GAION
http://www.gaion.kr

빅데이터 활용과 산업경쟁력

2012.10

GAION
강현섭 (hskkang@gaion.kr)



빅데이터란?

매달 **300억** 개의 콘텐츠가 **Facebook**에서 공유

- McKinsey (2011. 05) “Big Data: The next frontier for innovation, competition, and productivity”

전 세계 **데이터**는 매년 **40% 증가** vs. **IT지출**은 **5% 증가**

- McKinsey (2011. 05) “Big Data: The next frontier for innovation, competition, and productivity”

2009년과 2020년도 사이 디지털 **데이터**의 규모는 **44 X배로 증가**

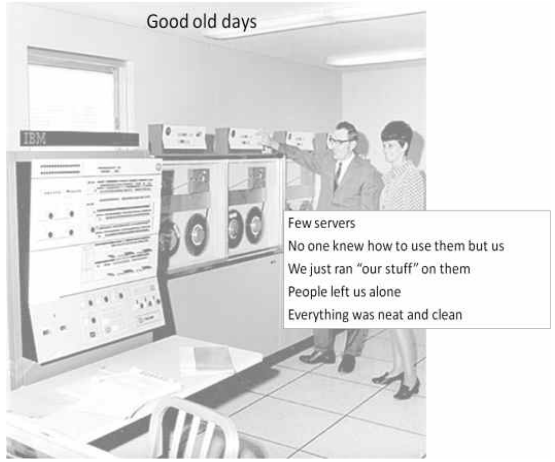
- IDC 2010 Storage Market View

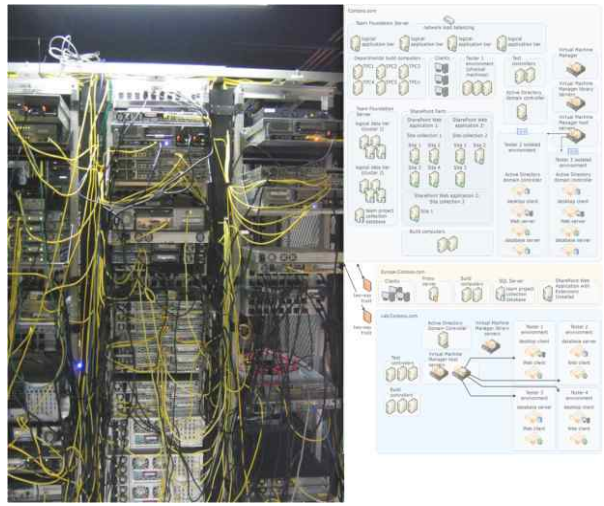
비정량 IT 데이터만의 순수 증가가 **90%** 의 전체 디지털 데이터 시장

- IDC 2011 Digital Universe Study: Extracting Value from Chaos

© 2007-2012 GAION. All rights reserved

빅데이터란?





45 억 건의 데이터(1980 년대 초반)

- IBM 3850 Mass Storage System - 100GB, $40,000/GB
- ***역시 그 당시에는 Big Data 였지요.***



68 억 건의 데이터 (현재)

- $10,000 Dell 128GB RAM 서버의 메모리 내에 처리될 수 있습니다.
- 모든데이터 처리 및 호출을 RAM 내에서 매우 빨리 처리 할 수 있습니다.
- ***당연히 Big Data 가 아니지요.***

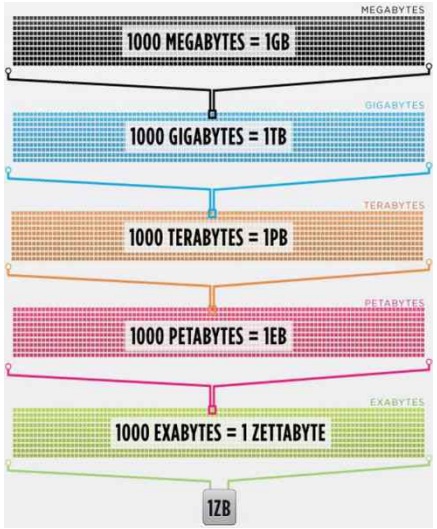


Scenario courtesy of ACM

© 2007-2012 GAION. All rights reserved

빅데이터란?

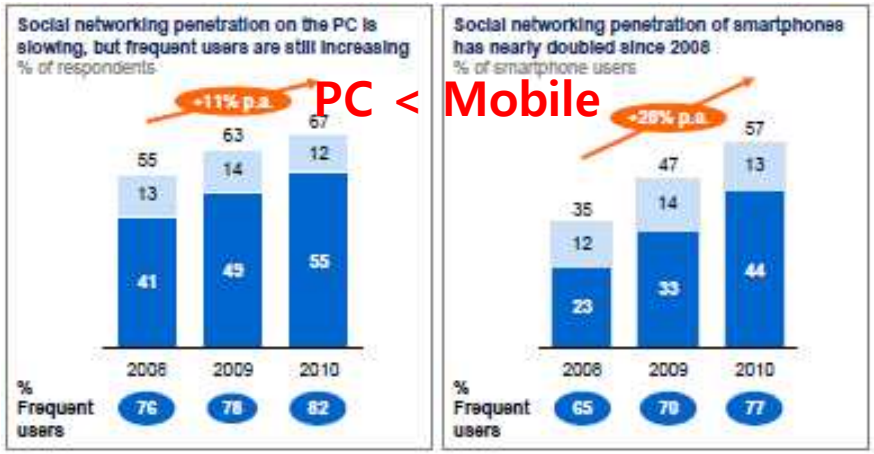




우리는 이미 제타(10^{21}) 시대에 돌입

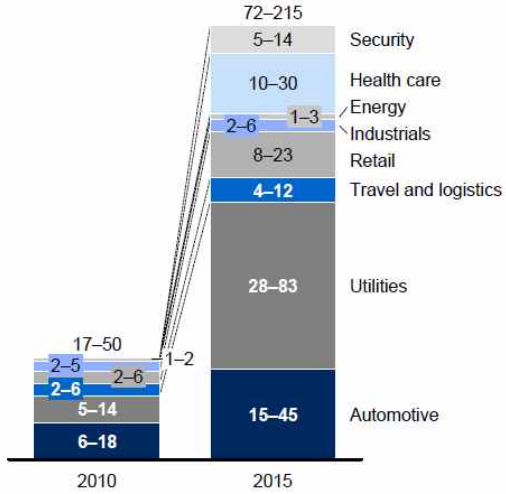
- iPad 16G를 축구장 크기로 쌓아도 대기권 2배 높이
- 美의회도서관 저장정보(235TB, 2011.04)의 4백만 배

모바일 활용의 증가



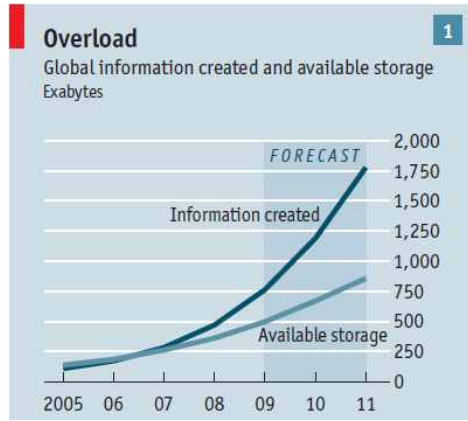
source : MacKinsey (2011.05)

M2M 센서의 지속적인 확산



source : MacKinsey (2011.05)

데이터 생산량의 증가



source: IDC (2011.02)

© 2007-2012 GAION. All rights reserved

빅데이터란?



고객 데이터 트래킹/ 수집 행위 증가

M2M 확산에 따른 센서 저변 확대

SNS의 급격한 확산과 비정형 데이터의 폭증

멀티미디어 콘텐츠와 콘텐츠 사용에 관한 정보의 증가

일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터

다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고
(데이터의) 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처

많은 데이터 + 관리 및 분석의 필요성 = Big Data

© 2007-2012 GAION. All rights reserved

빅데이터의 활용?

빅데이터는 초기 개념 정립 단계에서는 IT기술의 의미가 컸으나, 현재는 범위가 확대되어 인재·조직·기술·데이터의 모든 집합체를 통합하여 구성하는 것을 의미합니다.

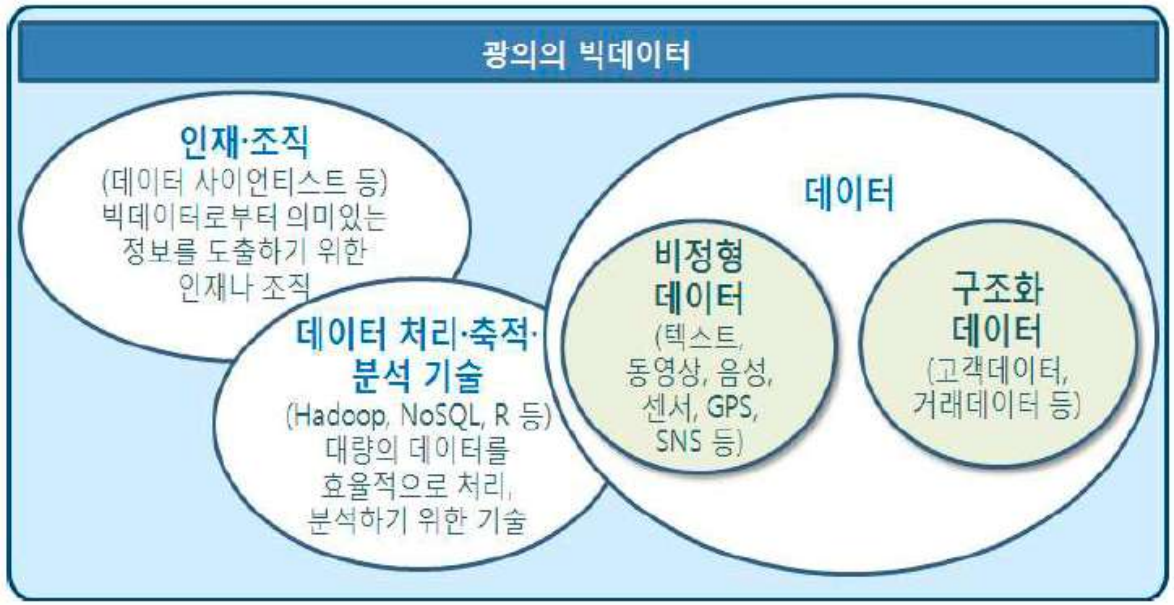
빅데이터 활용 기술

대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술

빅데이터의 범위

초기에는 기술 측면에서 접근하여 데이터만을 지칭 하였으나, 수집, 저장, 검색, 공유, 분석, 시각화 등 관련 기술·도구 및 인재·조직으로 확대되고 있음

< 시사점 >
기술은 중요하지 않다.
앞으로는 데이터 과학자의 역량과 그것을 뒷받침하는 조직이 중요하다.



© 2007-2012 GAION. All rights reserved

빅데이터의 활용 시 고려사항


Gartner

Why Gartner Analysts Research Events Consulting About

Gartner Newsroom

PRESS RESOURCES

Event Registration
Events Calendar
Gartner Blog Network
Videos
Podcasts
Photographs
Gartner Books
Analyst Coverage Areas
Press Release RSS Feed
Attribution Guide

ABOUT GARTNER

Products & Services
Investor Information
Executive Profiles
Board of Directors
Acquisition History
Gartner Policies
Office of the Ombuds

Press Releases CURRENT | VIEW ALL

Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data

Gartner Special Report Examines How to Leverage Pattern-Based Strategy to Gain Value in Big Data

STAMFORD, Conn., June 27, 2011—

Many IT leaders are attempting to manage "big data" challenges by focusing on the high volumes of information to the exclusion of the many other dimensions of information management, leaving massive challenges to be addressed later, according to Gartner, Inc.

Big data is a popular term used to acknowledge the exponential growth, availability and use of information in the data-rich landscape of tomorrow. The term "big data" puts an inordinate focus on the issue of information volume (in every aspect from storage through transform/transport to analysis). Big data is also heavily weighted toward current issues and can lead to short-sighted decisions that will hamper the enterprise's information architecture as IT leaders try to expand and change it to meet changing business needs.

Information managers may be tempted to focus on volume alone when they are losing control of the access and qualification aspects of data at the same time. Gartner analysts warn that too narrow a focus will force massive reinvestment in two to three years to address the other dimensions of big data.

