

스마트 데이터 := 빅데이터 X 인공지능

사례를 통해 살펴 본 Big Data의 진화

2015. 04. 03

안 준 형 / 본부장

사례를 통해 살펴 본 빅데이터의 진화

I

빅데이터의 기술과 시장전망

II

빅데이터의 분석 성공 제5원칙

III

빅데이터 구축사례

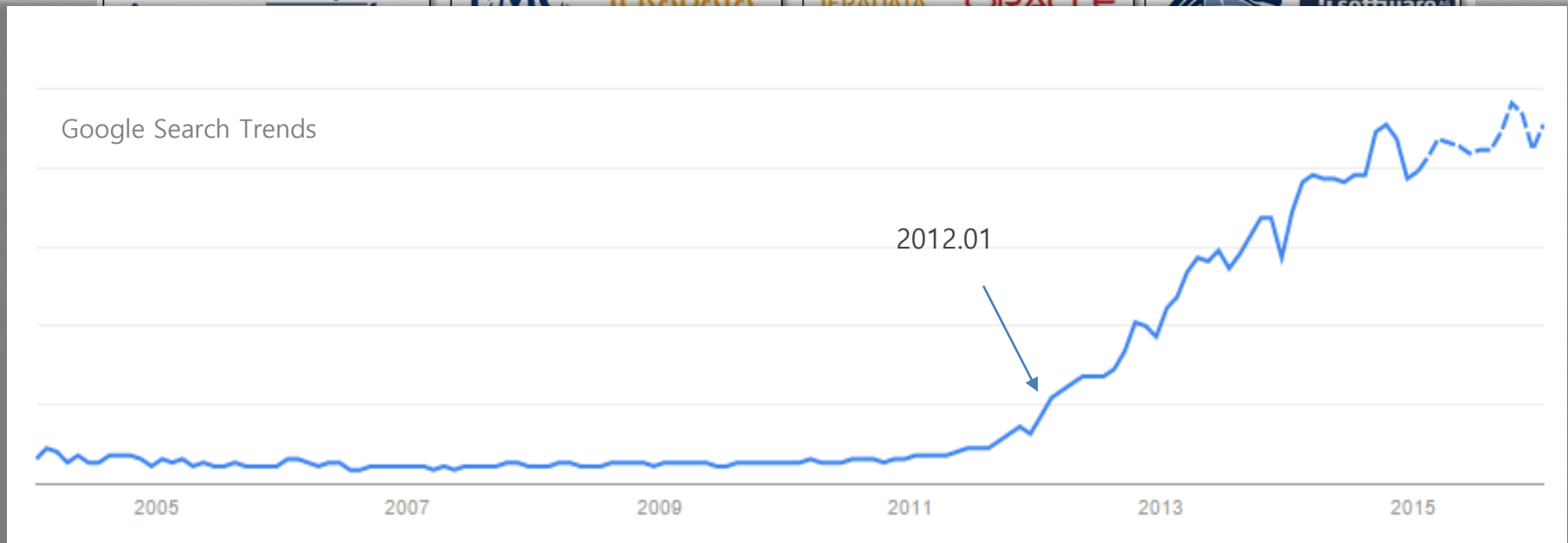
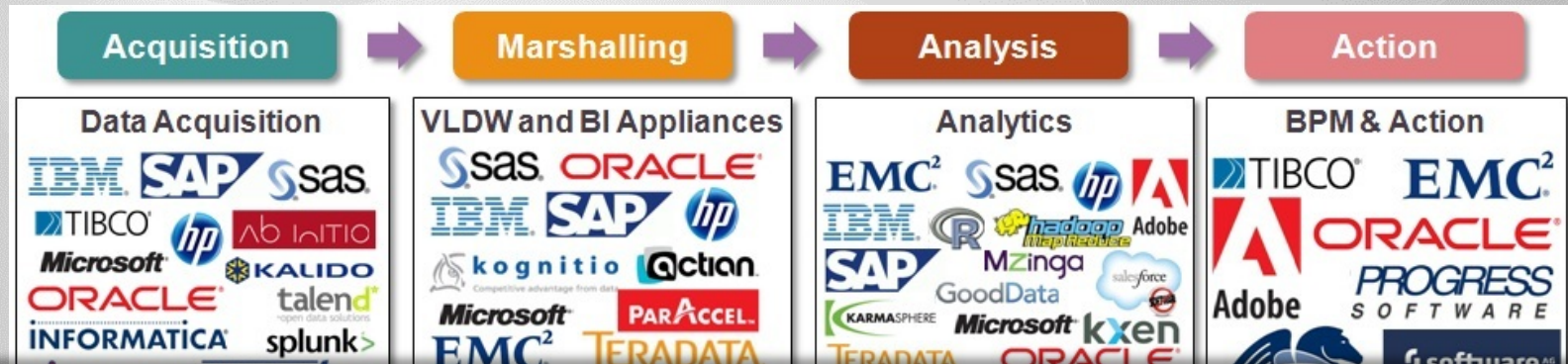
IV

