2012-11-21
김영석 차장, Business Analytics Software, IBM Korea

Big Data 기반 예측 분석(Predictive Analytics)의 중요성과 활용방안

CONTENTS

1. Big Data란?
2. IBM Business Analytics for Big Data
3. Use Case

Big Data

is about finding the right needle in a stack of needles

Big Data란....

- 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석 을 지원하도록 고안된 차세대 기술 및 아키텍처

  - IDC, '11

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

  - McKinsey, '11

- 빅데이터는 당초 수십-수천 테라바이트에 달하는 거대한 데이터 집합 자체만을 지칭하였으나 점차 관련 도구,플랫폼, 분석 기법까지 포괄하는 용어로 변화

  - 삼성경제연구소, '10

- Data growth challenges (and opportunities) as Being three-dimensional, ie increasing volume(amount of data), velocity(speed of data in/out), and variety(range of data types, sources). Gartner continues to use this model for describing "big data"

  - 2000년 초 MetaGroup

Big Data는 새로운 기회를 제공합니다

Data at Rest

테라 바이트에서
제타 바이트까지
확장

12 terabytes

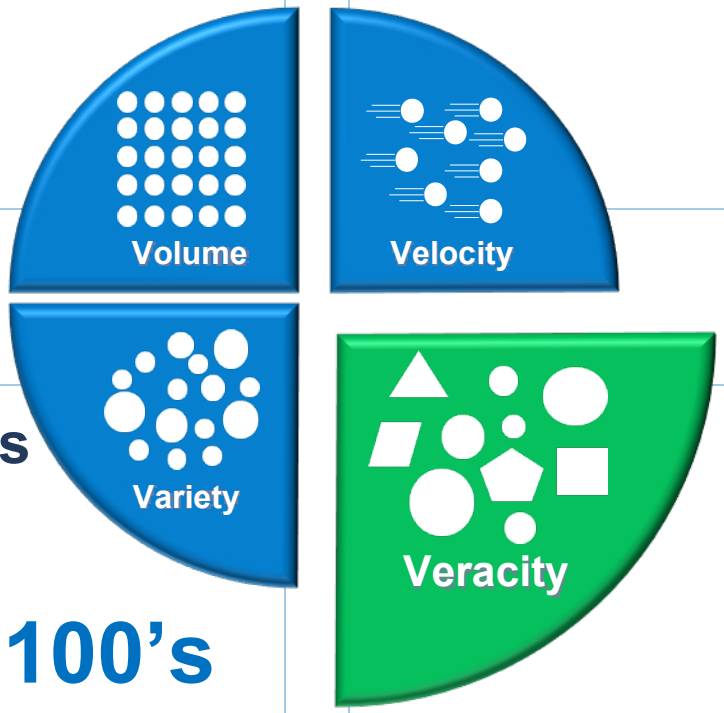
매일 생성되는 tweet

Data in Motion

5 million

초당 거래 이벤트 수

스트리밍 데이터/
대용량 데이터의
이동



Data in Many Forms

다양한 관계형 및
비 정형 구조의 데
이터에 대한 분석

100's

감시 카메라에서 쏟아지
는 다양한 동영상 피드

Data in Doubt

일관성 없음, 불충
분, 모호함, 대기시
간, 착시, 근사치로
인한 불확실성

다양한 유형의 모든 데이터에서 더 많은 Insight 발굴

Transactional & Application Data

- Volume
- Structured
- Throughput

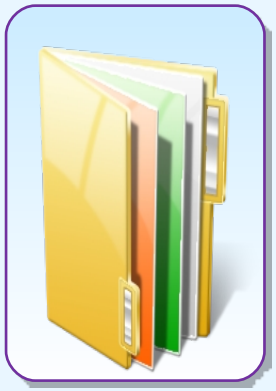
Machine Data

- Velocity
- Semi-structured
- Ingestion

Social Data

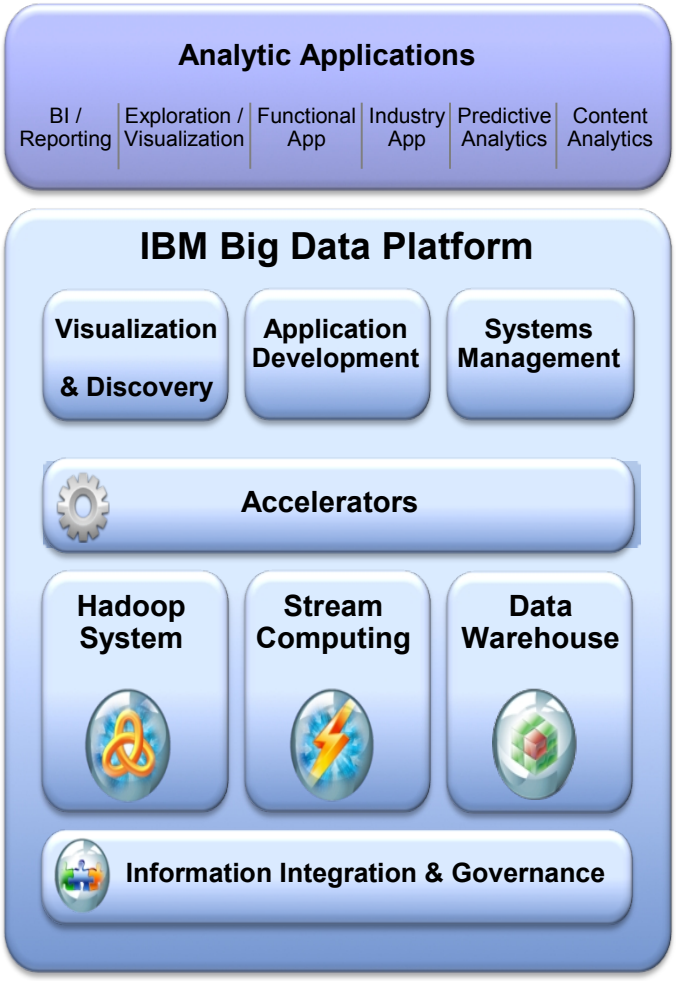
- Variety
- Highly unstructured
- Veracity

