

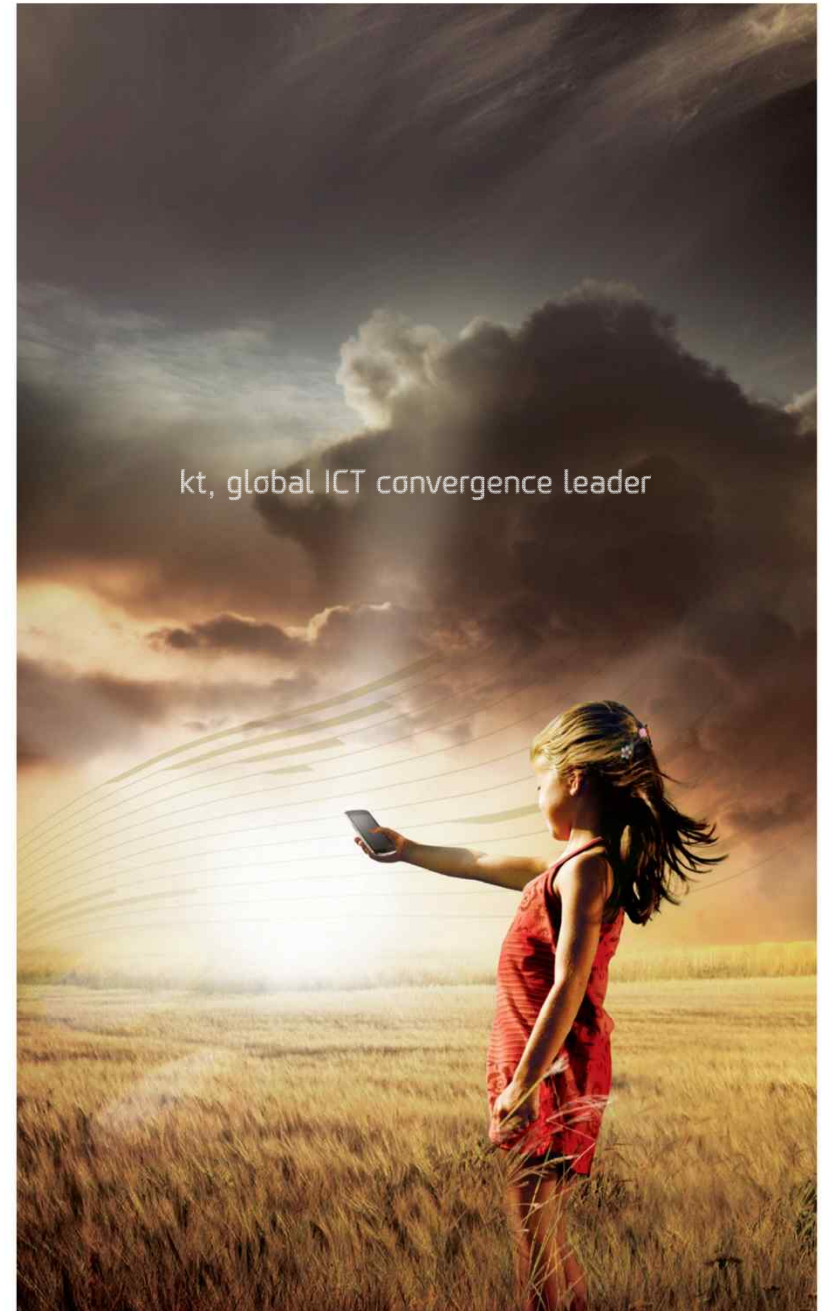
<창조인재 육성을 위한 빅 데이터 구현 사례 및 비즈니스 인사이트> 세미나
한국과학기술단체총연합회 국제회의실, 2013년 5월 2일, 13:45~17:40 (발표 13:55~14:45)

