



빅데이터 구축 사례

오태현

paranmul@gmail.com

paranmul@inbrein.com

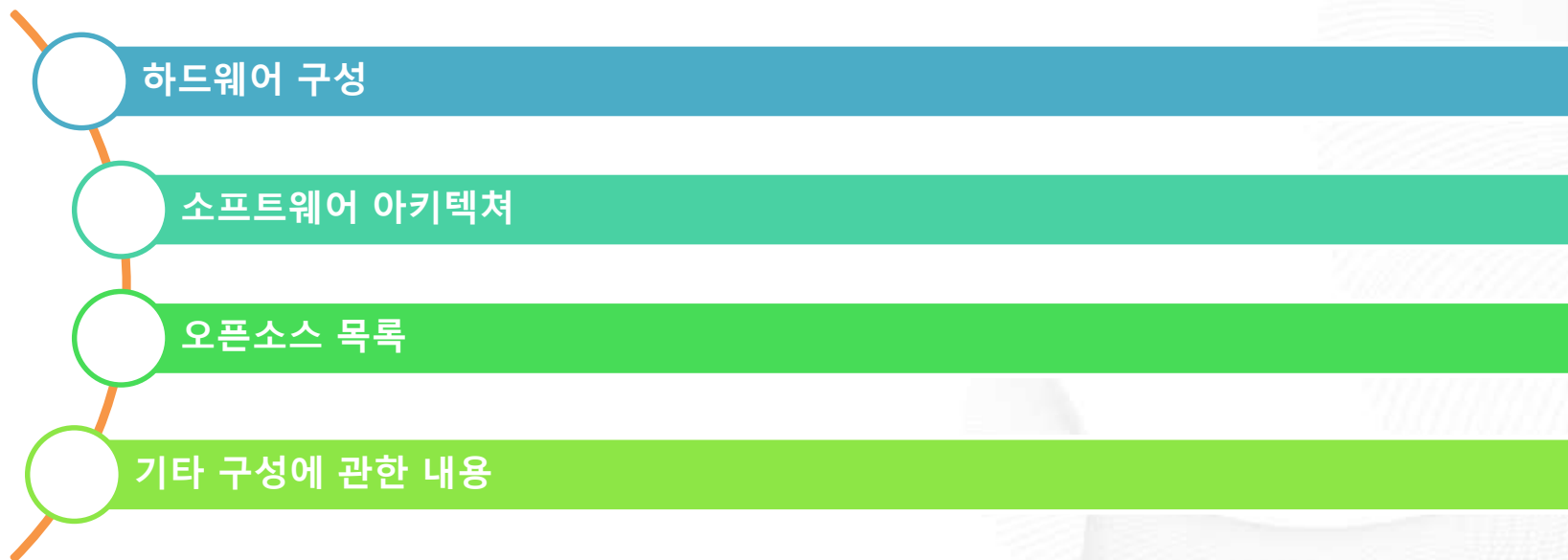
| I | N | B | R | E | I | N |



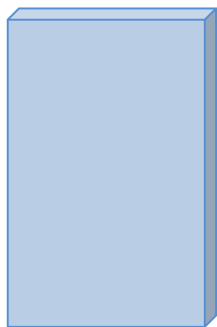
❖개요

- 빅데이터 구축 사례 소개

❖주요 진행 내용

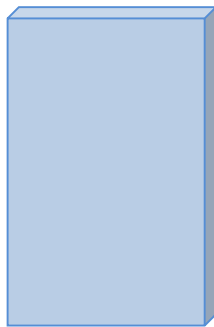


하드웨어 구성



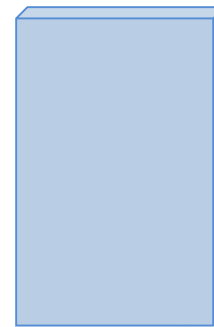
x 3

Master server



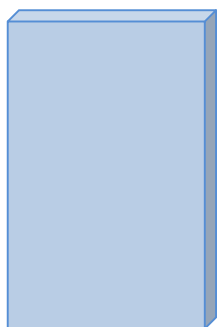
x 7

Hadoop
Hbase



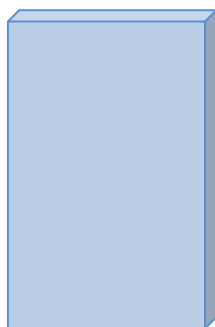
x 2

Kafka



x 4

Storm

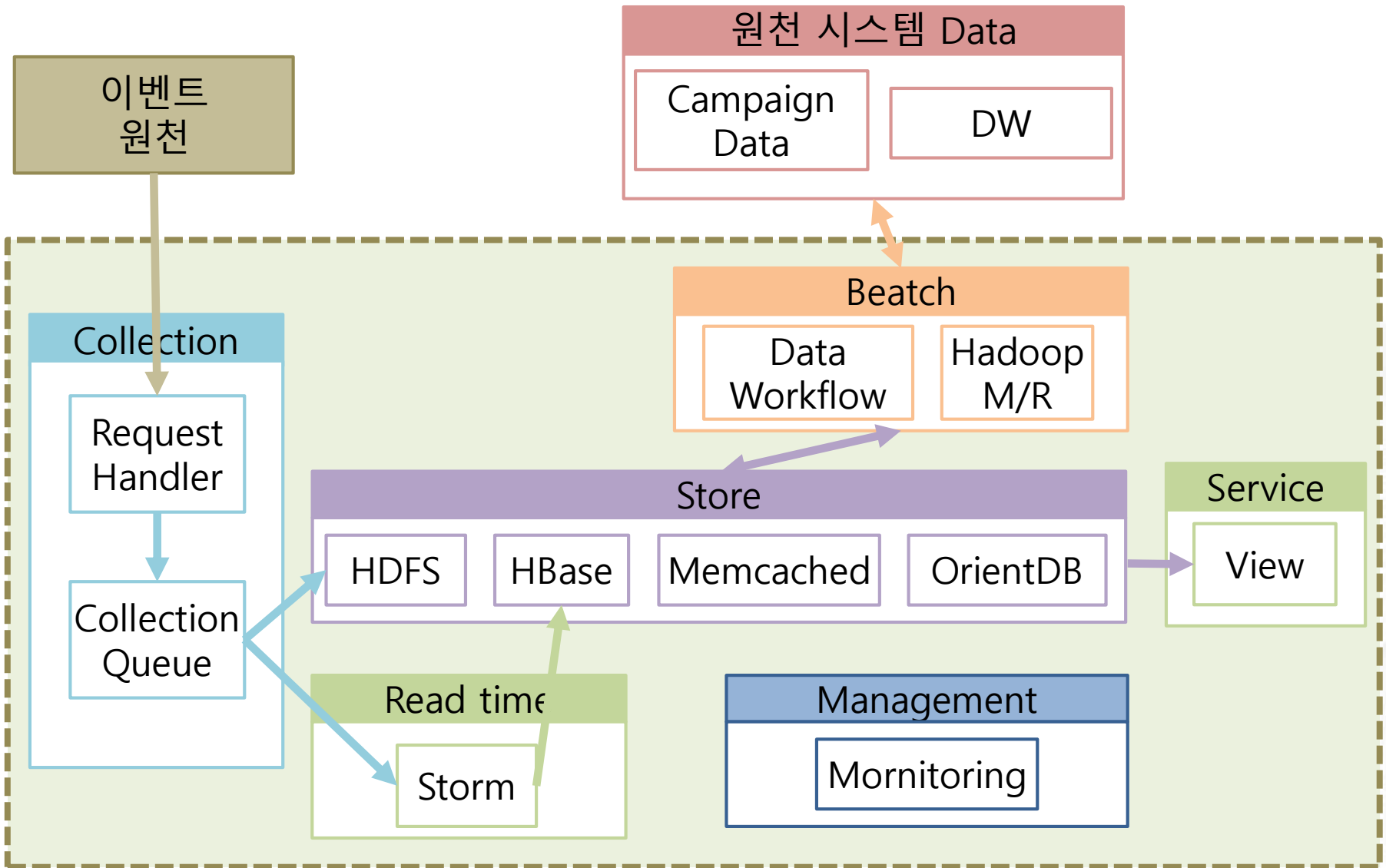