Enterprise Content



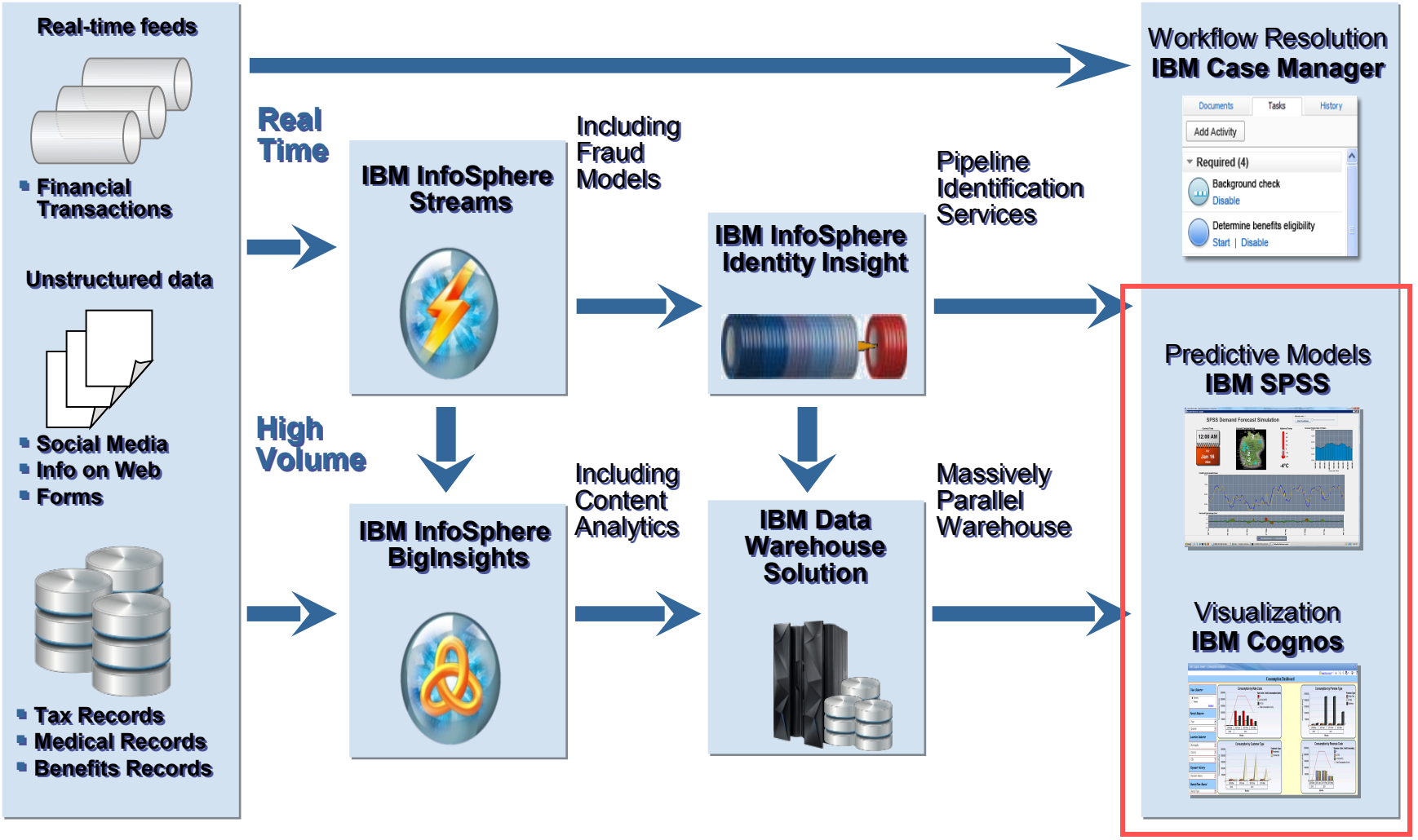
- Variety
- Highly unstructured
- Volume

Big Data 활용을 위한 다양한 Platform



- 연계된(federated) 빅 데이터 소스의 이해와 탐색
- 엄청난 양의 데이터를 저장하고 관리
- 모든 데이터의 구조화와 관리
- 스트리밍 데이터의 관리
- 비 정형 데이터의 분석
- 모든 데이터 소스의 통합(integration)과 관리(governance)

Big Data 활용을 위한 개념적 구성도



Big Data 통한 새로운 기회

Know everything about your customer
모든 데이터 소스의 분석

Run zero-latency operations
실시간 실행

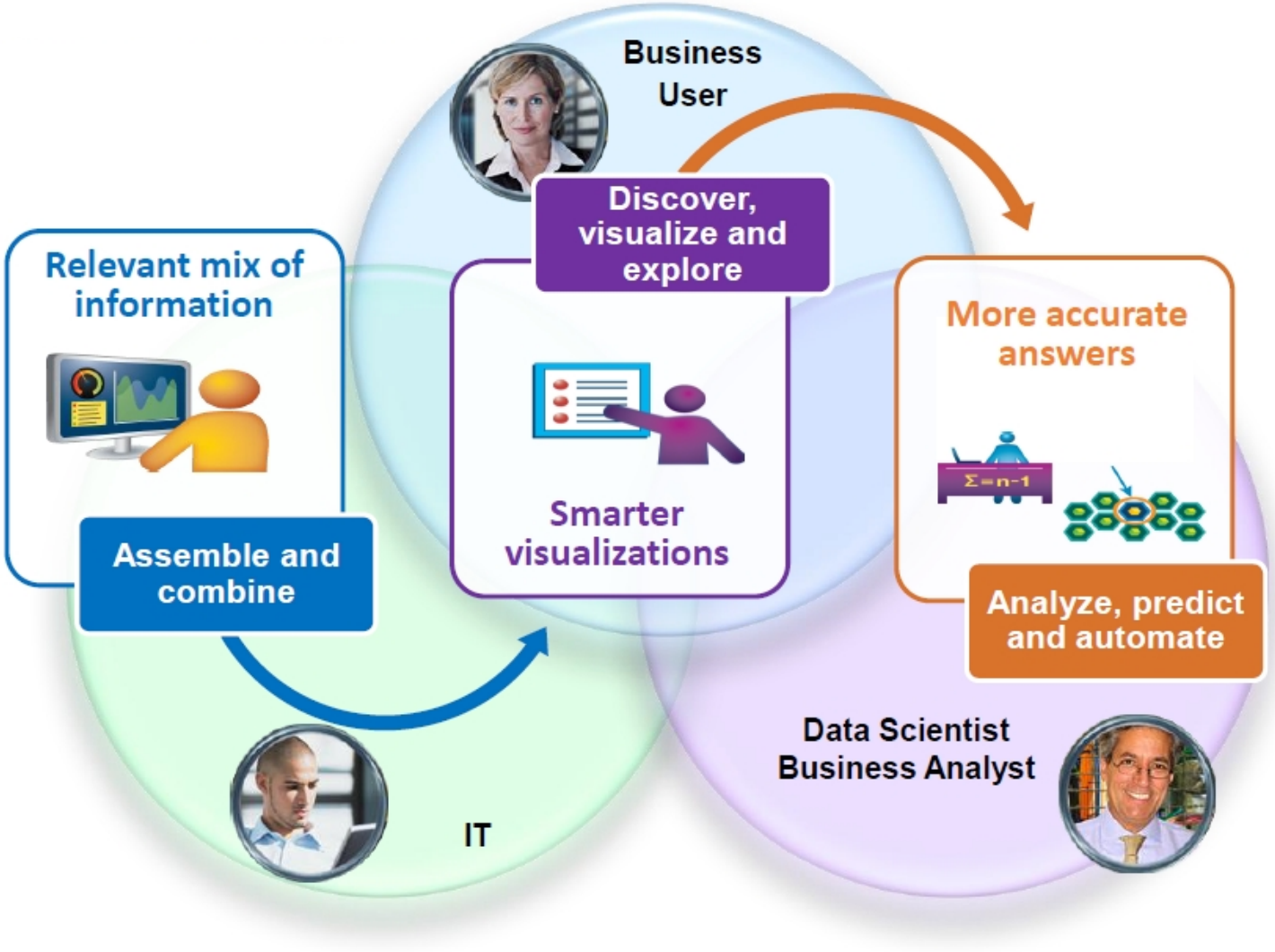
Innovate new products at speed and scale
모든 피드백을 수집하고 분석