빅 데이터와 비즈니스 인사이트

송민정

| 2013. 5. 2

kt



Contents

1 구글과 빅 데이터

2 빅 데이터와 생활/사회/산업의 변화상

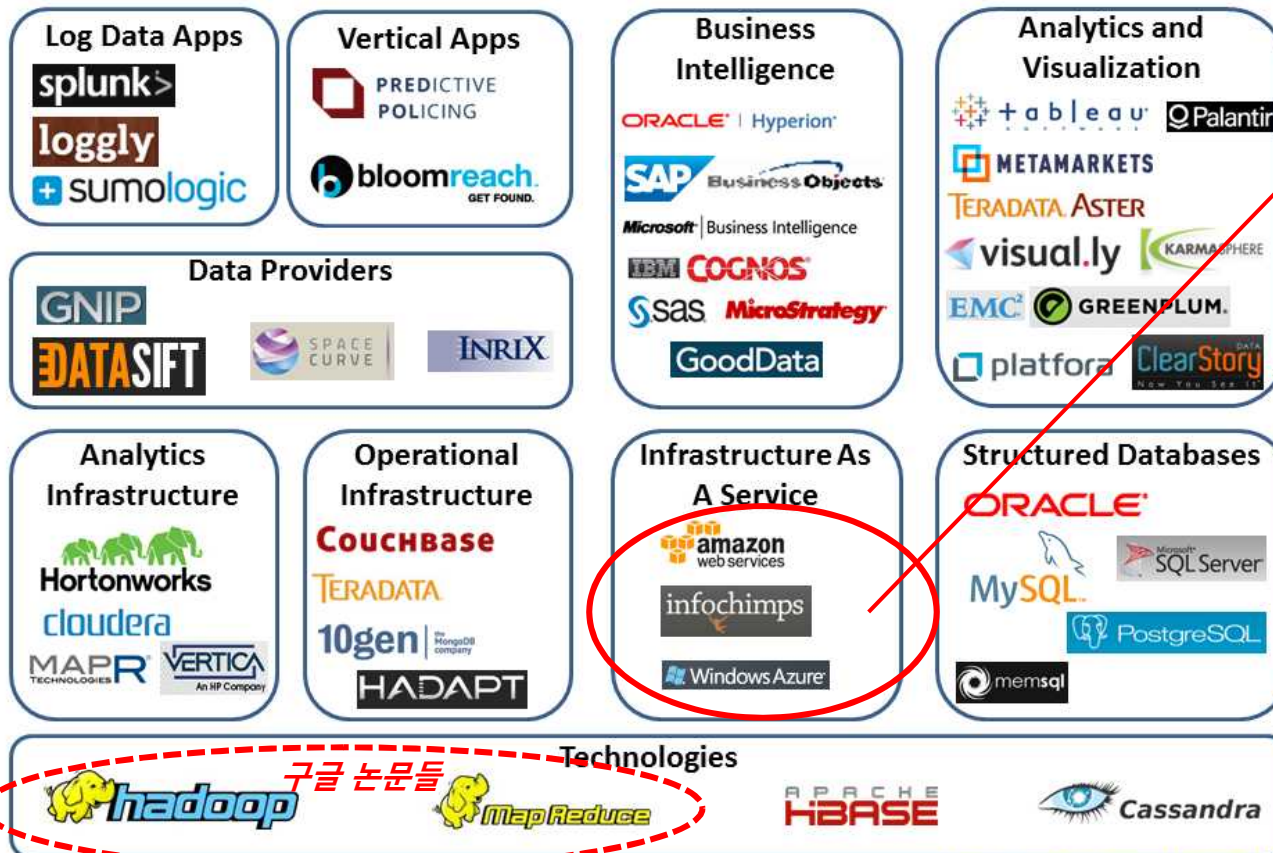
3 예측분석을 통한 비즈니스 인사이트

4 종합 요약

01 빅 데이터 업계지도에서 관찰된 구글위치는 인프라 & 분석툴

- 포스트닷컴의 기술 전문 칼럼니스트, 데이브 파인라이브(Dave Feinleib)가 만들어 공개한 빅 데이터 관련 업계지도를 보면, Google BigQuery와 hadoop이 구글 관련

