

게임 빅 데이터와 정보계의 결합 구축 전략

2013.5

윤종완

목 차

- 주변 이야기
 - 빅 데이터 돌풍
 - 현업의 혼동
- 빅 데이터와 정보계의 결합
 - 게임 빅 데이터로 인한 문제
 - 최종 모습
 - 걸림돌과 해결 노력
- 맺음말
- [별첨] 빅 데이터에 관한 오해

빅데이터 Big Data 세상을 바꾸다

2011. KBS 시사기획 <창>

빅 데이터 돌풍

- 언론 매체에 부쩍 많이 언급되는 빅 데이터
네이버 뉴스 2,044건('11), 8,838건('12), 4,588건(~ '13.4)
- 가트너 2013년 기업이 전략적으로 대응해야 하는 10대 기술
“전략적 빅 데이터 기술” 선정
- 솔루션업계의 빅 데이터 세미나 급증

“빅 데이터 = 3V (Volume, Velocity, Variety)”

현업의 혼동

- 빅 데이터가 뭐지?
- 우리도 빅 데이터를 처리하는데 뭐가 다르다는 거지?
- 내가 뭘 잘못하고 있나?
- 아무튼 빅 데이터를 도입하면 도대체 무엇이 좋지?
- 어디에 도입해야 하는데?