Instant awareness of fraud and risk
실시간 사기 감지

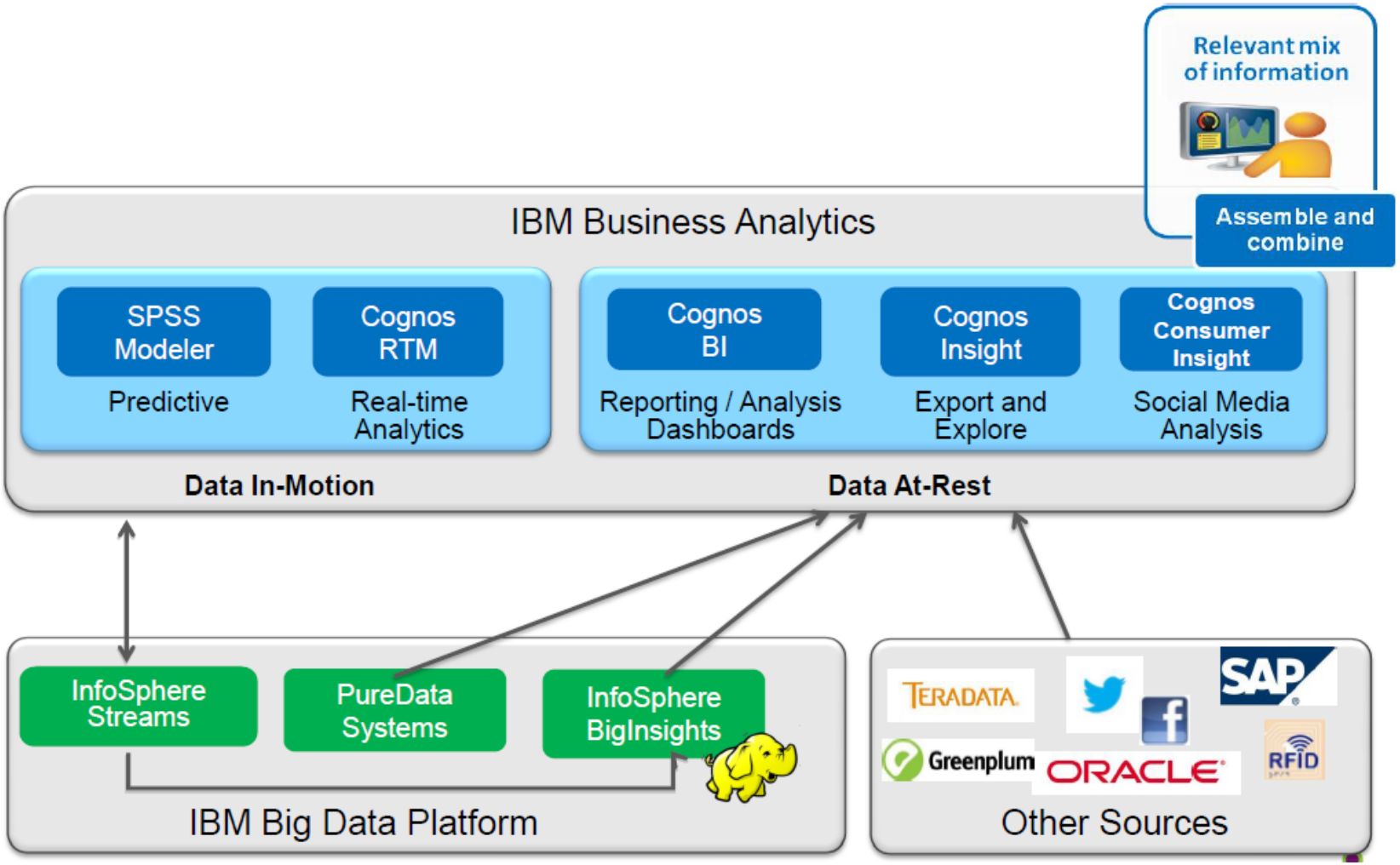
Exploit instrumented assets
새로운 제품과 서비스의 개발과 유지보수

