

# 빅 데이터 기반 예측분석(Predictive Analytics)의 중요성과 활용 방안

김지관 과장, Business Analytics Software, IBM Korea

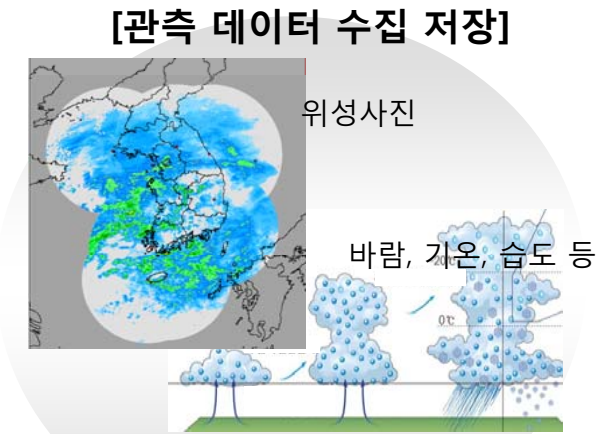
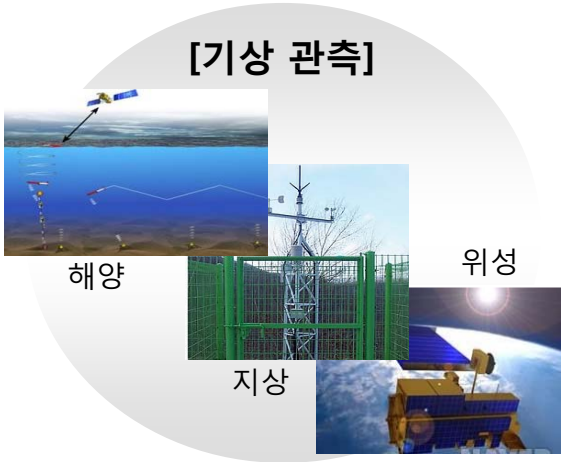


# CONTENTS

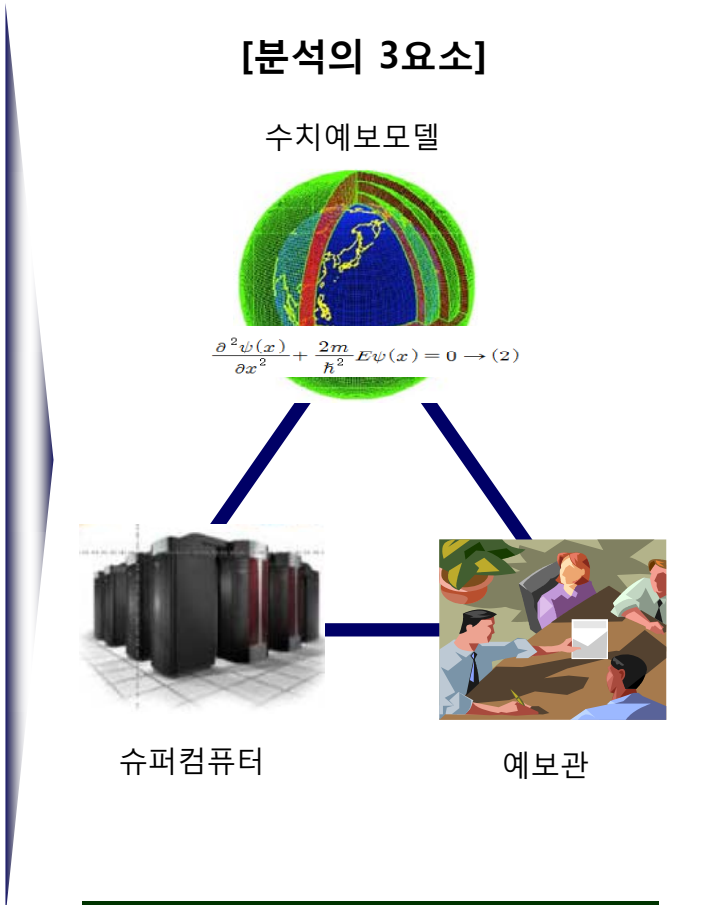
1. Big Data란?
2. Predictive Analytics with Big Data
3. Use Case