Big Data Landscape



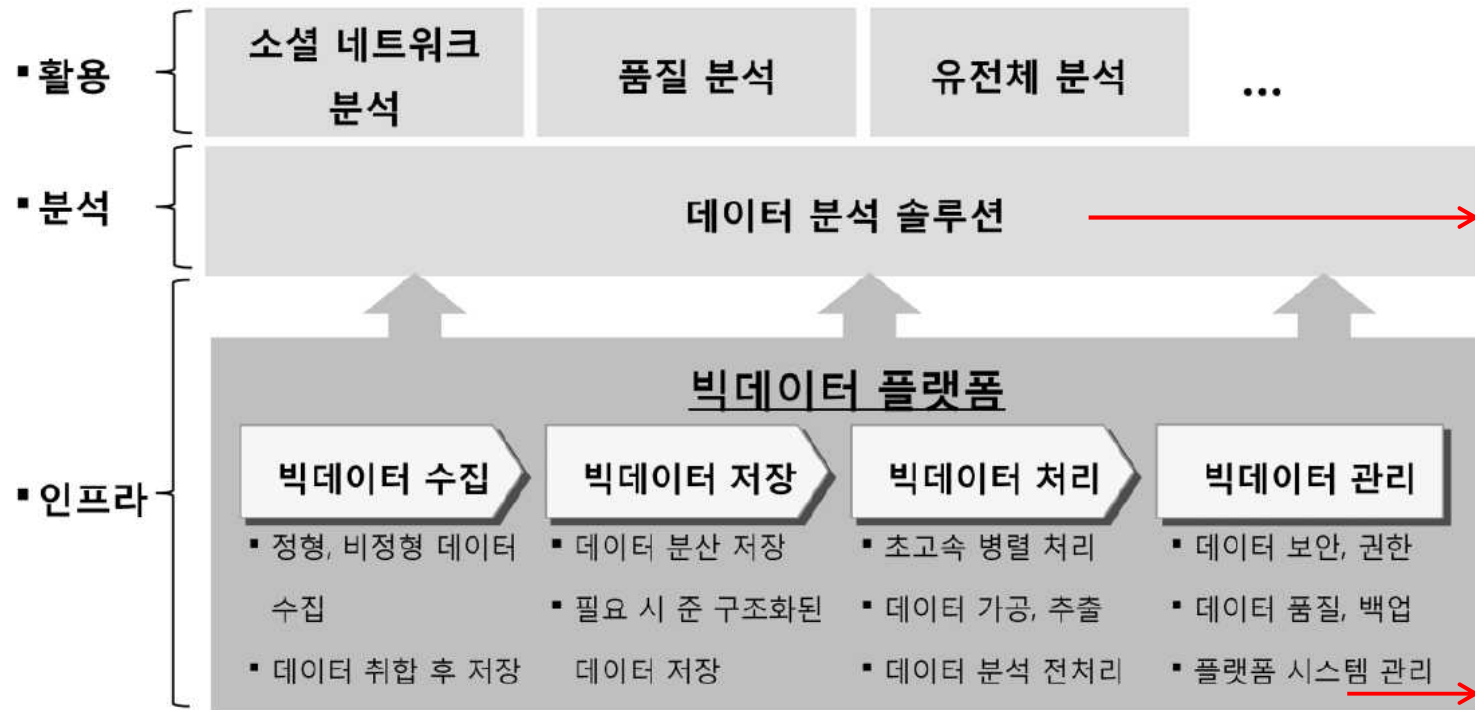
Google bigquery

클라우드 기반의 빅데이터 분석 툴인 '빅쿼리'는 고객에게 그래픽 UI 기반 SQL* 분석솔루션을 제공하며, 특히 분산된 개별 데이터 분석 결과 (fine-grained analysis)를 요약이나 통합 과정 없이 모두 제공하므로 고객이 직접 각 분석 결과를 파악하고 판단할 수 있음
*SQL(structured query language) : 데이터베이스에 접근할 수 있는 하부 언어

02 구글은 인프라 개선 통해 빅 데이터 플랫폼 진화 방향 제시

- 빅 데이터 플랫폼은 빅 데이터 기술의 집합체, 기술을 잘 활용하도록 준비된 환경
- 오픈소스 하둡(hadoop)의 근간이 된 논문을 발표한 구글은 빅 데이터 플랫폼의 대부
- 빅 데이터 플랫폼 인프라의 발전을 유도한 구글은 데이터 분석 솔루션 개발에 집중

[빅 데이터 플랫폼의 역할과 기능]*



[구글의 인프라 활용]

웹 검색 순위 평가를 위해 웹 페이지 링크 분석에 기반한 페이지 랭크 (Page Rank; 웹과 같은 링크 구조를 가지는 문서의 연결 분석으로 상대적 중요도에 따라 가중치를 부여하는 방법) 알고리즘을 사용하며, 매일 1조가 넘는 웹 페이지를 분석하기 위해 그래프 연산을 수행함*** SNS 분석에 그래프연산 적용

Spanner(New SQL) 자체 개발 고속병렬연산에 MPI 적용

스팸 분석, 웹문서 분석, 구글플레이 데이터 분석에 활용**

자료: *이주열, 빅 데이터 플랫폼의 미래(2013. 1)

**Google, Dremel: Interactive analysis of Web-scale datasets, 2010

***Google, Pregel: A System for Large-scale Graph Processing, 2009



MPI(message passing interface)

일괄처리방식 극복

03 구글과 빅 데이터 분석 대표 사례인 독감 트렌드, 보조역할만

- 빅 데이터 대표사례로 구글의 자사 검색 데이터 활용 독감 유행 수준 예측이 거론됨
- 미국 질병통제예방센터 데이터와 비교 결과 구글 예측의 신뢰성 높게 평가(NYT 등), 구글은 2008년 부터 '독감 트렌드'라는 독감 예측 서비스를 선보이고 있음.
- '13년 초 미 50개 주 중 47개에 퍼진 '살인독감'에 대해 구글 트렌드가 질병통제국 (CDC) 수치의 두 배 넘는 예상치를 발표하자, 2009년 논문 발표를 통해 상관관계를 인정했던 네이처가 빅 데이터 분석 결과가 항상 옳은 건 아님을 인정(Nature, '13.2)

nature International weekly journal of science

Access
To read this story in full you will need to login or make a payment (see right).
nature.com > Journal home > Table of Contents

Letter

Nature 457, 1012-1014 (19 February 2009) | doi:10.1038/nature07634; Received 14 August 2008; Accepted 13 November 2008; Published online 19 November 2008; Corrected 19 February 2009

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg¹, Matthew H. Mohebbi¹, Rajan S. Patel¹, Lynnette Brammer², Mark S. Smolinski¹ & Larry Brilliant¹

1. Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA
2. Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA

Correspondence to: Matthew H. Mohebbi¹ Correspondence and requests for materials should be addressed to J.G. or M.H.M. (Email: flutrends-support@google.com).

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year¹. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities². Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza^{3,4}. One way to improve early detection is to monitor health-seeking behaviour

ARTICLE LINKS

- Figures and tables
- Supplementary info

SEE ALSO

- Editor's Summary

ARTICLE TOOLS

- Send to a friend
- Export citation
- Export references
- Rights and permissions
- Order commercial reprints
- Bookmark in Connotea