IBM Business Analytics
for Big Data

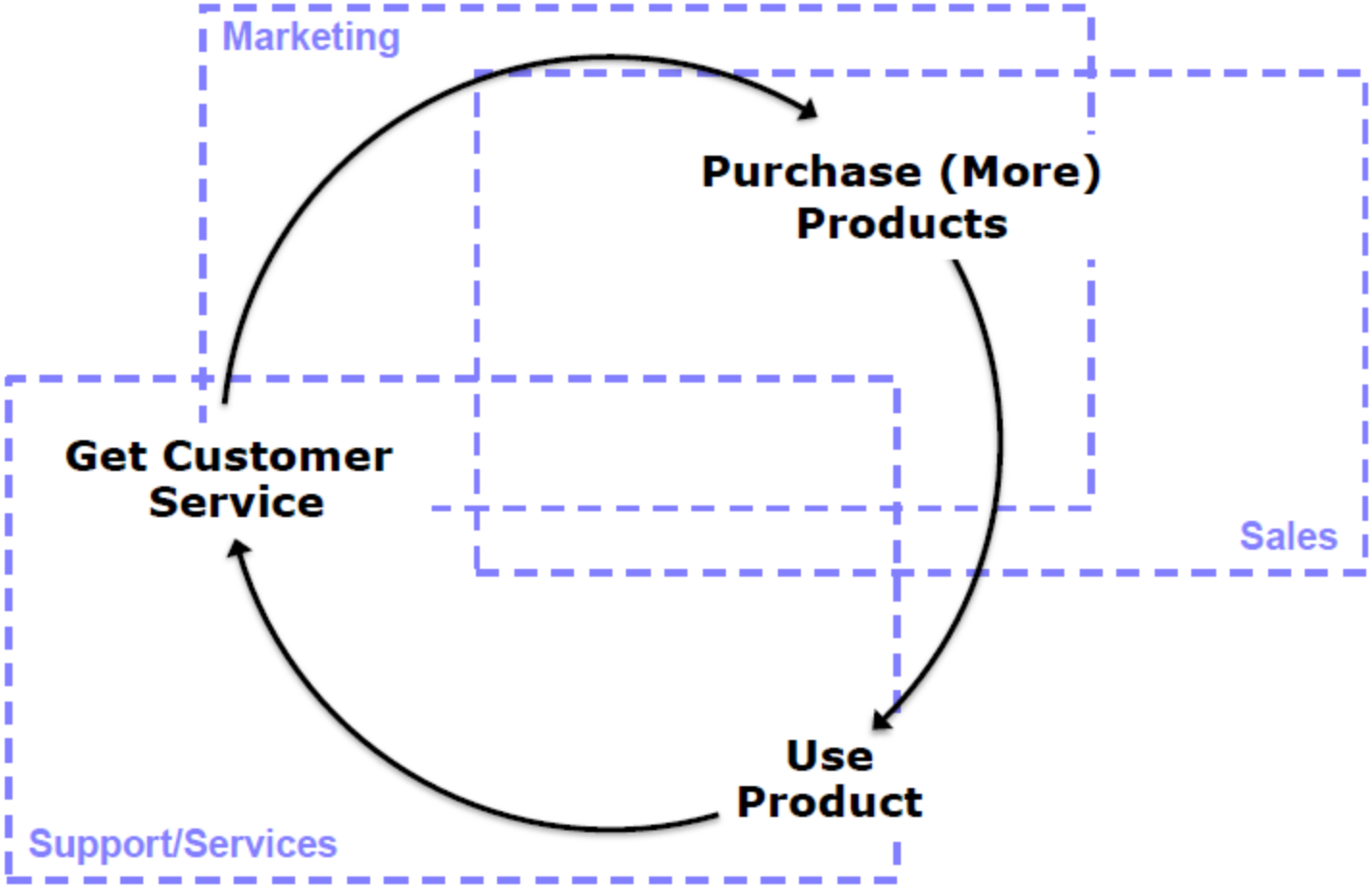
Big Data 기반의 IBM 분석 역량



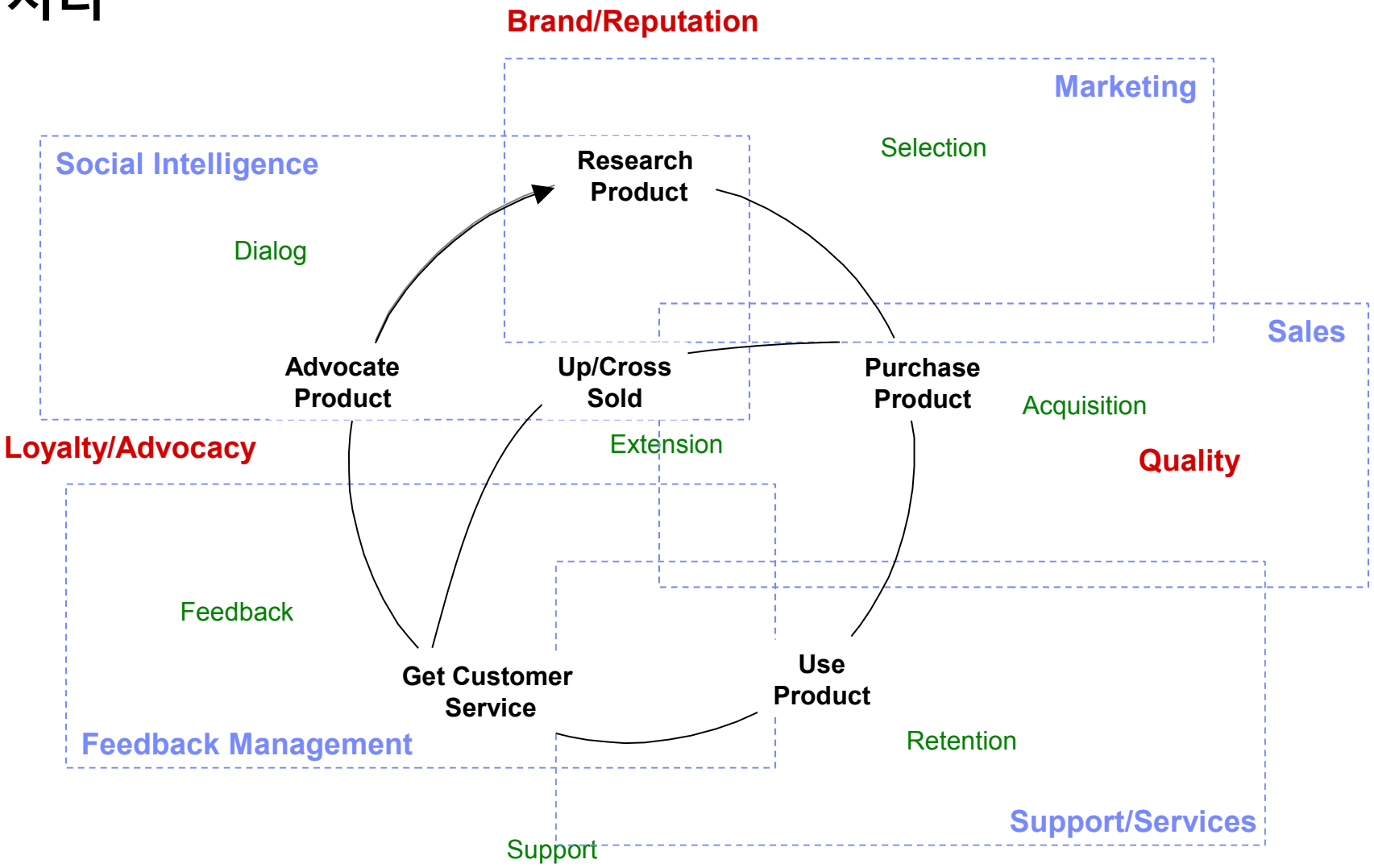
Big Data와 Enterprise Data의 결합을 통한 Business Analytics



