- ✓5.너무 (7432회)

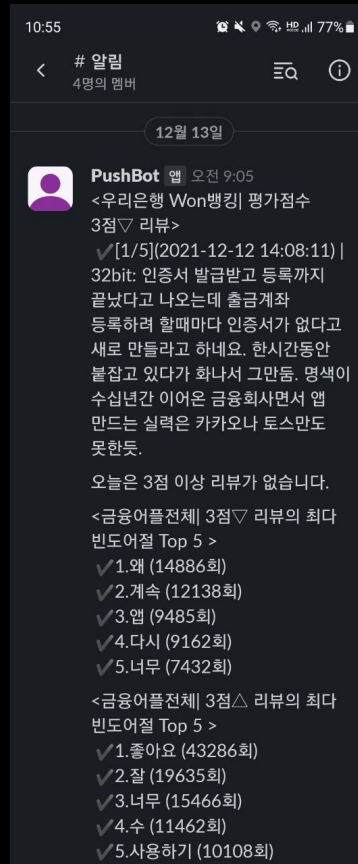
<금융어플전체3점 이상 리뷰의 최다 빈도어절 Top 5 >

- ✓1.좋아요 (43286회)
- ✓2.잘 (19635회)
- ✓3.너무 (15466회)
- ✓4.수 (11462회)
- ✓5.사용하기 (10108회)

B I @ # | 1:1 | 1:1 | 1:1 | 1:1

#알림에게 메시지 보내기

+ @ Aa



10:55

#알림 4명의 멤버

12월 13일

PushBot 앱 오전 9:05

<우리은행 Won뱅크| 평가점수 3점 미만 리뷰>

- ✓[1/5](2021-12-12 14:08:11) | 32bit: 인증서 발급받고 등록까지 끝났다고 나오는데 출금계좌 등록하려 할때마다 인증서가 없다고 새로 만들라고 하네요. 한시간동안 붙잡고 있다가 화나서 그만둬. 명색이 수심년간 이어온 금융회사면서 앱 만드는 실력은 카카오톡 토스만도 못한듯.

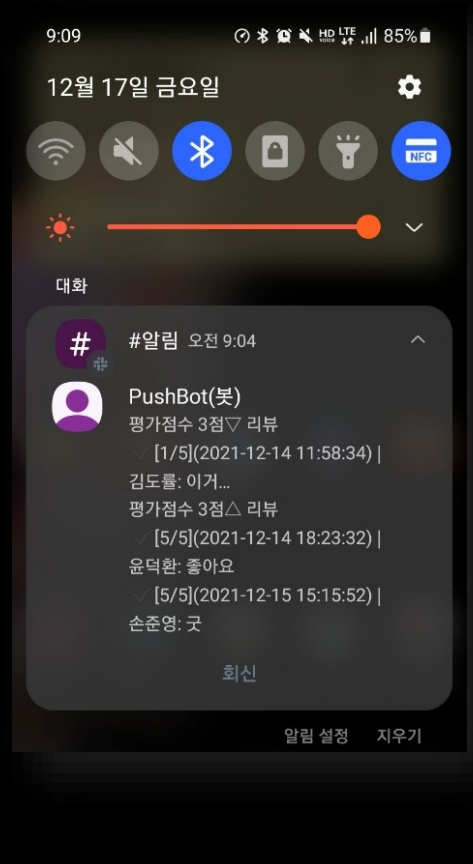
오늘은 3점 이상 리뷰가 없습니다.

<금융어플전체 3점 미만 리뷰의 최다 빈도어절 Top 5 >

- ✓1.왜 (14886회)
- ✓2.계속 (12138회)
- ✓3.앱 (9485회)
- ✓4.다시 (9162회)
- ✓5.너무 (7432회)

<금융어플전체 3점 이상 리뷰의 최다 빈도어절 Top 5 >

- ✓1.좋아요 (43286회)
- ✓2.잘 (19635회)
- ✓3.너무 (15466회)
- ✓4.수 (11462회)
- ✓5.사용하기 (10108회)