빅데이터 관련 제품 및 기술소개



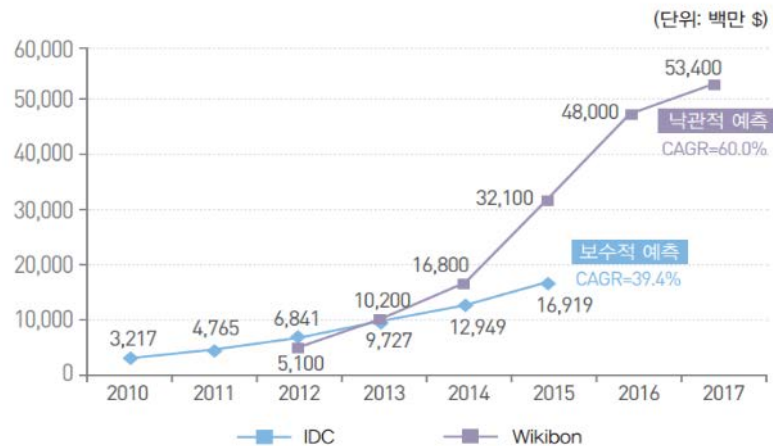
1

빅데이터의 기술과 시장전망



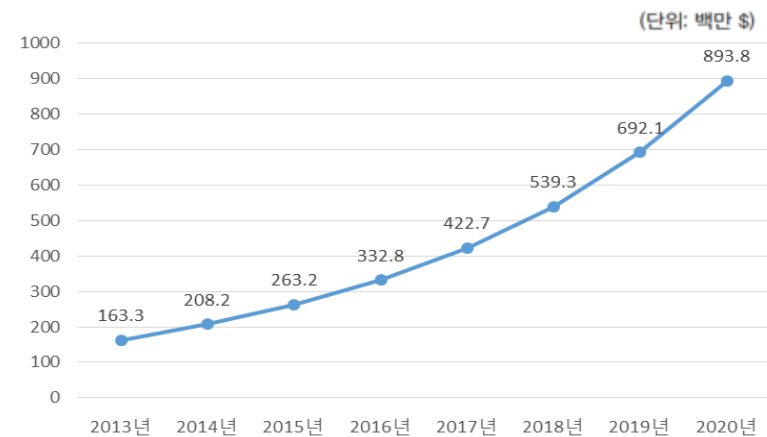
국내외 빅데이터 시장 규모는 기관마다 차이를 보이고 있으나, **2015~2020년** 매년 **40% 이상 성장**할 것으로 전망되며, **스마트 데이터의 실시간 처리**가 중요해집니다

빅데이터 세계 시장 규모



- IDC(International Data Corporation)는 전세계 빅데이터 시장이 매년 **39.4% 성장**하여 2015년 169억 달러 규모로 증가 전망
- 위키본(Wikibon)은 전세계 빅데이터 시장 규모가 2012년 51억 달러에서 2015년 321억 달러로 보다 높은 성장률(**연평균 60%**)에 이를 것으로 예상

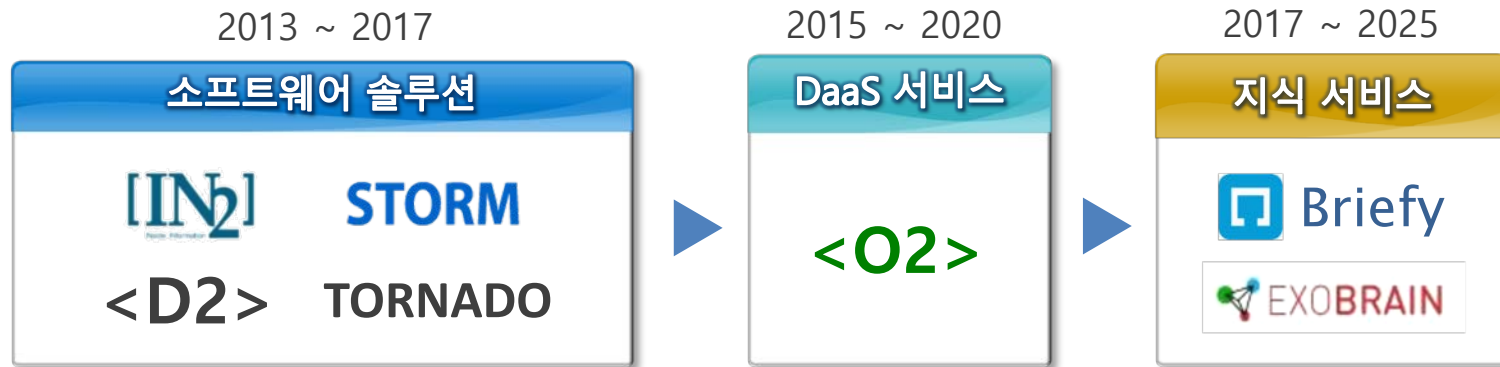
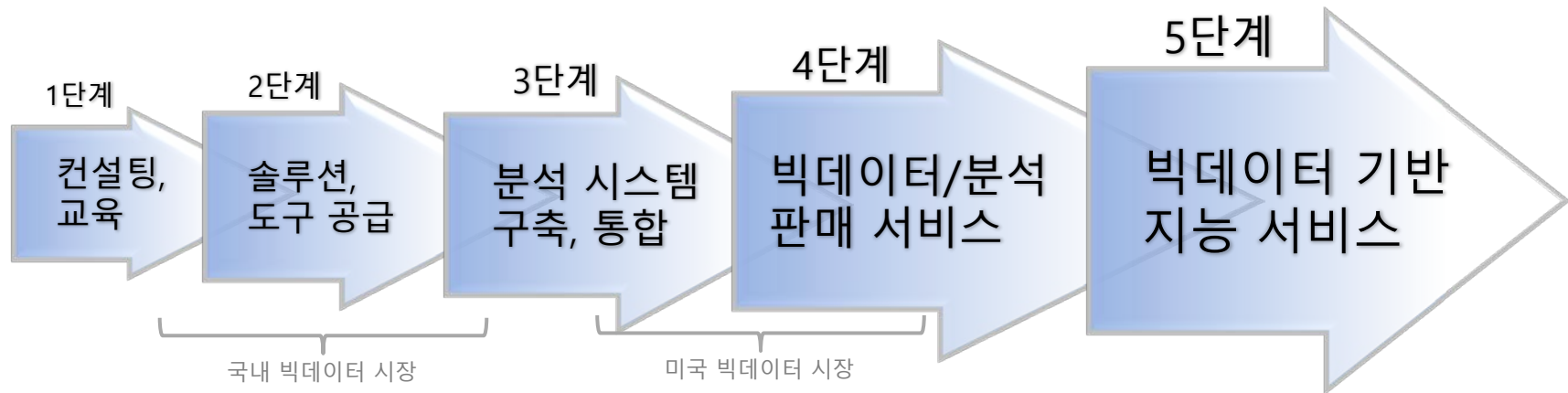
빅데이터 국내 시장 규모



- 국내 빅데이터 시장은 2015년 약 **263백만 달러**, 2020년 약 **900백만 달러**(한화 1조 원)로 예상
- 국내 ICT 관련 산업에서 빅데이터 분야가 차지하는 비중은 2013년 0.6%에서 지속적으로 증가하여 2020년에는 약 2.6%에 이를 것으로 전망

빅데이터 산업이 **솔루션 시장**과 빅데이터 **DaaS 서비스 시장**을 거쳐 결국에는 빅데이터 기반 **지능 서비스 시장**으로 성장할 것으로 예측됩니다.