"생경한 것에 대한 두려움"

** 빅 데이터 원류는 웹 서비스 인프라"*

** 일반 기업 IT 부서에 전파/ 확산 진행 중"*



2011. Strata Conference. New York.

돌풍과 혼돈의 이면

- 빅 데이터 요람인 Strata 분위기
- 3V는 META Group에서 2001년 언급한 단어
- WEB 2.0의 “Data is the next Intel Inside”
- 현업이 모르는 문제를 풀려는 솔루션 회사들
- 빅 데이터 솔루션은 오픈 소스 진영이 리딩

“데이터는 별반 차이가 없다. 조금 더 커졌을 뿐이다.

오히려, 오픈소스로 데이터를 다룬다는 점이 제일 생경함”

* 이걸 웹 포탈 실무자가 잘 안.

* D/W 실무자는 웹 포탈 경험이 적음.

Wrap up

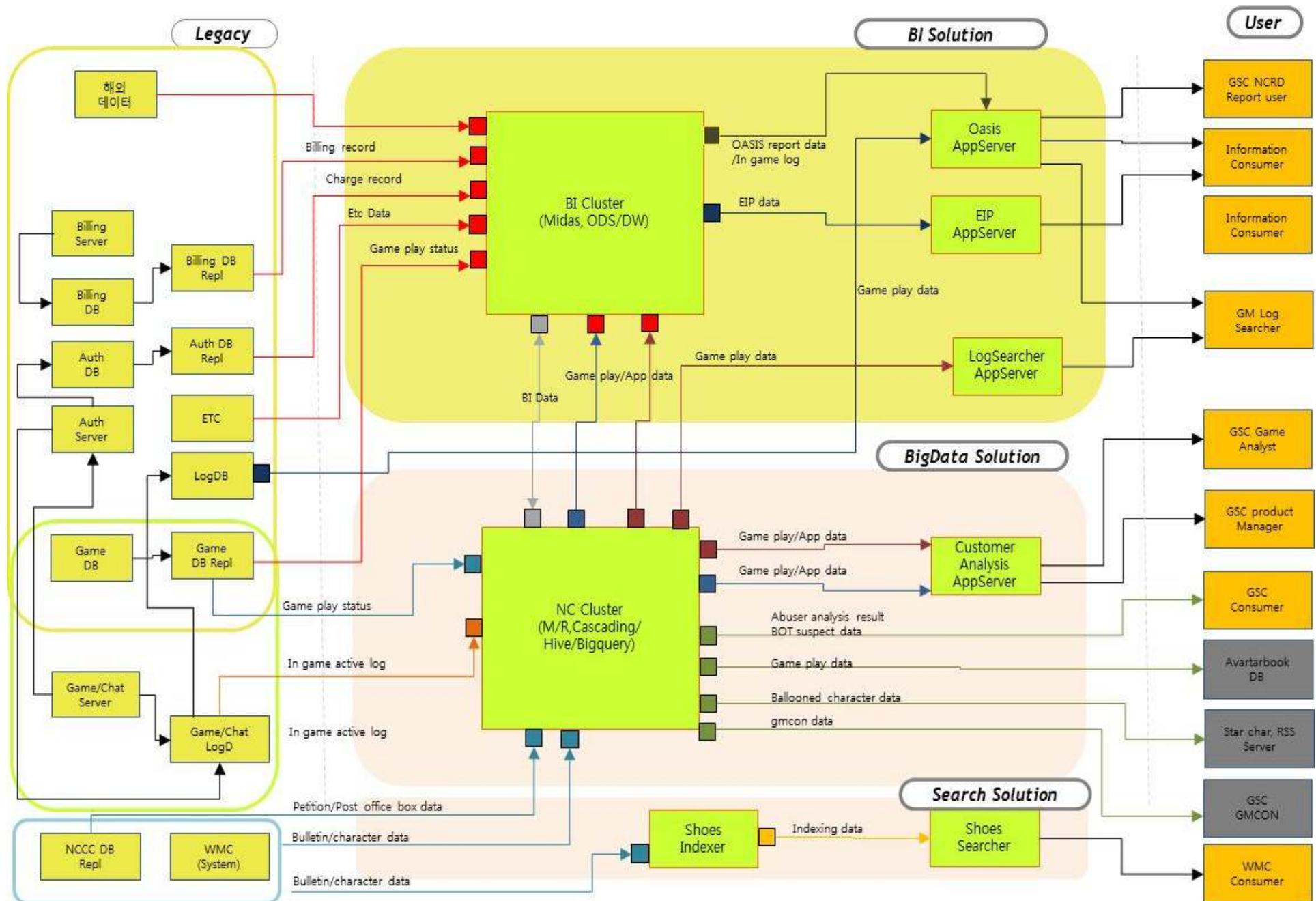
- 빅 데이터는 새로운 것이 아니다!
- 빅 데이터를 다루는 방식이 새롭다!
- 오픈 소스의 IT 도입을 유념해서 보자!
- (특히 경영진의 관심을 끄는 +alpha가 있다!)



