

Issue Monitor

# 빅 데이터 분석을 통한 기업 미래 가치 창출

삼성KPMG 경제연구원

October 2012



# 빅 데이터 분석을 통한 기업 미래 가치 창출

The contacts at KPMG in connection with this report are:

**정연상 경제연구원장**  
*President, 삼정KPMG*

Tel: + 82 2 2112 0033  
Fax: + 82 2 2112 7441  
[y chung@kr.kpmg.com](mailto:y chung@kr.kpmg.com)

**이광열 상무**  
*Partner, 삼정KPMG*

Tel: + 82 2 2112 0062  
Fax: + 82 2 2112 7441  
[kwangryeolyi@kr.kpmg.com](mailto:kwangryeolyi@kr.kpmg.com)

**김영균 책임연구원**  
*Manager, 삼정KPMG*

Tel: + 82 2 2112 7046  
Fax: + 82 2 2112 7441  
[y kim1@kr.kpmg.com](mailto:y kim1@kr.kpmg.com)

**유가형 선임연구원**  
*S. Analyst, 삼정KPMG*

Tel: + 82 2 2112 6812  
Fax: + 82 2 2112 7441  
[kahyungyu@kr.kpmg.com](mailto:kahyungyu@kr.kpmg.com)

## Executive summary

### 빅 데이터 시대 도래

- 쏟아지는 데이터와 삶의 변화
- 정형과 비정형데이터의 결합
- 빅 데이터 분석을 통한 가치 창출

### 빅 데이터 분석을 통한 미래가치 창출

- 왜 빅 데이터를 사용하는가?
- 유통 산업
- 레저 산업
- 서비스 산업
- 제조 산업
- 의료 산업
- 공공 서비스

### 빅 데이터 시대 준비사항

## Page

3

4

5

6

7

7

9

10

12

13

14

15

# Executive Summary

인터넷, 스마트폰과 소셜네트워크 서비스(SNS)의 확산은 빅 데이터의 시대를 열었다. 전 세계 9억 5,000명이 가입한 대표적인 SNS인 페이스북을 통해 매일 평균 3억 장의 이미지 파일, 25억 개의 콘텐츠가 생성된다. 빅 데이터는 이와 같이 생성 규모와 속도로 인하여 과거의 데이터베이스 저장 및 분석 방식으로는 관리가 불가능한 데이터를 통칭한다.

2012년 세계경제포럼에서 올해 가장 주목해야 하는 기술로 선정된 빅 데이터는 기업의 미래 경쟁력을 좌우하는 핵심 요소로 떠오르고 있다. 국내 기업들은 기업의 지속 성장을 위해 빅 데이터 도입에 관심을 가지고 있으나 실질적인 활용 방법 및 가시적 성과에 대한 확신을 갖지 못한 상태이다. 이에 삼정KPMG 경제연구원은 국내 기업들에 앞서 빅 데이터를 도입한 해외 기업들이 어떻게 신사업을 창출하고 리스크를 관리하며 서비스를 강화시켰는지 각각의 사례를 들어 제시함으로써 빅 데이터 도입을 고민하는 국내 기업들의 행보에 도움이 되고자 한다.

## ■ 빅 데이터의 시대, 미래를 예측한다.

- 인류는 자신의 미래에 대해서 제한적으로 예측해왔다. 그러나 빅 데이터의 시대의 新인류는 내일의 세계 경제와 주식 시장, 자연 재해, 범죄를 예측한다. 캘리포니아주 샌터크루즈 경찰청은 매일 아침, 범죄가 발생할 장소와 시간을 예측해서 시민들에게 알려준다. 범죄 발생을 예측해서 미리 범죄자를 검거한다는 영화 “마이너리티 리포트”가 현실이 됐다.
- 2011년 페이스북에 '우울하다' '열 받는다' '불안하다'는 멘션이 증가한 후 미국의 실업률이 증가했다. 트위터에 부정적으로 언급되는 기업의 주가는 88% 이상 하락했다. 빅 데이터는 경제의 흐름을 미리 읽어낸다. 이러한 빅 데이터를 비즈니스에 활용하려는 기업이 증가하고 있다.

## ■ 빅 데이터를 활용하는 기업은 미래 경쟁력을 얻는다

- 기업은 신사업 창출, 생산성 향상, 신속한 의사결정, 전문 서비스의 강화, 자연재해 대비 등을 목적으로 빅 데이터를 분석하고 미래의 리스크에 대비한다. 정부 역시 재난 관리, 테러 및 범죄 예방을 목적으로 빅 데이터를 활용하고 있다.
- 빅 데이터는 두 가지 형태의 데이터로 구성된다. 기업의 데이터베이스에 지정된 형태로 저장되는 정형 데이터와 워드 문서, PDF, 이미지 파일, mp3, 유튜브 영상, 트위터와 페이스북 등에서 생성되는 콘텐츠를 포괄하는 비정형 데이터이다. 빅 데이터를 활용하기 위해서는 빅 데이터 분석 기술 확보가 우선이 되어야 한다.

## ■ 빅 데이터는 기회이자 동시에 위협이다.

- 빅 데이터 분석은 기업에게 新시장, 新고객군의 기회를 제공한다.
- 인터넷과 SNS는 특정 기업 고객들의 결속을 강화시킨다. 온라인상에서 고객들은 기업별, 제품별 자발적인 커뮤니티를 형성하고 제품에 대한 의견을 활발히 교환한다. 특정 자동차 브랜드나 카메라 브랜드 사용자 커뮤니티가 대표적이다. 빅 데이터 시대의 고객들은 기업의 광고보다 SNS에서 교환되는 고객들의 후기를 더욱 신뢰한다. 고객의 작은 불평 하나가 기업의 존폐를 가르는 위기로 작용할 수 있다. 빅 데이터 시대의 기업들이 고객의 음성에 주의를 기울여야 하는 이유이다.