빅데이터 산업 발전 5단계



빅데이터 분석은 합리적 의사결정과 미래예측에 대한 시도와 수요 증가함에 따라 데이터에 기반한 논리적 발견과 통찰력 확보를 통해 올바른 의사결정 추구

빅데이터 역할과 기대가치

미래 사회 특징

빅 데이터의 역할과 가치

불확실성

통찰력

- 현실 세계 데이터 기반의 패턴 분석, 전망
- 다각적 상황 고려 큰 그림 이해, 통찰 확보
- 사회 현상 이해와 시나리오 시뮬레이션

리스크

대응력

- 환경, 소셜 데이터 분석과 이상 징후 감지
- 이슈 사전 인지와 실시간 의사 결정 지원
- 국가, 기업 경영 투명성 제고와 비용 절감

스마트

경쟁력

- 평판, 트렌드 분석 통한 기업 경쟁력 확보
- 상황 인지, 인공지능 기반 대국민 서비스
- 개인화, 지능화 기반 차세대 사업 모델

융합

창조력

- 이질적 지식의 융합 분석과 신 가치 창출
- 상관 관계 이해를 통한 시행착오 최소화
- 컨버전스 패턴 분석을 통한 융합 시장 창출

“금융 서비스 기관들의 **운영 비용 중 92%가 데이터 처리를 위해 사용**”

“일반적인 금융 서비스 회사를 기준으로, 2006년에 순수익 100만 달러 당 초당 129만 개의 명령어 처리가 필요했는데, 2010년 말에 되자, 179만개 명령어로 38% 증가했고, 물리적 서버는 46% 증가했다. 같은 기간 동안 순수익은 훨씬 저조한 증가세(19% 미만)를 보였다. **컴퓨팅 파워의 니즈는 수익보다 2~5배 더 빠르게 성장하는 경향이 있다.**”

(Wall Street & Technology, Howard Rubin)



Why the finance world should care about big data and data science

Roger Magoulas on data's potential to improve finance systems and create new businesses.

by Mac Stocum | @macstocum | +Mac Stocum | Comment | 31 August 2011

Finance experts already understand that data has value. It's the lifeblood of their industry, after all. But as O'Reilly director of market research [Roger Magoulas \(@rogerm\)](#) notes in the following interview, some in the financial domain may not grasp all that data has to offer. Data science and big data have led to an expansion of data types, Magoulas says, and the associated influx of information could very well shape investment strategies and create new businesses.



How does big data apply to the financial world?

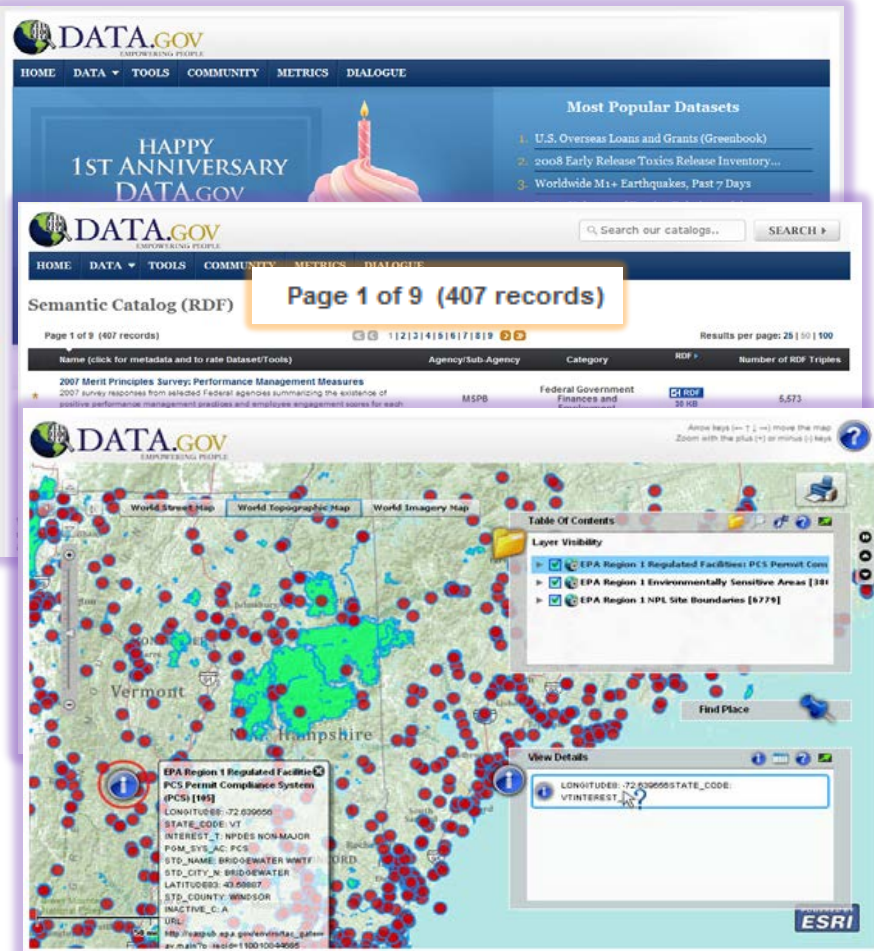
Roger Magoulas: There are two flavors of it. One is analyzing things like your investments, econometrics, trading activity, and longer-term data analysis. That's clearly part and parcel of the finance business, and people in the space already have great familiarity with this side of data.

The second flavor is the integrated approach to data in all facets of how organizations do business. This involves understanding customers, understanding competitors, [understanding behavior](#), taking advantage of the [world of sensors](#), and using a computational and quantitative mindset to make sense of a very confusing world.

Is there a disconnect between the finance world and terms like "data science" and "big data"?

Roger Magoulas: Everyone is [struggling with the semantics](#), so finance isn't worse off than others. They're actually making an effort to understand it. Adding to the semantic confusion, the terms "data science" and "big data" are sometimes co-opted by organizations trying to show how they embody these attributes. That's fine, but the finance ecosystem has a responsibility to learn as much as it can about these areas. The best way to do that is directly from the data science practitioners: see the tools data scientists use and how they approach their work. That firsthand experience will help finance experts inform their investment strategies and see where the data space is heading.