2009년 부터 꾸준한 사랑을 받는 AION 게임



위해 행위를 찾고 관리하는 <위해환경관리>

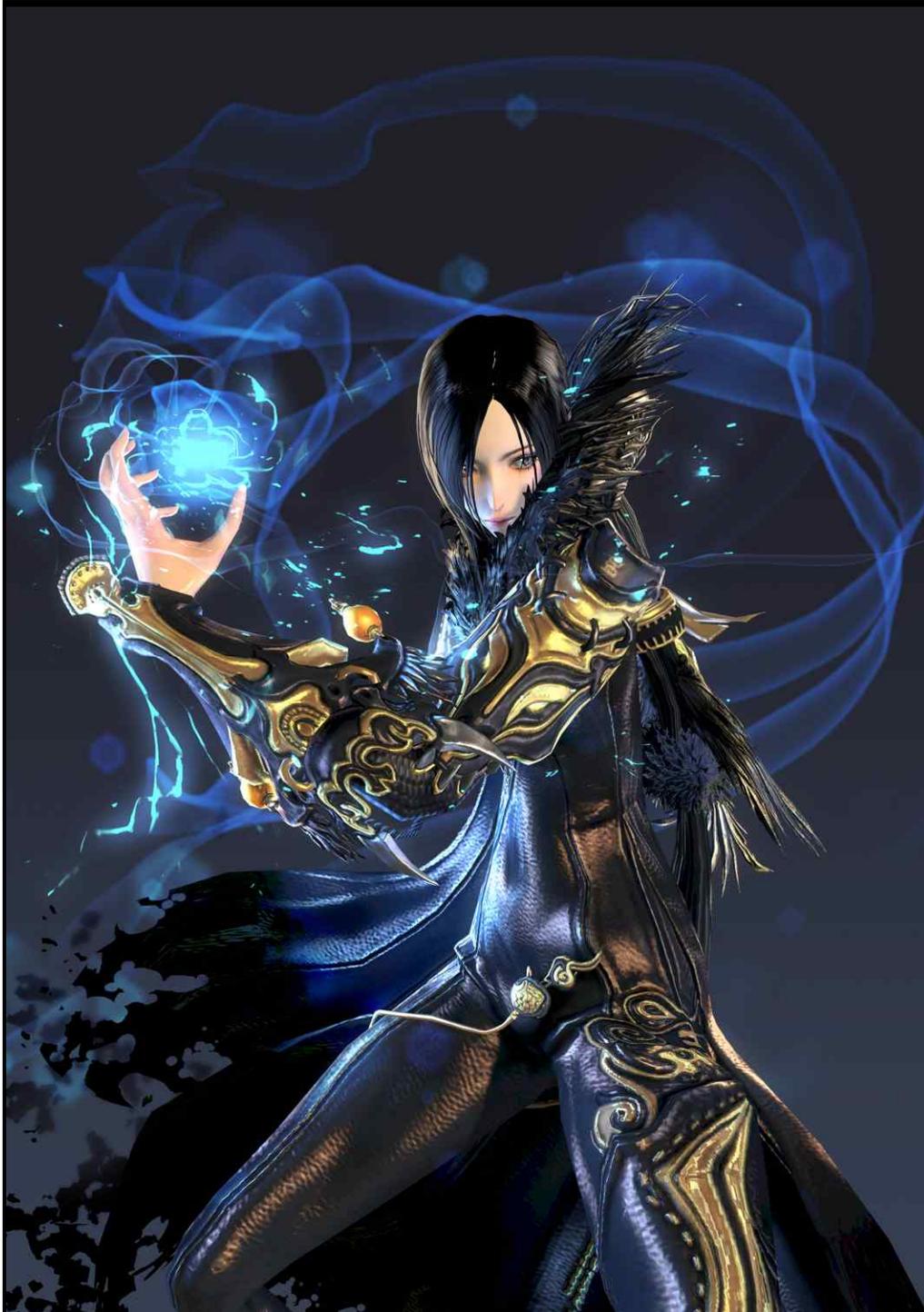


기존 D/W 모습: RDB 4 Node, Hadoop 300 Node

게임 고객의 변화

언제 이탈할지 모르는 고객, 점점 복잡해지는 고객 행태





내부 고객의 변화

'2010 ~ '2011
Blade & Soul 출시 준비



더 빠르게, 더 세밀히 보자!



(일 1Teta Data 발생, 로그 OLAP 개발)



더 큰 볼륨, 더 큰 컴퓨팅 요구!

기존 D/W의 문제점

- D/W의 컴퓨팅 Scalability
 - 상업용 ETL 솔루션의 병렬 처리 옵션 억 단위 가격
 - 상업용 솔루션은 고가의 고성능 소수 서버 요구
- D/W의 Disk 공간 Scalability
 - 수 Tera 데이터 증설에 억 단위 투자.
 - 가용성 확보를 위해 back up 솔루션 도입. 수억 투자.
- 정형 데이터 중심의 D/W의 한계 직면
 - 기존 정형 D/W와는 근본적으로 처리 방식이 다름.
 - 비정형에 굳건한 상업용 솔루션이 빈약함.

빅 데이터 정보계 모습

200여명 사용자, 40명 분석가, 800 여종 보고서

분석/요약

HiveR
(Hive, Shark)

로그조회
(LogR)

IBM Cognus
(MS-SQL)