# 빅 데이터 시대 도래

## 쏟아지는 데이터와 삶의 변화

우리의 하루는 정보 수집에서 시작된다. 기상 후 TV를 켜고 날씨를 확인한다. 출근길 지하철에서는 스마트폰으로 신문을 읽고 이슈가 되는 키워드를 검색한다. 출근 후에는 점심과 저녁 식사 장소를 인터넷으로 검색하고 동시에 회식 장소를 추천해달라는 타인의 질문에 답변을 달아주기도 한다. 퇴근 후 집의 PC를 통해 다음 휴가지로 계획중인 체크 프라하의 숙박 시설에 대한 자료를 찾아본다. 예매순위 1위인 영화에 대한 감상평을 나의 페이스북에 링크를 걸어둔다. 최신 유행 가요를 검색하고 음반에 대한 타인의 리뷰를 읽은 후 다운을 받기도 한다. 우리는 눈물 때부터 잠들 때까지 쉬지 않고 정보를 접하고, 쉬지 않고 정보를 생산한다.

### 폭발적인 데이터 증가량



*"현존하는 전체 데이터의 90%가 지난 3년 동안 생성됐다"*

출처 : IDC (2010)

Note: 기가(Giga) = 1024 megabyte, 테라(Tera)=1024 gigabyte,

페타(Peta)=1024 terabyte, 엑사(EXA)=1024 petabyte, 제타(Zeta)=1024 exabyte

현존하는 전체 데이터의 90%가 지난 3년 동안 생성됐다. 인터넷과 스마트폰의 확산, 양방향 의사교환이 가능한 페이스 북과 트위터 등의 소셜 네트워크 서비스의 등장, 저장매체 가격의 하락은 디지털 정보의 양을 천문학적으로 증가시켰다. 시장조사 기관 IDC는 전 세계 디지털 생산량은 2009년 기준 0.8 제타 바이트에서 2020년 35.2 제타 바이트로 44배 이상 증가할 것으로 전망하고 있다.

*"기업이 필요로 하는 빅 데이터는 충분하다. 문제는 빅 데이터를 분석할 수 있는 기술력 확보가 관건이다"*

실제 전 세계적으로 9억 명의 회원을 보유하고 있는 대표적 소셜 네트워크 서비스 업체 페이스 북에는 하루 1억 건의 메시지와 한달 3억 건의 사진이 게시된다. 페이스북을 통해 좋아하는 이성의 나이, 직업 등의 신변 정보는 물론 생각, 자주 찾는 장소, 좋아하는 색상과 디자인, 갖고 싶어하는 의류, 자동차 브랜드, 최근 구매한 상품, 향후 구매를 원하는 상품, 같은 취향으로 연락을 주고받는 그룹을 파악하는 것은 그리 어려운 일이 아니다. 기업이 필요로 하는 고객에 대한 중요 정보들은 이미 주워 담기 어려울 정도로 넘치고 있다.

불확실성과 복잡성이 증가하는 비즈니스 환경에서 기업의 외부정보의 수집과 분석력 확보, 예측 능력의 중요성이 더욱 부각되고 있다. 과거에는 중요하게 생각되지 않던 데이터가 새로운 가치를 창출할 수 있다는 가능성에 힘입어 새롭게 평가 받고 있다. 미래 시장 분석은 지속성장을 추구하는 기업에겐 피할 수 없는 과제이며 생존의 문제이기 때문이다. 빅 데이터를 통해 기업은 인간의 행동을 보다 정확하게 예측 할 수 있는 심미안을 갖게 될 것이다. 빅 데이터가 인공지능(Artificial Intelligence)기능과 결합된다면, 경영자는 직원의 도움 없이 기업의 경쟁자 분석, 운영 전략, 재무 관리, 고객 관리에 대한 정확한 해결책을 얻을 수도 있을 것이다. 문제는 누가 먼저 빅 데이터를 분석하는 기술력을 확보하는가에 있다.



# 빅 데이터의 시대 도래

## 정형 · 비정형 데이터의 결합

빅 데이터는 기업의 데이터베이스에 지정된 형태로 저장되는 정형 데이터와 워드 문서, PDF, 이미지 파일, mp3, 유튜브 영상, 트위터와 페이스북 북에서 생성되는 콘텐츠를 일컫는 비정형 데이터로 구성된다.

빅 데이터의 85%가 비정형 데이터이다. 즉 빅 데이터의 증가는 비정형 데이터의 증가를 의미한다. 그러나 비정형 데이터는 기존의 정형 데이터 분석 방법을 활용할 수 없다는 한계를 지녔다. 빅 데이터를 위한 새로운 처리, 분석 방법이 절실한 상황이다. 세계 경제 포럼(다보스포럼)은 2012년 가장 주목해야 할 과학기술로 “빅 데이터 처리기술”을 선정했다. 가트너(Gartner) 또한 2011년 IT 기술 트렌드 전망 보고서에서도 “빅 데이터 분석에 필요한 차세대 분석기법과 소셜 분석기법”을 주요 기술로 설명하고 있다. 무의미하게 생각했던 데이터 사이에 금맥이 있음을 알아차린 기업들 사이의 빅 데이터 분석 기술 선점이 치열하다. 빅 데이터 분석기술을 보유한 기업과 그렇지 못한 기업간의 경쟁력 격차는 더욱 커질 것으로 전망된다.

“정형 데이터와 비정형 데이터의 결합은 기업에게 효과적인 전략을 제시해준다.”

고객의 정보, 상품, 매출과 관련해 규칙적으로 생성, 저장되는 데이터 분석을 통해 지금까지 제품의 생산, 판매를 효과적으로 관리해온 기업의 입장에서 페이스북, 트위터 등에서 생성되는 실시간 대화의 내용과 반응, 선호도, 순간 순간의 감정 데이터들이 기업 수익성 증진의 얼마만큼 기여할 수 있는가에 대해 회의적인 시선을 지닐 수 있다. 그러나 아래 사례를 살펴보면 정형 데이터에 의존하는 기업은 정형 데이터와 비정형 데이터로 무장한 기업에 비해 빈틈을 지닐 수 밖에 없음을 알 수 있다.

국내 유명 등산복 브랜드가 도심에 매장을 오픈한다. 등산복 경쟁자 분석, 상권 분석, 유동인구 조사, 지금까지 런칭된 등산복 경쟁자들의 판매 분석까지 정형 데이터를 통한 완벽한 준비를 마쳤다. 고급스러운 매장 인테리어를 갖추고 최고급 품질의 등산복을 색상, 품질별로 완벽하게 라인업 한 후 고객을 맞는다.