“국가재정 강화, 지능형 교통, 질병치료, 범죄예방, 국가안전관리 강화 등을 위한
정부 빅데이터 활용”



- 미국 국세청, 탈세 방지 시스템 통한 국가 재정 강화
빅데이터 기반 통합형 탈세, 정부사기 방지 시스템을 통해 연 3,450억 달러 절감
- 일본, 센서데이터를 활용한 지능형 교통안내 시스템
노무라 연구소가 시행한 실시간 최적 경로 안내와 에너지 절감 시스템
- 미국 국립보건원, 유전자 데이터 공유를 통한 질병치료
1700명의 유전자 정보를 아마존을 통해 개방, 질병 진단과 예측 연구비 절감
- 보험회사, 웹포인트의 효율적 환자 치료 대응
IBM 왓슨 도입을 통해, 환자에게 적절한 정보와 최신 치료법을 제시
- 싱가포르, 국가위험관리시스템을 통한 국가안전관리
빅데이터 기반의 테러, 재난, 전염병 등의 위험을 관리하는 RAHS 운영
- 샌프란시스코, 범죄 예방 시스템으로 안전 지역사회 구축
8년간 범죄 데이터 분석을 통한, 범죄 발생 예측 시스템 운영 (71% 정확도)

2

빅데이터 분석 성공 제 5 원칙

솔트룩스는 지난 5년간 수많은 빅데이터 분석 플랫폼 구축 및 분석 컨설팅 사업 경험을 통해 빅데이터 분석 시스템 구축을 위한 5가지의 핵심 성공 원칙을 도출하였습니다.

1 이해 관계자와의 비전/목표를 합의하라

2 작게 시작하고 빠르게 움직여라

3 끝에서부터 시작하라

4 미들 아웃 방법이 중요하다

5 기술은 단지 사람을 거들 뿐

이해 관계자와의 비전/목표를 합의하라

- 의사결정자, 데이터 제공자들, 결과 수요자들, 개발자들
- 결과에 대한 기대 수준을 정의하고 합의하라
- 데이터와 기술의 현실과 한계를 직시해야만 한다



작게 시작하고 빠르게 움직여라



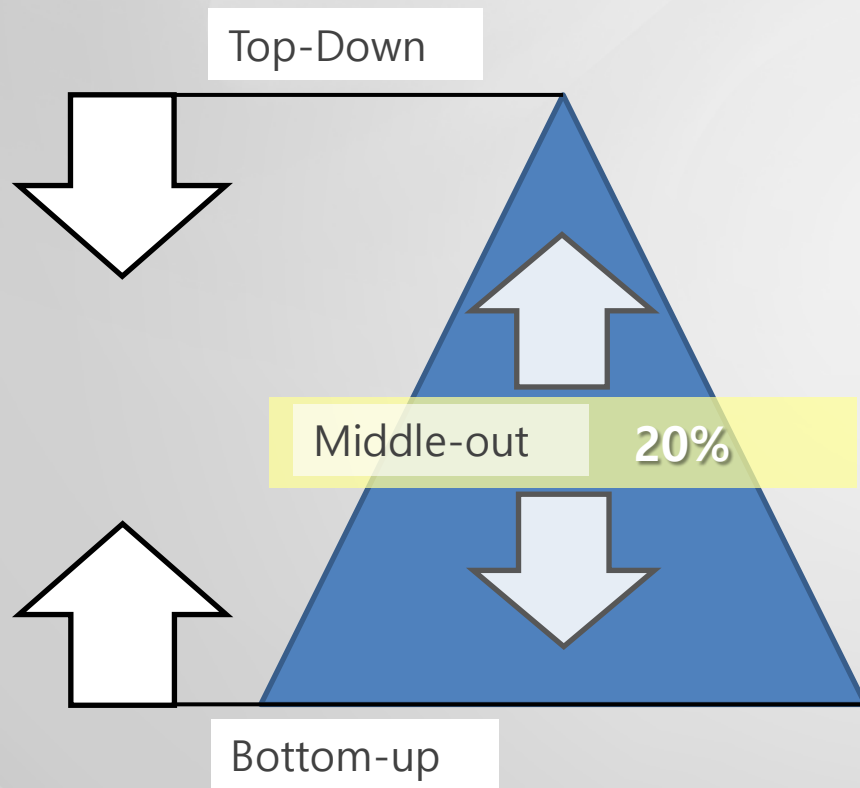
- 잘못하다가 소 잡는 칼 만들어 닭 잡게 된다
- 작은 데이터와 시스템으로 시작해 단계적으로 확대하라
- 단계별 검증을 통해 데이터 크기, 종류, 시스템 규모를 확장하라

끝에서부터 시작하라



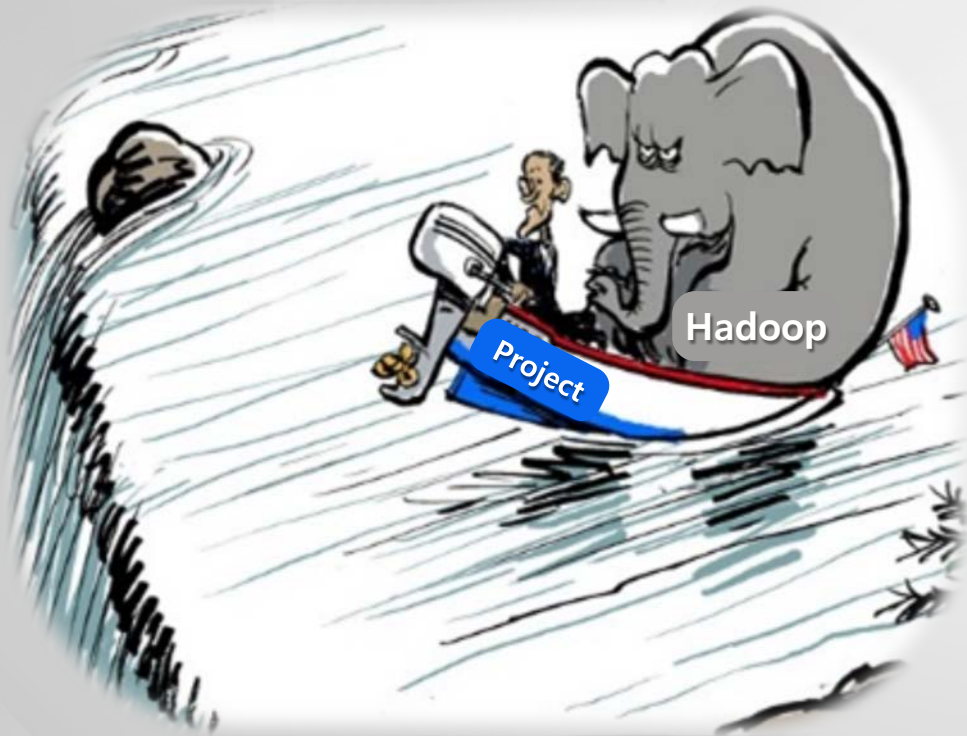
- 시작 때 분석 결과가 어떻게 보일지 시각적으로 상상할 수 있어야 한다
- 분석 체계, 결과의 활용과 발전이 지속가능할지 시작할 때 점검하라 (응용/기술/사람/예산)