9:09

12월 17일 금요일

대화

#알림 오전 9:04

PushBot(吳) 평가점수 3점 미만 리뷰

- ✓[1/5](2021-12-14 11:58:34) | 김도률: 이거...
- 평가점수 3점 미만 리뷰
- ✓[5/5](2021-12-14 18:23:32) | 윤덕환: 좋아요
- ✓[5/5](2021-12-15 15:15:52) | 손준영: 굿

회신

알림 설정 지우기

Data Lineage

Level	Task(For)	Data(Do)	Source(From)	Destination(To)	Parallel Proc(How Many)	ETL Type(How to)	ETL Method(What)	Note.(Memo)	Loop
0	데이터 스키마 정의	Business	Pseudo Code	Logical Objects	1	Abstraction to Knowledge	Realize	Create Table	Once
0	테이블 파티셔닝	Split	Table	Partitioning Table	1	-	-	for Performance	Once
0	인덱스 생성	B*Tree	RootNode	LeafNodes	1	-	-	for Performance	Once
0	모니터링 환경 구축	Performance Data	Resource Usage	WebService	1	text To Visual	Data Gathering	Prepare for DataLineage	Once
1	웹 데이터 수집	Web Documents	GooglePlayStore	ServerStorage	20	Text To File	Web Crawling	Incremental Data	Daily
2	원천 테이블 적재	ReviewTextFile	ServerStorage	DBMS	20	File To DBMS	DBMS Injection	RDBMS(MySQL)	Daily
3	데이터 클렌징	Partitioned Tables	Raw Table	Raw Table	20	-	-	Cleansing Duplicated Data	Daily
4	요약 테이블 생성	Partitioned Tables	Raw Table	Summary Table	1	-	-	Summary(Score/Perf/MainDept)	Daily
5	보고서 테이블 생성	Partitioned Tables	Summary Table	Report Table	1	-	-	Report(Score/Perf/MainDept)	Daily
6	테이블 최적화	Statistics	Raw Data	Metadata	1	-	-	27 Tables	Daily
7	WordCloud 이미지 생성	Image Files	Data	Image	1	Data to MediaFiles	DataProcessing	Renewal Imgs.	Daily
8	Text 추출	TextFile	Data	TextFile	1	Data to Text	DataProcessing	Extract Text	Daily
9	HDFS 업로드	put HDFS	TextFile	HDFS	1	put HDFS	put HDFS	TextFile to HDFS	bit15
10	WordCount	Word Count	HDFS(input)	HDFS(output)	1	WordCount	hadoop WordCount	hdfs jar	Daily
11	word Ranking	Ranking	HDFS(output)	Data	1	HDFS to Data	Store Ranking Data	Data	Daily
12	Push Msg	Push Msg	Data	Push Msg	1	Data to Push Msg	Push	Slack	Daily

Batch TimeLine

```
[root@localhost ~]# crontab -l
5 0 * * * sh /Data/Batch/1_removeCSV.log > /Data/Batch/1_removeCSV.log 2>&1 # start At. 00:05 # 1 min
10 0 * * * sh /Data/Batch/3_getNewFiles.sh > /Data/Batch/3_getNewFiles.log 2>&1 # start At. 00:10 # 15 min
25 1 * * * sh /Data/Batch/4_loadAllFiles.log > /Data/Batch/4_loadAllFiles.log 2>&1 # start At. 01:25 # 5 min
30 1 * * * sh /Data/Batch/9_deleteDupRows.sh > /Data/Batch/9_deleteDupRows.log 2>&1 # start At. 01:30 # 5 min(No.9 is faster after others.)
35 1 * * * sh /Data/Batch/6-2_summary.sh > /Data/Batch/6-2_summary.log 2>&1 # start At. 01:35 # 10 min(parallel with 6-2,3,4)
45 1 * * * sh /Data/Batch/6-3_getPerfData.sh > /Data/Batch/6-3_getPerfData.log 2>&1 # start At. 01:45 # 10 min(parallel with 6-2,3,4)
55 1 * * * sh /Data/Batch/6-4_genWordCloud.sh > /Data/Batch/6-4_genWordCloud.log 2>&1 # start At. 01:55 # 60 min(parallel with 6-2,3,4)
00 2 * * * sh /Data/Batch/6_OPTIMIZE_TABLE.sh > /Data/Batch/6_OPTIMIZE_TABLE.log 2>&1 # start At. 03:00 # 30 min
00 4 * * * sh /Data/Batch/8_getRows.sh > /Data/Batch/8_getRows.log 2>&1 # start At. 04:00 # 1 min
30 4 * * * /IDEs/anaconda3/bin/python /Data/Batch/modelTrainer.py > /Data/Batch/modelTrainer.log 2>&1 # start At. 04:30 # 2h 10 min
00 8 * * * /IDEs/anaconda3/bin/python /Data/Batch/B01_word_count_hadoop.py > /Data/Batch/B01_word_count_hadoop.log 2>&1 # start At. 08:00 # 30 min
00 9 * * * sh /Data/Batch/A01_enqueue_avg_msgs.sh > /Data/Batch/A01_enqueue_avg_msgs.log 2>&1 # start At. 09:00 # 1 min
05 9 * * * /IDEs/anaconda3/bin/python /Data/Batch/A02_dequeue_avg_msgs.py > /Data/Batch/A02_dequeue_avg_msgs.log 2>&1 # start At. 09:05 # 1 min
```