바로 옆 건물에는 해외의 유명 등산복 브랜드가 매장을 오픈한다. 최고급 품질, 최고급 매장, 유동인구 등의 조건은 동일하다. 그러나 이 기업은 정형 데이터 외에 비정형 데이터를 비즈니스에 이용한다. 서울 지역 SNS 이용자들이 “봄철 날씨 변덕”, “올 봄철 바람 많이 불것, 비가 많이 올것”, “봄철 비 많이 온다는데 어찌지...”라는 내용의 글을 올린다는 본사의 분석 결과를 듣고 이 매장은 방수 기능이 있는 등산복을 전면에 배열하고 방수 기능 등산복에 대한 특별 세일을 실시한다. 어느 매장으로 향하는 고객의 수가 더 많을 것인가에 대한 대답이 빅 데이터를 활용해야 하는 이유이다.



# 빅 데이터의 시대 도래

## 빅 데이터 분석을 통해 경제적 가치를 창출

빅 데이터 분석은 처리할 데이터 양이 방대하고 비정형 데이터 비중이 높아 처리 복잡도가 높다. 방대한 데이터를 실시간에 처리하는 분석 기술로 비정형 문장들에 내포된 의미를 추출하며 추출된 정보 사이에 가정을 수립하는 텍스트 마이닝(Text Mining), 특정 서비스 및 상품에 대한 사용자의 의견을 판별하는 오피니언 마이닝(Opinion Mining), 입 소문의 중심인 사용자(Influencer)를 파악하는 소셜 네트워크 분석(Social Network Analytics), 유사성이 높은 대상집단과 타 군집에 속한 객체간의 상이성을 분석하여 새 사용자 그룹을 도출해내는 군집분석(Cluster Analysis) 등이 사용된다.



이를 패키지화하여 상용화시킨 것이 야후에서 개발한 하둡(Hadoop)이다. 하둡은 정형 데이터와 비정형 데이터 모두를 분산 클러스터 기술을 통해 빠르고 신뢰성 있게 분석을 해주는 장점을 지녔다. 이 밖에 구글의 빅 쿼리(Big Query), 아마존의 다이나모(Dynamo)가 빅 데이터 분석을 위해 활용된다.

주요 연구기관들이 발표한 빅 데이터를 통해 얻을 수 있는 경제적 효과는 다음과 같다

빅 데이터 활용을 통해 얻는 가치	
기관명	경제적 효과
Economist (2010)	데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입자본으로 비즈니스의 새로운 원자재 역할을 한다
MIT Sloan (2010)	데이터 분석을 잘 활용하는 조직일수록 차별적 경쟁력을 갖추고 높은 성과를 창출한다
Gartner (2011)	데이터는 21세기의 원유이며 미래 경쟁 우위를 결정한다. 기업은 다가올 "데이터 경제시대"를 이해하고 정보고립을 경계해야 한다
McKinsey (2011)	빅 데이터는 혁신, 경쟁력, 생산성의 핵심요소이다

출처 : Economist, MIT Sloan, Gartner, McKinsey

맥킨지는 의료와 건강, 소매, 제조 분야에서 1%의 추가 생산성 향상을 가져올 것이며 적게는 \$1,000억에서 \$7,000억 규모의 경제적 효과 창출할 것이라고 내다봤다.

빅 데이터가 제공하는 경제적 가치		
의료·건강	소매업	제조업
		
<ul style="list-style-type: none"><li>▪ 매년 3,300억 가치</li><li>▪ 0.7%/year 생산성 증가</li></ul>	<ul style="list-style-type: none"><li>▪ 이윤 60% 증가 가능</li><li>▪ 0.5~1% 생산성 증가</li></ul>	<ul style="list-style-type: none"><li>▪ 제품개발비 50% 감소</li><li>▪ 운전자본 7% 절감가능</li></ul>

출처 : McKinsey, Big Data, 2011

# 빅 데이터 분석을 통한 미래 가치 창출

**왜 빅 데이터를 사용하는가?** 2011년 매일유업의 대표 상품 애플루트 분유에서 식중독을 유발하는 황색포도상구균이 검출됐다는 뉴스가 확산되자 네티즌들과 아기 엄마들의 강한 항의가 인터넷을 뜨겁게 달구었다. 논란이 확산되자 매일유업의 CEO는 신문사 방송사 등의 기존의 주요 언론 매체 대신 유튜브를 찾아가서 사과 동영상을 올렸다.

한국의 대기업들이 인터넷과 SNS의 파급력을 인정하고 있음을 의미한다. 이제 그 파급력을 비즈니스의 기회로 삼기 위한 시도가 진행되어야 한다. 이미 빅 데이터 시대가 제공하는 비즈니스 기회를 효과적으로 활용하고 있는 해외 기업들의 사례를 소개한다.

## 유통 산업

빅 데이터의 시대는 SoLoMoMe(Social, Local, Mobile, Personalized)의 시대이다. 소비자들은 소셜 미디어에 대한 의존도가 높고, 지역 정보가 풍부해지며, 모바일 스마트폰을 사용하고, 개인화된 맞춤 서비스를 원한다. 빅 데이터 분석 기술을 지닌 유통기업, 특히 온라인 유통 기업들에게 성장의 기회가 되고 있다.

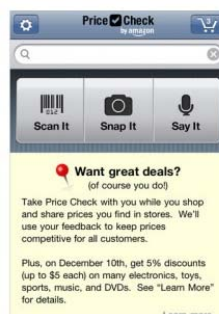
### 아마존, 고객 스마트폰을 이용한 데이터 수집

2011년 12월, 아마존의 웹사이트에 “아마존을 위해 기업 스파이가 되시겠습니까?”는 흥미로운 광고를 게재했다. 아마존의 스파이가 준비해야 하는 도구는 모든 금고를 열 수 있는 만능 열쇠가 아니라 아마존 어플리케이션이 탑재된 스마트폰이다. 스파이는 안드로이드 마켓에서 아마존의 “Price Check” 어플을 다운로드 받은 후 월마트나 베스트 바이와 같은 오프라인 상점을 방문하여 스마트폰으로 상품의 가격을 스캔하는 임무를 부여 받는다.

아마존 경쟁사의 오프라인 상점을 방문한 사람들은 상품의 바코드를 스캔하고 사진을 찍고 해당 상품에 대한 단어를 검색하고 여러 인터넷 업체들이 내세운 가격과 비교한다. 상품의 디자인과 기능은 상점에서 확인하고 빈 손으로 걸어나와 아마존에서 스파이의 대가로 제공한 5달러 쿠폰을 이용하여 인터넷에서 같은 상품을 구매한다.