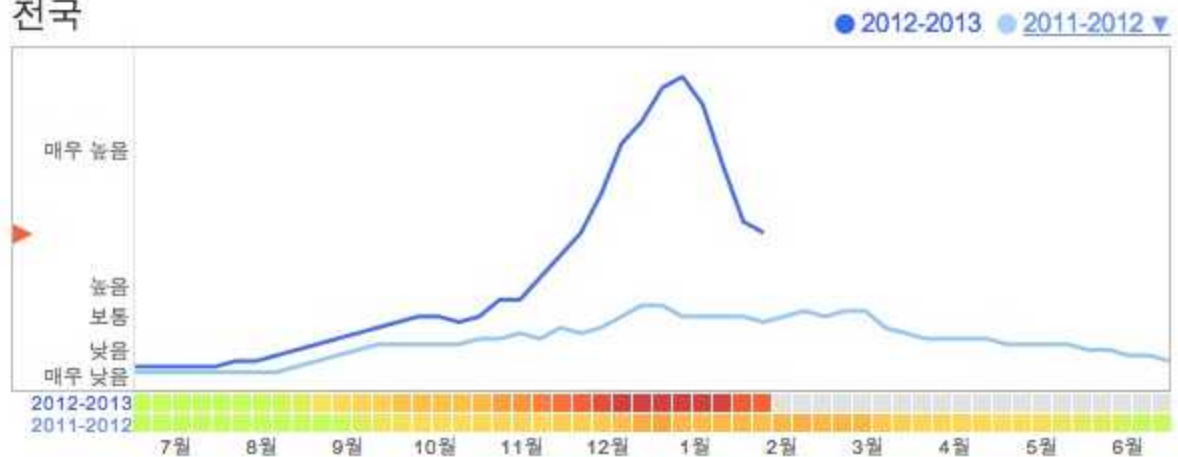
SEARCH PUBLISHED FOR

- Jeremy Ginsberg
- Matthew H. Mohebbi
- Rajan S. Patel
- Lynnette Brammer
- Mark S. Smolinski

독감 트렌드 - 미국

독감 유행 수준을 파악할 때 검색어가 좋은 지표가 될 수 있습니다. Google 독감 트렌드는 집계된 Google 검색 데이터를 사용하여 독감 유행 수준을 예측합니다. [자세히 알아보기 »](#)

전국



자료: Nature(2009); Google.org(2013), Nature(2013. 2.13)

Contents

1 구글과 빅 데이터

2 빅 데이터와 생활/사회/산업의 변화상

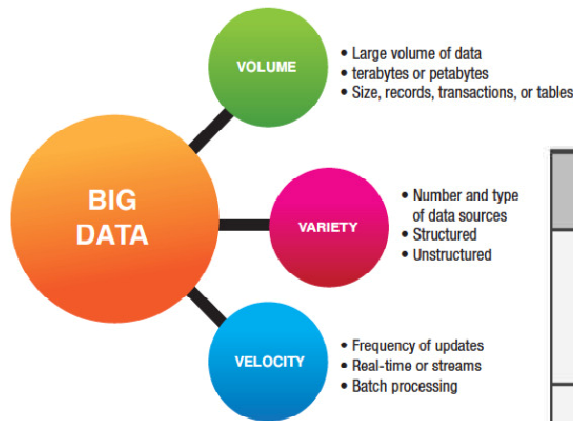
3 예측분석을 통한 비즈니스 인사이트

4 종합 요약

01 3V 특성 가진 빅 데이터의 현재 한계 극복을 통한 진화 기대

- 3V의 진화 방향은 실시간 처리 속도, 처리 방식의 다양화, 관계형 데이터 모델 지원
 - 구글에서 보듯이, 일괄 처리에 의한 결과 대기 아닌 즉각적 실시간 빅 데이터 처리, 분할 병렬 처리 및 병합 후 단순 데이터 처리 아닌 대규모 계산 등의 일반 연산 처리, 스키마 없는 단순 모델 아닌 관계형 데이터모델 등
- 현재의 오픈소스 하둡의 한계를 일부 극복해나가는 상용 빅 데이터 플랫폼 등장 기대

The Three Vs of Big Data



[미래 빅 데이터 플랫폼의 진화 방향: 3V 특성을 충분히 충족]

빅데이터 특성	한계 및 제약	진화 방향
용량 (Volume)	· 데이터센터 규모의 관리 한계	· 범 지구적 규모로 확장 가능 (지역 거점 별 데이터센터 확장)
다양성 (Variety)	· 특정 분산병렬 처리 방법 제공 · 관계형 데이터 모델 미지원	· 다양한 분산병렬 처리 방법 제공 · 관계형 데이터베이스 및 트랜잭션 제공
속도 (Velocity)	· 일괄처리(실시간 처리 미지원)	· 실시간 데이터 조회와 처리

02 빅 데이터로 인한 생활 변화상: 건강하고 안전한 행복 추구

- 건강(Healthcare)을 돕는 빅 데이터("Big data in health care hype & hope" 2012)
 - 천식 환자를 트래킹하고 정보를 수집, 이를 활용하는 Asthmapolis (천식 나라)
 - : 모바일 센서를 천식 흡입기에 부착하고 iOS/Android 앱을 통해 증상, 시간, 위치를 모니터링.
 - 제공자는 위험 환자 식별 외에 역학 연구 및 공중 보건에 활용할 수 있는 실시간 데이터를 수집
 - SNS를 통한 질병 관측 지도를 보여주는 Sickweather (아픈 날씨)
 - : SNS를 스캔하여 질병 발생을 추적하고 날씨 예보처럼 사용자들에게 예보해줌. 사용자들은 자신의 지역에 어떠한 일이 일어나는지를 입력할 수도 있고, 저렴한 가격으로 약품 광고를 지원(현재 영어권만 지원 중)



자료: [http://hongiv.tistory.com/737\(2012.10\)](http://hongiv.tistory.com/737(2012.10))

03 빅 데이터로 인한 사회 변화상: 개인 맞춤형의 사회 생활

- 사회 생활, 남녀 만남 등을 돕는 빅 데이터(함유근·채승병, 빅데이터 경영을 바꾸다 2012)
 - 태그드닷컴은 특히 관계정보 데이터를 분석, 실시간 맞춤형 데이팅 서비스를 제공. 2004년 시작하여 2011년 현재, 220개국, 3억의 멤버를 가지게 됨(2008년 흑자) 서비스는 새로운 사람들과의 만남에 초점, 소셜게임, 가상의 선물 등을 활용하게 함