x 2

Backup node



소프트웨어 아키텍처



사용된 오픈소스 목록

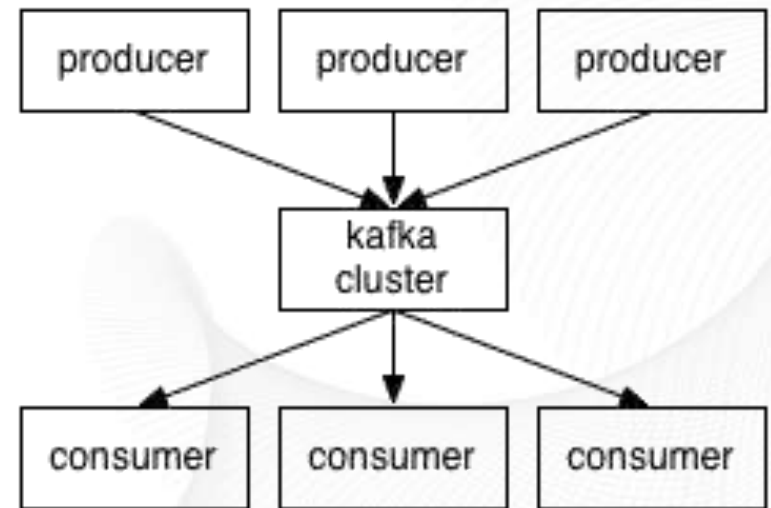


- Verte.x
- Flume
- Kafka *
- Camus
- Storm *
- Hive
- Hadoop
- jCascalog *
- Sqoop *
- Hbase
- Azkaban *
- Graphite *
- Ganglia *
- Memcached
- MariaDB





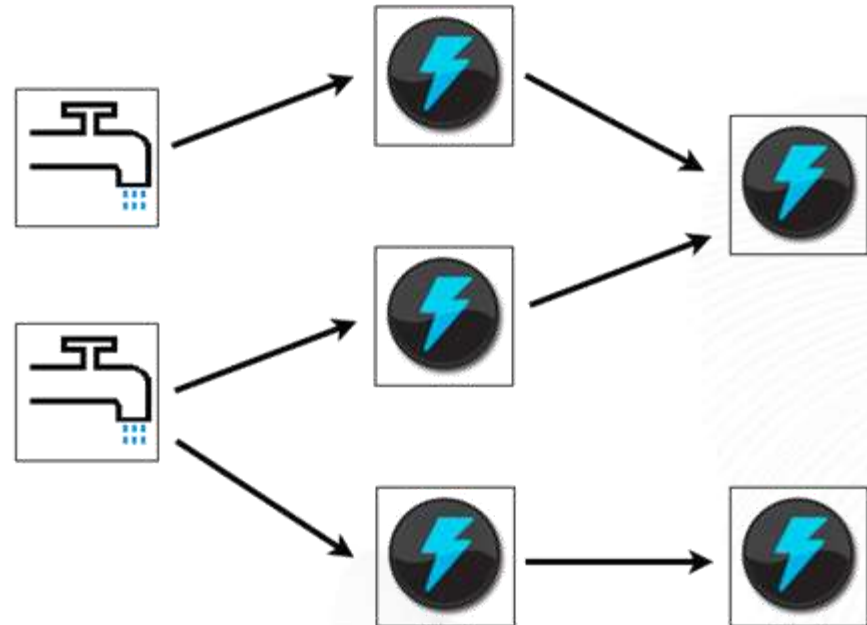
- 데이터를 분산, 파티션, 복제 커밋 로그서버
- Topic
- 초당 18000개, 로그 하나 크기 1k,
- 1초에 17M(한 서버당), 1분에 840M, 1시간에 50G
- 두대의 Kafka server를 사용



Storm



- Realtime
- Nimbus
- Topology
- Spout
- Bolt
- -> Hbase





- Hadoop 의 mapreduce 코딩을 쉽게 하는 오픈소스
- Cascading을 java로 사용할 수 있게함

mbrid	pcid
101	001
102	002
103	003
104	004
105	005
106	006

```
Tap inputTab = new Hfs(new TextDelimited(false, "\\t"), inputPath);  
Tap outputTap = new Hfs(new TextDelimited(false, "\\t" , "\\n"), outputPath);  
  
Subquery base = new Subquery("!mbr-id-long", "!pcid-cnt")  
    .predicate(inputTab, "!mbr-id", "!pcid-cnt")  
    .predicate(new ConvertData(), "!mbr-id").out("!mbr-id-long");  
  
Api.execute(outputTap, base);
```