'미들 아웃' 방법이 중요하다



- 20%의 중심/핵심 데이터가 전체의 80%를 설명한다
- 목적에 부합되는 핵심 데이터는 반드시 현업 전문가, 수요자 인터뷰를 통해 파악한다 (CQ)

기술은 단지 사람을 거들 뿐



- 최신 기술 보다는 조직과 규모에 적합한 기술을 적용하라
- 시작부터 분석 품질 메트릭스와 검증 데이터를 확보, 끝이 아닌 과정에서 수집 데이터와 적용 기술을 검증해야 한다

3

빅데이터 구축사례

KT, 한화, 도로공사, 건강보험공단 등에 **고객 목소리(VOC) 통합 분석 시스템** 제공

KT Vital – 고객 목소리 분석

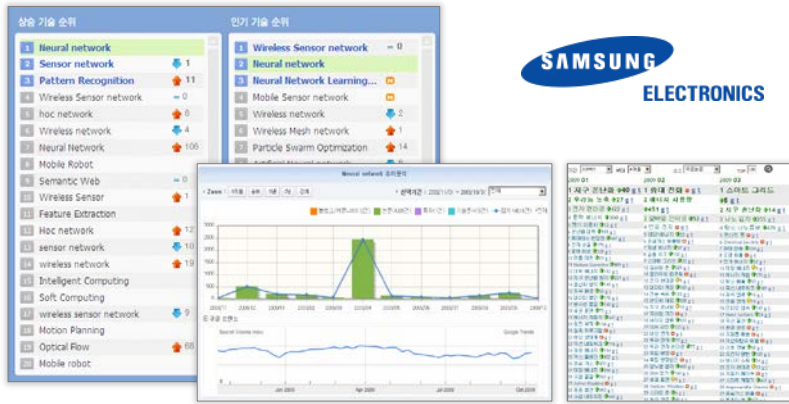


한화 그룹 – 고객 목소리 분석



기업의 방대한 기술 문서와 공공 기관의 기술/특허/정책 빅데이터 분석과 미래 예측

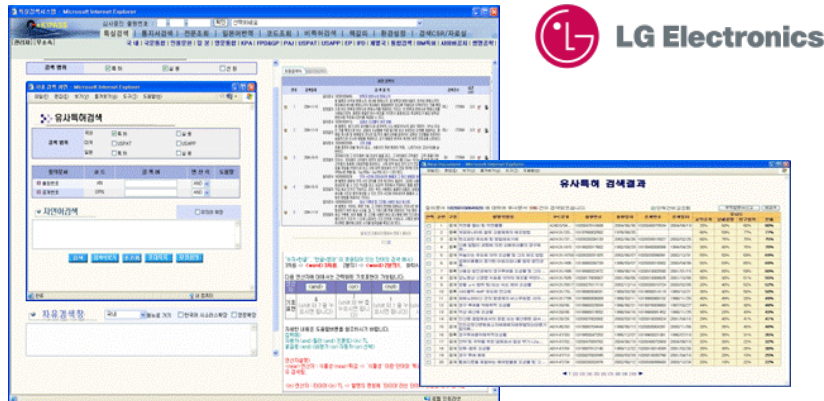
삼성전자 신기술 센싱 (한/영/일 5천만 기술문서)



KISTEP R&D 성과분석 및 예측



LG전자 특허 자동 분석 및 추천



특허청 특허 데이터 분석

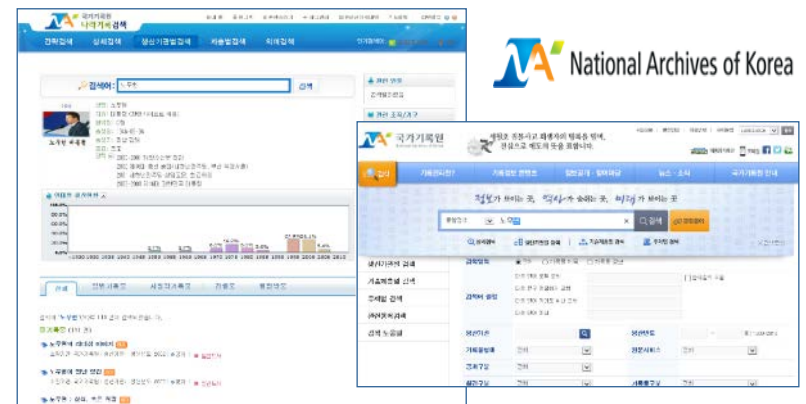


공공 빅데이터의 개방과 활용, 분석을 위한 솔루션 제공과 통합 포털 구현

안행부 정보공개 포털 (15억 건, 국내 최대 규모)