자료: tagged.com 홈페이지; Crunchbase(2011. 11.23)

04 빅 데이터로 인한 산업의 변화상: 실시간 대응 체제의 일상화

- 산업의 효율성을 돕는 빅 데이터(WSJ, 2012. 6.14; 박성민 외 재인용 2013. 4. 10)
 - 미국의 유통업체인 시어스(Sears)는 실시간으로 점포별 재고, 경쟁업체 가격, 날씨 예보, 고객 정보 등을 분석하여 최적 가격 책정 및 재고 관리

The screenshot shows the Sears website interface. At the top, there's a search bar with the text 'Enter keyword or item number' and a 'Search' button. Below the search bar is a navigation menu with categories like 'LocalAd', 'Deals', 'Parts & Services', 'Gifts', 'Gift Cards', and 'Gift Registry'. A 'LOCAL AVAILABILITY' pop-up window is open, asking for a ZIP code or city/state to check for local availability. The main content area features a large banner for 'columbus day sale & values' with a 'shop all deals' button. Below the banner, there are three promotional tiles: 'extra 15% off tools' with images of power tools, 'extra 10% off mattresses \$499 & up' with an image of a mattress, and 'extra 5% off GPS, MP3 players, ultrabooks, cameras & camcorders' with an image of a camera. The left sidebar contains a 'Departments' menu with various product categories.

CTO:
“우리 회사는
실시간
대응력을 갖춘
디지털 기업으로
변신하고 있다.”
(WSJ, 2012. 6.14)

Contents

1 구글과 빅 데이터

2 빅 데이터와 생활/사회/산업의 변화상

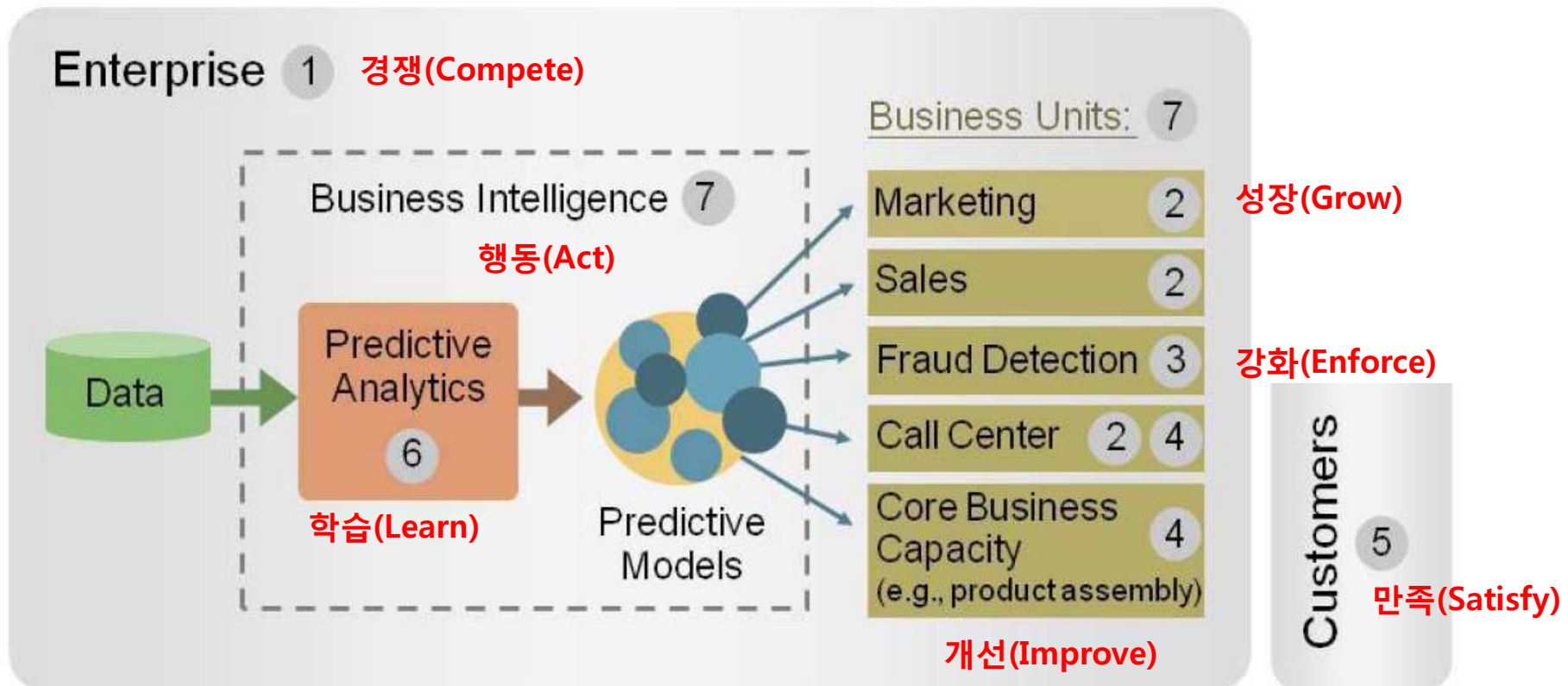
3 예측분석을 통한 비즈니스 인사이트

4 종합 요약

01 빅데이터 기반 예측분석은 기업 경쟁전략의 필수적 핵심역량

- 기업은 고객이나 조직에 대한 예측 스코어를 산출하는 예측분석 모델들을 활용할 필요
- 기업에게 빅 데이터 기반 예측분석이 필요한 7가지 이유는 곧 기업의 전략 목표(그림)