Sqoop



- RDB 와 HDFS사이에 데이터를 전달 하는 오픈소스
- pwd, mssql, mariaDB, Oracle, netezza, mysql, hbase


```
import
-D      transaction.isolation.level=READ_UNCOMMITTED
--num-mappers  1
--driver      com.microsoft.sqlserver.jdbc.SQLServerDriver
--connect      jdbc:sqlserver://ipaddress:17001;databaseName=dbname;
--username     user
--password     user
--query
[
SELECT * FROM
(
SELECT      *      FROM      tablename
) B WHERE $CONDITIONS
]
--split-by      ID
--fields-terminated-by  \t
--target-dir     /data/db
```



Azkaban



• 스케줄러



Azkaban

Moneyball Azkaban

Projects

Scheduled

Executing

History

HDFS

azkaban

1

Scheduled Flows

Flow	Project	Submitted By	First Scheduled to Run	Next Execution Time	Repeats Every	Has SLA	Action
etl-kafka	moneyball-etl-kafka	azkaban	2014-01-05 21:00:00	2014-01-09 18:00:00	1 hour(s)	false	<div>Remove Schedule</div> <div>Set SLA</div>
nploader-export-load-to-table-realtime-item-view	moneyball-ten-minutes	azkaban	2014-01-05 21:00:00	2014-01-09 18:00:00	10 minute(s)	false	<div>Remove Schedule</div> <div>Set SLA</div>
daily-logdata-job-end	moneyball-daily	azkaban	2013-12-24 01:00:00	2014-01-10 01:00:00	1 day(s)	false	<div>Remove Schedule</div> <div>Set SLA</div>
search-keyword-rt-fd-ent-event-hdfs-dist-cp-job-runner	moneyball-backup-data	azkaban	2014-01-07 12:00:00	2014-01-10 12:00:00	1 day(s)	false	<div>Remove Schedule</div> <div>Set SLA</div>

Flow Execution 3596

Project: moneyball-daily / Flow: daily-logdata-job-end

Status: SUCCEEDED

Submit User: azkaban

Start Time: 2014-01-05 01:00:00s

End Time: 2014-01-05 02:01:03s

Duration: 1h 00m 03s

[Graph](#) / [Job List](#) / [Flow Log](#)

Prepare Execution

Name	Timeline	Start Time	End Time	Elapsed	Status	Logs
delete	<div></div>	2014-01-05 01:00:30s	2014-01-05 01:00:30s	0 sec	Succeeded	Log
import-dbg-ctg	<div></div>	2014-01-05 01:00:30s	2014-01-05 01:00:26s	21 sec	Succeeded	Log
load-to-hdfs-dbg-ctg	<div></div>	2014-01-05 01:00:26s	2014-01-05 01:00:49s	22 sec	Succeeded	Log
tbl-convert-event-fst-fst	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:00:49s	0 sec	Succeeded	Log
convert-item-around-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:17:25s	16m 36s	Succeeded	Log
convert-ent-keyword-item-view-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:12:34s	11m 45s	Succeeded	Log
convert-cat-item-qt-ctg-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:10:57s	10m 8s	Succeeded	Log
convert-cat-view-purchase-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:17:26s	16m 37s	Succeeded	Log
convert-ent-keyword-ft-order-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:05:16s	4m 27s	Succeeded	Log
convert-campaign-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:07:28s	6m 49s	Succeeded	Log
convert-ent-keyword-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:13:36s	12m 47s	Succeeded	Log
convert-itemview-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:09:54s	9m 5s	Succeeded	Log
convert-order-view-event	<div></div>	2014-01-05 01:00:49s	2014-01-05 01:18:21s	17m 42s	Succeeded	Log
convert-entry-click-event	<div></div>	2014-01-05 01:18:31s	2014-01-05 01:34:13s	15m 42s	Succeeded	Log
tbl-f-table	<div></div>	2014-01-05 01:34:13s	2014-01-05 01:34:13s	0 sec	Succeeded	Log
f-table-customer-itemview	<div></div>	2014-01-05 01:34:13s	2014-01-05 01:37:51s	3m 38s	Succeeded	Log