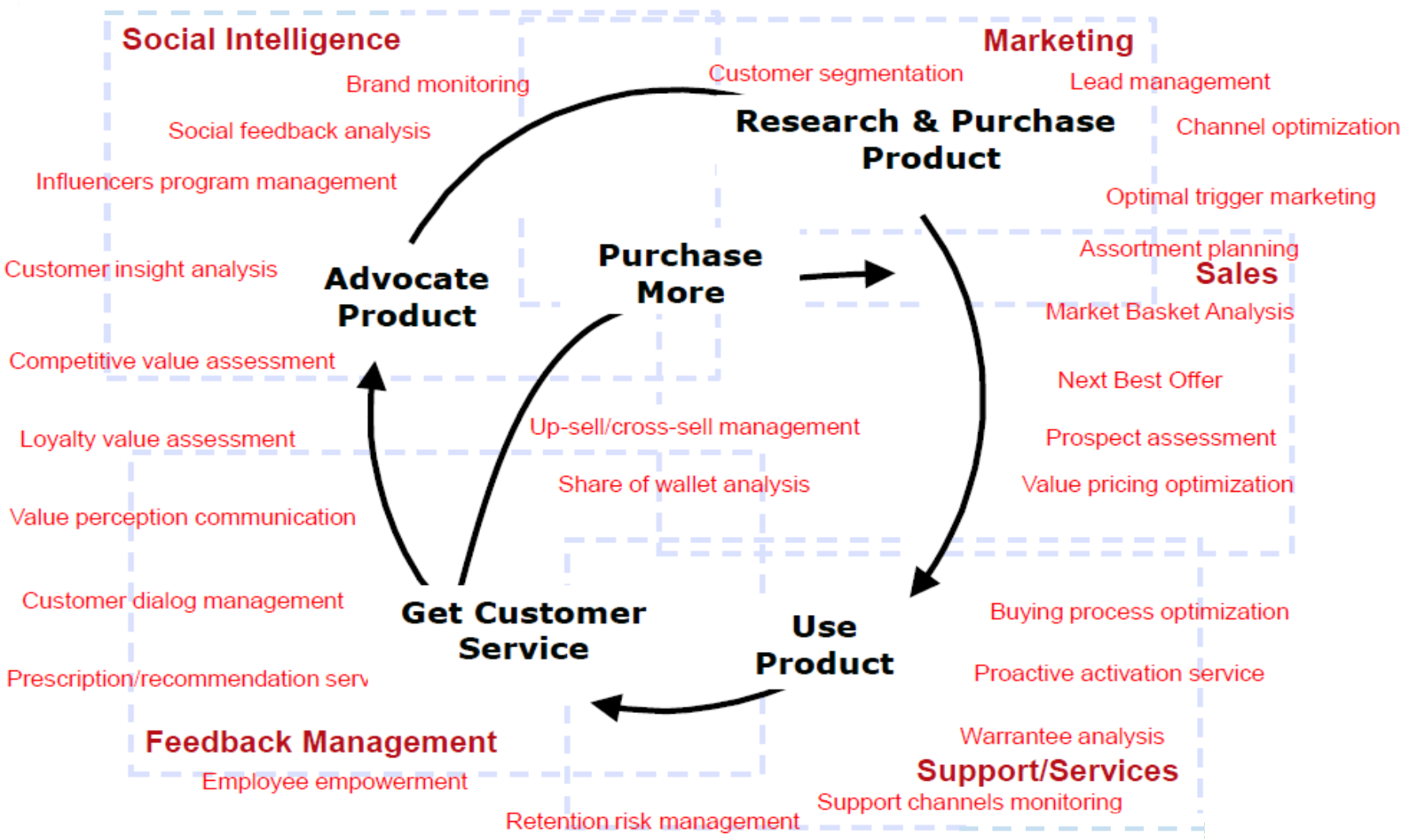
과거의 고객분석 접근 방향



보다 진화된 고객 만족(Customer Intimacy) – 모든 접점에서 처리



Big Data 관점에서 접근 가능한 모든 고객 분석 영역

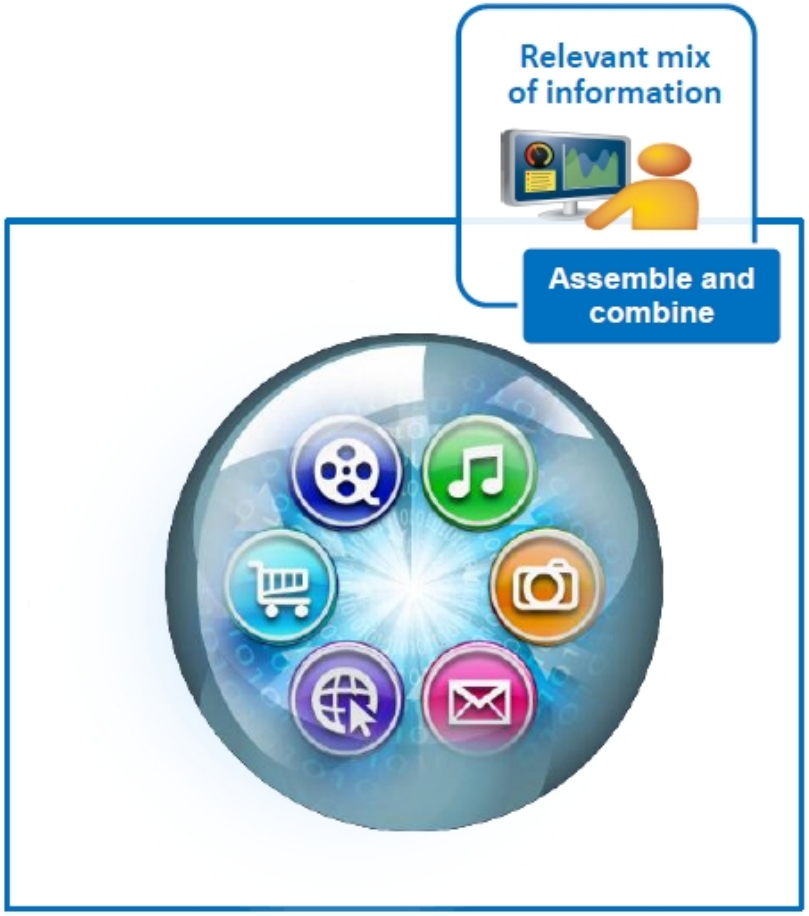


관련된 모든 정보를 연계 분석

다른 소스의 정보를 기업 내부의 정보와 함께 혼합하여 분석

- 빅 데이터 소스에 접근
- 기업 내부 데이터와 통합
- 대용량 데이터의 최적화

XO Communications



Benefit

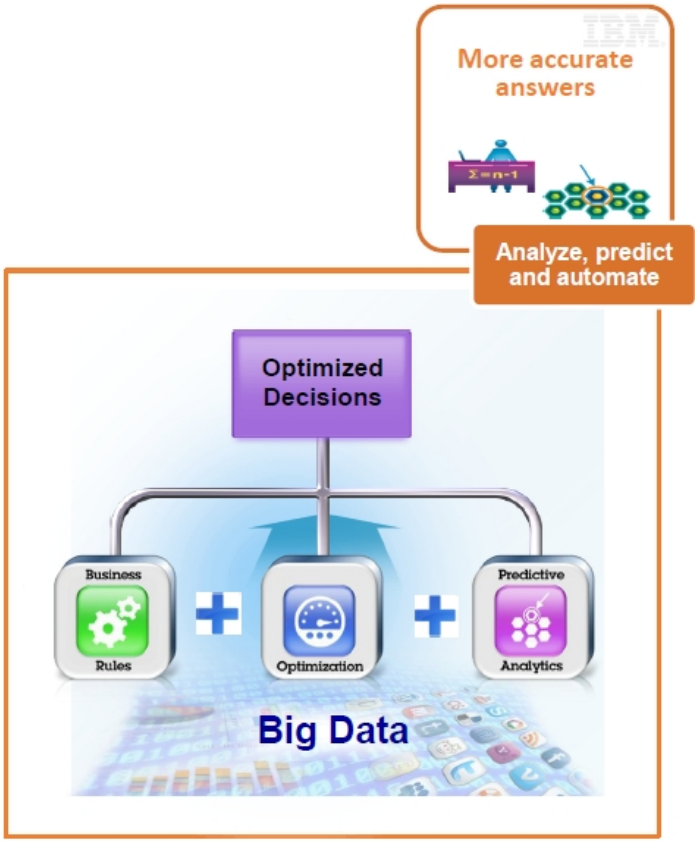
- 고객 서비스를 위한 획기적인 인사이트를 제공함으로써 고객 유지를 개선
- 매출 유지율 ***60% 개선***
- 다양한 소스 데이터를 활용한 고객분석