[Siegel이 제시한 예측분석이 필요한 7가지 이유=전략 목표]



02 비즈니스 인사이트가 필요한 전략목표는 성장, 강화, 개선, 만족

- 사업부서는 비즈니스 인사이트를 수시로 필요로 하며, 예측모델들을 통해 파악된 잠재고객 명단을 차별화를 위한 전략 요소로 활용할 수 있음
- 성장, 강화, 개선, (고객)만족과 관련한 사업부서의 핵심 업무는 마케팅과 고객 대응

[사업부서에서 필요로 하는 주요 전략 목표 별 예측 예시]



이유	전략 목표 내용	업무	예측 예시
성장	비용 절감	마케팅, 영업, 콜센터	잠재고객, 고객이탈 등 평가
강화	부정행위 적발	사기 방지	의심후보군 파악
개선	생산, 검사, 정비 개선	콜센터, 핵심 사업 역량	리스크 스코어 평가
만족	나은 제품, 나은 가격, 안전 구매	고객 대응	마케팅의 적절성, 향상 등

자료: : Siegel, Eric, "Seven Reasons You Need Predictive Analytics Today," PredictionImpact Inc., 2010, 재구성

03 성장전략 목표(매출↑비용↓) 달성을 위한 응용 확대, ROI 증대

[유유제약]

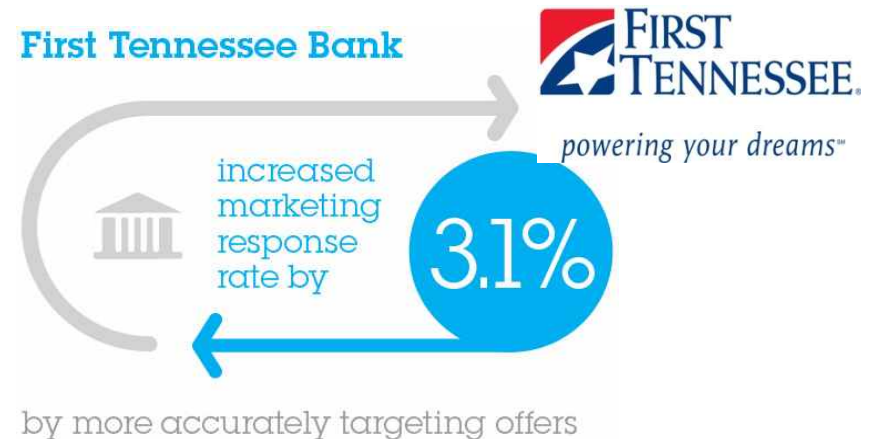


응용의 확대 사례: 유유제약은 멍연고 '베노플러스' 마케팅 위해 잠재고객 정보 평가에 초점

- 기존 대상인 어린이에서 여성들로,
- 기존 시기인 여름에서 겨울(수험생 성형수술)로,
- 기존 용도인 치료에서 미용으로 확대 적용 결과,
- 매출이 50% 이상 증가하는 경험을 하게 됨.

이후 한국모델협회 등과 업무 계약을 체결 (2013.8예정) (중앙일보, '13. 1. 18)

[퍼스트 테네시 은행]



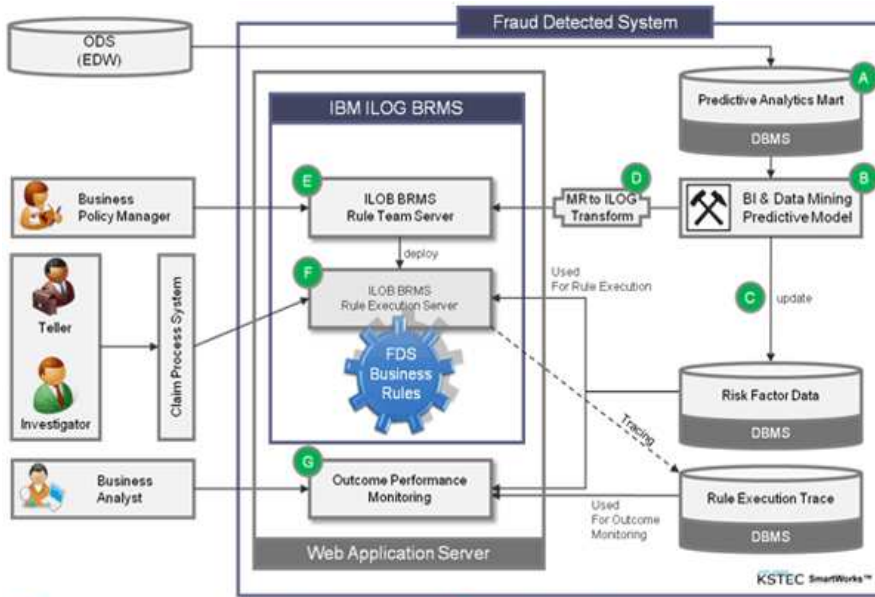
ROI 증대사례: 퍼스트 테네시 은행은 마케팅 투자비 최적화를 위해 P&L 데이터와 세분화된 고객 요구 사항을 결합 분석, 최고 ROI 제공 프로그램에 집중

- 고부가가치 고객군에 타겟팅 한 결과, 마케팅 응답률을 3.1% 증가시키고,
- 매력적 고객타겟팅으로 특별 상품 제공에도 성공, 통신판매 비용 20%, 인쇄비용을 17% 절감