FACT 생성

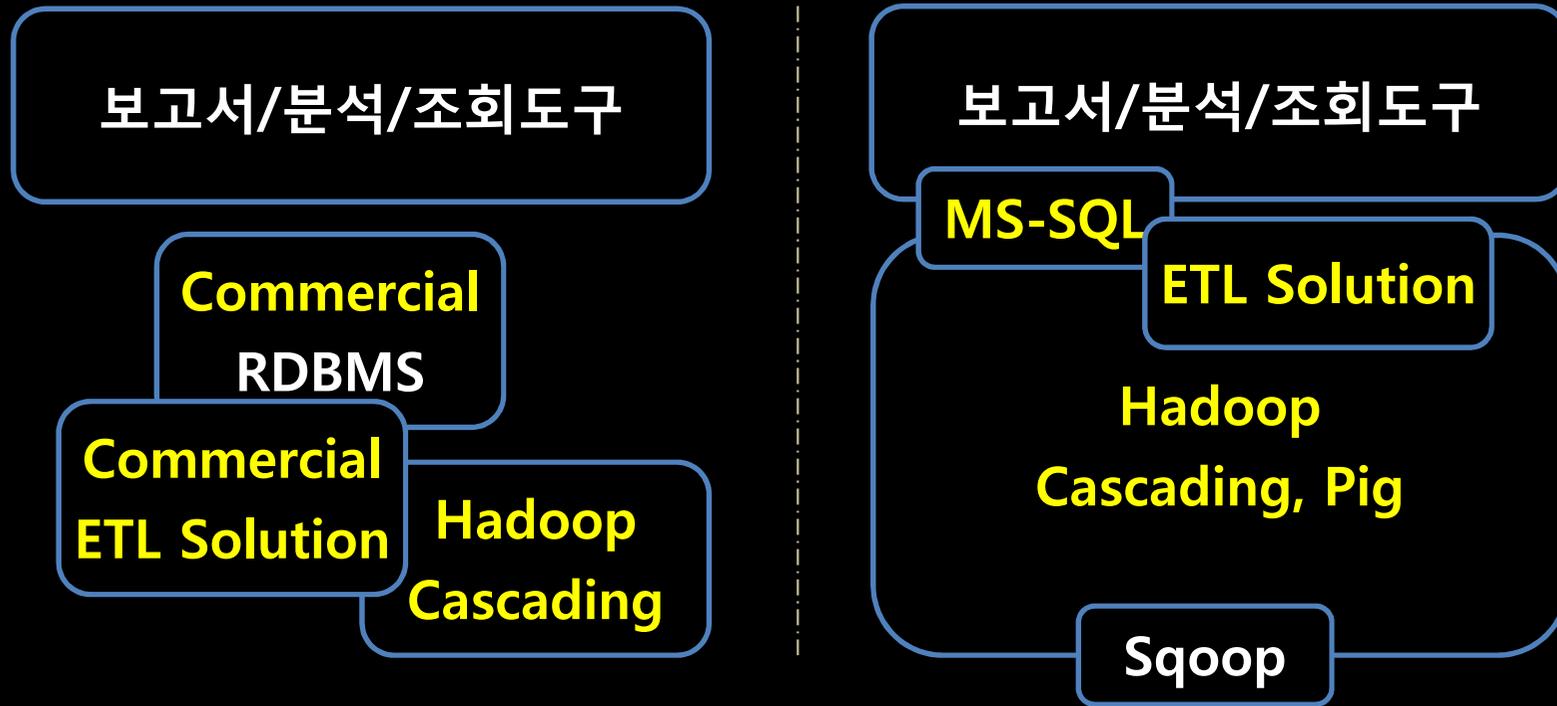
Cascading, HQL, Pig

수집/적재

Sqoop, HDFS (LAR)