자동화된 예측 분석

자동화되고 최적화된 의사 결정을 위한 분석

INFINITY
PROPERTY & CASUALTY CORPORATION

- Big Data를 사용한 예측 모델 생성을 위한 알고리즘 사용
- Scoring과 Deployment을 위한 데이터 프로세싱

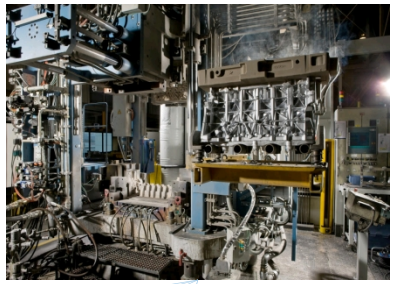


Benefit

- 더 똑똑한 사기 감지
- 사기 조사를 위한 시간의 ***95% 절감***
- 사기 청구 방지에 대한 성공율 ***50~88% 증가***

제조산업에서의 Big Data



Performance logs

Sensor Data

Production Logs

Environment Data

Volume

Velocity

Variety

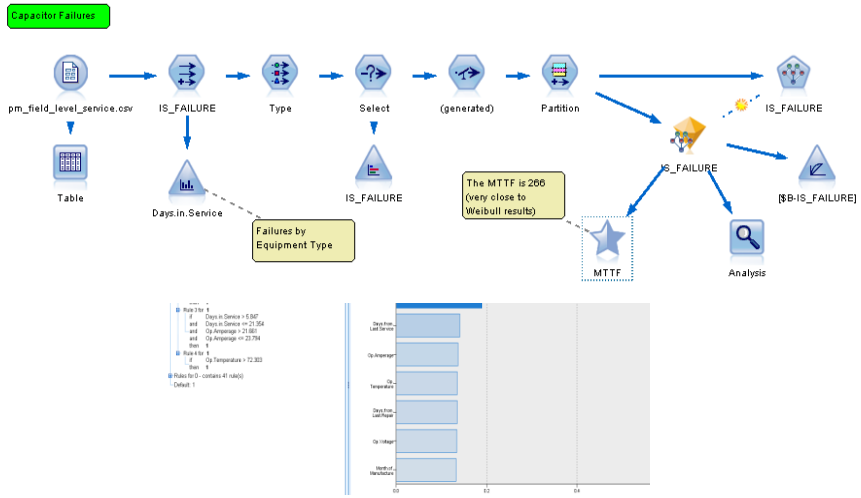
Veracity

결함요인 식별 및 이상감지 on Big Data

❖ SPSS 솔루션은 결함정보를 통하여 예측모델을 생성하며, 결함정보는 결함근본원인 분석을 통하여 결함이 발생하는 요인을 파악하여 개선함으로써 결함발생을 예방할 수 있습니다.

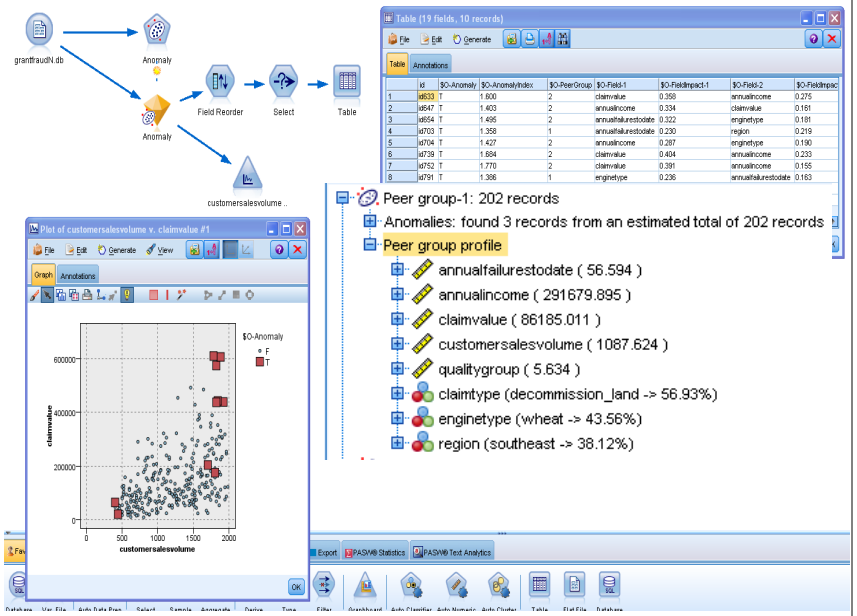
결함요인 식별

- 센서 로그, 유지 보수 기록, 컨디션 모니터링 데이터 등을 포함한 모든 데이터를 활용 가능.
- 예측 모델(**Neural Nets, Logistic Regression, Decision Trees, SVM**(기계학습알고리즘), **SLRM**(표준선형회귀모형), etc.)을 활용하여 모든 장비에 대한 미래의 결함확률을 추정
- 과거 결함이 발생한 정보를 통하여 결함이 발생하는 요인을 파악
- 결함이 발생하는 요인 정보를 예측모델 생성에 제공
- 결함이 발생하는 요인을 개선함으로써 결함이 발생하는 건수를 감소시킴



이상감지

- 예측모델을 바탕으로 실시간으로 모니터링 할 수 있는 기능을 제공함으로 써 신뢰성 있게 관리 이상 현상이 발생시 빠르게 대처함
- 예측모델을 바탕으로 각각의 이상현상의 범위를 설정하고, 이상을 사전에 감지하여 담당자에게 제공함으로써, 결함을 사전에 예방함



Energy & Utilities에서의 새로운 성과

Pacific Northwest NATIONAL LABORATORY

- 전체 Peak load의 ***15% 감소***
- 소비자 전기료의 ***10% 절약***
- 더 나은 자산관리를 통해 20년 동안 인프라 비용에서 ***$ 700 억 감소 예상***

ONCOR

- 스마트 그리드 투자 및 운영
- ***176달러 만의 비용 절약*** 을 통한 최종 고객에게 추가 요금 절감을 제공

Vestas

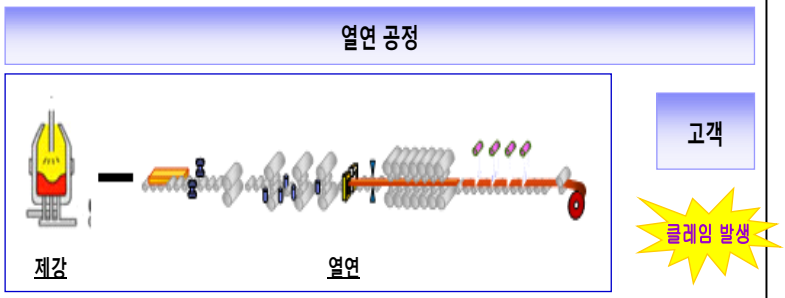
- 각 지역을 위해 공공 및 민간 기상 데이터의 ***2.8 petabytes 분석***
- ***97%까지 감소*** - 바람 예측 정보를 위한 모델링 시간 감소 (주 단위에서 시간 단위로)

Use Case
(Steel Industry)

비즈니스 이슈 및 개선목표

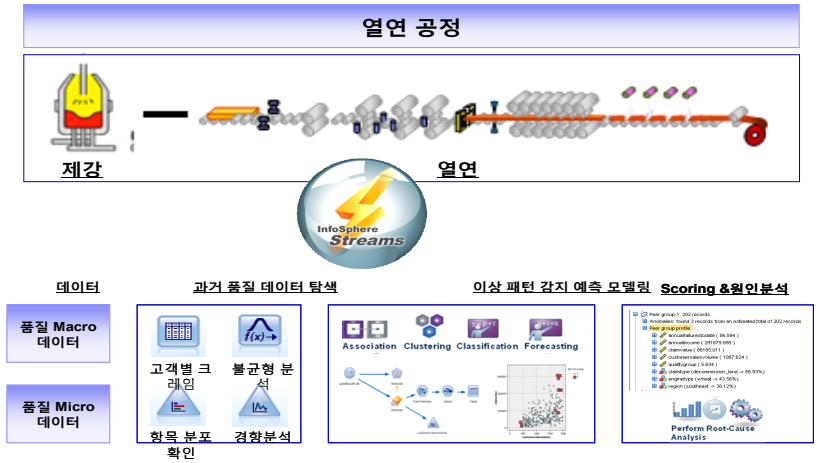
❖ 고객 Claim을 통해 결함을 인지하는 현황구조에서 xxxx의 결함을 공정 중 예측하여 xxxx 제거작업을 통해 품질 향상을 목적으로 합니다.