국가기록원/대통령기록관 시맨틱 검색



부산시 지식 네트워크 분석 / 전문가 추천



경기도 공공데이터/LOD 포털

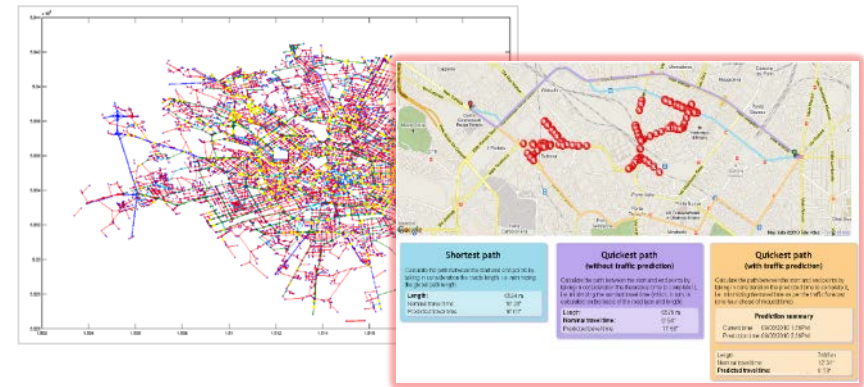


센서네트워크 기반의 실시간 빅데이터 모니터링/예측과 미디어 빅데이터 분석과 추천

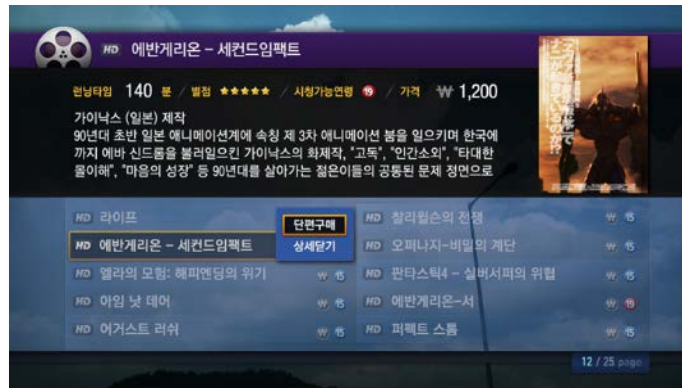
국토부 스마트시티 지하매설물 관리



EU/밀라노 스마트시티 교통 예측/최적화



KT IPTV VOD 검색과 맞춤형 추천



KBS 차세대 미디어 플랫폼 / 추천 엔진



국방 부문에서 공통 사용 가능한 국가 **안보 빅데이터 분석** 공통 플랫폼 제공



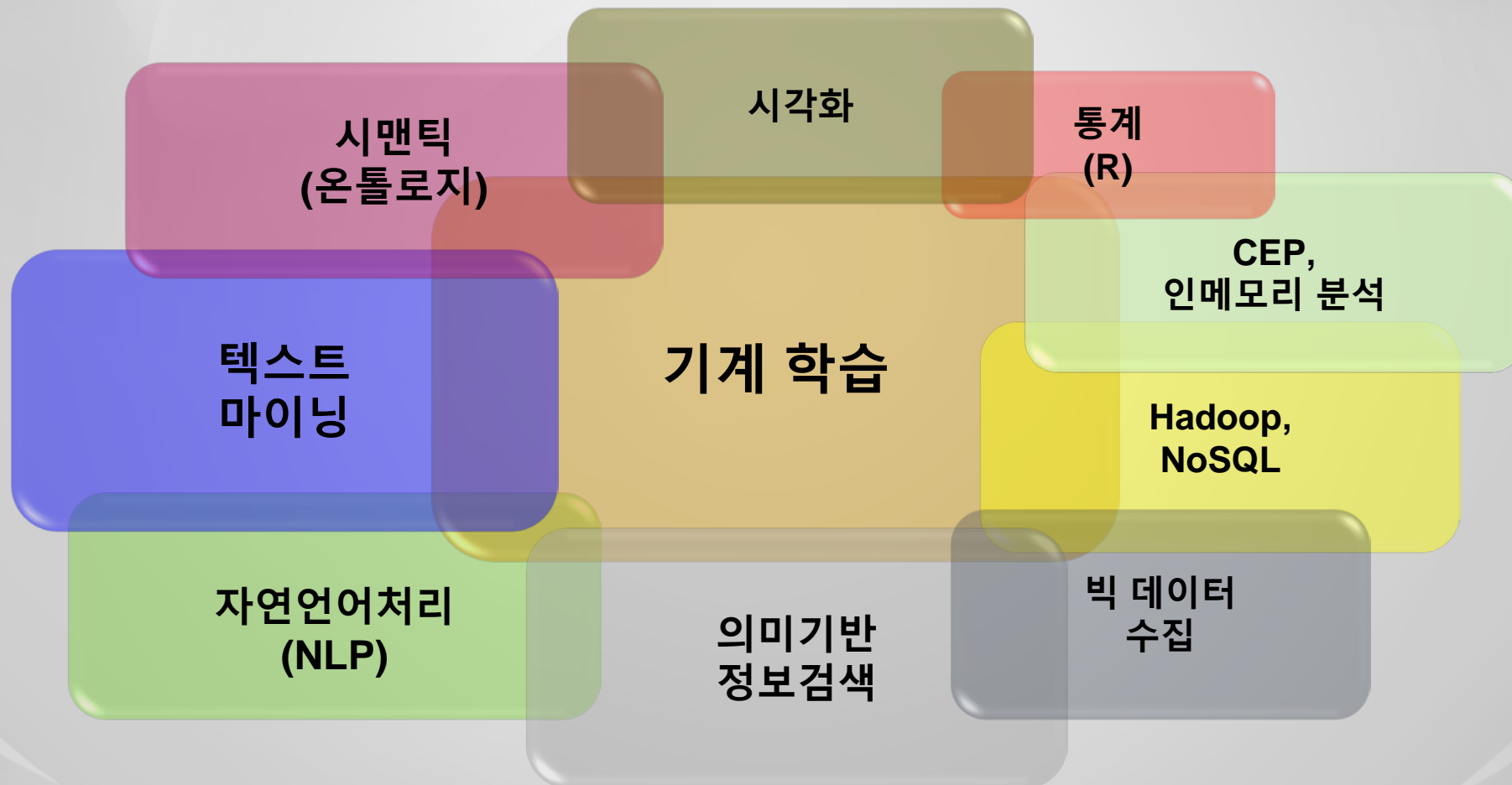
플랫폼 구성도



4

빅데이터 관련 제품 및 기술소개

데이터 수집, 변환, 필터링, 저장, 검색, 분석, 검증, 시각화 순의 실행 체계로 구성되며
데이터 종류와 분석 목적, 범위에 따라 적용되는 기술이 상이



빅데이터의 수집, 변환, 분석, 시각화, 의사결정 지원에 이르는 **빅데이터 분석 가치사슬 전체를 커버하는** 최고의 플랫폼 기술과 시스템 구축역량 보유하고 있습니다.