마케팅이 비용 요소가 아닌 수익 촉진 요소로 혁신

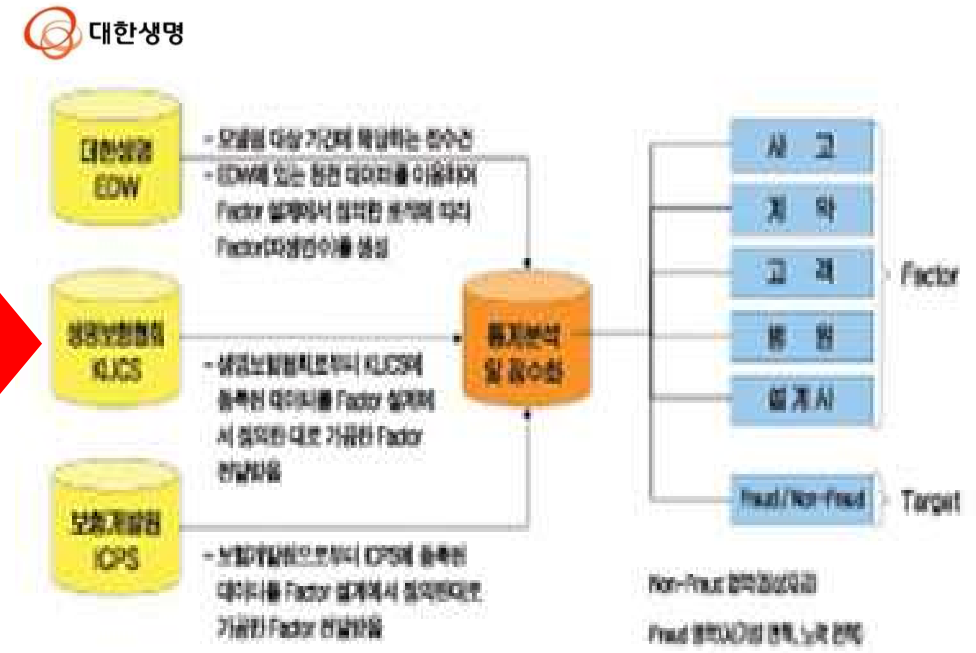
04 강화전략 목표 달성을 위한, 사기의심 거래 후보군의 정밀 파악

[KSTEC의 보험사기방지시스템 'SmartWorks FDS']



- A EDW 등을 통해 Legacy 데이터 다운로드
- B 데이터 분석 및 FDS 예측 스코어 모델 생성
- C 도출된 리스크 팩터/집계성 팩터 로 실행 DB 이관
- D FDS 예측 스코어 모델(MR)을 ILOG으로 자동 전환
- E 현업 사용자의 BR 작성
- F FDS 기반 지급 심사 수행(로 엔진에 의해 자동 실행)
- G 시스템 성과(KPI) 및 추이 모니터링(대시보드)

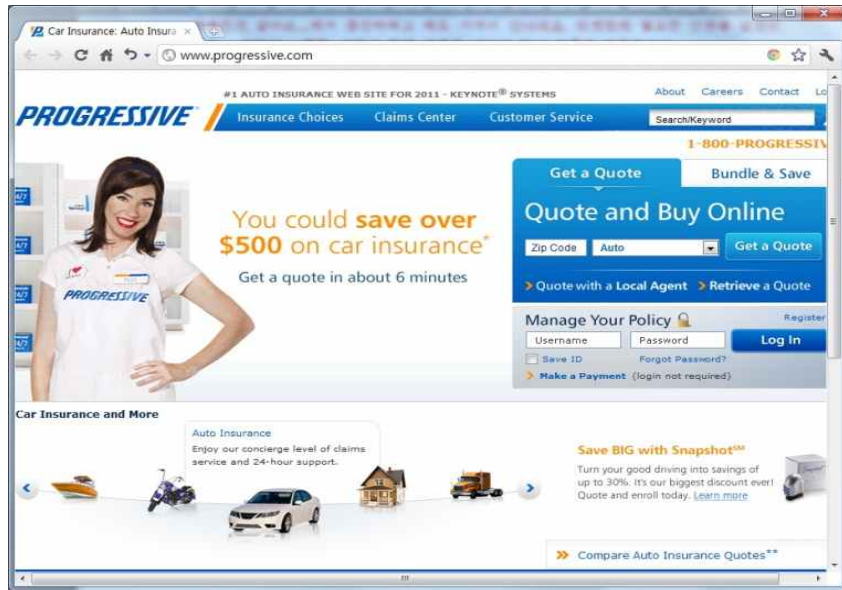
[대한생명의 보험사기 방지시스템 분석 흐름도]



사기 의심 거래 후보군 파악 사례: 대한생명은 2007년 KSTEC에 의해 개발된 보험사기시스템을 구축했고, 2010년 정형 데이터 기반의 예측분석을 보완하여 보험개발원, 보험협회 등이 제공하는 공공 데이터를 활용해 생명, 손해보험 전체 계약을 토대로 세부적인 항목을 평가할 수 있는 K-CES (Korealife claim expert search system)로 업그레이드 → 그 결과, 연간 50억 원의 사기방지 효과 경험 (출처: fntime.com, 2010. 7.19)