2천 개 테이블, 수 백개 파일, 600 Tera 로그 데이터

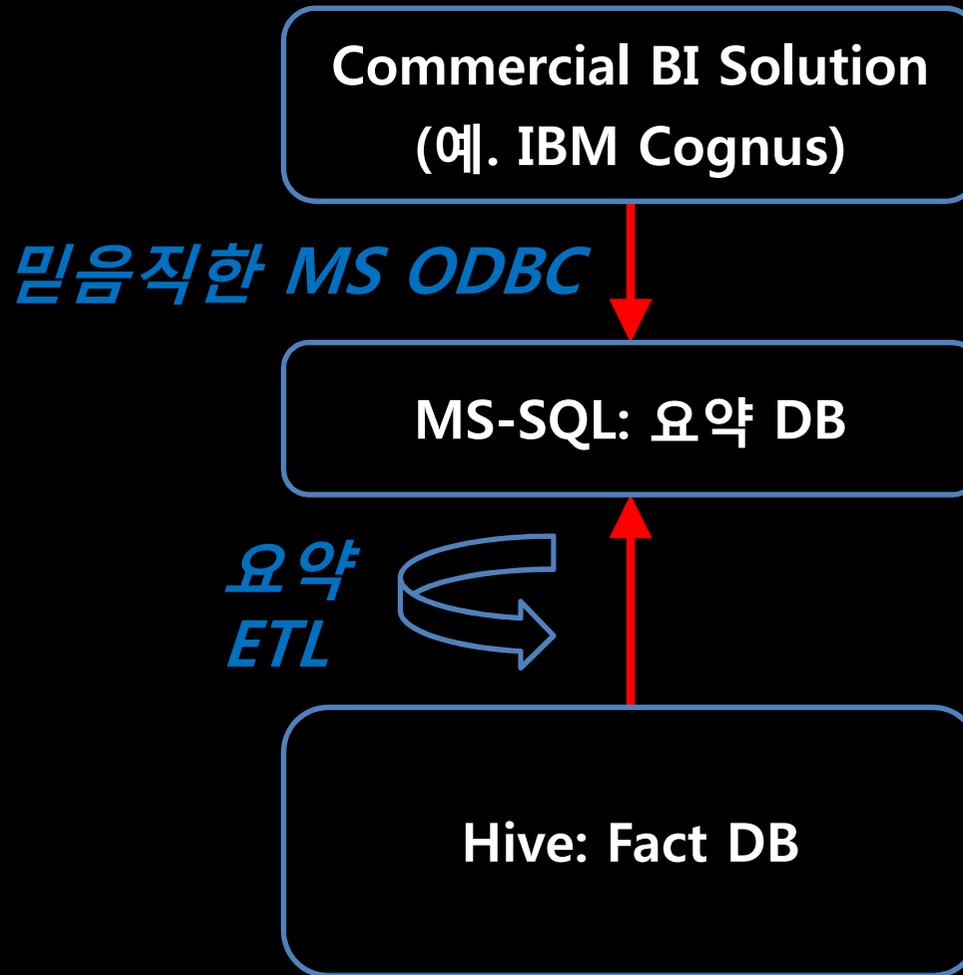
(1/4) 혼합 or 결합



2천 개 테이블, 수 백개 파일, 600 Tera 로그 데이터

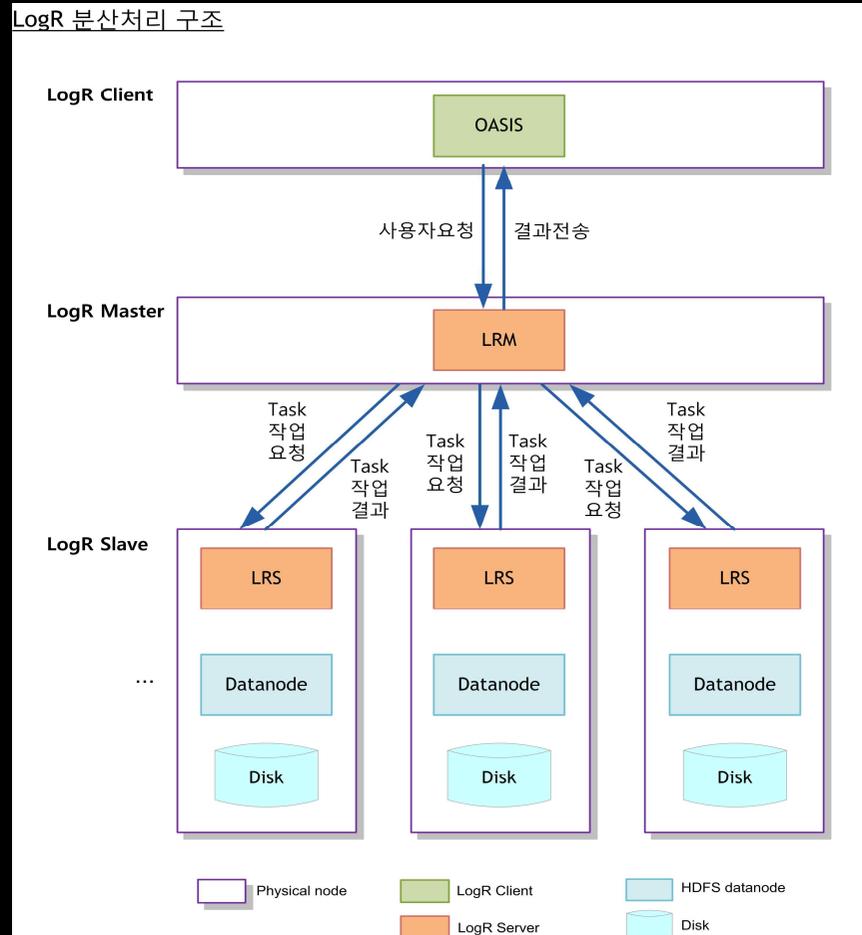
“낮은 비용, 큰 확장력을 가진 결합 구조의 채택”