Gartner **에 의하면 빅데이터 관리의 요구사항은 단순 많은 양의 데이터 만이 아닙니다.**

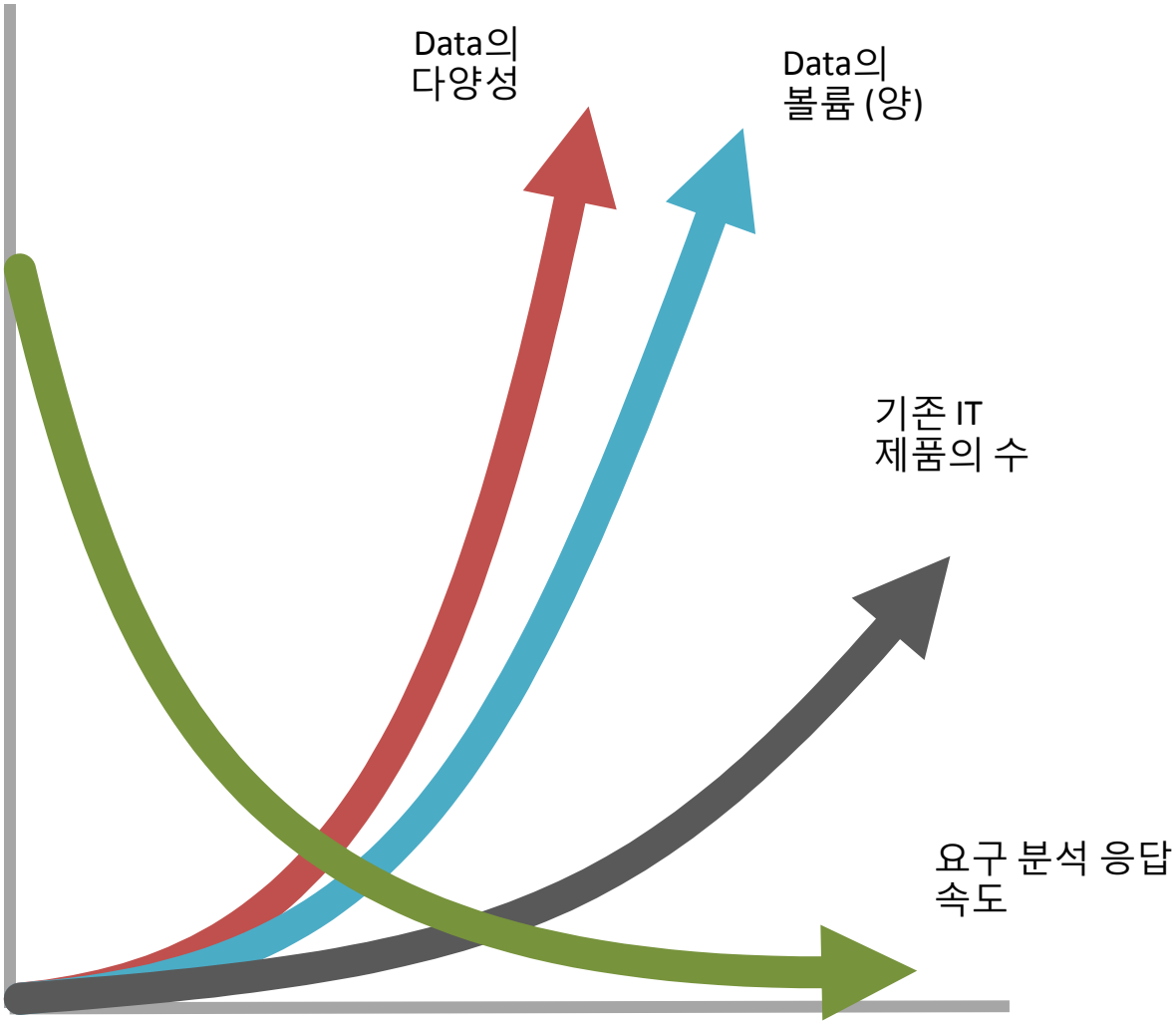
Volume **양** :
많은 양의 데이터 트렌젝션

Variety **종류** :
여러 이종의 데이터

Velocity **속도** :
쉬운 데이터의 접근과 분석 호출

© 2007-2012 GAION. All rights reserved

빅데이터의 활용 시 고려사항



"2009년과 2020년도 사이 디지털 데이터의 규모는 44 X배로 증가"

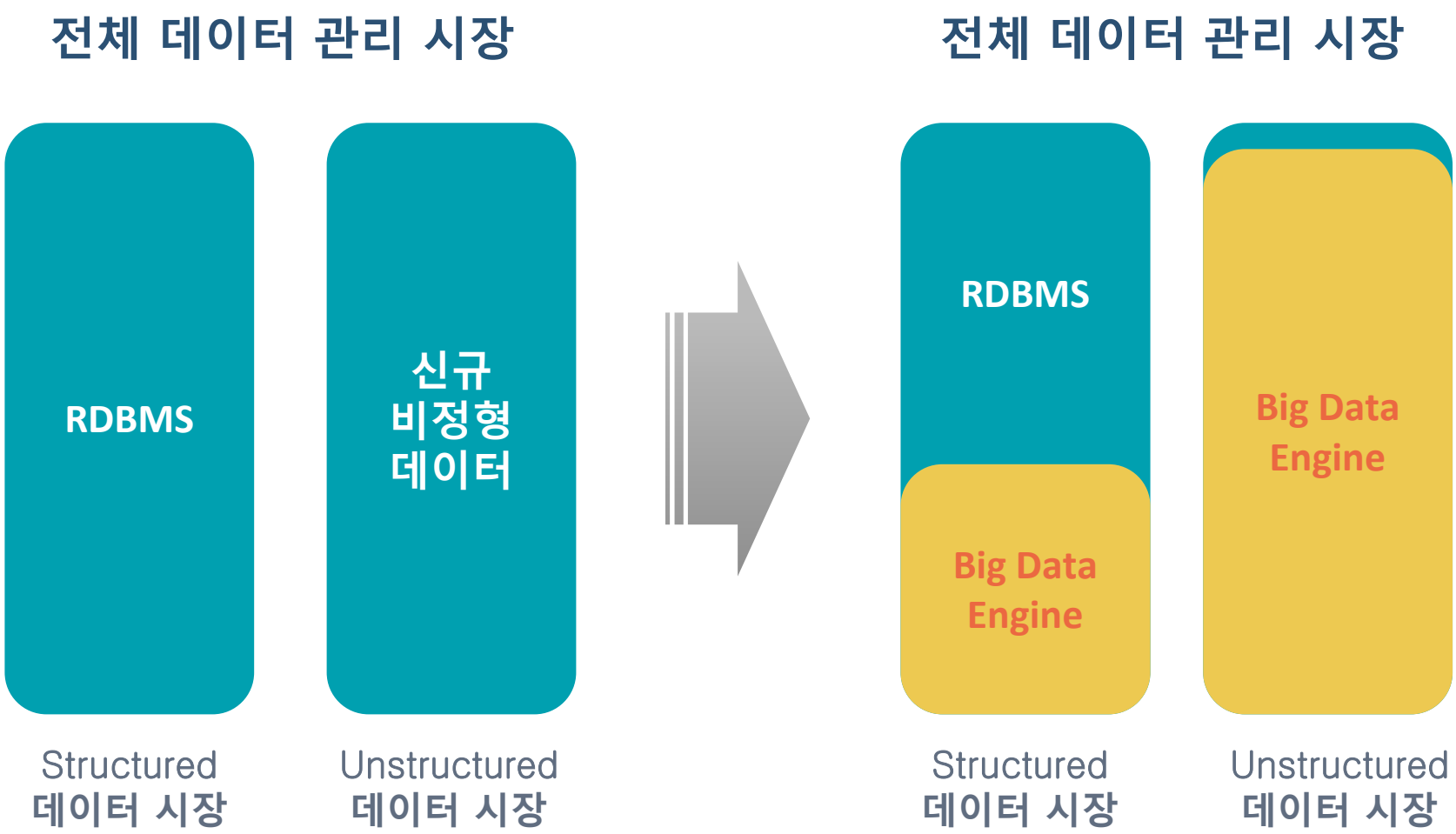
IDC 2010 Storage Market View

"비정량 IT 데이터만의 순수 증가가 90% 의 전체 디지털 데이터 시장"