■ As is



- 결함 발생 여부를 고객사 전달 전에 알 수 없음
- 결함 제거를 위하여 모든 코일에 결함 제거 작업을 적용할 수 없음 (비용 낭비 발생)
- 결함 발생 여부를 고객이 확인
- 고객 불만 발생

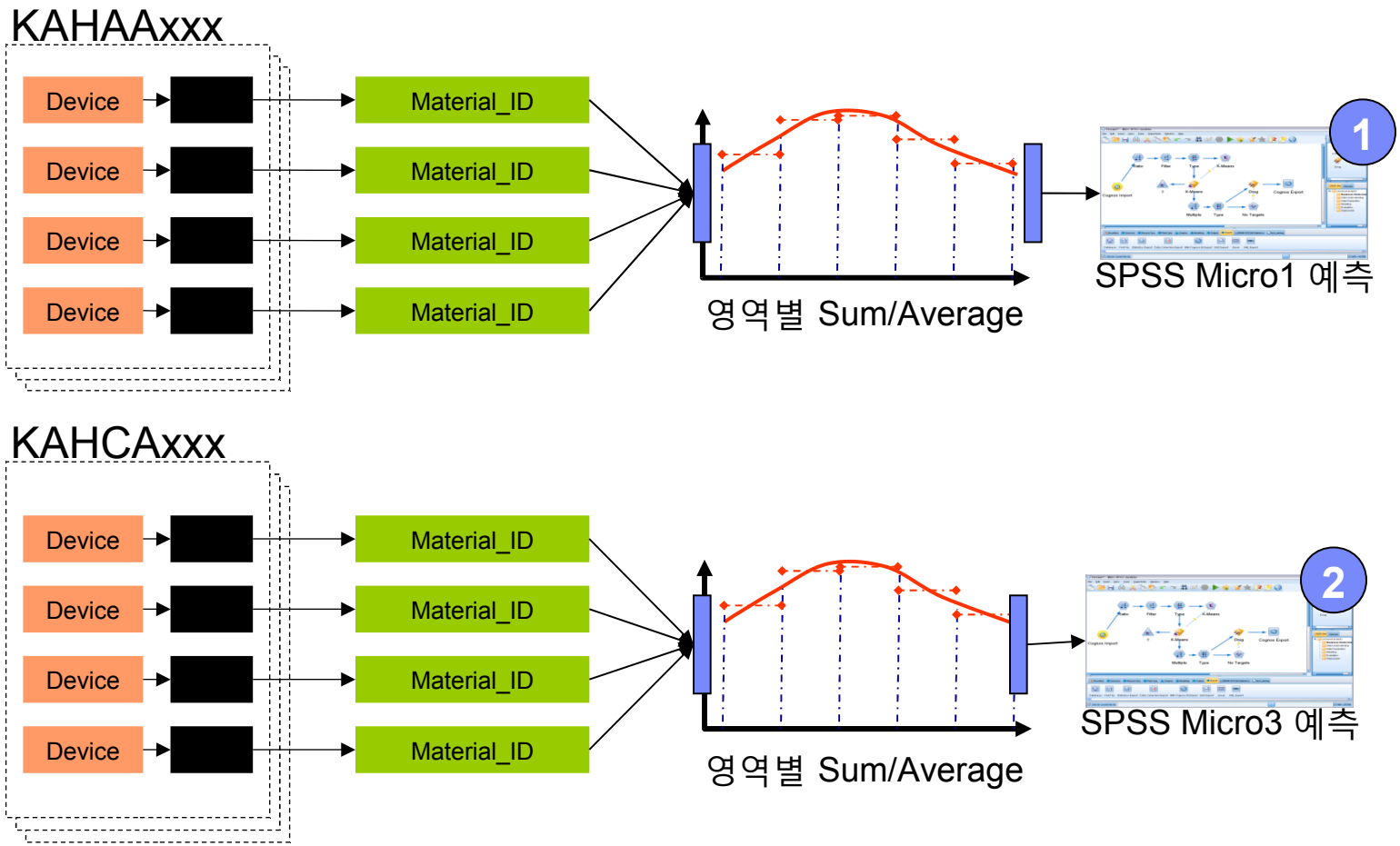
■ To be



- 실시간 결함 발생 감지 솔루션 적용
- 결함 발생 여부를 조업자가 확인
- 결함 발생 예상되는 코일만 결함 제거 작업을 적용
- 고객 만족
- 고 품질이 요구되는 열연 시장에서 경쟁력 확보

데이터 전처리 및 예측 모델링

- 품질 및 설비 데이터 파일을 대상으로,
  - Micro 품질: Material ID별로 데이터에 대한 영역별 SPSS 예측 모델 수행

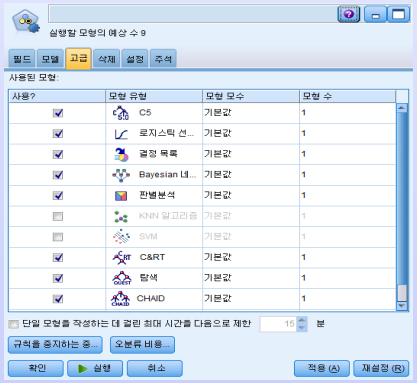


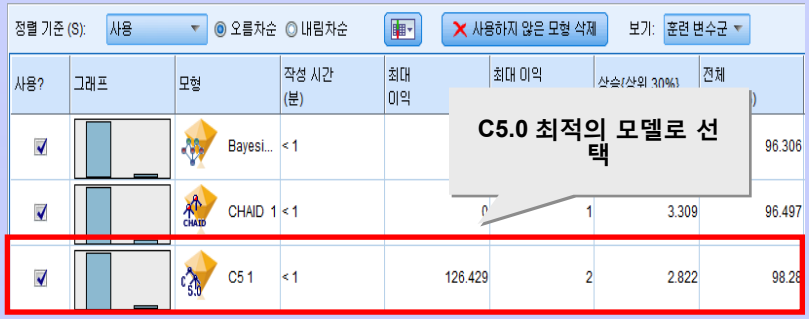
결함 탐지 모델

결함 탐지 모델

Step1. 자동화 모델링 기능을 통해 솔루션에 탑재 해 있는 모든 모델링을 돌려 최적의 모델 선택

[자동 분류자]



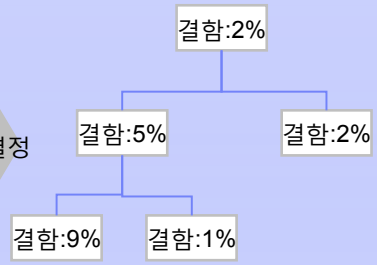


Step2. 최적의 모델(의사결정나무분석)을 사용하여 결함에 영향을 주는 인자간의 상호관련성을 분석

[C5.0]