(2/4) 당겨 쓰기 or 말아 올리기



“생산성 좋은 BI 상업용 솔루션과의 결합 고려”

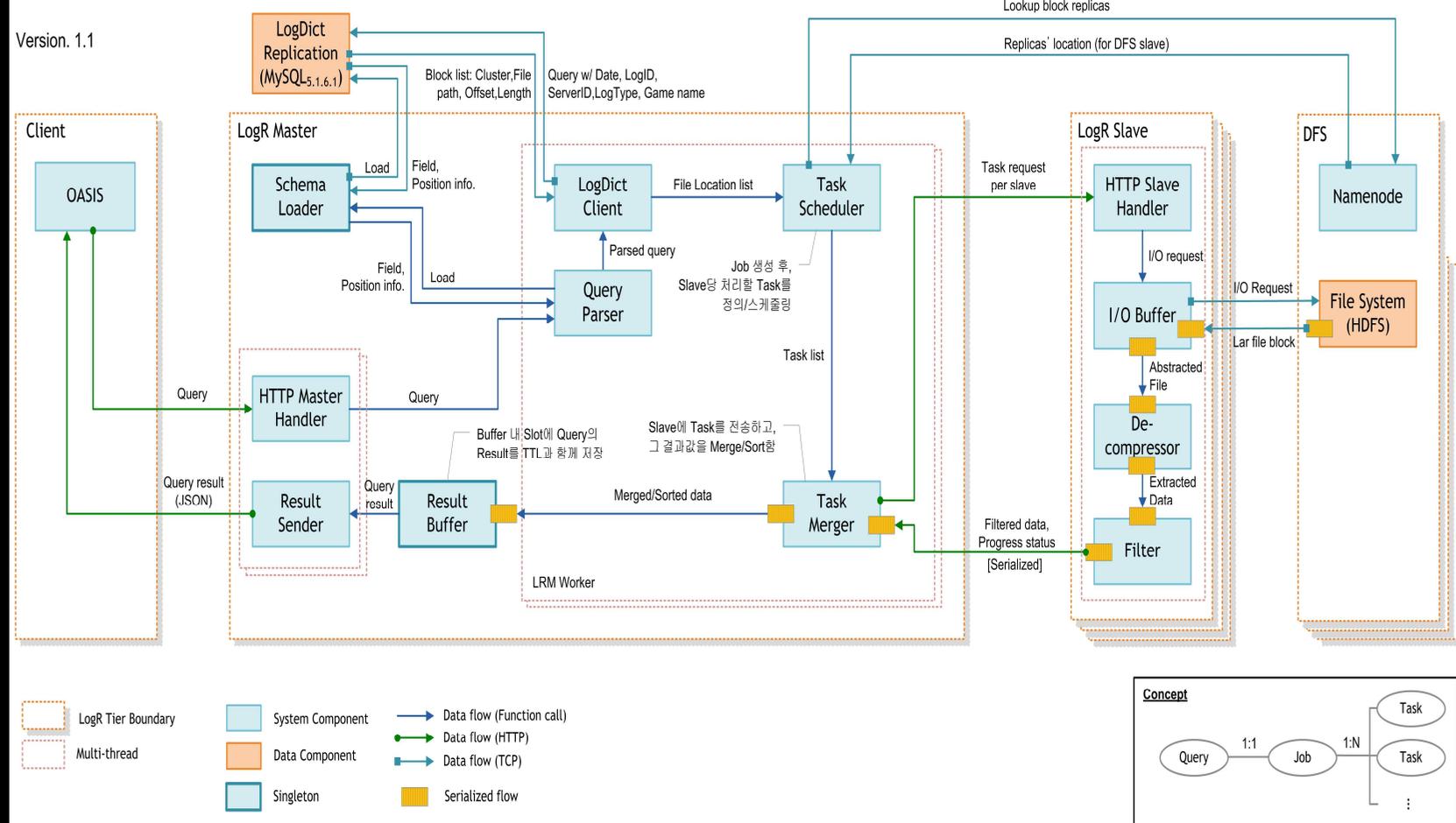
(3/4) 준 실시간 구현



“30분 미만 지연, 빠른 조회 속도”

준 실시간 구현

LogR 시스템 Data Flow Diagram



“오픈 소스 선정 포기. LogR 자체 프로젝트 개발”

(4/4) Hive 튜닝

- Hive D/W가 일반 기업의 분석 D/W로 부족한 4가지 점
 1. Toad와 같은 친숙한 에디터가 없음
 2. Hive는 다중 사용자 고려 부족
 3. Data 접근 및 권한 관리 기능 없음
 4. 작은 규모의 테이블 쿼리 속도가 매우 느림

“중심 D/W로 Hive를 채택하면 대대적인 소스 수정이 필요함!”