빅 데이터 플랫폼을 개발한 빅 데이터 시대의 선두업체 아마존이 이러한 프로모션을 진행하는 표면적 이유는 자사의 가격비교 어플리케이션의 홍보이다. 아마존의 숨은 목적은 경쟁업체가 내건 최저가를 고객들을 통해 실시간으로 수집하고 경쟁사보다 더 좋은 조건의 가격을 제시하는 것이다. 아마존은 수 백만 고객들의 스마트폰을 통해 수천여 개 경쟁사의 가격 정보를 입수하고 분석하여 자사의 가격 정책에 반영하고 있다.

### 아마존의 타사 가격정보 확보 전략



출처: Amazon

# 빅 데이터 분석을 통한 미래 가치 창출

## 월마트랩, 소셜 게놈으로 빅 데이터 분석 착수

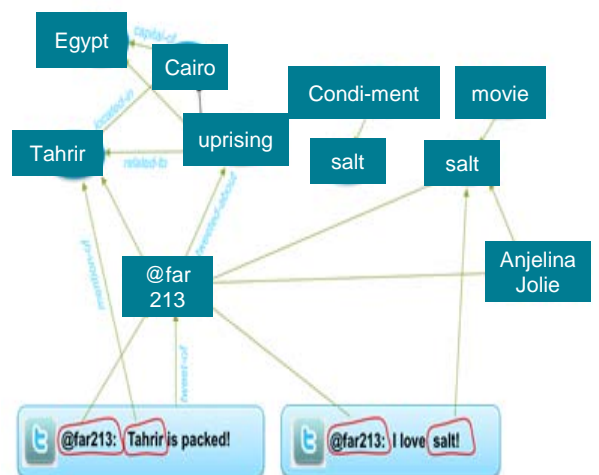
“언제나 최저가격(Everyday Low Prices)”을 모토로 내세운 월마트는 빅 데이터 시대에 대한 부담이 크다. 마이크 튜크 월마트 CEO는 오프라인 상점들이 아마존을 위한 쇼룸(Show room), 온라인 유통업체들의 들러리로 전락하는 상황을 우려하며 자체적으로 빅 데이터 분석기관인 월마트랩을 오픈하고 이어 2011년 소셜 게놈 플랫폼을 개발한 코스믹스(Kosmix)사를 인수했다.

월마트랩은 매일 수십억개씩 발생하는 트위터와 페이스북의 멘션, 블로그 포스팅, 유튜브 비디오를 살아있는 유기체로 설명한다. 끊임없이 진화하는 유기체를 분석하기 위해 수억만개의 엔터티와 엔터티들의 관계를 토대로 헤아릴 수 없는 양의 빅데이터를 저장하고 있는 소셜 게놈(Social Genome) 플랫폼을 활용한다.

트위터를 사용하는 A는 “나는 솔트(SALT)가 좋아”라고 적었다. 몇초 후 월마트랩에 A씨가 적은 트윗의 내용이 전달되고 소셜 게놈은 평소 A씨가 트위터에 영화에 대한 평을 자주 쓴다는 점을 근거로 A씨가 언급한 솔트는 요리에 사용되는 소금이 아닌 안젤리나 졸리의 영화 제목이라는 결론을 내린다. 몇 분 후 A씨의 가까운 지인에게 다음과 같은 메일이 발송된다. “안녕하세요. 당신의 친구 A씨의 생일이 다가옵니다. 마침 A씨가 영화 솔트에 관심이 있네요. 영화 솔트와 관련한 생일 선물을 추천해드립니다.” 월마트랩은 페이스북을 사용하는 B씨에게 이런 메시지를 보낸다. “당신의 페이스북에 당신이 좋아하는 커피에 대해 작성해주시겠어요? 월마트에서 판매하는 드롱기 커피메이커를 45% 할인해 드리겠습니다.

미국 최대 유통기업인 월마트 조차 시대의 흐름을 무시하지 못했다. 기존의 마케팅과 유통 방식으로는 유통업계 1위의 자리가 위태로움을 깨달은 월마트는 빅 데이터 분석을 통한 유통 체계의 변화를 시도하고 있다.

## 월마트랩과 소셜 게놈 프로세스



출처: Walmart Lab



# 빅 데이터 분석을 통한 미래 가치 창출

## 레저 산업

여행 정보는 상품 정보와 함께 사람들이 인터넷에 가장 흔히 의존하는 분야이다. 세계 최대 규모의 여행정보 사이트 트립어드바이저의 비정형 데이터 수집 방법과 향후 빅 데이터 활용 계획을 알아본다.

### 여행업계의 아마존을 꿈꾸는 트립어드바이저

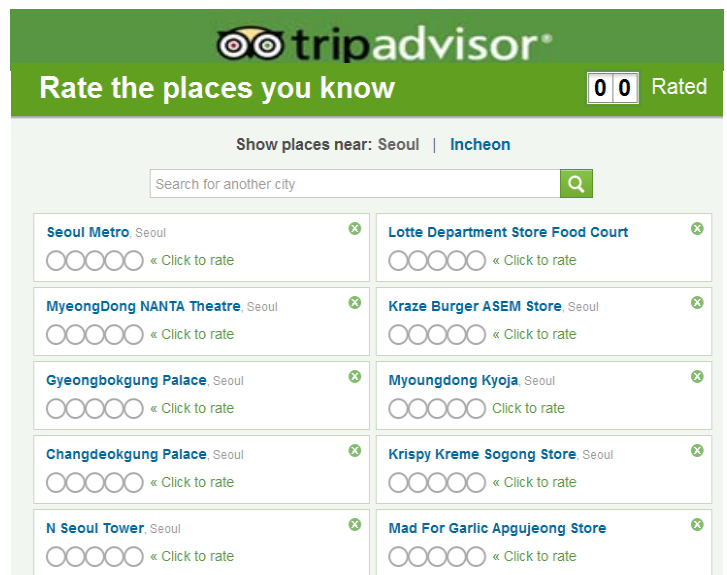
매달 5,000만 명의 여행객이 방문하는 트립어드바이저(Trip Advisor)는 100% 여행객들의 자발적인 참여와 리뷰로 유지되고 있는 소셜 사이트이다. 2012년 7월 트립어드바이저에 등록된 리뷰는 7,500만개를 넘어섰다. 여행 정보가 필요한 사람들은 트립 어드바이저의 웹사이트에 접속한 후 원하는 지역을 선택한다. 이미 그 지역을 다녀온 수많은 사람들의 상세한 설명과 사진을 읽게 된다. 해당 지역 숙소와 식당은 가격대별로 순위가 매겨져 제시된다. 트립어드바이저가 제공하는 정보를 통해 여행을 다녀온 사람은 다시 인터넷에 접속해서 자신의 리뷰를 업데이트한다. 이러한 방식으로 트립 어드바이저는 전세계 여행객들로부터 전세계 모든 호텔, 모든 식당, 모든 관광지의 사진과 상세한 정보를 모아왔다.