Data Interaction Process

RDBMS -> HDFS(WordCount) -> RDBMS(Rank)

Text Cleans

```
for index, row in meta.iterrows():
    sql=''
select trim(original_content) as original_content from {tab_full_nm} where score<3.0
and trim(original_content) is not null and length(trim(original_content))>0 ''.
format(tab_full_nm=row['full_name'])
temp_data = pd.read_sql_query(sql,connection)
temp_data['original_content'] = temp_data['original_content'].str.replace(
    "[^ㄱ-ㅎㅌ-ㅣ가-힣 1234657890]", "") #remove_chars
temp_data.to_csv(''/Data/Batch/word_count/source/low/{file_name}.csv''.format(
    file_name=row['full name']), index=None, header=None,encoding='utf-8-sig')
```

Put HDFS

```
os.system('su - hdroot -c "/data/hadoop-3.3.1/bin/hdfs dfs -put
/Data/Batch/word_count/source/low/*.csv /input/low/"')
os.system('su - hdroot -c "/data/hadoop-3.3.1/bin/hdfs dfs -put
/Data/Batch/word_count/source/high/*.csv /input/high/"')
```



Ranked Data Push Process

RDBMS -> HDFS(WordCount) -> RDBMS(Rank)

Data Ranking

```
sql='''update summary.under_avg_msg set send_flag=1 where send_flag=0 and
at >= date_add(now(), interval - 72 hour)'''
cursor.execute(sql)
connection.commit()
sql='''select score,at,username,original_content,reviewid from
summary.under_avg_msg where send_flag=1 order by at asc'''
low_rvs= pd.read_sql_query(sql,connection)
sql='''update summary.under_avg_msg set send_flag=2 where send_flag=1'''
cursor.execute(sql)
connection.commit()
```

Enqueue/Dequeue Message(Push)

```
msg_txt='''<금융어플전체 | 3점△ 리뷰의 최다 빈도어절 Top 5 >
'''
i=0
for index, row in high_words.iterrows():
    i=i+1
    msg_txt=msg_txt+'''  ✓{ranking}.{words} ({count}회)
''''.format(ranking=int(i),words=row['word'],count=int(row['cnt']))

if len(high_words)>0:
    post_message(myToken, chnl, msg_txt)
print(msg_txt)
```




Monitoring for Operation

Server(Linux)