(4/4) Hive 튜닝

The screenshot shows the HiveR web interface. The main window displays a SQL editor with the following queries:

```
select * from bns_action where time >= '2013-01-08 06:00:00.000' and time < '2013-01-08 06:00:00.000' limit 100;
1 select *
2 from temp_an_free_account
3 limit 200;
4
5 select *
6 from dbs_faction
7 limit 200;
8
9 select *
10 from dbs_acc
11 limit 400;
12
13 select *
14 from bns_action where time >= '2013-01-08 06:00:00.000' and time < '2013-01-08 06:00:00.000' limit 100;
```

A "Search and Replace" dialog box is open, showing search and replace options. The "Search in" section has "Selection" selected. The "Direction" section has "Backward" selected. The "Options" section has "Ignore Case" and "RegExp" selected.

Below the editor, a table of query results is displayed:

act_time	login_time	action	account	logout_time	worldid	billing	ip	pccafecd	accountid	free
2013-01-01 04:00:00	2013-01-01 02:57:17	SERVEROUT	X521cheng		111	1300	119.201.45.90	0000000	6570582	1
2013-01-01 04:00:00	2013-01-01 03:59:10	SERVERIN	Lelra14		112	1300	123.141.8.53	P028410	6651207	1
2013-01-01 04:00:00	2013-01-01 02:57:17	LOGOUT	X521cheng	2013-01-01 04:00:00	111	1300	119.201.45.90	0000000	6570582	1
2013-01-01 04:00:00	2012-12-31 23:23:23	SERVEROUT	Bluekisa87		8	0	211.213.80.41	0000000	11616435	0
2013-01-01 04:00:00	2013-01-01 03:59:57	SERVERIN	Silverfood		41	1	183.108.196.2	P025883	11355289	1
2013-01-01 04:00:00	2012-12-31 18:06:57	SERVEROUT	Gksrk76		111	1	210.120.49.136	P025332	9385695	1

“오픈 소스에 부족한 편리성! HQL 편집기 HiveR 자체 개발”

(4/4) Hive 튜닝

- ~~1. Toad와 같은 친숙한 에디터가 없음.~~
2. Hive는 다중 사용자 고려 부족
3. Data 조회 권한 관리 기능 없음
4. 작은 규모의 테이블 쿼리 속도가 매우 느림.

*"Hive Broker Server 오픈 소스 수정 개발 착수
Hadoop Scheduler 수정 개발
Hive Local 모드 적용과 Spark 도입 진행"*

Wrap up II

- NCSOFT의 온라인 게임은 빅 데이터를 가지고 있음. 고객을 더 잘 알려는 니즈 때문에 빅 데이터의 중요성은 높아지고 있음.
- 그러나, 상업용 솔루션 기반 기존 D/W는 공간, 컴퓨팅 확장이 매우 낮아, NCSOFT의 빅 데이터 솔루션에 부적합했음.
- 솔루션 비용 대신 개발자를 투입하여 오픈 소스를 직접 수정하는 방식으로 NCSOFT에 딱 맞는 D/W 구축에 노력 중임.

맺음말

- 생경함으로 오는 혼동이 좀 줄었나요?
- -1을 생각했지만, 혹시나 +1이 되진 않았나 싶습니다.
- 세부 질문은 jongwanyun@gmail.com로 해주세요.
- 이상! 윤종완@엔씨.TC.데이터플랫폼실 실장이었습니다.
- 감사합니다.

[별첨] 빅 데이터에 관한 오해

- 전사 데이터를 전무 모아 'Big'하게 만든다고, 가치가 생긴 **않음!**
- 'Big' 인 사이트를 찾으려 '대단한' 알고리즘을 찾지 말라!
데이터 분석이 'King'은 아니다.
- 오픈 소스를 깐다고 끝나진 않는다. 또, 공짜도 아니다!