Azkaban



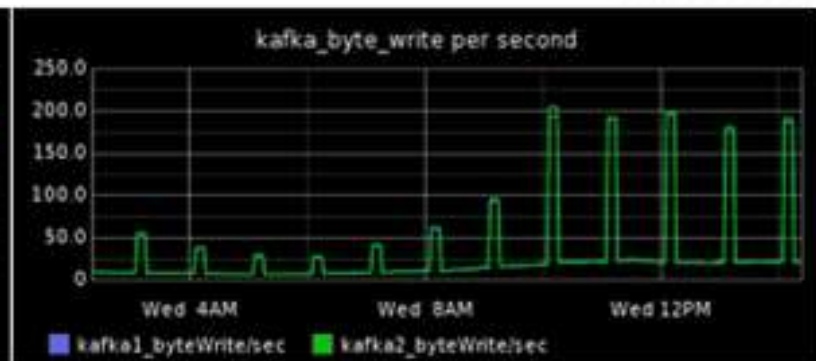
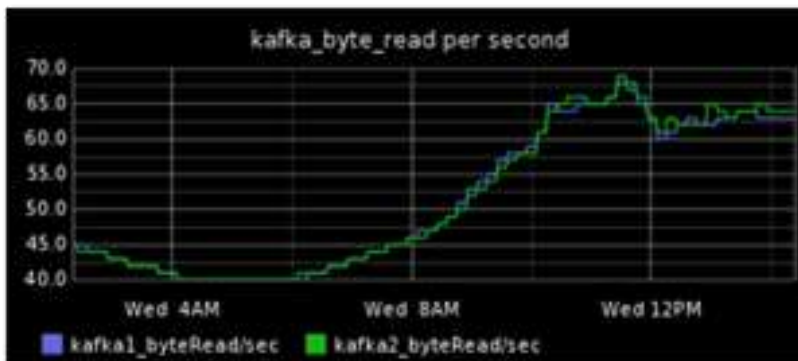
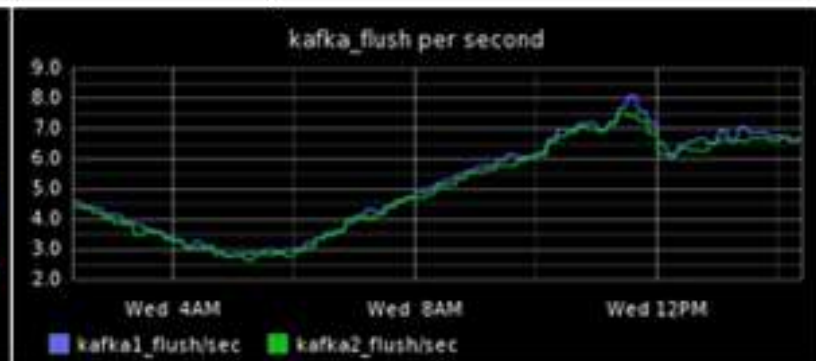
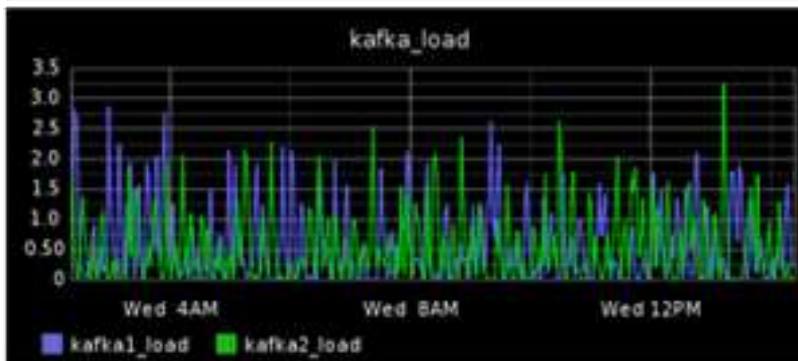
- 스케줄러