🔍 검색

시스템

개요

로그

스토리지

네트워킹

Podman Containers

계정

서비스

툴

소프트웨어 업데이트

애플리케이션

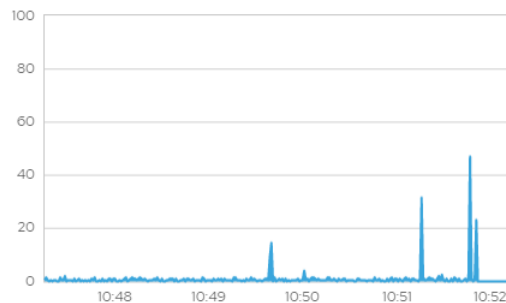
커널 덤프

터미널

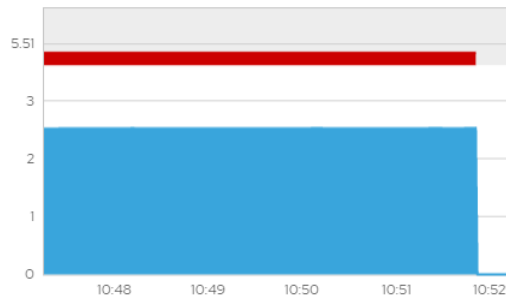
Diagnostic Reports

개요 사용 그래프

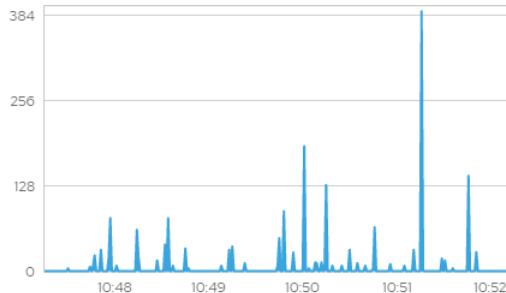
2 CPU



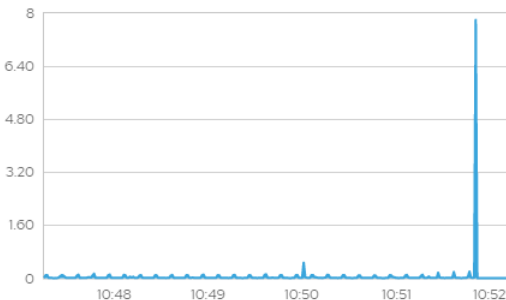
메모리 & 스왑