IDC 2011 Digital Universe Study: Extracting Value from Chaos

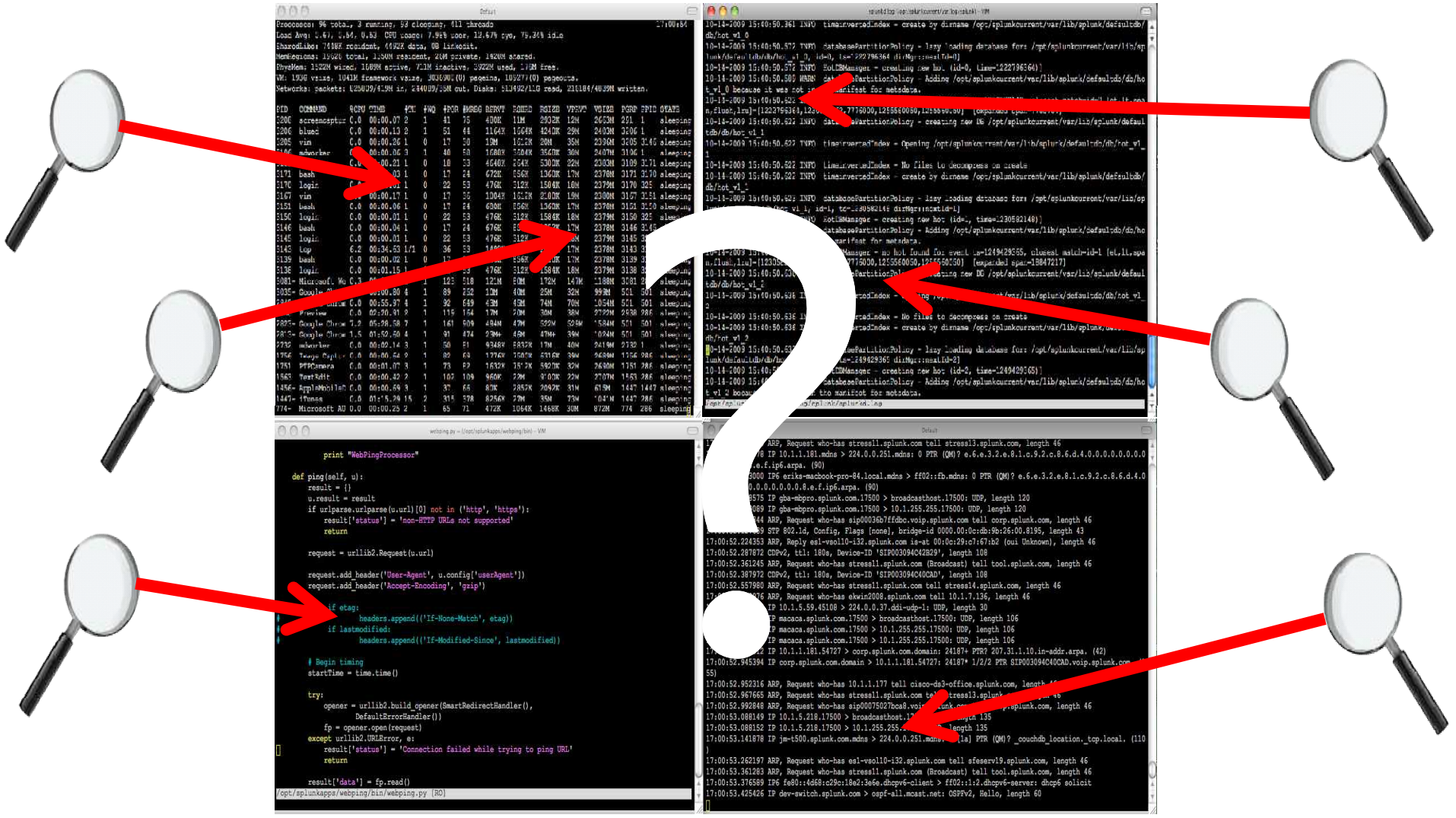
© 2007-2012 GAION. All rights reserved

빅데이터의 활용 시 고려사항



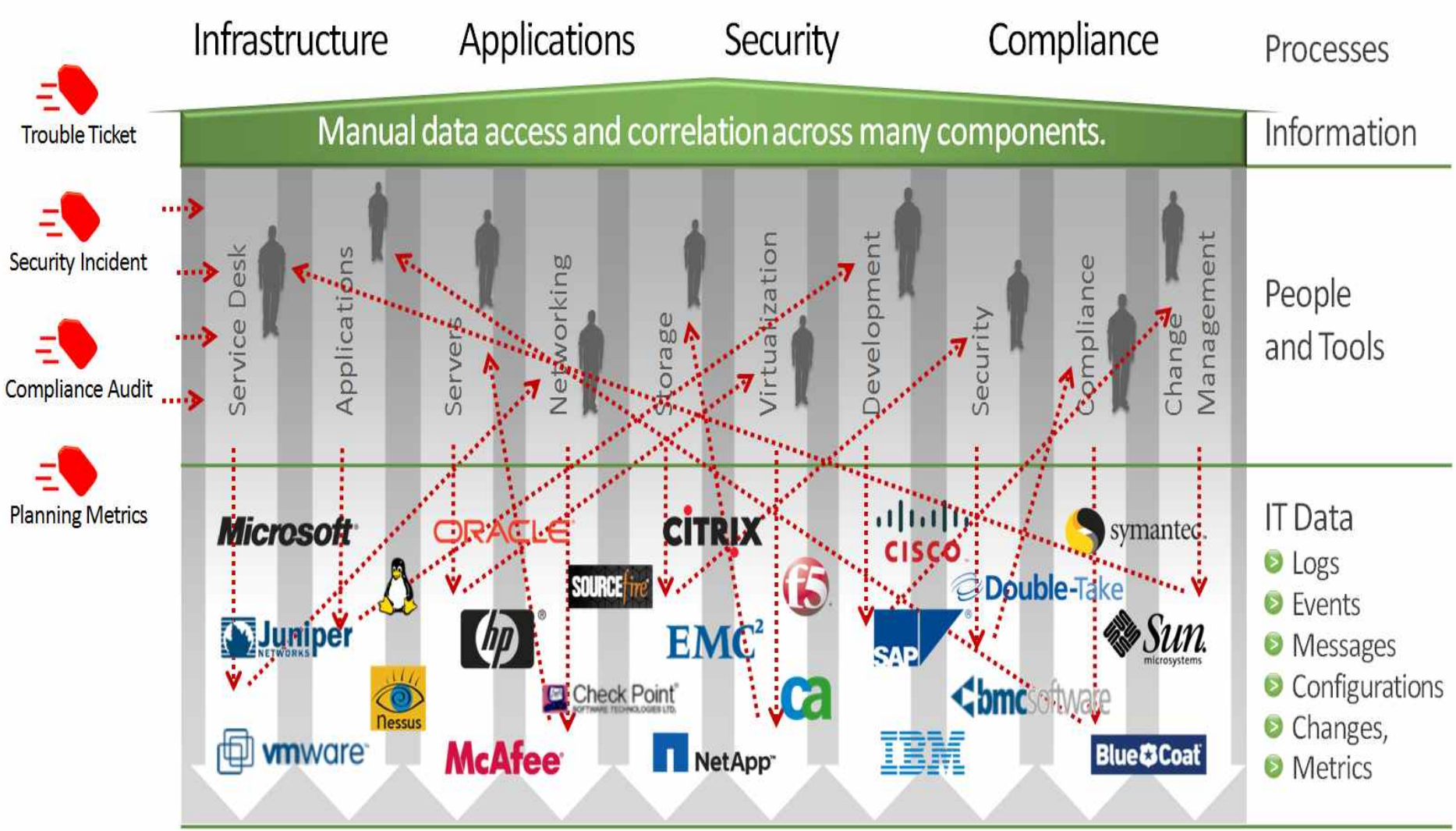
© 2007-2012 GAION. All rights reserved

데이터 관리의 현주소

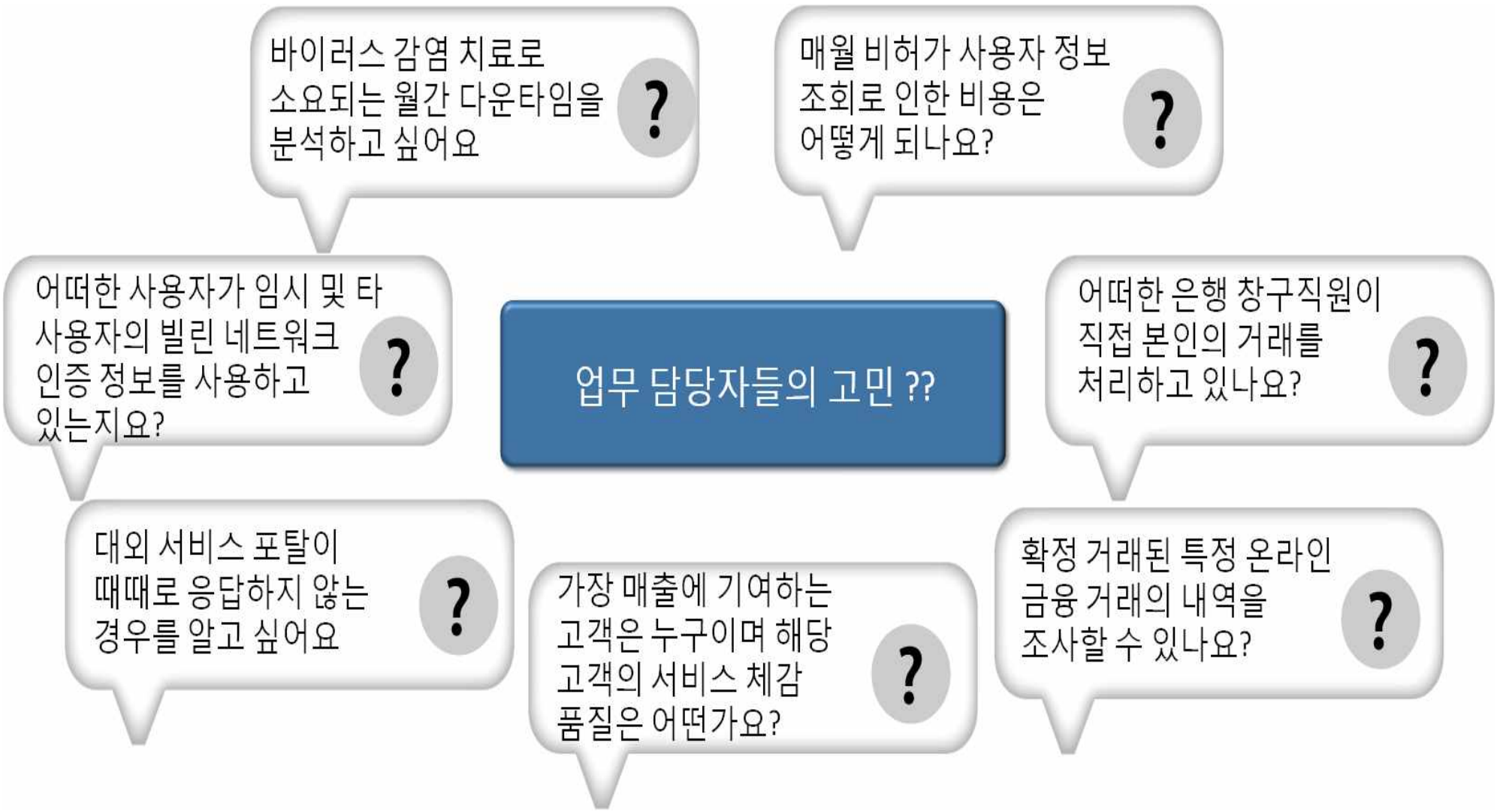


© 2007-2012 GAION. All rights reserved

IT 담당자들의 고민

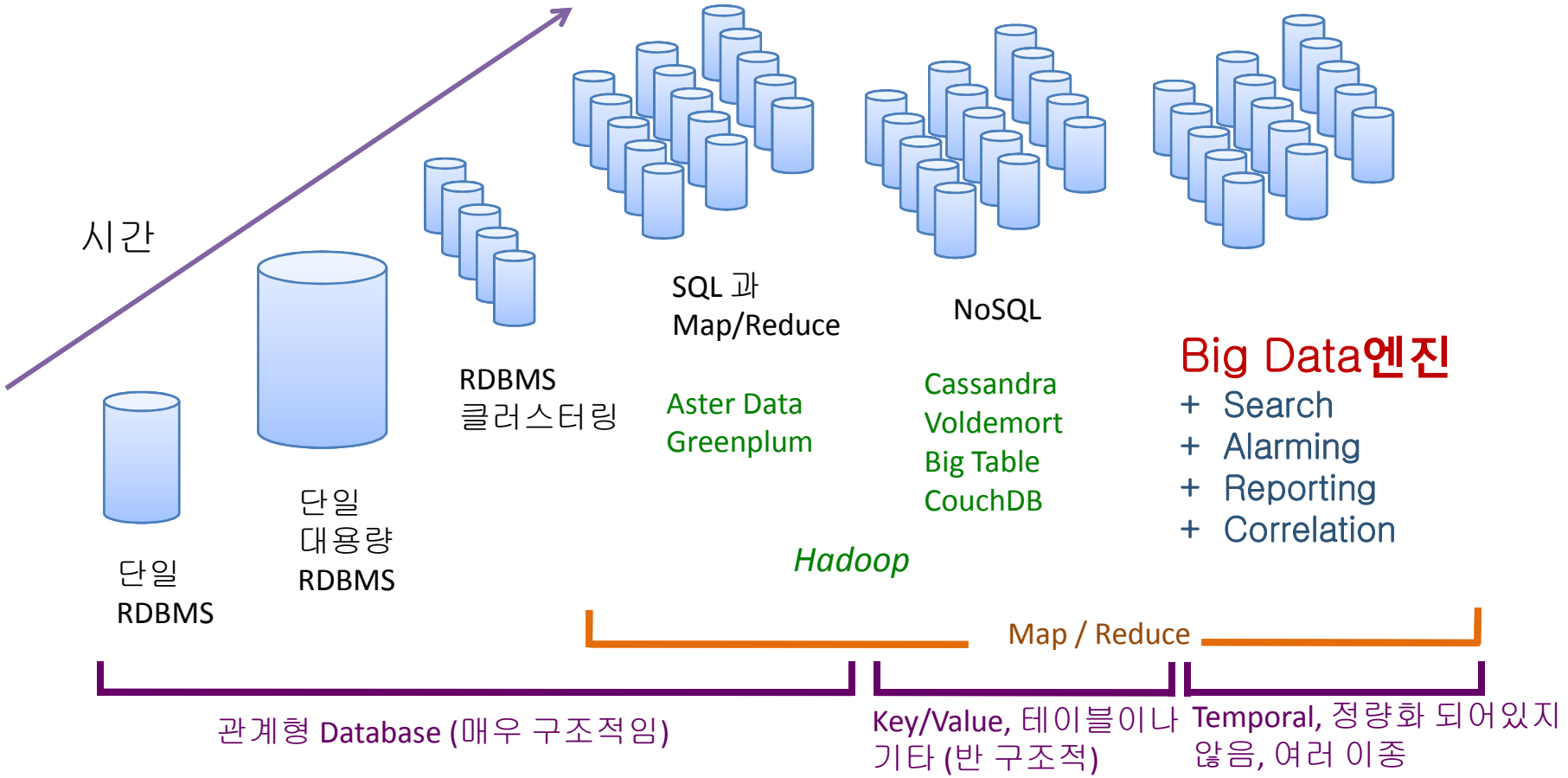

Infrastructure
Applications
Security
Compliance
Processes
Trouble Ticket
Manual data access and correlation across many components.
Information
Security Incident
Compliance Audit
Planning Metrics
Service Desk
Applications
Servers
Networking
Storage
Virtualization
Development
Security
Compliance
Change Management
People and Tools
Microsoft
ORACLE
CITRIX
CISCO
symantec.
Double-Take
Juniper NETWORKS
hp
EMC²
SAP
Sun microsystems
Nessus
Check Point
ca
bmcSoftware
vmware
McAfee
NetApp
IBM
BlueCoat
IT Data
Logs
Events
Messages
Configurations
Changes,
Metrics

업무 담당자들의 고민


바이러스 감염 치료로 소요되는 월간 다운타임을 분석하고 싶어요
매월 비허가 사용자 정보 조회로 인한 비용은 어떻게 되나요?
어떠한 사용자가 임시 및 타 사용자의 빌린 네트워크 인증 정보를 사용하고 있는지요?
업무 담당자들의 고민 ??
어떠한 은행 창구직원이 직접 본인의 거래를 처리하고 있나요?
대외 서비스 포탈이 때때로 응답하지 않는 경우를 알고 싶어요
가장 매출에 기여하는 고객은 누구이며 해당 고객의 서비스 체감 품질은 어떤가요?
확정 거래된 특정 온라인 금융 거래의 내역을 조사할 수 있나요?

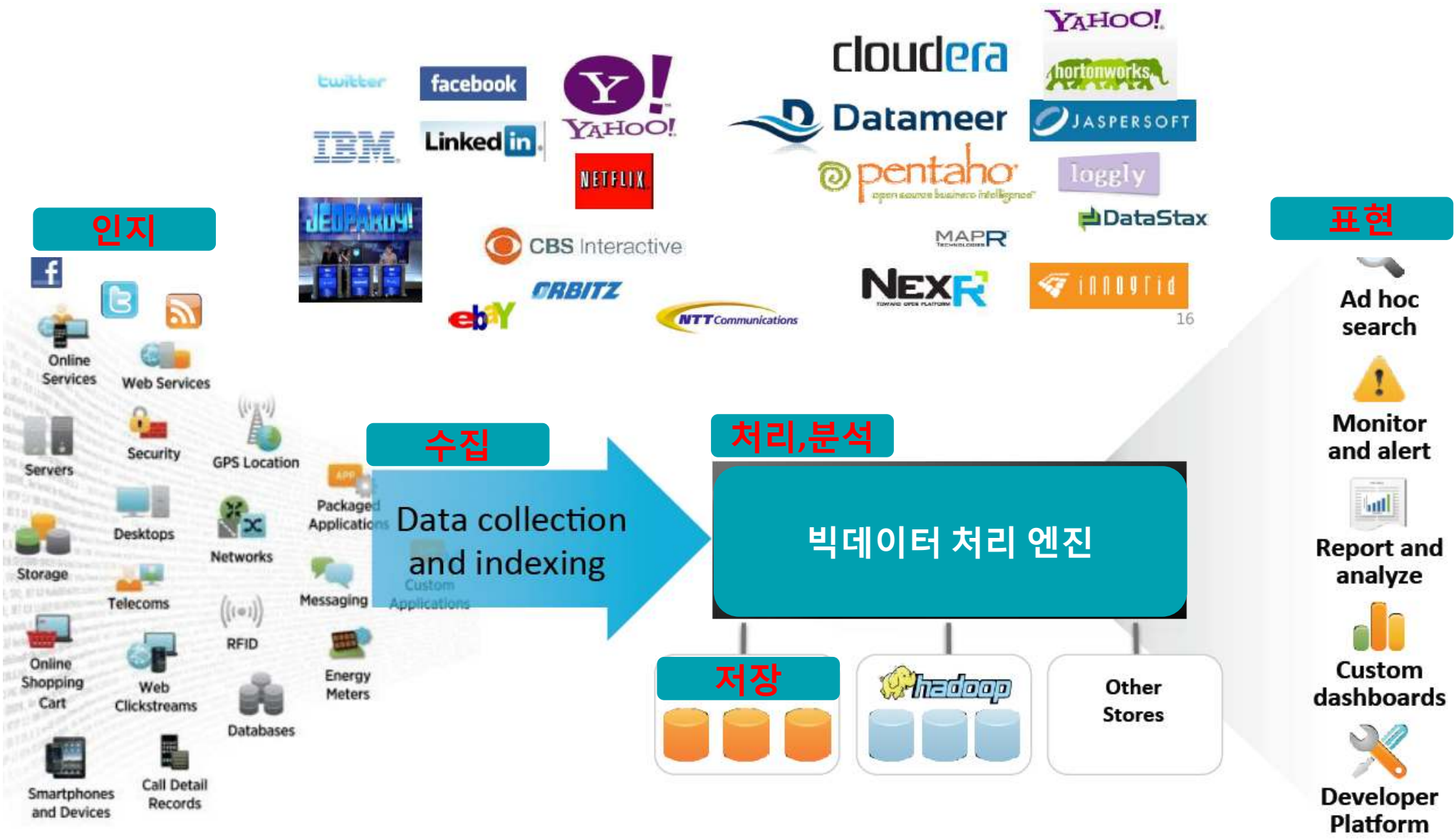
빅데이터의 출현에 따른 데이터 분석기술의 변화

빅데이터의 출현에 따라 단일 데이터 분석에서 비정형·대용량 데이터 분석 기술의 필요성이 대두되었고, 단순 데이터 분석을 넘어 실시간 분석, 상관관계 분석이 필수적으로 되었습니다.