빅 데이터 시대의 주축인 페이스북은 트립어드바이저와의 파트너십을 통해 “내가 방문한 도시” 어플리케이션 기능을 제공하기 시작했다. 사람들은 해당 어플을 통해 자신이 거쳐갔던 도시에 빨간 핀을 꼽아 표시하고 페이스북 지인들에게 그 곳의 여행 경험을 공유한다. 사이트 접속자의 위치를 파악하여 접속자 인근 장소들의 이름을 나열하며 대한 평가를 요청하기도 한다. 트립어드바이저는 페이스북과 트위터 이용자들에게 가장 인기있는 여행 어플리케이션으로 부상했다.

트립 어드바이저는 개개인의 텍스트와 이미지 리뷰, 즉 방대한 비정형 데이터의 분석을 통해 여행 목적별 (유적지 여행, 휴양지 여행, 도심형 휴양 여행, 쇼핑 여행, 가족 여행, 로맨틱 여행), 지역별(근거리 여행, 원거리 여행, 특정 국가 여행), 취향별 (시간이 걸리더라도 저렴한 여행, 또는 비싸더라도 편한 여행, 버스나 기차 선호 여행, 짧은 시간에 많은 곳을 방문하고자 하는 여행, 한 곳에서 오래 머무르는 여행)등에 대한 개인 선호 스타일을 분석해낸다. 비슷한 지역의 비슷한 연령대가 선호하는 여행 형태도 쉽게 파악할 수 있다. 빅 데이터 분석을 통해 각 여행자의 구미에 맞는 여행 상품을 제안하고 자세한 여행 정보를 함께 제공할 수 있다. 아직 본격적으로 관광업에 진출하지 않았지만, 이미 트립어드바이저의 성공 가능성에 대해 세계 여행업계가 긴장하고 있다. 여행사들이 현지

에 인력을 파견하여 얻는 한정적인 여행 정보와 전세계인들이 매일 쏟아 붓는 정보의 양과 질에 분명한 차이가 존재하기 때문이다.

여행업계의 아마존이 될 것으로 기대되는 트립어드바이저는 실제 아마존의 초기 멤버들을 영입하여 빅 데이터를 통한 수익사업 방안을 모색 중이다.



# 빅 데이터 분석을 통한 미래 가치 창출

## 서비스 산업

서비스 산업의 기업 가치는 고객의 평가에 의해 달라진다. 다른 어떤 산업보다 고객의 피드백에 주목해야 한다. 빅 데이터의 시대에 고객의 평가는 인터넷을 통해 기업에게 전달된다. 인터넷 상의 비정형 데이터를 수집해 자동차 셰어링 서비스를 선보인 신생 렌터카 기업 Zip car와 비정형 데이터를 적극적으로 분석해 기업 운영에 반영하는 세계적인 렌터카 기업 Hertz의 사례를 살펴본다.

### Zip car, 신사업 창출의 원동력

“뉴욕은 주차비가 너무 비싸서 차를 유지할 수가 없다”, “유가가 배럴당 \$100을 넘었다” 자동차를 소유한 사람들의 불만이 인터넷을 가득 채웠다. 그러나 사람들은 자동차가 주는 편리함을 포기할 수는 없다. 자동차 구매를 포기하고 렌터카를 이용하는 사람들에 불만은 있다. “렌터카를 빌리기 위해 렌터카 회사가 있는 곳까지 이동해야 하고 수 많은 서류에 사인을 해야 하고, 차량이 준비될 때까지 기다려야 하며 이런 저런 보험을 선택해야 한다. 세 시간 차를 쓰기 위해 하루 요금을 지불하는 것이 억울하다”. 언론과 인터넷 커뮤니티를 달구던 사람들의 불만은 신종 자동차 서비스 사업을 탄생시키는 단초가 됐다. 자동차 셰어링 서비스이다.

사람들은 자신과 가까운 곳에 렌터카가 준비되어 있고, 인터넷으로 간편히 예약해서 자신이 원하는 시간에 사용할 수 있기를 바란다. 이러한 사람들의 아이디어를 반영한 Zip car가 2000년 미국에서 서비스를 시작했다. 자동차 구매 비용, 등취득세, 매년 내야 하는 세금, 기름값, 주차비, 차 수리비에서 해방된 사람들은 이웃과 함께 차를 나눠 쓰는 것을 기꺼이 감수한다.

Zip car는 빅데이터 개념이 나오기 전에 창업되었지만 인터넷, 커뮤니티와 블로그 등의 비정형 데이터를 통해 얻은 아이디어를 비즈니스에 접목한 사례로 구분할 수 있다. Zip car는 빅 데이터의 시대에도 자동차 소유 관심 있는 이들의 핫 이슈가 되고 있다. 트위터의 창업자 잭 도시는 “Zip car는 지구 환경 보존에 유익하고 매번 다른 차를 선택해서 탈 수 있어 재미있다”고 평가하며 Zip car에 대한 비정형 데이터를 생성하고 있다.

2010년 Zip car는 \$186 million (약 2,000억원)의 매출을 올렸지만 순익은 마이너스였다. 그러나 2011년 나스닥 상장 첫 날, 투자자들이 매긴 Zip car의 가치는 \$1.2 billion (약 1.3조원)이었다. 비록 2008~2010년 삼년 연속 적자를 면하지 못하였으나 자동차에 관심 있는 인터넷 유저들의 찬사를 받고 있기 때문이다.