디스크 I/O

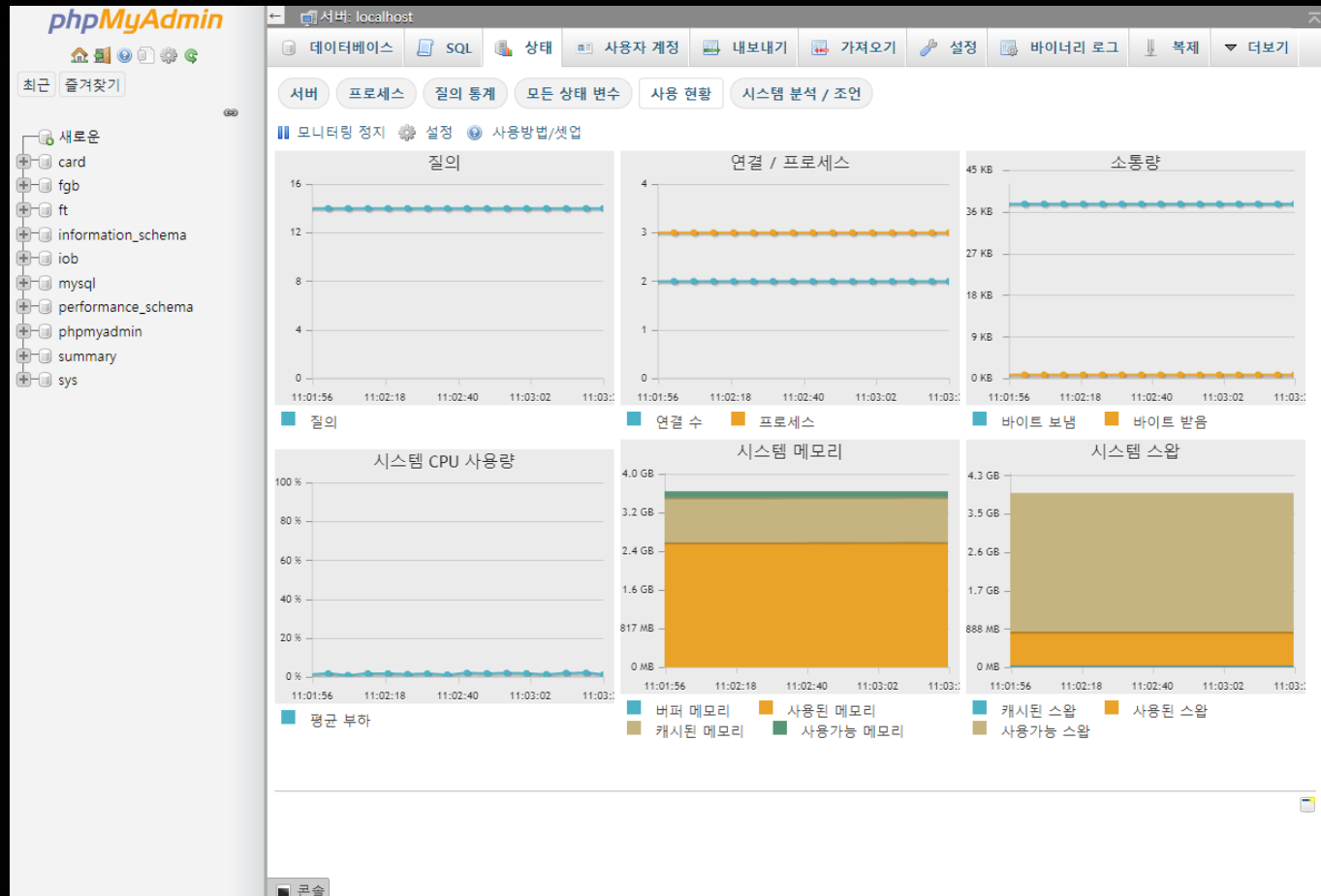


네트워크 통신량



Monitoring for Operation

DBMS(MySQL)



Monitoring for Operation

SVN



Repositories

Users

Administration

Extensions

Super Administrator (admin)

로그아웃

MAINTENANCE

Status

Server Logs

Software Updates

Statistics

Jobs

CONFIGURATION