© 2007-2012 GAION. All rights reserved

빅데이터 관리 기술 현황



© 2007-2012 GAION. All rights reserved

빅데이터 기술 적용 분야

IT 인프라 및 애플리케이션 에러 시 원인분석

Discovery 자료로의 활용

보안사고 시 대응(Forensic), 보안규제 준수 시스템(ESM,SIEM,PCI)

IT 인프라 관리(NMS, SMS, DB, WAS, Virtualization……)

마케팅 및 , BI

대용량 상관 분석 시스템, 대용량 DB 트랜젝션 관리

변경관리, 서버 관리자 작업 이력관리

내부 감사 시스템

기존 관제 제품과 연동하여 IT 통합 관제 센터 구축



© 2007-2012 GAION. All rights reserved

RFID 데이터 적용 방안

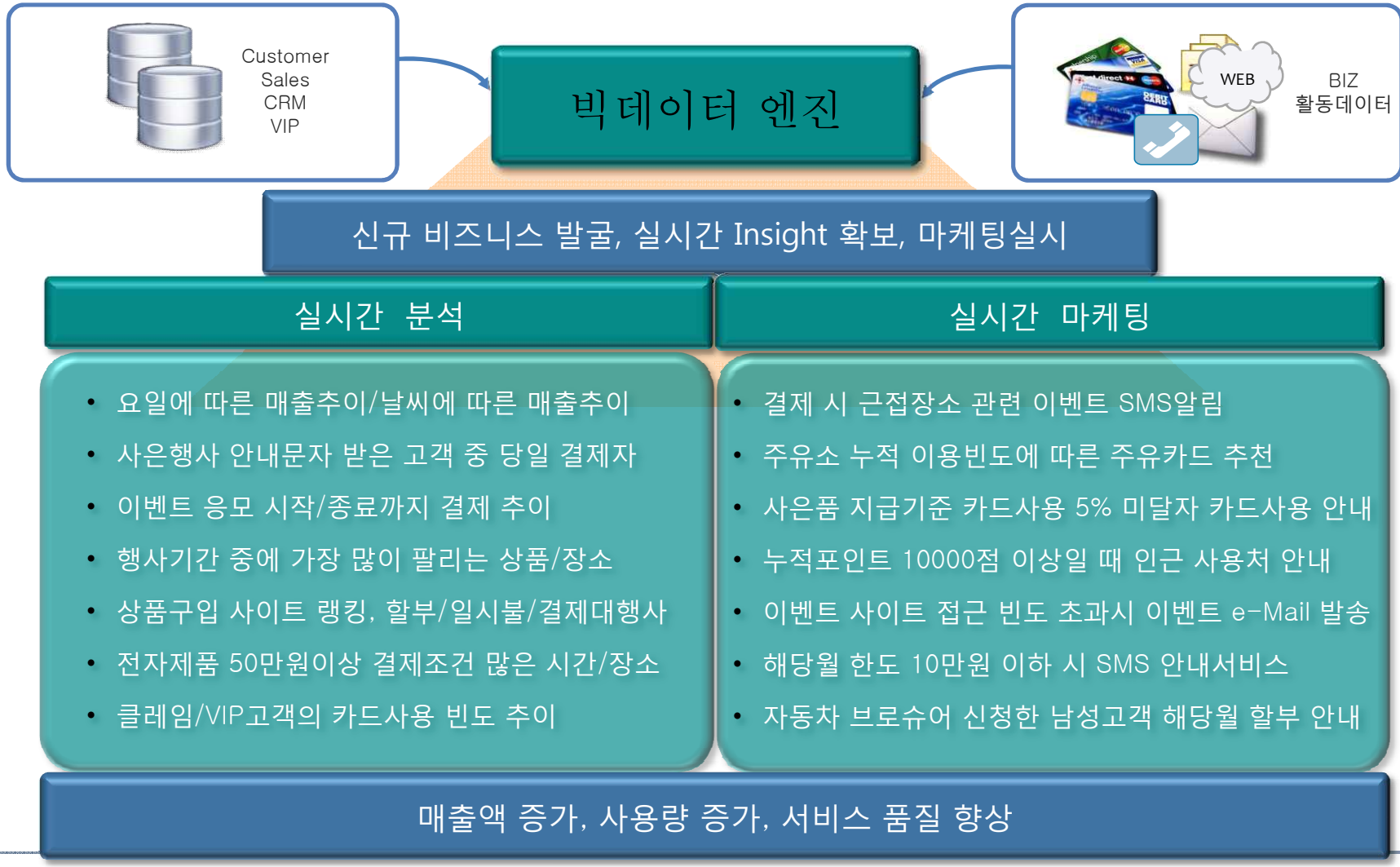
빅데이터 활용 기술을 RFID 센서 데이터에 활용하여 생산 및 물류 프로세스를 모니터링하고 생산 된 완제품에 대한 물류 가시성 확보,이동 및 유통단계의 물류 추적, 재고 관리를 통한 업무 효율 및 비용절감을 할 수 있도록 합니다.



© 2007-2012 GAION. All rights reserved

마케팅 분야 적용 방안

빅데이터 기술을 활용하면 실 시간성 비정형 데이터를 분석하여 신규 비즈니스 발굴, 실시간 인사이트 확보, 실시간 마케팅이 가능해집니다.

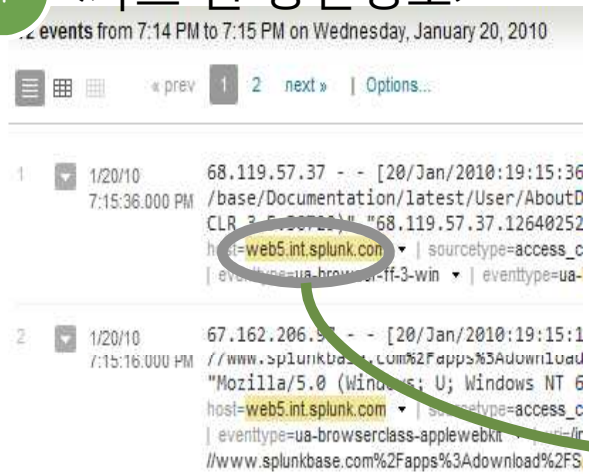


© 2007-2012 GAION. All rights reserved

마케팅 분야 적용 방안

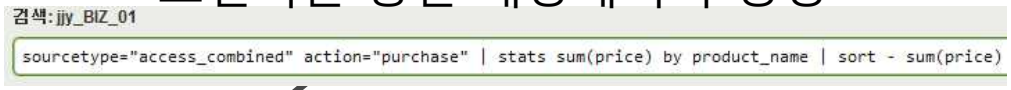
기업의 내부 축적데이터와 외부 축적데이터(Log)를 분석하면 새로운 비즈니스와 IT데이터 Intelligence 확보 및 자동화 마케팅 도구로 활용 할 수 있습니다.

1 <카드 원 승인정보>



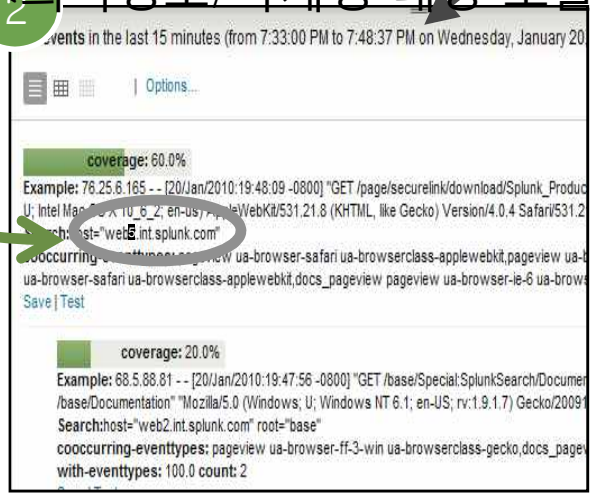
- Drill-down 기능을 통하여 고객 행동패턴 Log 정보 데이터를 분석

<조건식을 통한 대상데이터 생성>



<연계검색 조건>

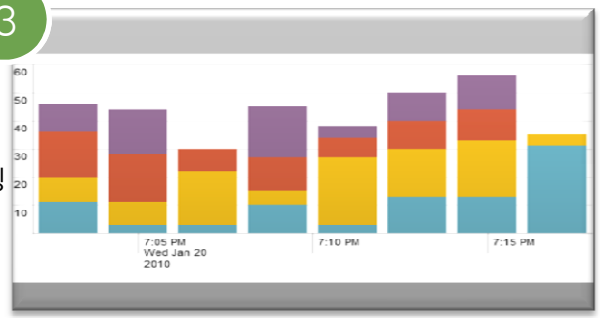
2 <의미정보/마케팅 대상 도출>



- 비쥬얼한 차트로 Dashboard로 현 트랜드를 인지 / SMS, 이메일등 다양한 방법의 마케팅 실시

3

대쉬보드 /모니터링



- 결제정보+고객정보, 결제정보+지역정보 등 여러 데이터의 상관관계를 선정한 후 연계

이벤트 / 알람



마케팅 분야 적용 방안

사용자UI, 검색 창을 통해 방대한 IT 데이터 중 여러 이종의 이벤트를 자유로이 추출 하고, 그 결과를 실시간으로 고객의 전자메일, SMS, Smartphone App 으로 프로모션 합니다.

BIZ System (고객DB)

Timestam | Card No | Customer info

2011-08-30 15:53:50 ID[66645] REGION [Seoul]Card No[9420902957798201 92109033 56228722] local1.infoINFORADOP(12) Tel [01053225255] SEX[M] TYPE[PROS] LEVEL[A0] Fav[CA2BB] Email[dilipjohn01@sprintpcs .com] via proxy 68.28.211.52[SprintPCS Group](ACK=1 RTT=83) recorded OK

- 고객정보(나이, 연락처, 지역, 성별 등)
- Sales , Finance, compliance, 요율, 영업정보, CRM

Opera-ting System (결제Log)

Timestam | Card No | Site

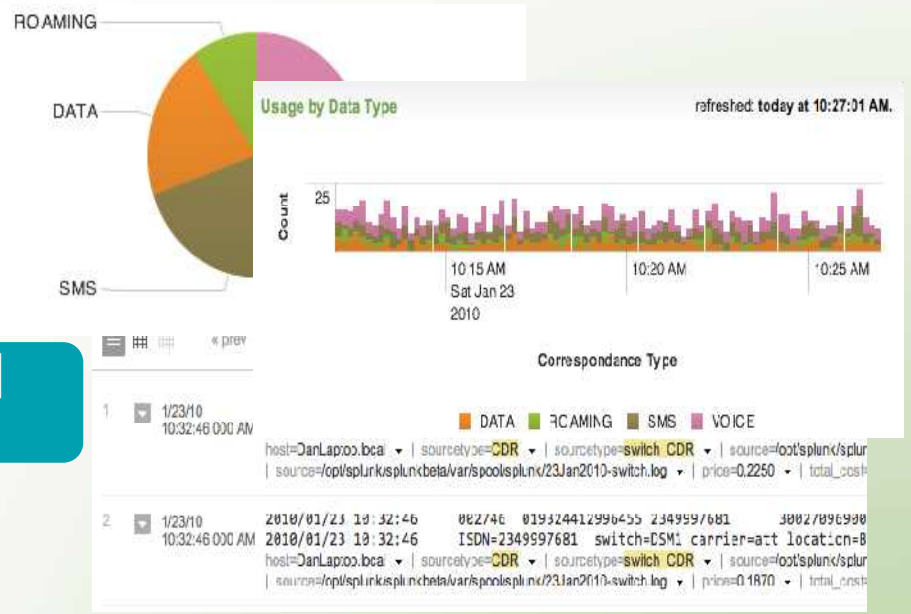
2011-08-30 15:53:50 SAT 9420902957798201 002 252 173.112.114.156 "신촌cgv&x=0&y =0&sm=top_hty&fbm=0&ie=utf8?product_id=K9-CW-01 &JSESSIONID=SD4SL3FF6ADFF3" -80 -10.157.118.161 BREW Applet/0x010A9D6D+(BREW/ 3.1.5.142;+DeviceId: ++2460;+Lang:+en) 200 0 0 679 2 80

Product ID

- 고객 성향 및 패턴 / 제품과 서비스 사용 내역 / 결제정보 / log
- End-to-end 트랜잭션 내역

빅데이터 엔진

실시간 매출 모니터링 및 마케팅 실시



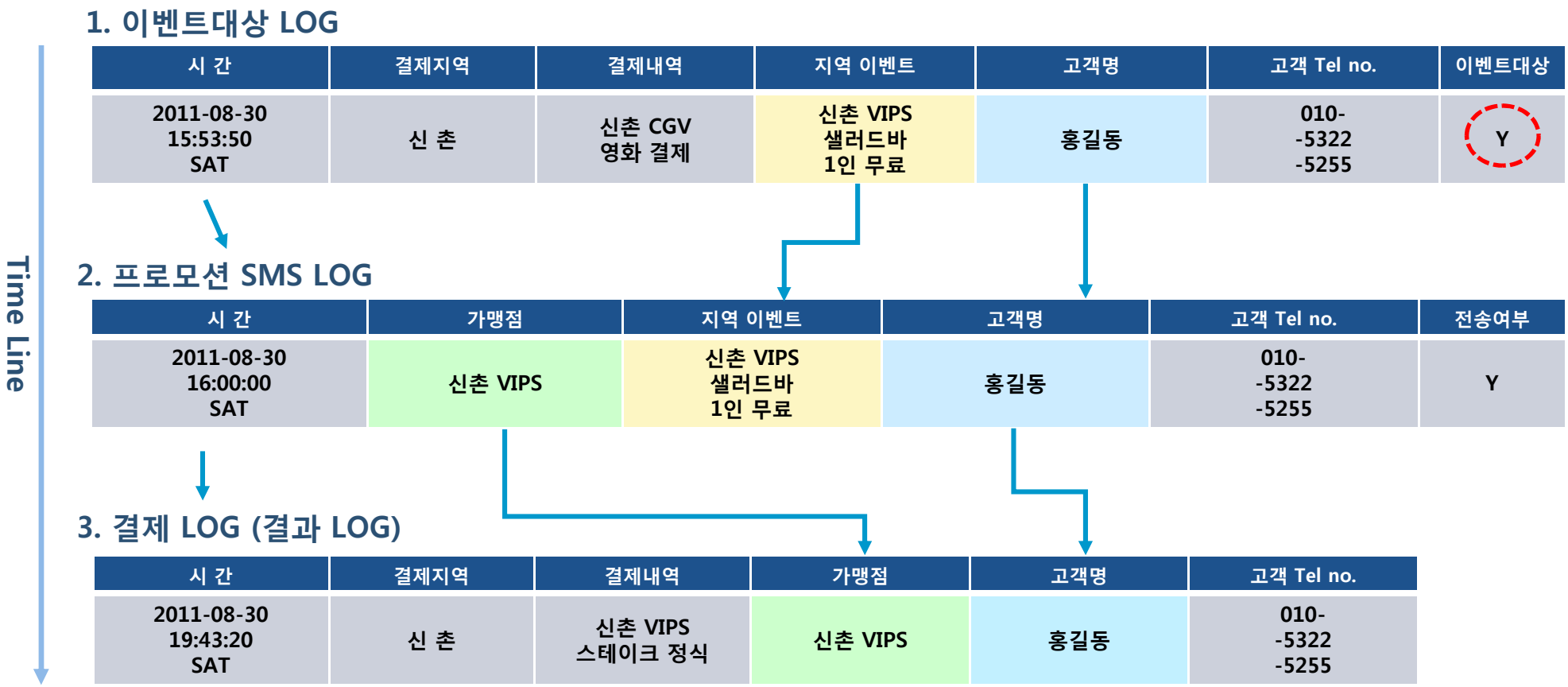
시 간	결제지역	결제내역	지역이벤트	고객 tel no.	대상
2011-08-30 15:53:50 SAT	신 촌	신촌 CGV 영화 결제	신촌 VIPS 샐러드바 1인 무료	010--5322 -5255	Y