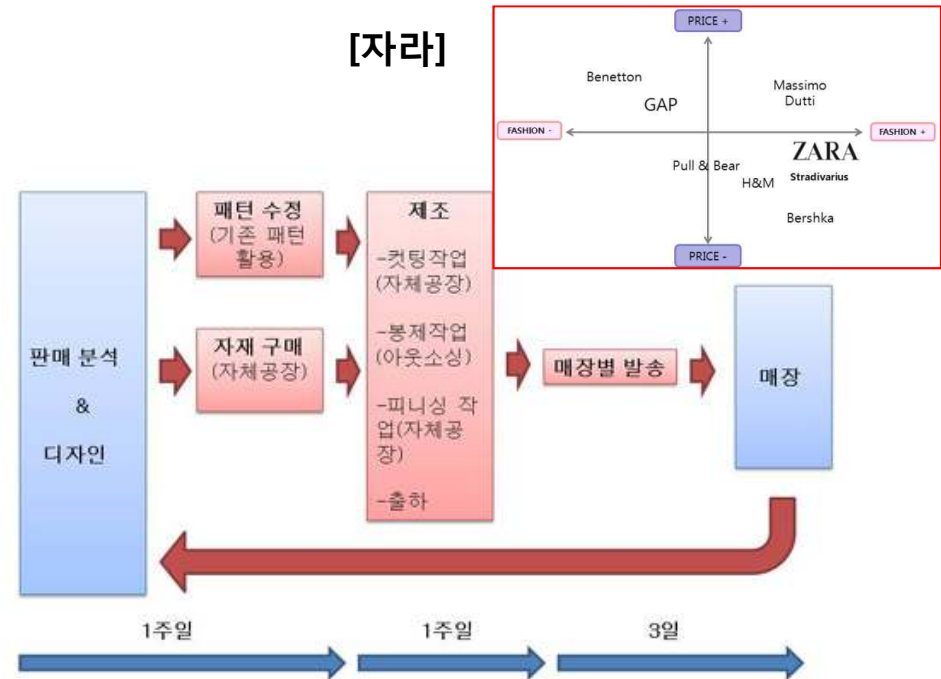
05 개선전략 목표 달성을 위한 생산과 공급의 효율성 증대

[프로그레시브]



손실률 개선 사례: 프로그레시브는 “보험료산정시스템” 통해 ‘고비용 고위험군’ 자동차 보험 손실 부문에 초점
 → 타 보험사들은 고위험군 운전자들을 몇 개 범주 안에 넣어 평균비용, 공통 보험료를 산정하는 반면, 프로그레시브는 경쟁사와 대비해 10배가 넘는 상세 분류 기준을 도입한 결과, ‘저위험군’ 가입자 대상 경쟁사 대비 낮은 보험료를 책정하는 선택적 결정

[자라]



생산과 공급의 효율성 제고 사례: 자라(Zara)는 현재 유행을 신속 포착, 기획에서 유통까지 3주 내에 완성하는 초스피드 SCM으로 일반 패션브랜드의 시즌 당 3천여 종 대비 4배 많은 1만1천여 종 상품을 선보임
 → 다품종 소량 생산방식 비용 문제 해결을 위해 빅데이터 분석 기술을 활용, 무재고 시스템 실현

06 고객만족 전략 목표 달성 위한 분석치의 셀링포인트화

[넷플릭스]

예측 분석 결과가 소비자에게
혜택이 된 사례: 넷플릭스(Netflix)
경우, 고객이 과거에 대여한
영화 목록과 시청한 영화에
부여한 평점 등의 데이터를
분석하는 영화 추천 시스템인
시네매치(Cinematch) 개발
→ 넷플릭스는 매출의 80%를
추천에 의해 발생시킴
→ 고객은 개인별 맞춤형 영화
콘텐츠를 제공받게 됨
→ 클라우드소싱으로 개발자에게
더 나은 빅데이터 분석 툴 공모

Netflix Prize

Home Rules Leaderboard Register Update Submit Download

Leaderboard **10.05%** Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52



Contents

1 구글과 빅 데이터

2 빅 데이터와 생활/사회/산업의 변화상

3 예측분석을 통한 비즈니스 인사이트

4 종합 요약

00 종합 요약



구글과 빅 데이터

- 빅 데이터 업계지도에서 관찰된 구글 위치는 인프라 & 분석틀
- 구글은 인프라 개선 통해 빅 데이터 플랫폼 진화 방향을 제시
- 구글과 빅 데이터 분석 대표 사례인 독감 트렌드, 보조역할만



빅 데이터와 생활/사회/산업의 변화상

- 3V 특성 가진 빅 데이터의 현재 한계 극복을 통한 진화 기대
- 빅 데이터로 인한 생활 변화상: 건강하고 안전한 행복 추구
- 빅 데이터로 인한 사회 변화상: 개인 맞춤형의 사회 생활
- 빅 데이터로 인한 산업의 변화상: 실시간 대응 체제의 일상화



예측분석을 통한 비즈니스 인사이트

- 빅 데이터 기반 예측분석은 기업 경쟁전략의 필수적 핵심역량
- 비즈니스 인사이트가 필요한 전략목표는 성장, 강화, 개선, 만족
- 성장전략 목표(매출 \uparrow 비용 \downarrow) 달성을 위한 응용 확대, ROI 증대
- 강화전략 목표 달성을 위한, 사기의심 거래 후보군의 정밀 파악
- 개선전략 목표 달성을 위한 생산과 공급의 효율성 증대
- 고객만족 전략 목표 달성 위한 분석치의 셀링 포인트화

Thank you