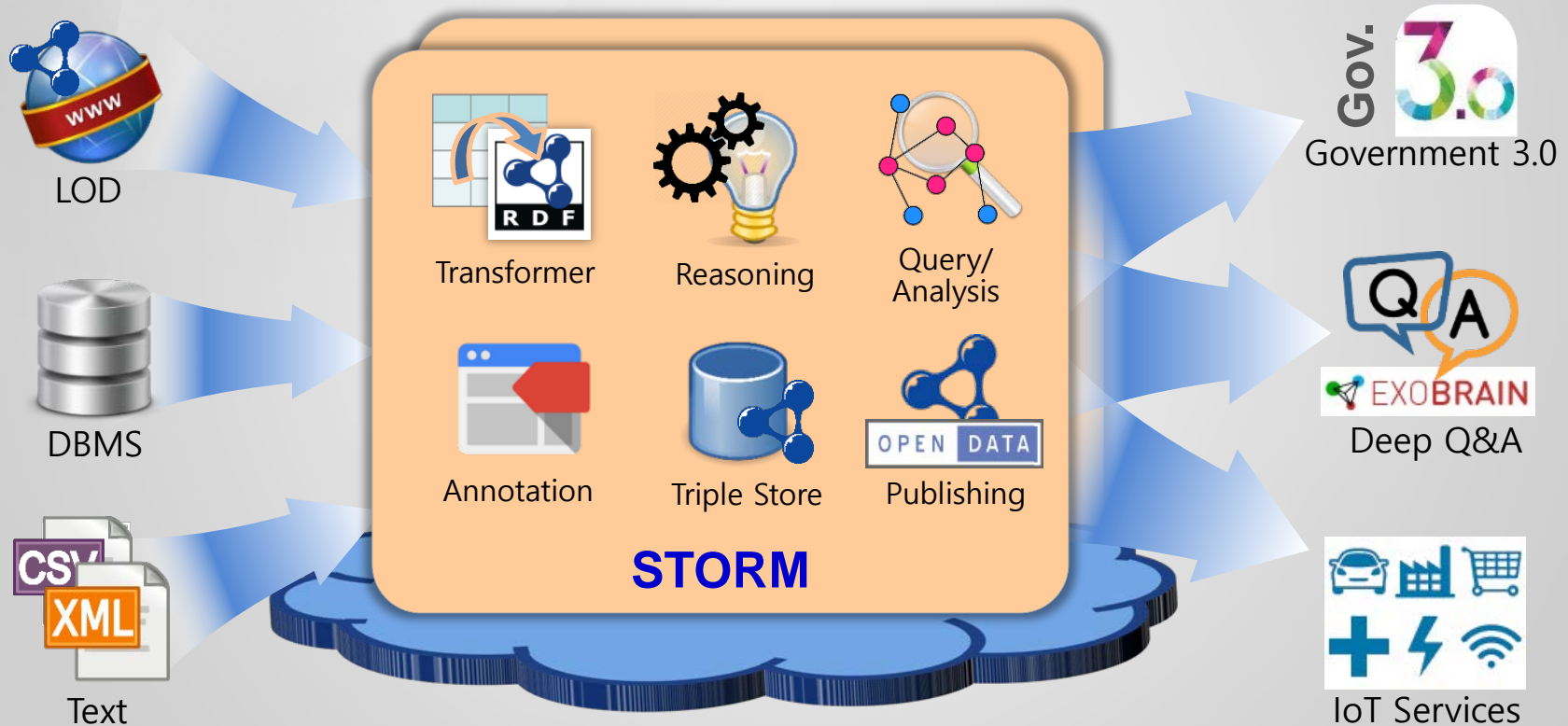


[IN2]는 클라우드 컴퓨팅 및 의미 기반 대용량 정보 검색, 텍스트마이닝과 지식표현, 추론 기술이 융합된 "클라우드 기반 **시맨틱 검색 및 마이닝 플랫폼**"입니다.