"토요일 오후 영화관람 고객, 인근 음식점 이벤트 실시간 SMS 안내 "

© 2007-2012 GAION. All rights reserved

마케팅 분야 적용 방안

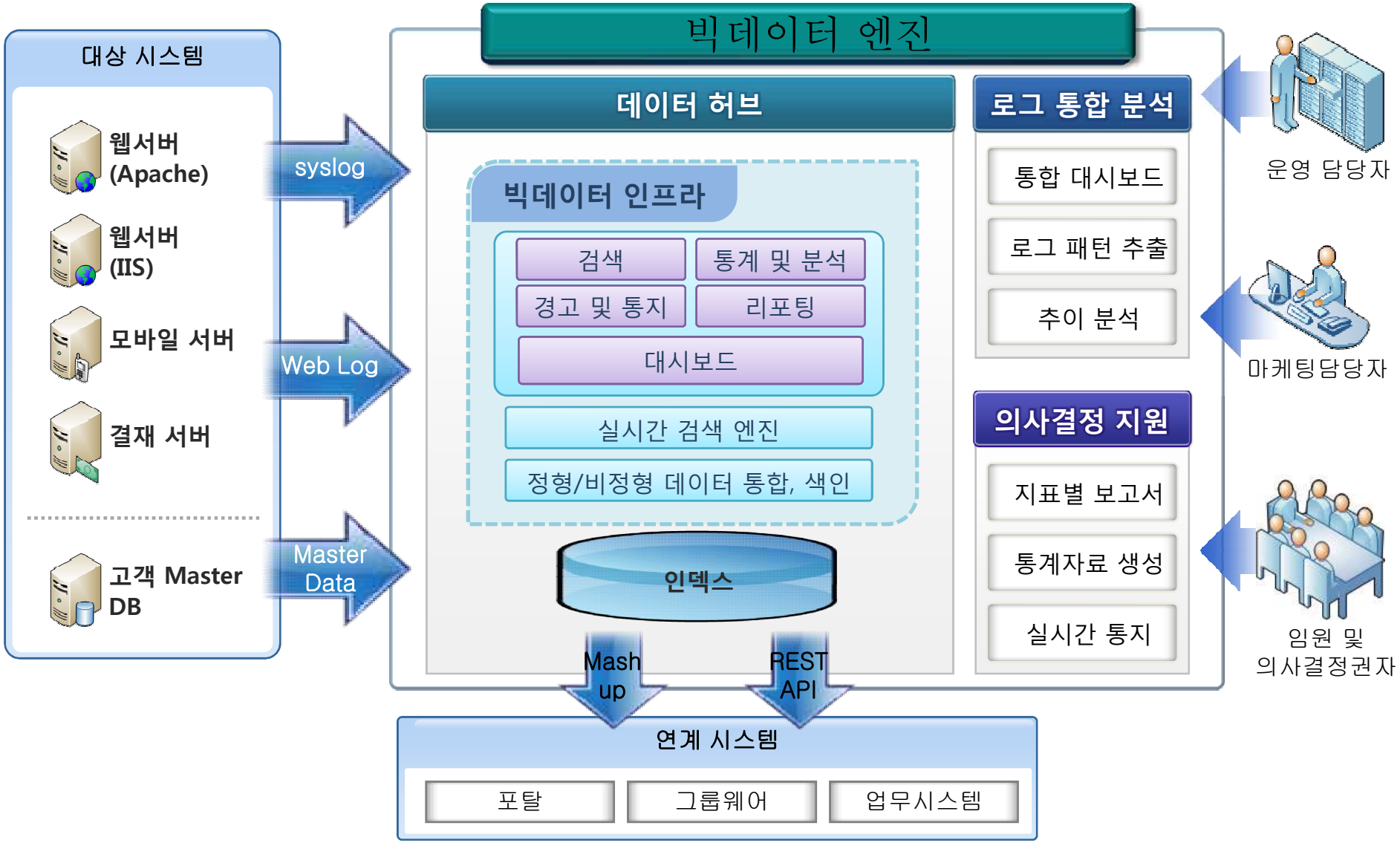
트랜잭션 데이터의 연결고리를 통해 마케팅효과를 실시간으로 확인이 가능합니다. 측정효과는 비주얼하게 다양한 형태의 대쉬보드 그래프로 표현될 수 있습니다.



"VIPS 이벤트 대상이었던 홍길동씨가 신촌VIPS에서 결제했음을 확인" → 이벤트 성공

© 2007-2012 GAION. All rights reserved

마케팅 분야 적용 방안



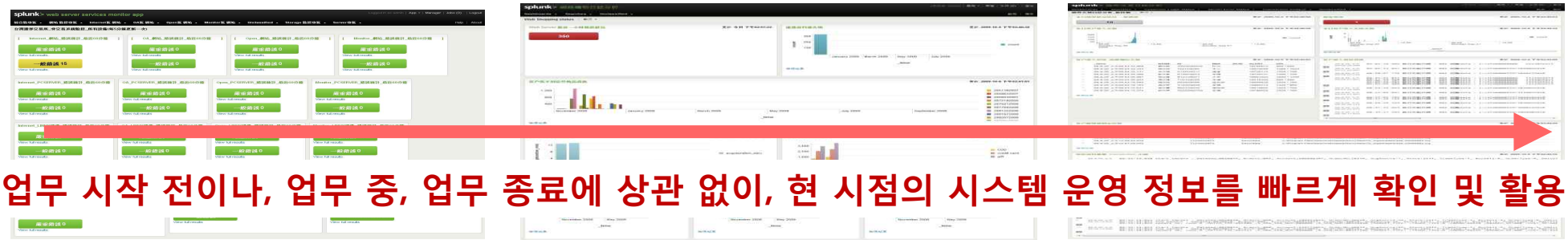
© 2007-2012 GAION. All rights reserved

IT 운영부문 적용 방안 – 관제

업무가 시작되기 전에서 종료될 때까지, 시스템 운영 상 발생되는 데이터에 대해 빠르게 분석하고 활용할 수 있습니다.

거래 또는 승인 데이터 등 하나의 정보를 확인 또는 사용하기 위해 많은 시간이 소요





업무 시작 전이나, 업무 중, 업무 종료에 상관 없이, 현 시점의 시스템 운영 정보를 빠르게 확인 및 활용

거래 시작 전:

전체 시스템, Network,
Application에 대한 **Health 분석**

거래 업무 중:

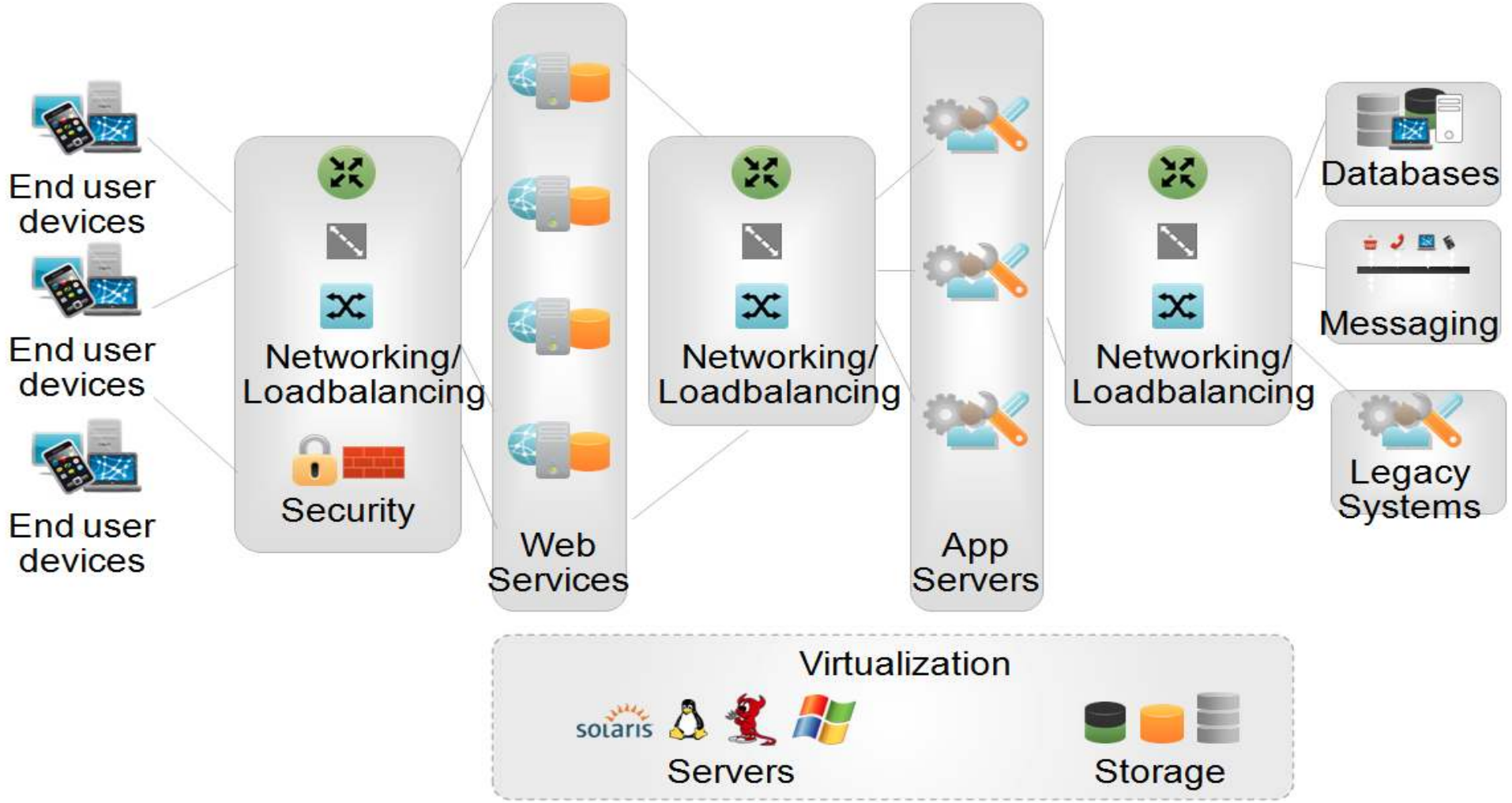
실시간 Transaction, 시스템 현황,
Application 로그 분석 등의
실시간 운용 분석

거래 종료 후 :