### Zip car의 예약화면과 스마트 키



# 빅 데이터 분석을 통한 미래 가치 창출

## Hertz, 빅 데이터 분석 통해 고객 평가 리뷰

전 세계 8,300개의 지점을 운영하고 있는 Hertz는 매일 수 천여 개에 달하는 고객의 의견을 웹 설문, 이메일, 문자 메시지를 통해 수집한다. 또한 “허츠 차량 청결(Hertz Vehicle Cleanliness)” “허츠 직원교육(Staff Courtesy)” “허츠 고장(Mechanical Issues)” 등의 단어를 포함한 인터넷 상의 비정형데이터들을 주기적으로 수집하여 Hertz의 인텔리전스 분석 시스템에 투입한다.

Hertz는 이렇게 수집한 방대한 양의 비정형 데이터를 정형 데이터로 변형시켜 전사적 전략 수립과 지점별 전략 수립을 위한 자료, 오퍼레이션 비용 감소를 위한 목적으로 활용하고 있다.

고객의 불만을 분석하여 고객 만족도를 높이는 시도는 기존 CRM 시스템의 기능과 다르지 않다. 그러나 CRM이 정형 데이터를 분석한다면, 빅 데이터는 주로 비정형 데이터 분석을 통해 고객의 요구를 더욱 구체적으로 파악한다는 방법론적인 면에서 차이가 있다.

빅데이터의 텍스트 마이닝 분석 방법이 도입되기 이전, 고객 설문지의 주관적 답변 분석은 각 지점 매니저들의 역할이었다. 비효율적인 시간 소비, 인력 낭비와 더불어 매니저 각자의 기준에 의해 고객들의 답변이 해석되고 분류되는 일관성의 부재가 문제였다. 이후 주관적 답변에 대한 시스템 분석을 통해 일관성과 정확성, 처리 속도가 개선되었다는 평가를 받고 있다.

렌터카 회사들의 고민은 고객의 대기 시간이다. 렌트차량이 신속하고 정확하게 준비되지 않으면 하루에도 수 백건의 고객 불만 사항이 인터넷을 뒤덮일 각오를 해야 한다. 빅 데이터 분석을 통해 Hertz 필라델피아 지점에서 하루 중 특정 시간대에 렌트차량 대기 시간과 반환시간이 지연되고 있음을 알아냈다. Hertz 필라델피아 지점은 즉시 해당 시간대에는 숙련된 직원으로 교체하고 지점 매니저가 참관하는 운영 형태로 바꾸는 조치를 취했다.



소비자 편의를 전적으로 반영한 Zip car의 등장으로 SNS는 Hertz, AVIS, Enterprise 등 기존 렌터카 업체의 서비스에 대한 불만이 고조되었다. Hertz는 고객들의 요구를 최대한 반영하여 Zip car의 비즈니스 모델과 흡사한 Hertz on Demand 상품을 출시했다.

또한 지구 온난화 및 환경 오염에 대한 인식이 확산 및 Hertz 이미지 제고를 위해 Hertz on Demand 용 전기차를 사용한다. 전기차 구매 사실을 홍보하기 위해 허츠가 선택한 수단은 트위터였다. 트위터에 허츠 전기차가 주차된 장소를 수수께끼 형식으로 제공하여 이용자들이 허츠 전기차를 찾아내는 이벤트를 기획했다.

소비자와 직접적인 인터랙션을 주고받는 서비스 업계의 기업에게 있어서 빅 데이터 분석의 중요성은 더욱 크다. 기업은 빅 데이터 분석을 통해 인터넷 상에서 기업에 대한 사용자 의견을 긍정(Positive), 부정(Negative), 중립(Neutral)으로 판별하고 기업에게 비우호적인 입소문의 중심, 허브(Hub) 역할을 하는 소비자 (Influencer)를 찾아서 적극적인 구매 정책을 펼치는 등의 기업 리스크 관리를 가능하게 한다.

# 빅 데이터 분석을 통한 미래 가치 창출

## 제조 산업

철강 제조 기업 포스코와 풍력 터빈 제조기업 베스타스는 인터넷 곳곳에 산재된 정형, 비정형 데이터를 끌어와서 기업의 목적에 맞게 활용하고 있다. 포스코는 철강의 재료가 되는 원료의 구매 타이밍과 가격을 결정하기 위해, 베스타스는 최적의 풍력 발전단지 선정과 지역별 터빈의 전력 생산량 예측을 위해 빅 데이터를 분석한다.

### 포스코, 원자재 구매 시기와 가격 결정

세계 제일의 철강 기업인 포스코는 해외 20여 개 광산을 소유하고 있다. 원재료 가격 상승에 대비한 자급책의 일환이다. 포스코는 원재료 확보를 위한 노력은 빅 데이터 투자로 이어졌다. 철강산업은 국제적인 원자재 투기 세력의 개입으로 원료 가격의 변동이 크기 때문에 위험 관리 차원에서 빅 데이터 분석이 필요하다.

포스코는 남미와 호주 광산의 생산 상황과 해당국의 사회 경제적 상황, 런던금속거래소(LME)의 광물가격, 해외 지역별 광물가격을 수집하고 분석함으로써 미래의 철광석 가격을 예측한다. 또한 고객사의 수요 데이터와 전 세계 철광석 및 현물거래소의 가격 데이터를 조합한 뒤 철광석 구매의 최적 타이밍을 결정한다. 그 외에도 생산과정 중 수십만 가지의 공정별 온도와 습도, 압력, 성분 등에 대한 데이터를 분석함으로써 공정별 불량률 감소를 꾀하고 있다.

### 베스타스, 최적의 풍력 발전단지 선정과 전력 생산량 예측

세계 1위의 풍력 터빈 제조 및 풍력단지 설비 업체인 베스타스는 빅데이터 분석을 통해 풍력 터빈의 에너지 효율성을 높였다. 풍력 터빈은 바람의 힘을 전기로 변환시키는 핵심 부품이다.

베스타스는 날씨, 조수 간만의 차, 위성 이미지, 지리 데이터, 날씨 모델링 조사, 산림지도, 해상 지도 등 페타바이트 규모의 정형, 비정형 데이터를 수집한다. 날씨 부문 하나만 보더라도 4년 째 20 페타바이트(petabyte) 이상의 전 세계 날씨 데이터를 집적하고 있다. 베스타스는 실시간으로 변하는 바람의 방향, 높이에 따른 변화, 기후 요소 데이터를 수집하고 분석을 통해 최적의 풍력 발전소 부지를 선정해낸다.