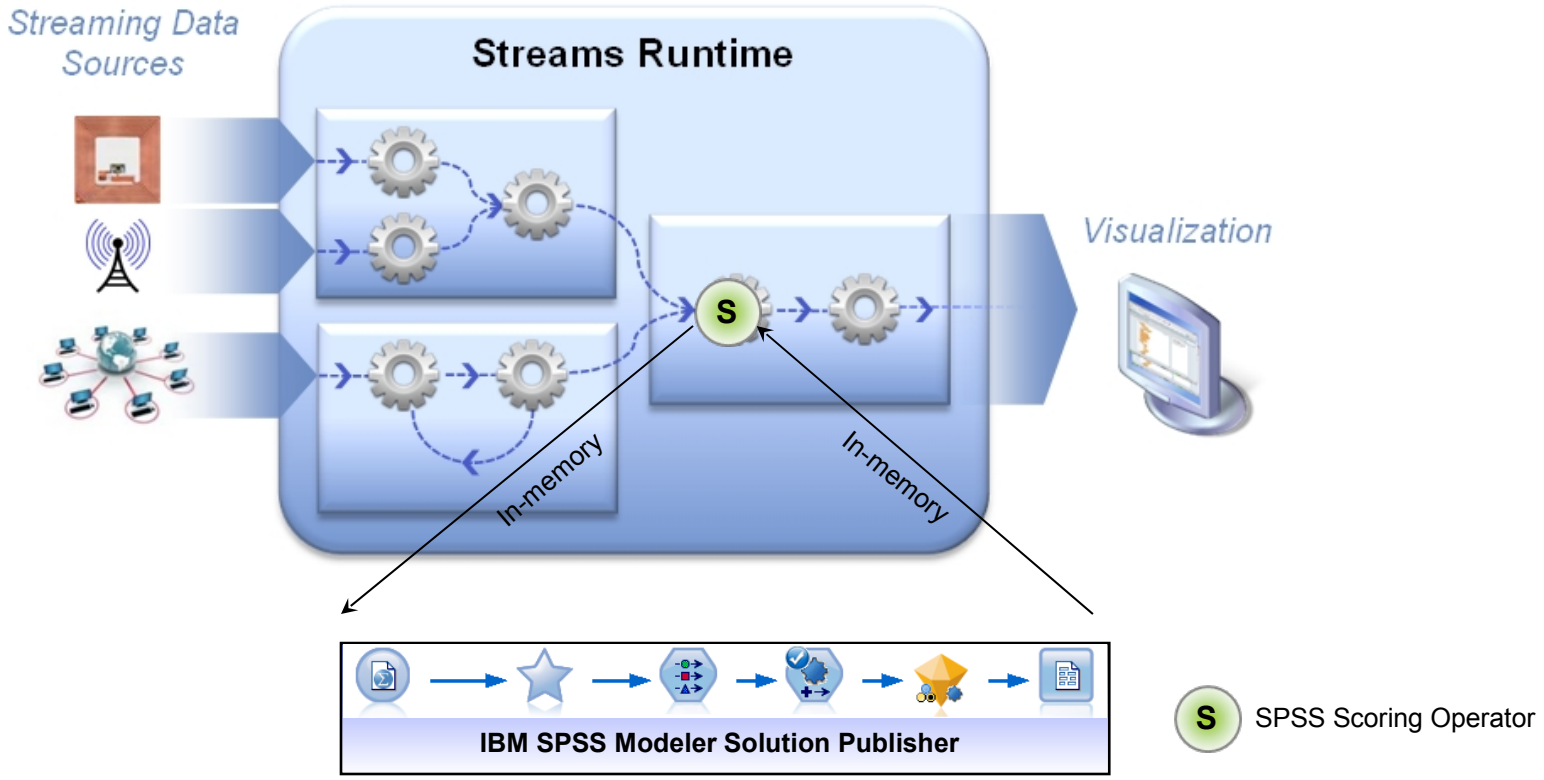
DC입측폭
EL
FM출측두께
길이방향_FM
길이방향_FW
FM폭
YP

의사결정나무

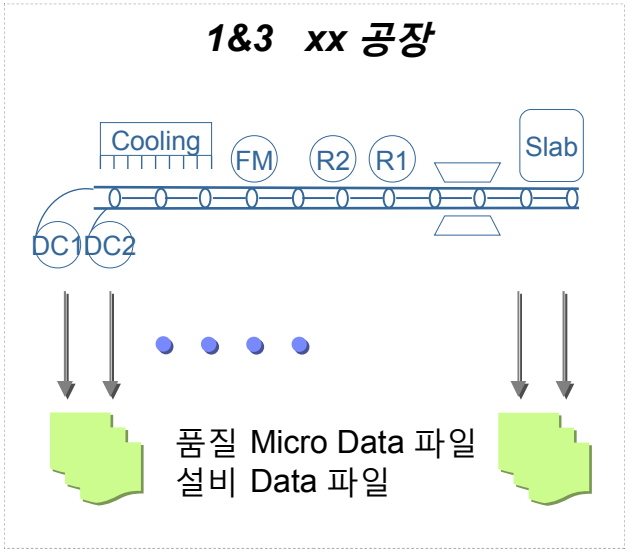


24

InfoSphere Streams & SPSS Modeler 연계



Architecture(안)

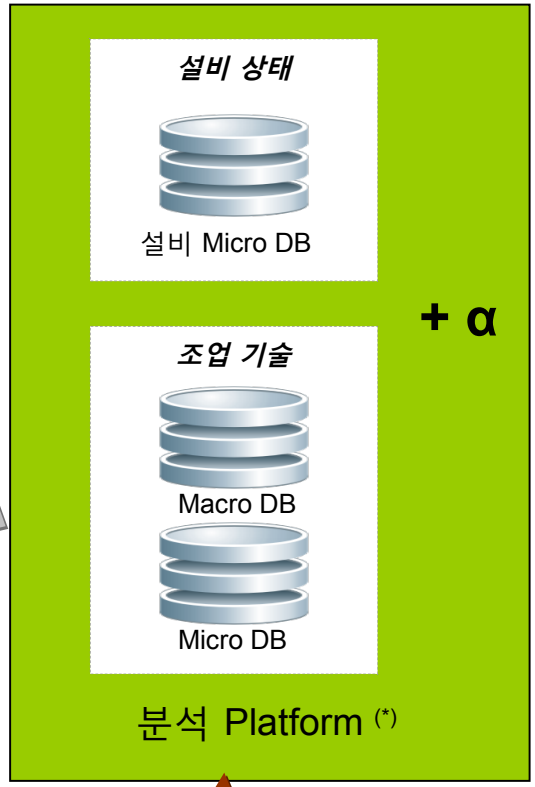


품질 Micro
품질 Macro
설비 Data

I/F

품질 Macro Data

품질 Micro Data 파일
설비 Data 파일



Cognos

실시간 이상 패턴 분석

SPSS

- **설비 데이터 이상패턴 감지 시 즉시 (10초~30초) 결과 도출**
- **품질 데이터 이상 패턴 감지 시 다음 코일이 공정 입고 전(1분~2분) 결과 도출**

(*) 과거 data을 포함하여 품질과 관련된 유의미한 분석을 가능하게 하는 Platform