시스템 성능 및 보안
컴플라이언스 **리포트 생성**

© 2007-2012 GAION. All rights reserved

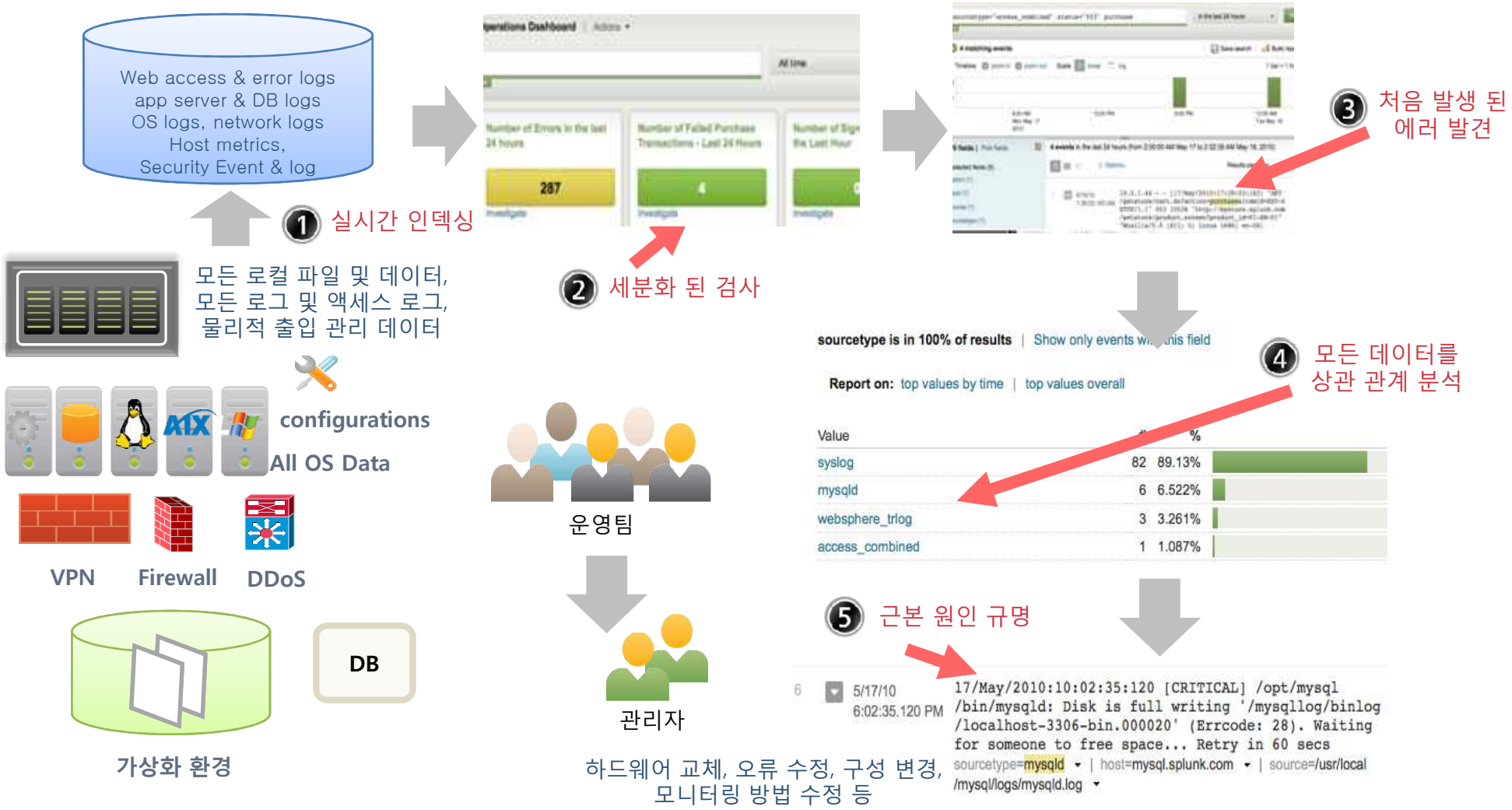
IT 운영부문 적용 방안 – 성능관리



© 2007-2012 GAION. All rights reserved

IT 운영부문 적용 방안 – 문제해결

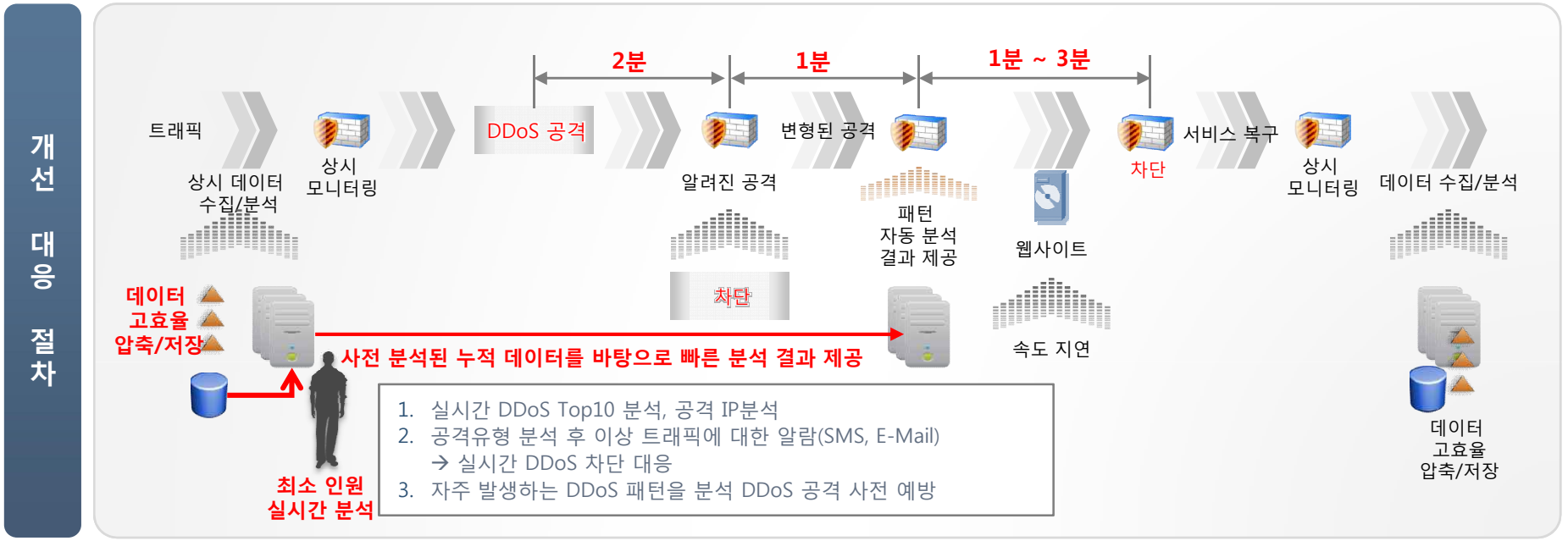
모든 정보 시스템 로그 통합을 통한 분석 효율화로 빠르고 정확한 원인 규명 및 조치가 가능합니다.



© 2007-2012 GAION. All rights reserved

보안 부문 적용 방안

빅데이터 기술을 활용하여 많은 양의 데이터를 분석해야 하는 보안 부문에서 사전 분석/대응으로 DDoS공격에 빠른 대처가 가능합니다.

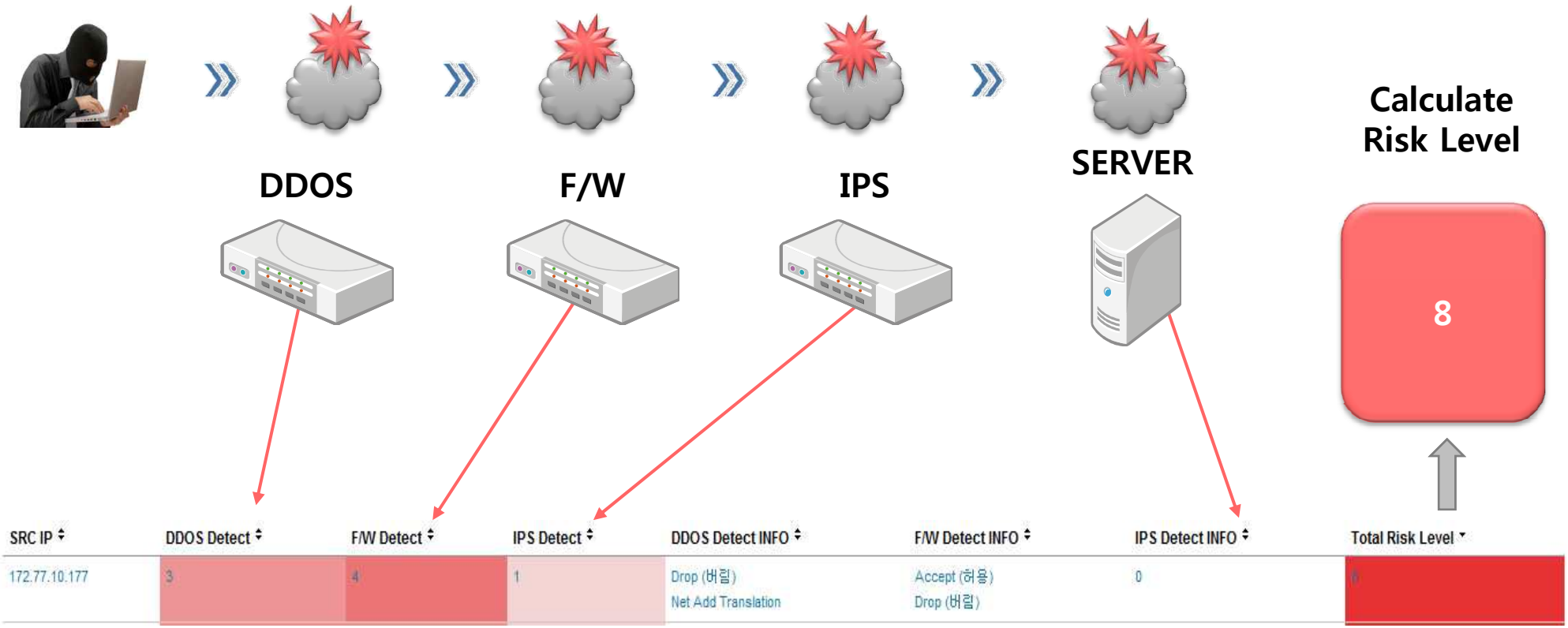


공격 패턴 변화	시스템을 활용한 분석	분석시간 50% 단축
맞춤형 화면 구성	신규 이슈 발생 시 화면 구성 즉시 변경	직관적 대응
상시 로그 저장	사전 공격 유형 분석	사전 분석/차단

© 2007-2012 GAION. All rights reserved

보안 부문 적용 방안

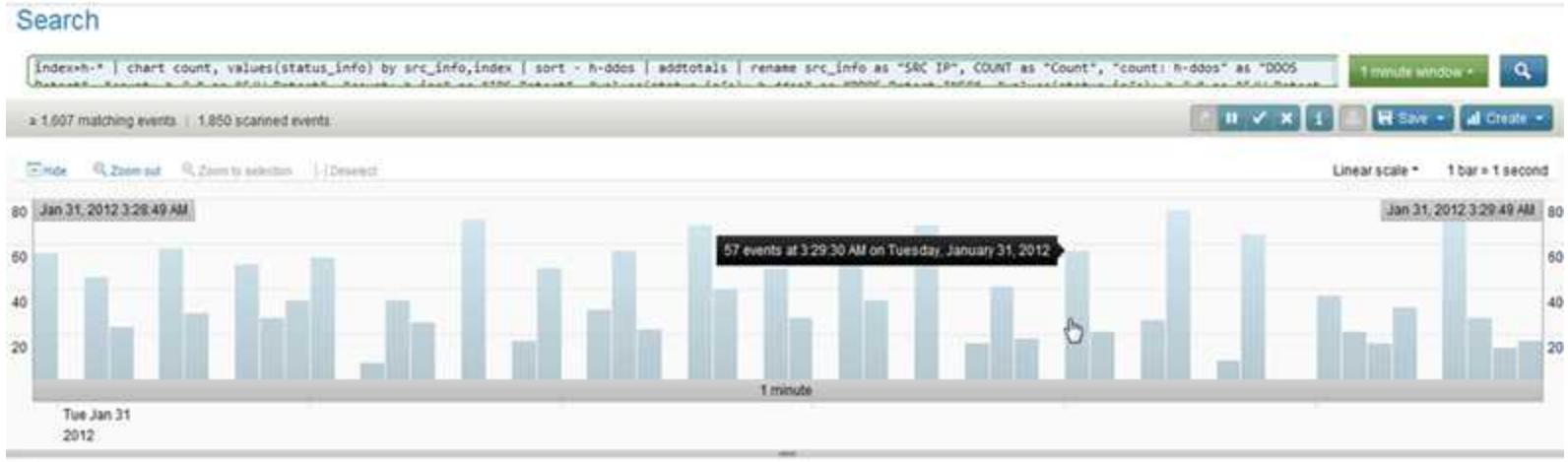
빅데이터 분석 기술을 활용하여 여러 보안장비들의 데이터를 분석하고 상관관계를 파악합니다.



© 2007-2012 GAION. All rights reserved

보안 부문 적용 방안

네트워크 데이터 및 호스트 파일 무결성 관리 데이터 등의 조합으로 상관관계 분석을 통하여 Risk Level 을 정의하여 모니터링 합니다.



	SRC IP	DDOS Detect	F/W Detect	IPS Detect	DDOS Detect INFO	F/W Detect INFO	IPS Detect INFO	Total Risk Level
1	172.77.11.153	3	2	2	Drop (버림) Reject (거부)	Accept (허용) Drop (버림)	Drop (버림) Reject (거부)	
2	172.77.12.190	4	1	2	Accept (허용) Reject (거부)	Accept (허용)	Reject (거부)	
3	172.77.13.72	2	1	4	0	Reject (거부)	Accept (허용) Reject (거부)	
4	172.77.10.41							
5	172.77.10							
6	172.77.11							

10 next » 20 per page

Total Risk Level

APT Matrix 상관관계 분석

- 공격자 SRC IP를 바탕으로 , 다양한 보안 모니터링 데이터에서의 동일한 공격자를 탐지 되었는지에 대해 분석
- F/W, IPS, DDos 시스템 등의 보안 서브 시스템에서 공격자 가중치를 정의하여 Risk Level 계산

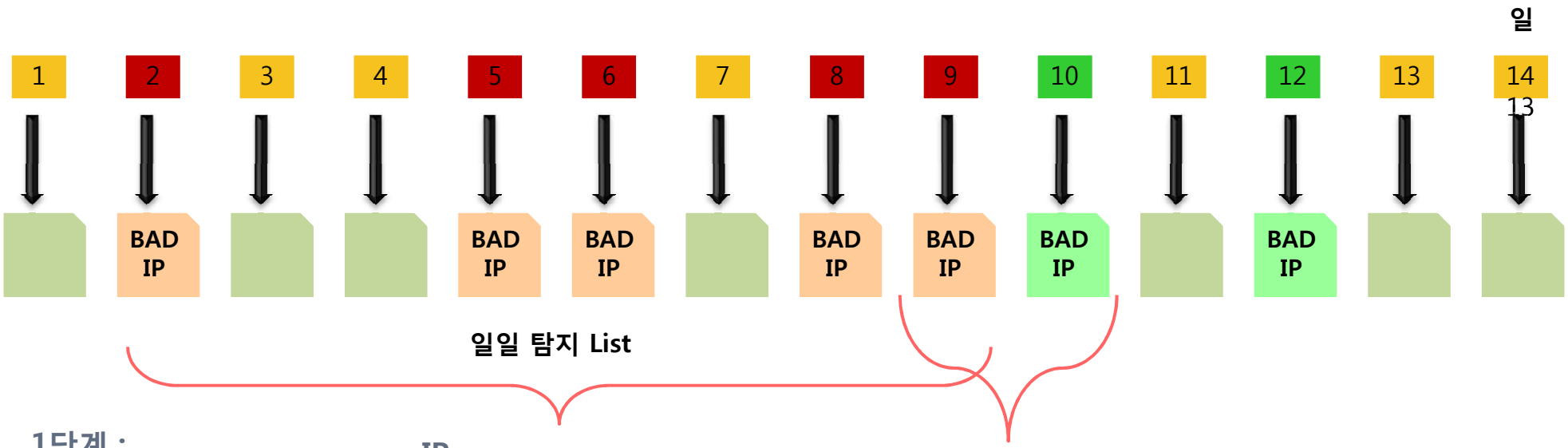
APT : Advanced Persistent Threat

© 2007-2012 GAION. All rights reserved

보안 부문 적용 방안

빅데이터 기술을 활용하여 장기간의 보안 관련 데이터를 분석하는데 적용할 수 있습니다.