Graphite



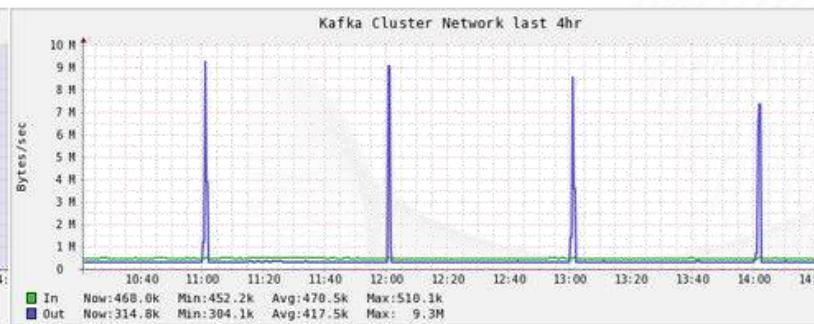
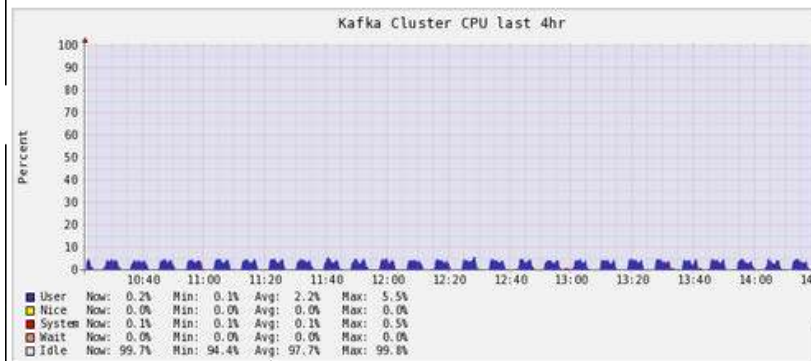
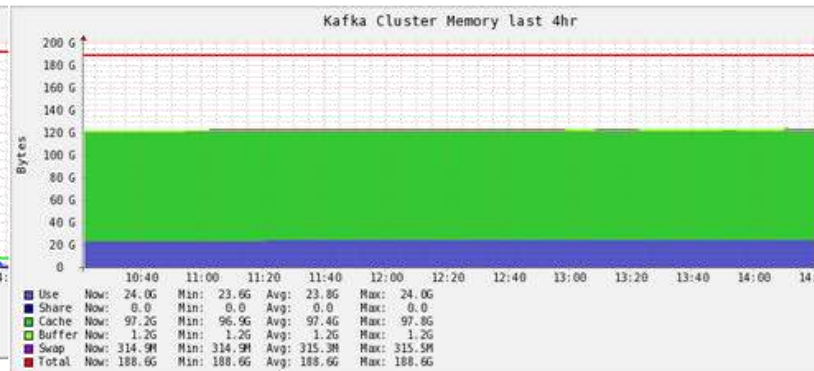
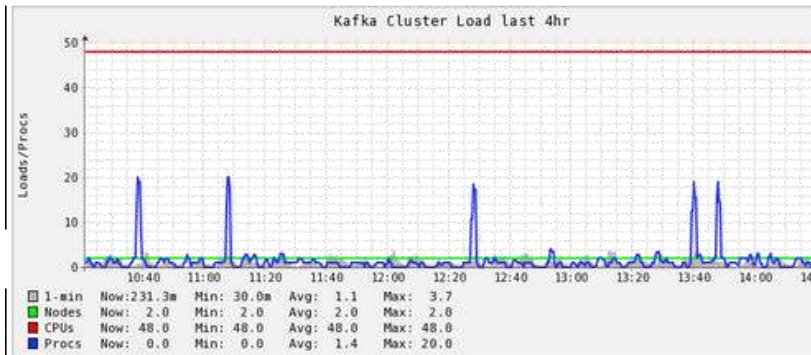
- Python 기반
- Java 프로세스 모니터링



Ganglia



- 서버 자원 상황 모니터링
- 네트워크, 메모리, CPU



환경 구성에 관해서



- 기간, 인원
- 구성 중 문제점들
 - Hortonworks
 - Oozie 관리 페이지 로딩
 - Zookeeper 사용하는 오픈소스 증가 – 분리
 - 한 서버에 다양한 오픈소스 공생
 - 많은 수의 포트 오픈
 - MR job 증가
 - 개발서버 부재
 - 스톰 트라이던트 & 카프카
 - Hbase 키 설계



The background features a light gray and white color scheme with several interlocking gears of different sizes. A prominent blue horizontal band spans the middle of the image. On the left side of this band, there is a faint, illegible list of text. The Korean text '감사합니다' is centered in the blue band in a large, white, sans-serif font.

감사합니다