Server Settings

Authentication

Proxy Server

Mail Server

Logging

Server Monitoring

EXTENSIONS

TeamForge

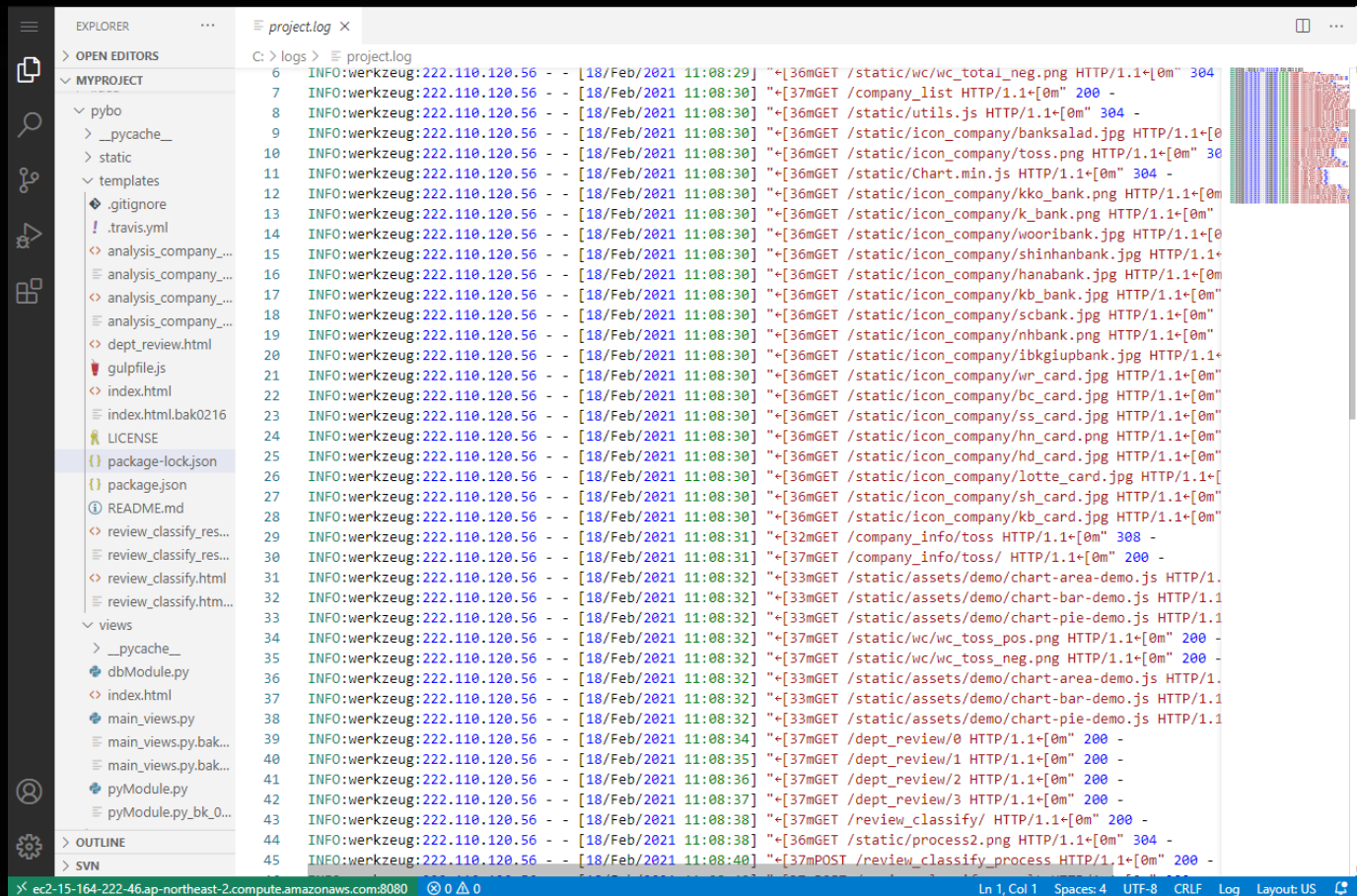
Tip: Need help configuring or managing your Subversion environment? CollabNet provides support, training and services. [Learn more on the Extensions page.](#)