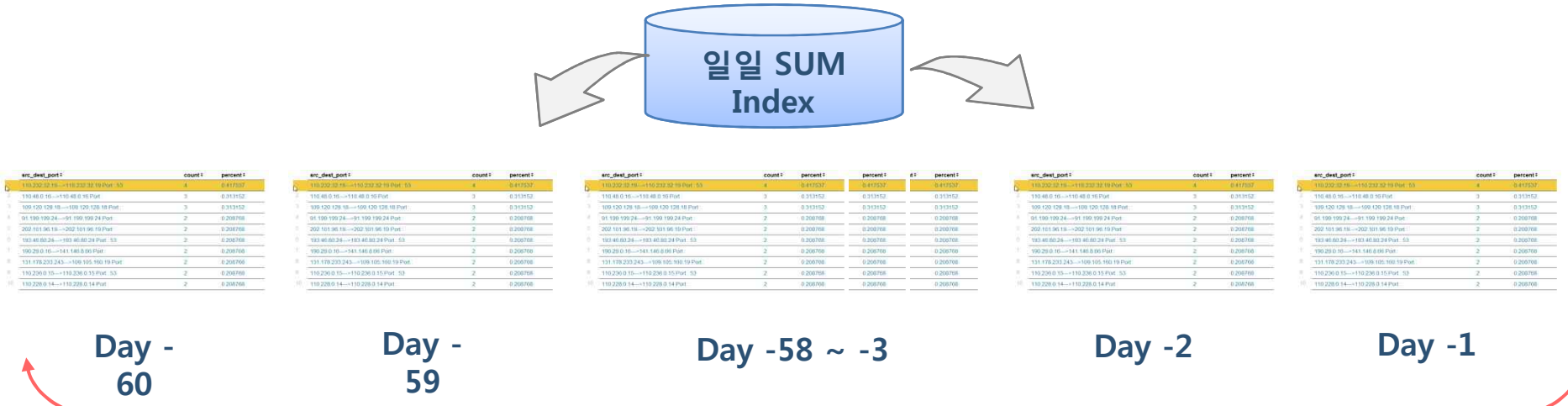
- **공격자의 행동을 이해하기 위해 모든 Allow/Deny 트래픽에 대하여 6개월 이상의 빅데이터를 처리**



© 2007-2012 GAION. All rights reserved

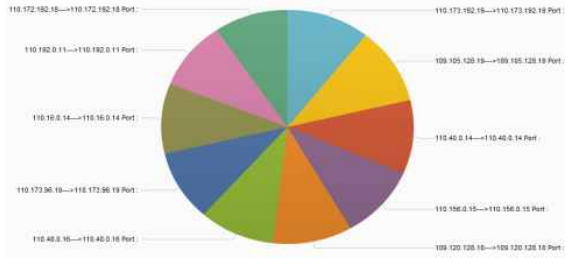
보안 부문 적용 방안

일일 단위의 분석을 통합하여, 구체적인 공격자의 행동을 찾아 임계값 이상 접근한 대상을 감지합니다.



상위 장기간 특정 HOST 및 서비스 접근 위험군

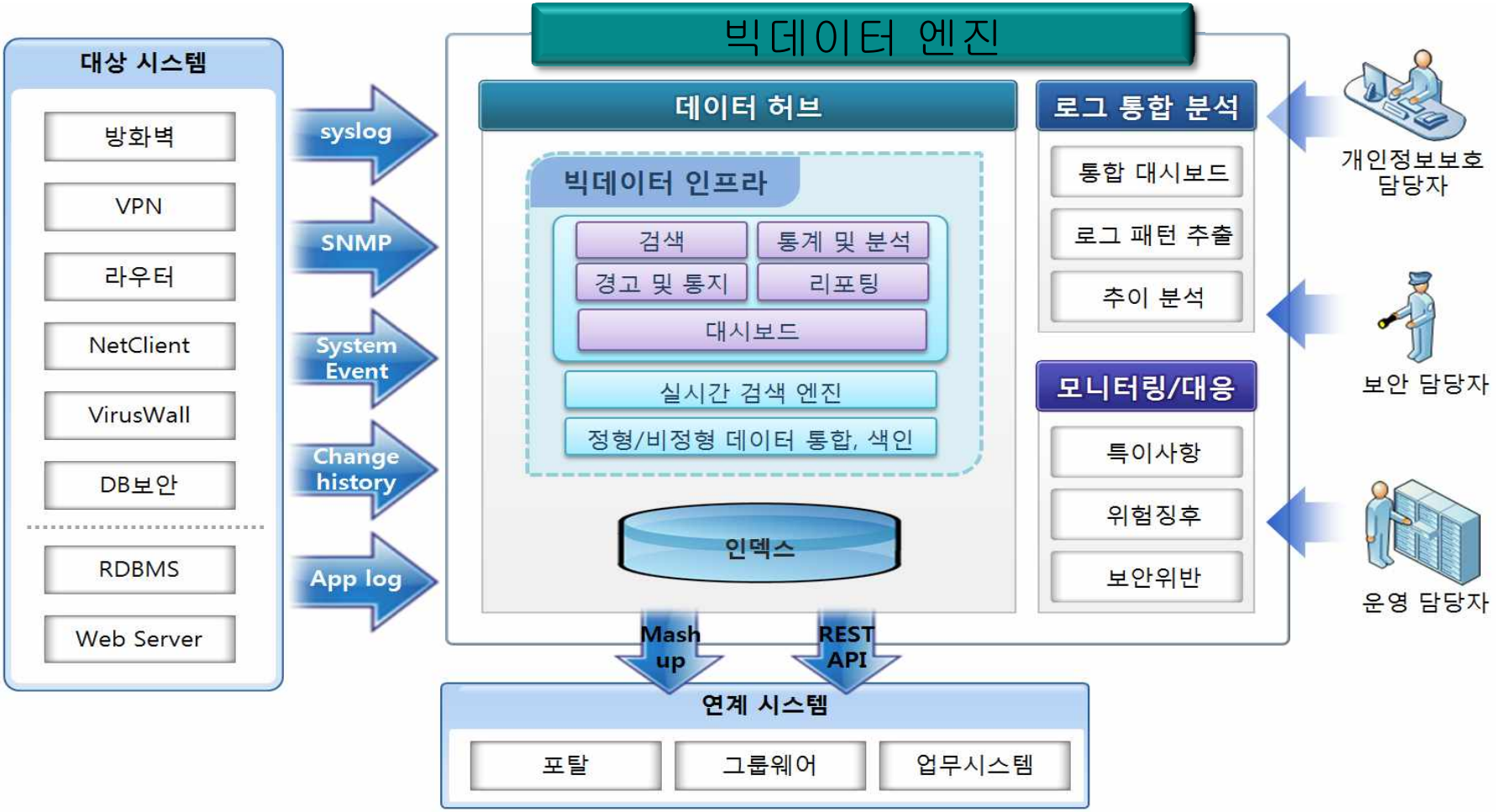
	src_dest_port		percent
1	110.173.192.19--->110.173.192.19 Port :	24	0.115196
2	110.228.0.14--->110.228.0.14 Port :	23	0.110396
3	110.184.0.13--->110.184.0.13 Port :	23	0.110396
4	110.152.0.14--->110.152.0.14 Port :	22	0.105597
5	110.48.0.16--->110.48.0.16 Port :	21	0.100797
6	109.120.128.18--->109.120.128.18 Port :	20	0.095997
7	110.240.0.12--->110.240.0.12 Port :	19	0.091197
8	110.40.0.14--->110.40.0.14 Port :	18	0.086397
9	110.236.0.15--->110.236.0.15 Port :	18	0.086397
10	110.232.32.19--->110.232.32.19 Port : 53	18	0.086397



© 2007-2012 GAION. All rights reserved

보안 부문 적용 방안

빅데이터 기반의 솔루션을 통해 기업은 체계적인 내·외부 현황분석을 통한 정보 보호 대응 모니터링 체계 구축 및 정보보안 업무의 시스템화, 체계화 기반을 확립할 수 있습니다.



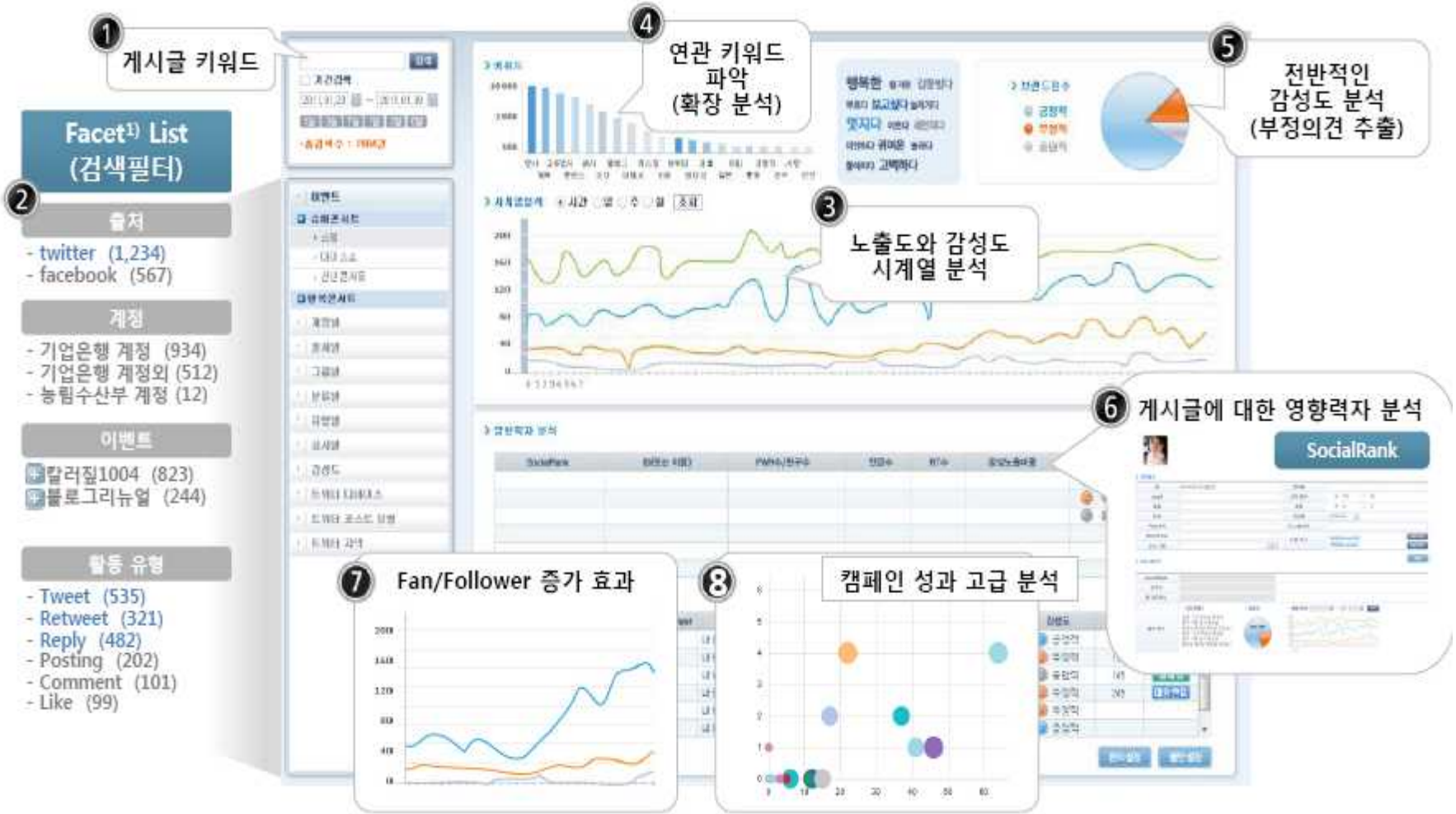
© 2007-2012 GAION. All rights reserved

소셜 분석 부문 적용 방안


STEP1 Establish Goals
STEP2 LISTEN!
STEP5 Analyze and Optimize
Protection
Promotion
PR
Brand Mgt
Civil Complaint
Ads
PR/Verbal
Policy
이슈는 무엇인가?
만족도나 선호도는?
어떤 이야기가 지속되나?
누가 부정적 영향력자인가?
누가 민원신청자인가?
민원의 원인은 무엇인가?
얼마나 펴졌는가?
누가 관심을 갖는가?
어떤 반응인가?
누가 영향력자인가?
누가 정책 대상인가?
누가 대중을 주도하는가?
1차 분석
노출도
사람 유형
관계 분석
키워드/속성 추출
감성 분석
성과 지표
2차 분석
지속성 (전파율)
도달율
영향력자 분석
경쟁 비교
성과 분석

소셜 분석 부문 적용 방안 – 게시판 분석

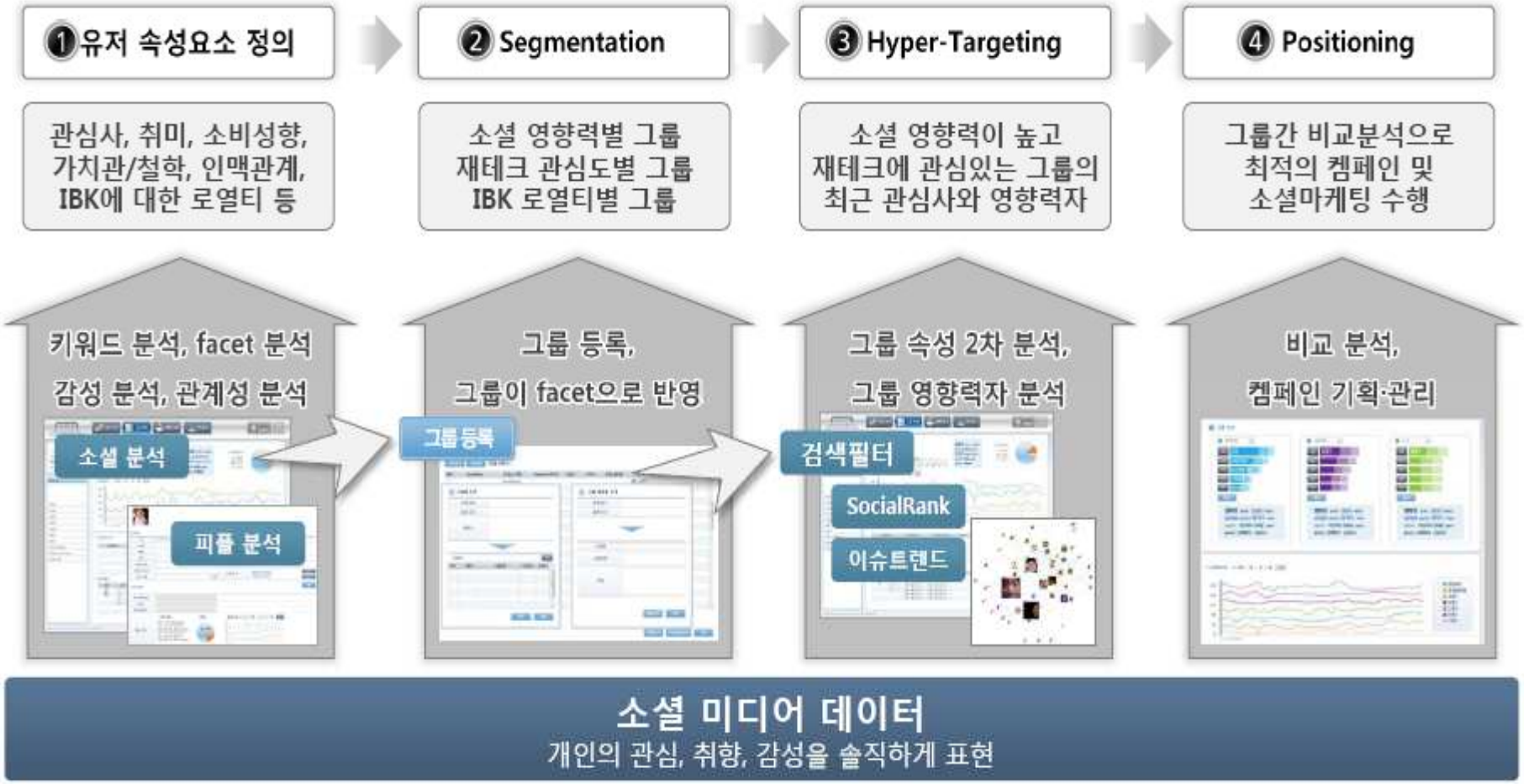
빅데이터 기술로 자사 게시판뿐만 아니라 경쟁사의 게시판 글 분석에 활용합니다.