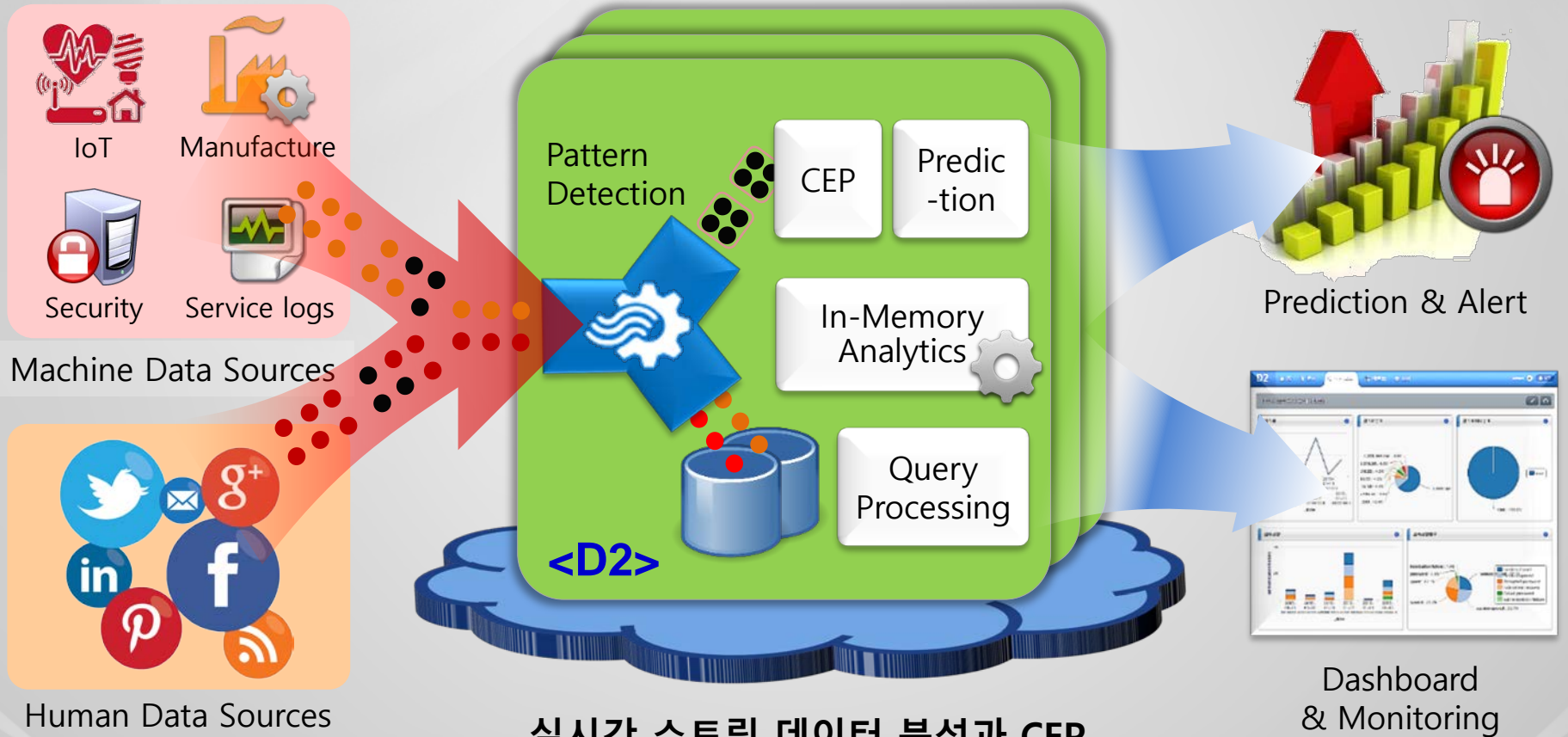
실시간 비정형 분석과 시맨틱 검색

STORM은 LOD 대응이 가능하도록 시맨틱 메타데이터의 저장, 관리, 질의와 하이브리드 추론기술 내장한 **초대용량 시맨틱 데이터 웨어하우스** 입니다.



시맨틱(그래프) 빅데이터의 저장과 하이브리드 추론

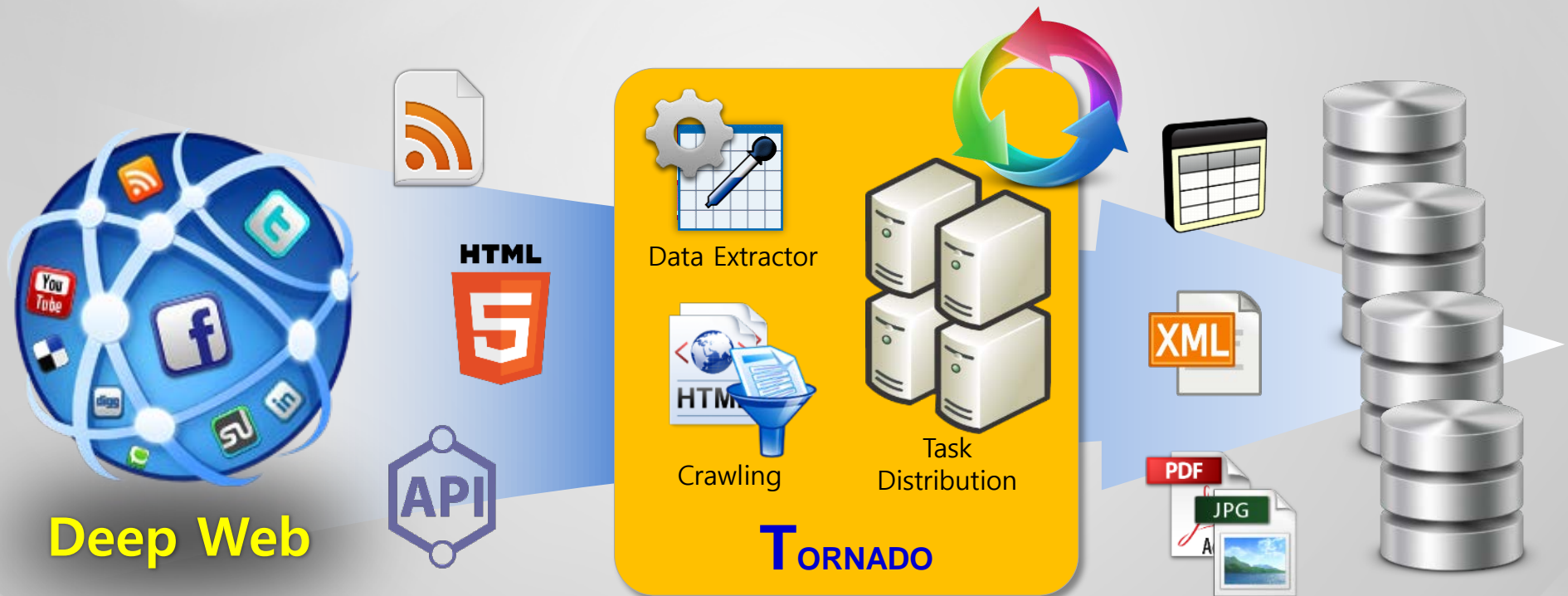
<D2>는 머신 데이터를 포함한 다양한 형식의 스트림데이터(비정형, 반정형)를 실시간 수집/통합/ 분석/예측 및 CEP 처리가 가능한 빅데이터 분석 플랫폼입니다.



실시간 스트림 데이터 분석과 CEP

TORNADO 는 동적 딥웹(Deep Web)에서 사용자가 원하는 방대한 데이터를 실시간으로 자동, 병렬 수집 가능한 **다목적 빅데이터 수집 시스템**입니다

소셜미디어, 전자상거래 등 웹 상의 모든 빅데이터를 자동 수집, 추출, 변환, 저장



웹으로부터 빅데이터를 자동 수집과 변환

Rainbow는 다양한 소스로부터 지속적으로 발생하는 대용량 데이터를 사용자가 쉽게 이해할 수 있도록 돕는 강력한 **시각화 및 BI엔진** 입니다.



데이터 시각화와 퍼블리싱

솔트룩스는 세상 사람들이 자유롭게
지식 소통하는 세상을 꿈꿉니다!!

감사합니다.