또한 풍력 터빈에 연결된 각각의 날개가 날씨 변화에 어떻게 반응하는지를 분석하고 세계 각 지역에 공급된 풍력 터빈들이 향후 얼마나 더 많은 풍력 에너지를 생산해 낼 것인지 전망한다. 기공급한 풍력 터빈의 유지보수 일정을 위한 최적의 시간을 계산하는데 있어서도 빅 데이터를 사용한다.



# 빅 데이터 분석을 통한 미래 가치 창출

## 의료 산업

미국 의료산업에서 빅 데이터 분석은 “킬러 앱”으로 통한다. 킬러 앱이란 등장과 동시에 시장을 완전히 재편하는 제품이나 서비스를 뜻한다. IBM은 3분 이내에 2억 페이지 분량의 정형, 비정형 데이터를 분석해 환자의 치료법을 제시하는 왓슨을 출시했으며 대형 병원들을 중심으로 빅 데이터 분석을 통한 환자의 질병진단과 치료방법에 대한 자체적인 연구가 진행되고 있다.

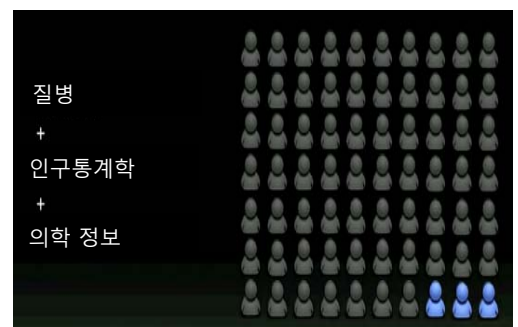
### 클리브랜드 클리닉, 환자의 질병 진단 및 치료에 활용

건강에 대한 관심이 늘어나면서 “나와 같은 병을 지닌 사람들([www.patientslikeme.com](http://www.patientslikeme.com))”과 같은 소셜 네트워크가 인기이다. 한 가지 병에 대한 수십가지의 치료법이 올라와 있다. 사람들이 각자 받은 치료에 대한 정보를 업데이트 하기 때문이다. 한 가지 약에 대한 수십가지 정보도 공유된다. 약의 효과와 부작용, 증상별 복용량, 복용 중지 사유, 가격과 경험 등이 제공된다. 100mg, 200mg 등 약 복용량에 따라 환자들을 그룹핑하여 그들이 해당 질병 이외에 어떤 질병에 공통적으로 노출되어 있는지 보여준다. 인터넷 상에서 일반인들의 질병과 약에 대한 지식 공유가 확대됨에 따라 의료행위와 치료에 관한 비정형 데이터도 급속히 증가하고 있다.

의료업계의 빅뱅으로 불리는 빅 데이터에 주목한 클리브랜드 클리닉은 아예 2009년 빅데이터 분석 회사인 Explorys를 창업했다. 세계 최고의 심장질환 병원, 미국 3대 병원으로 알려져 있는 클리브랜드 클리닉은 Explorys를 통해 1,500만 명 환자들의 의료 데이터를 분석, 관리, 검색할 수 있는 소프트웨어 플랫폼을 St. Joseph 의료 시스템, 메트로 헬스 (Metro Health), 순마의료시스템 (Summa Health system), 메디스타 헬스 (Medstar Health), 카톨릭의료협력시스템 (Catholic Health partners), 랑케스터 종합 헬스 (Lancaster General Health), 퀸즈메디컬센터 (Queens' medical center) 등의 미국내 144개 주요 병원 및 보험사, 의료연구 기관 등에 제공하고 있다.

소정의 수수료를 지불하는 병원과 보험사, 의료연구기관들은 동일 질환에 노출된 환자들의 식생활 패턴과 거주환경 분석, 질병 정보와 유전자 통계 정보, 자신의 병원의 환자 차트 및 면담내용 뿐 아니라 타 병원 환자들의 증상 및 치료 방법, 임상 자료를 공유 받게 된다. 또한 인터넷의 최신 의학 기술 등의 정보를 얻게 된다. 또한 의료연구 기관의 최신 의료 논문, 인터넷 상에 등재된 최신 의학 기술, 의료 소셜 네트워크의 정보를 통합하고 문맥을 유추하여 해당 환자에게 필요한 의학적 치료법을 제공받는다.

의료업계의 경쟁력은 빅 데이터를 중심으로 재편되고 있다. 의료진이 확보할 수 있는 정확한 최신 의료 정보와 수많은 임상 데이터 자료가 곧 의료 서비스의 질을 결정하게 된다.





# 빅 데이터 분석을 통한 미래 가치 창출

## 공공 서비스

빅 데이터 분석을 사용하여 교통 흐름을 통제하는 국가가 증가하고 있다. 실제 운행중인 수많은 차량을 통해 방대한 양의 교통 데이터를 수집·분석하여 최적의 교통 안내 서비스를 제공하는 것도 빅 데이터 분석의 대표적인 활용 사례이다

### 빅 데이터 분석을 통한 지능형 교통정보서비스

스웨덴의 로알 기술 대학(Royal Institute of Technology)는 운행중인 자동차들의 GPS 데이터, 도로의 레이더 센서를 통한 교통 유량과 운행차량의 속도 데이터, 교통혼잡부담금 데이터, 기후 정보, 교통 사고 및 도로 공사 정보를 토대로 실시간 교통정보를 제공하고 있다. 모든 데이터는 대용량의 정형, 비정형 데이터를 리얼타임으로 분석해냄으로써 현 교통상황 및 목적지까지의 경로별 소요 시간을 계산해 요청자에게 제공하고 있다.

일본 역시 노무라연구소와 건설성, 통산성, 운수성, 우정성, 경찰청 5개 정부 기관이 협력하여 지능형 교통안내 시스템(Intelligent Transport System) 서비스를 제공하고 있다. 일본 전역의 택시 11,000대와 교통정보 제공에 동의한 일반 운전자들의 GPS를 통해 자동차의 주행 스피드를 계산하는 방식으로 실시간 교통정보를 수집하고 이를 사용자들의 스마트폰으로 송신한다. 2011년 이와테현의 대지진이 발생했을 때 일본의 교통안내 시스템은 구조 차량 및 지원자원 수송 차량은 지진 발생 지역의 교통 체증을 피해 신속하게 현장에 접근할 수 있게 함으로써 지진 피해를 최소화시키는 데 일조했다는 평가를 받고 있다.