1) Facet: 검색 필터와 동적인 분포 조회, 두 가지 기능을 제공하는 다양한 속성 정보

소셜 분석 부문 적용 방안 – 관계성 분석

STP 및 소비자행동, 영향력자 등의 분석을 정확한 데이터를 기반으로 하여 진행할 수 있습니다.



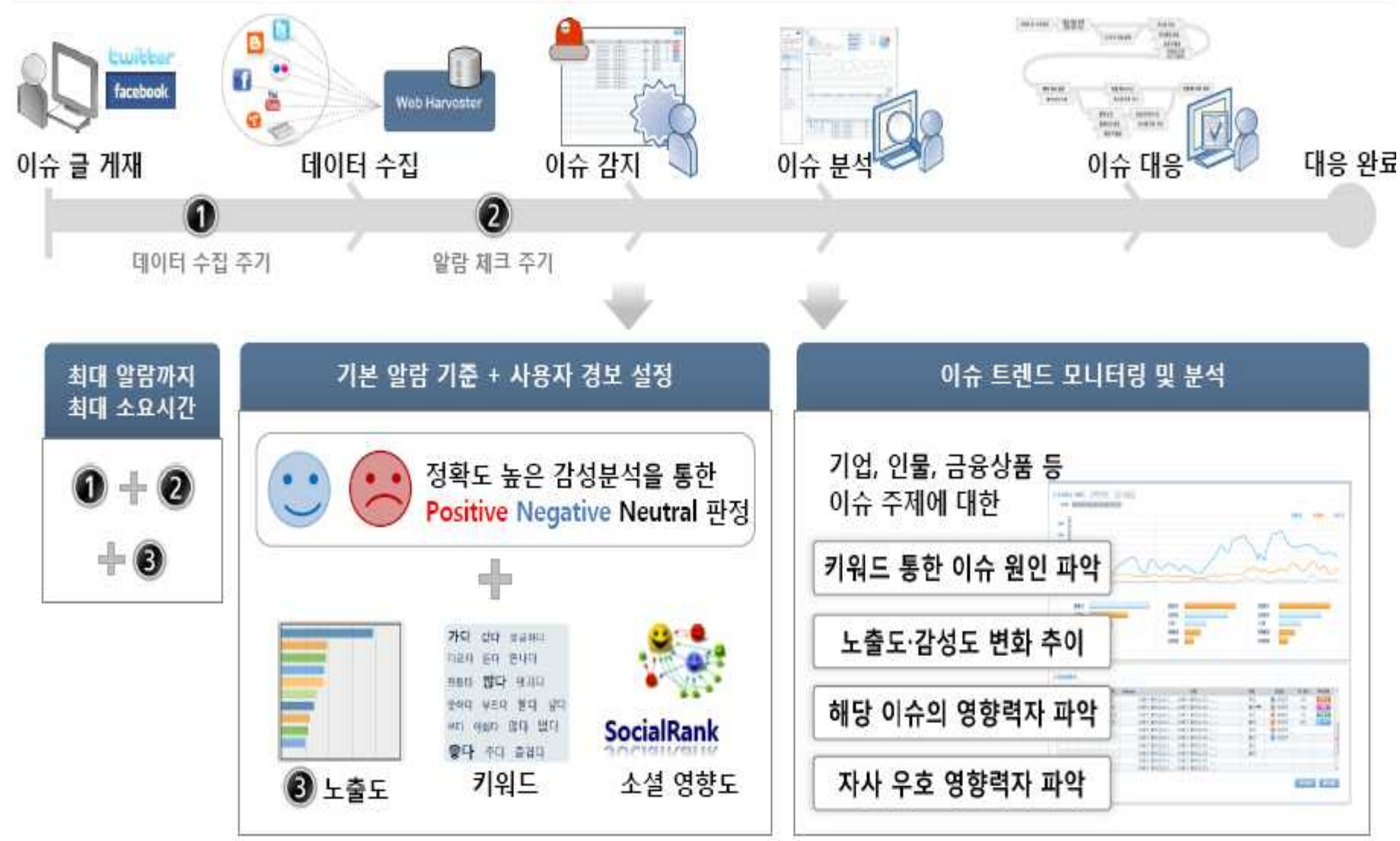
1) STP: Segmentation, Targeting, Positioning

소셜 분석 부문 적용 방안

정확한 데이터를 통해 마케팅 Insight를 얻을 수 있습니다.



소셜 분석 부문 적용 방안


twitter
facebook
Web Harvester
이슈 글 게재
데이터 수집
이슈 감지
이슈 분석
이슈 대응
대응 완료
1
데이터 수집 주기
2
알람 체크 주기
최대 알람까지
최대 소요시간
1 + 2
+ 3
기본 알람 기준 + 사용자 경보 설정
정확도 높은 감성분석을 통한
Positive Negative Neutral 판정
+
3 노출도
키워드
SocialRank
소셜 영향도
이슈 트렌드 모니터링 및 분석
기업, 인물, 금융상품 등
이슈 주제에 대한
키워드 통한 이슈 원인 파악
노출도·감성도 변화 추이
해당 이슈의 영향력자 파악
자사 우호 영향력자 파악

기대효과 – 비즈니스 측면

빅데이터 분석을 통해 데이터에 기반한 비즈니스 의사결정 가능



분석체계 표준화

- 빅데이터 분석기법 및 체계 표준화를 통한 활용 능력 향상
- 실시간 분석 역량 확보를 통해 기업의 대내외 경쟁력 강화

기업 경쟁력 강화

- 경쟁업체 정보 및 추이 분석을 통한 기업 경쟁력 향상
- 고객의 기업이미지 모니터링을 통한 개선 활동 수행

마케팅 업무 활용

- 사용자 패턴 및 지표 분석을 통한 마케팅 전략 수립
- 객관적인 지표를 통한 마케팅 활용도 제고

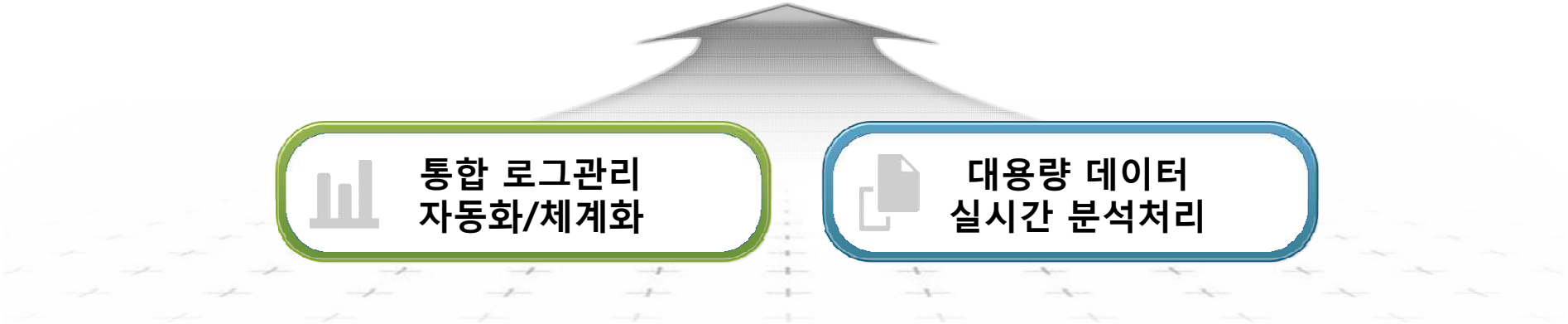
의사결정 능력 향상

- 광범위한 기업 내·외부 데이터에 근거하여 의사결정의 정확도 향상

© 2007-2012 GAION. All rights reserved

기대효과 – IT 측면

데이터의 생성-수집-검색-통계/분석-보고의 표준화로 빅데이터 관리체계 수립



실시간 처리

- 실시간 데이터 수집 시 수집대상 시스템 영향 최소화
- 실시간 이벤트 탐지 및 모니터링 지원

유연한 대용량 처리

- 대용량 데이터(빅데이터) 처리에 무한한 확장성 지원
- 안정적인 처리속도 제공

용이한 상관관계 분석

- 다양한 Rule기반의 상관관계 분석 가능
- 이벤트 상관관계 적용을 통한 확장성 제공

편리한 검색

- 정형/비정형 데이터 분석 및 검색 기능
- 검색어 조합을 통한 상관 검색 가능

© 2007-2012 GAION. All rights reserved

빅데이터를 활용한 가치 창출 방법

기업에서는 빅데이터를 이용해서 어떠한 가치를 만들어 갈 수 있는지 고민할 수 밖에 없습니다.
해답이 보이지 않을 때 아래의 3단계 과정을 밟아감으로써 실마리를 찾을 수 있습니다.

☆ **빅 데이터 가치 창출을 위한 3단계**

1. 기업 데이터 저장소에 적용할만한 가치가 있는 데이터 소스를 확인하라
(데이터 소스가 내부인지 외부인지는 중요하지 않다)

2. 강력한 확장형 데이터 관리 기반을 설립하라
(데이터 관리의 표준화 – 빅데이터 허브 구축)

3. 데이터 상부에 새로운 분석 툴을 적용해 새로운 시각을 확보하라
(빅데이터 분석 솔루션을 통한 통계,분석, 레포트화)

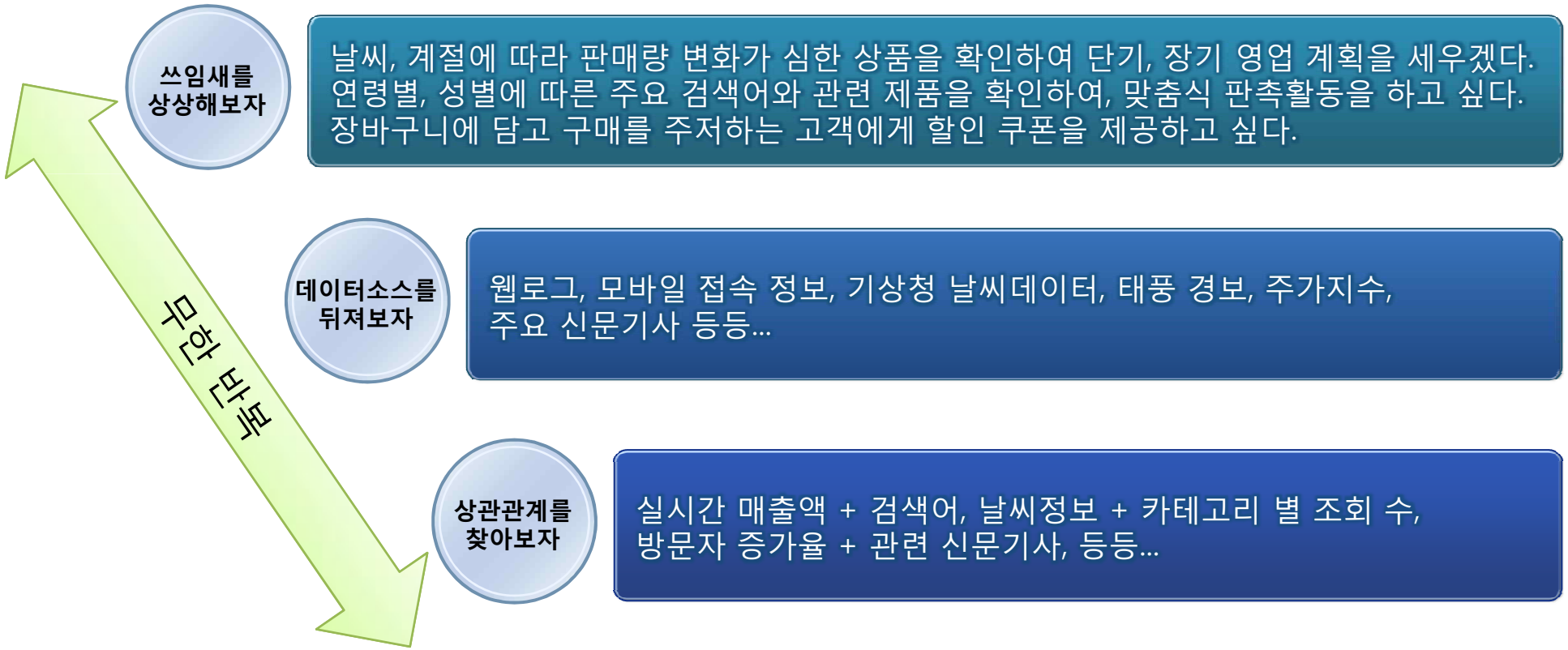
© 2007-2012 GAION. All rights reserved

빅데이터 가치 창출 시나리오

빅데이터 기반 기술은 중요하지 않습니다.

이를 활용할 수 있는 무궁무진한 아이디어와 창의력, 그리고 그것을 즉시 실행에 옮길 수 있는 힘이 중요한 것입니다.

☆ **빅 데이터 가치 창출을 시나리오**



© 2007-2012 GAION. All rights reserved

감사합니다.

GAION
강현섭 (hskkang@gaion.kr)