정보

Software version	5.2.4-4429.59
Subversion version	1.8.19-4429.59
Running since	02/18/2021 09:51:15 KST
Repo health	Total repositories: 1 - Go to the Repositories tab to create or discover repositories.
Throughput on primary interface	0.0 B/s IN (over about 5 minutes); 0.0 B/s OUT (over about 5 minutes)
Disk Usage	as of 02/18/2021 10:53:16 KST
Used space on root volume	27.32 GB
Used space by repositories	0.0 B
Free space on repository volume	36.51 GB

Monitoring for Operation

Web Server(Flask)



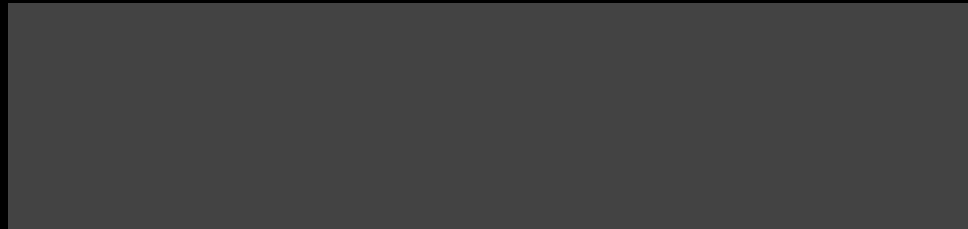
The screenshot displays the Visual Studio Code interface. On the left, the Explorer sidebar shows a project structure with folders like 'pybo' and 'templates', and files such as 'index.html' and 'package-lock.json'. The main editor area is open to 'project.log', which contains a series of log entries. Each entry follows a pattern: a timestamp (e.g., '18/Feb/2021 11:08:29'), a log level (e.g., 'INFO'), a source IP (e.g., 'werkzeug:222.110.120.56'), and a request details (e.g., '"/static/wc/wc_total_neg.png HTTP/1.1+[0m" 304 -'). The status bar at the bottom indicates the current file is 'ec2-15-164-222-46.ap-northeast-2.compute.amazonaws.com:8080' and shows various icons for window management and encoding.

```
6 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:29] "c[36mGET /static/wc/wc_total_neg.png HTTP/1.1+[0m" 304 -
7 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[37mGET /company_list HTTP/1.1+[0m" 200 -
8 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/utills.js HTTP/1.1+[0m" 304 -
9 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/banksalad.jpg HTTP/1.1+[0m" 304 -
10 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/toss.png HTTP/1.1+[0m" 304 -
11 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/Chart.min.js HTTP/1.1+[0m" 304 -
12 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/kko_bank.png HTTP/1.1+[0m" 304 -
13 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/k_bank.png HTTP/1.1+[0m" 304 -
14 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/wooribank.jpg HTTP/1.1+[0m" 304 -
15 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/shinhanbank.jpg HTTP/1.1+[0m" 304 -
16 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/hanabank.jpg HTTP/1.1+[0m" 304 -
17 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/kb_bank.jpg HTTP/1.1+[0m" 304 -
18 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/scbank.jpg HTTP/1.1+[0m" 304 -
19 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/nhbank.png HTTP/1.1+[0m" 304 -
20 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/ibkgiupbank.jpg HTTP/1.1+[0m" 304 -
21 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/wr_card.jpg HTTP/1.1+[0m" 304 -
22 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/bc_card.jpg HTTP/1.1+[0m" 304 -
23 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/ss_card.jpg HTTP/1.1+[0m" 304 -
24 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/hn_card.png HTTP/1.1+[0m" 304 -
25 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/hd_card.jpg HTTP/1.1+[0m" 304 -
26 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/lotte_card.jpg HTTP/1.1+[0m" 304 -
27 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/sh_card.jpg HTTP/1.1+[0m" 304 -
28 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:30] "c[36mGET /static/icon_company/kb_card.jpg HTTP/1.1+[0m" 304 -
29 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:31] "c[32mGET /company_info/toss HTTP/1.1+[0m" 308 -
30 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:31] "c[37mGET /company_info/toss/ HTTP/1.1+[0m" 200 -
31 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-area-demo.js HTTP/1.1+[0m" 200 -
32 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-bar-demo.js HTTP/1.1+[0m" 200 -
33 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-pie-demo.js HTTP/1.1+[0m" 200 -
34 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[37mGET /static/wc/wc_toss_pos.png HTTP/1.1+[0m" 200 -
35 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[37mGET /static/wc/wc_toss_neg.png HTTP/1.1+[0m" 200 -
36 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-area-demo.js HTTP/1.1+[0m" 200 -
37 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-bar-demo.js HTTP/1.1+[0m" 200 -
38 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:32] "c[33mGET /static/assets/demo/chart-pie-demo.js HTTP/1.1+[0m" 200 -
39 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:34] "c[37mGET /dept_review/0 HTTP/1.1+[0m" 200 -
40 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:35] "c[37mGET /dept_review/1 HTTP/1.1+[0m" 200 -
41 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:36] "c[37mGET /dept_review/2 HTTP/1.1+[0m" 200 -
42 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:37] "c[37mGET /dept_review/3 HTTP/1.1+[0m" 200 -
43 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:38] "c[37mGET /review_classify/ HTTP/1.1+[0m" 200 -
44 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:38] "c[36mGET /static/process2.png HTTP/1.1+[0m" 304 -
45 INFO:werkzeug:222.110.120.56 - - [18/Feb/2021 11:08:40] "c[37mPOST /review_classify_process HTTP/1.1+[0m" 200 -
```



기대 효과

프로젝트 기대효과 및 활용방안





기대 효과

- 입력 한 금융사의 **기능/비기능적** 강/약점 정보 제공
- 타 금융사의 **장·단점**을 **분석**하고 **비교**하여 우리금융의 **개선방향** 도출

활용방안

개선점 자동 알림

우리WON뱅킹 외 우리 금융 앱(optional)
개선의견 **자동 알림**으로 서비스 품질 개선
-> 기존 데이터 전처리 과정 축약을 통한
불만사항 인지 시간 감소(최대 1일)

리뷰 이력 조회

멀티 플랫폼 서비스로
내/외부망에 국한되지 않는
알림 이력 조회

신규 서비스 수요 예측

고객 **선호 서비스** 학습
-> 신사업 성공여부 예측



Fin.