### 빅 데이터 분석을 통한 스마트 워크 사례

한국 역시 고속도로 교통정보, 실시간 교통정보를 인터넷과 스마트폰을 통해 제공되고 있다. SK텔레콤은 개인 스마트폰 사용자에게 무료로 제공하던 실시간 교통정보에 건설사의 PMIS (Project Management Information System : 건설공사 기획부터 설계, 구매, 시공, 유지보수까지 건설 프로젝트 단계에서 생성되는 정보를 통합적으로 관리하고 정보를 공유하는 시스템)을 연계하여 수익사업을 시작했다. 전국 2만 여 대의 레미콘차량 운전기사는 T-Map을 통해 제공되는 실시간 교통정보로 운행시간 및 근로시간을 단축시킬 수 있고 레미콘 회사는 각 건설현장별 배차, 출하 현황을 실시간으로 파악할 수 있어 레미콘 차량의 회전율을 극대화하고, 유류비도 절감할 수 있다. 유류비 5~10% 절감과 차량 회전율도 10~20% 상승효과를 기대하고 있다.

### SK텔레콤이 제공하는 실시간 교통정보 서비스



# 빅 데이터 시대 준비사항

## 빅 데이터 시대 준비사항

국내 기업들이 빅데이터를 활용하여 가치를 창출하기 위해서는 우선 현재 가지고 있는 데이터를 활용할 수 있는 방안을 찾아야 한다. 앞에서 다양한 빅데이터 활용 사례를 통해 설명한 바와 같이 기업의 특성에 맞게 다양한 방안을 고려할 수 있다.

SNS 데이터 분석에 관심 있는 기업이라면 한글 자연어 처리 기술에 대한 비교적 장기적인 투자가 요구 된다. 고품질의 SNS 데이터 분석에는 한글 처리에 대한 깊은 이해와 기술적 노하우가 요구된다. 따라서 SNS 데이터 분석을 하려는 기업은 단계적으로 SNS데이터 수집 및 정량적 측정방법을 도입한 후 점진적 한글 자연어 처리 기술 고도화 과정이 요구된다.

이미 해외에서 하둡 기반의 솔루션이 개발되고 있으므로 보유하고 있는 솔루션을 하둡 기반으로 전환하여 구축하는 방안도 고려해 볼 필요가 있다. 하둡기반 솔루션 구축을 위해서는 다음 세가지 단계가 필요하다.

첫 번째, 빅데이터 분석을 하기 위해서는 우선 분석의 대상이 되는 데이터를 수집해야 한다. IDC에 따르면 기업 데이터의 85%정도가 구조화되지 않은 데이터라고 한다. 즉 현재 기업들이 보유한 데이터 중 15%만 저장하고 있고 85%의 자료는 자산화되지 않기 때문에 다양한 종류의 데이터를 지속적으로 수집하는 것이 요구된다. 분석할 대상 데이터를 지속적으로 발굴하고 수집하면서 활용방안을 검토해야 한다.

두 번째는 지속가능한 운영 인프라 구축이다. 하둡 시스템은 단순히 하둡 배포판을 다운로드하고 설치하는 데서 완성되지 않는다. 적합한 하드웨어 및 하드웨어와 하둡 배포판을 최적화하여 구성해야 솔루션의 성능 최적화 및 운영환경으로 이관시 안정적인 서비스 품질을 제공할 수 있다. 국내에서 운영서비스 수준의 인프라 구축이 가능한 인력이 매우 적은 것이 현실이다. 따라서 인프라를 구축하는 데 시간과 노력을 들이기 보다는 기존에 이러한 서비스를 제공하는 회사를 통해 인프라 구성을 마친 후 서비스를 제공하는 것이 비용 효율적인 대안이 될 수 있다.

마지막으로 전담조직의 구성이다. 현재 대부분의 기업은 기업 내외의 데이터가 조직별로 분산되어 빅 데이터의 효율적인 운용이 어렵다. 데이터 접근성, 인프라 구축, 분석 역량 등을 뒷받침할 수 있는 Data 중심의 조직이 필요하다. 다양한 교육과정들을 잘 활용하여 빅 데이터 비즈니스 활용 기획자, 빅데이터 기반 기술 엔지니어 등을 육성한다면 기업의 미래가치 창출을 통해 신성장 동력을 발굴하는 기회가 될 것이다.

빅 데이터 활용을 위한 조건	핵심 이슈	기업 준비사항
1 Data 접근성	<ul style="list-style-type: none"> <li>외부의 제 3자 데이터 활용 가능성</li> <li>내/외부 데이터의 체계적 결합 및 전사적 이용 가능성</li> </ul>	<ul style="list-style-type: none"> <li>프라이버시, 보안, 지적재산권, 법적 책임관련 사전 준비</li> <li>외부 DB의 내부 활용방안</li> </ul>
2 빅 데이터 인프라	<ul style="list-style-type: none"> <li>클라우드 기반 통합 분석 시스템</li> <li>전사적 데이터 통합 활용 체계</li> </ul>	<ul style="list-style-type: none"> <li>분산된 데이터의 클라우드 기반 통합 데이터 공유 프로세스 정립</li> </ul>
3 분석역량	<ul style="list-style-type: none"> <li>대용량 데이터 분석기술(하둡 등)</li> <li>실시간 기반분석, 시각화 S/W 등</li> </ul>	<ul style="list-style-type: none"> <li>내부 DB와 결합분석을 통한 Warning System 구축(VoC: Voice of Customer 등)</li> <li>실시간 의사결정 지원 방안</li> </ul>
4 Data 중심 조직	<ul style="list-style-type: none"> <li>전문적 분석조직 및 전문 인력 양성</li> <li>데이터 기반 의사결정 조직 구조</li> </ul>	<ul style="list-style-type: none"> <li>빅 데이터 분석 전문 조직 검토</li> <li>Insight를 끌어낼 수 있는 전문가 채용</li> </ul>

출처: Big Data, 미래를 여는 비밀 열쇠, KT 경제경영연구소, 2011.7



[www.kr.kpmg.com](http://www.kr.kpmg.com)

© 2012 Samjong KPMG ERI Inc., the Korean member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative ("KPMG International"), a Swiss entity. All rights reserved. Printed in Korea.

The KPMG name, logo and "cutting through complexity" are registered trademarks or trademarks of KPMG International Cooperative ("KPMG International").