# 우리 생활 속의 빅 데이터와 예측분석 - #1



**Big Data**



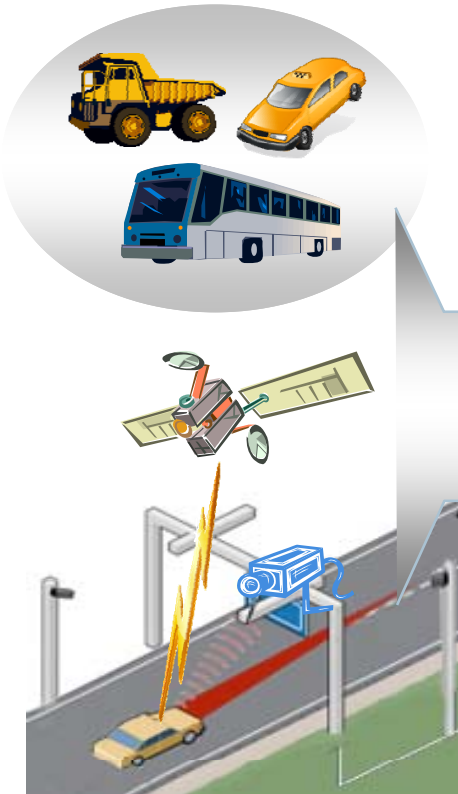
**Predictive Analytics**



**Better Life**

# 우리 생활 속의 빅 데이터와 예측분석 - #2

## Capture of Big Data



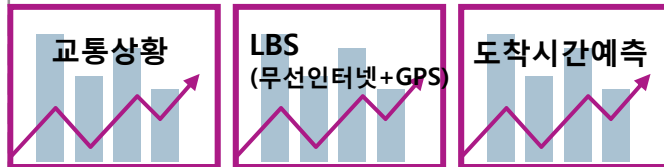
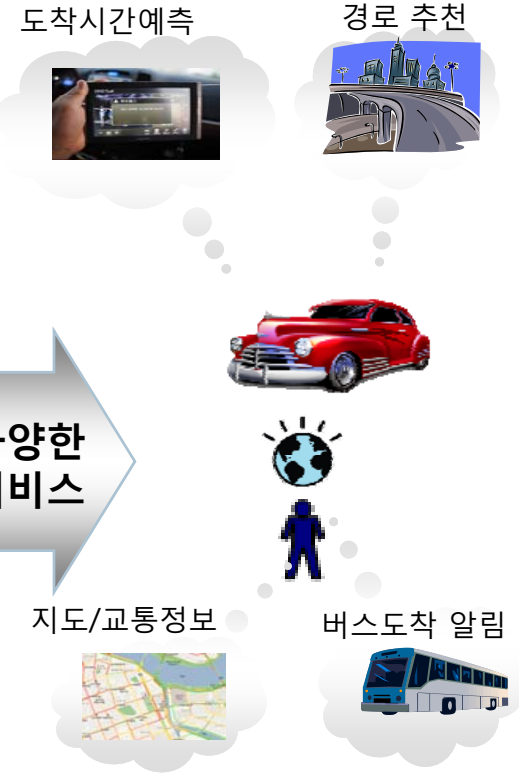
## Predict

by Predictive Analytics



## Act

for Better Life



# Big Data란?



## 빅 데이터란....

- ❖ 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 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



# 빅 데이터는 새로운 기회를 제공합니다

기존에는 불가능했던 일들이  
다양한 형태의, 대용량 데이터를 신속하게 활용한다면  
이제는 가능해집니다.



**Variety:** 다양한 관계형 및 비 정형 구조의 데이터에 대한 분석

**Velocity:** 스트리밍 데이터/대용량 데이터의 이동

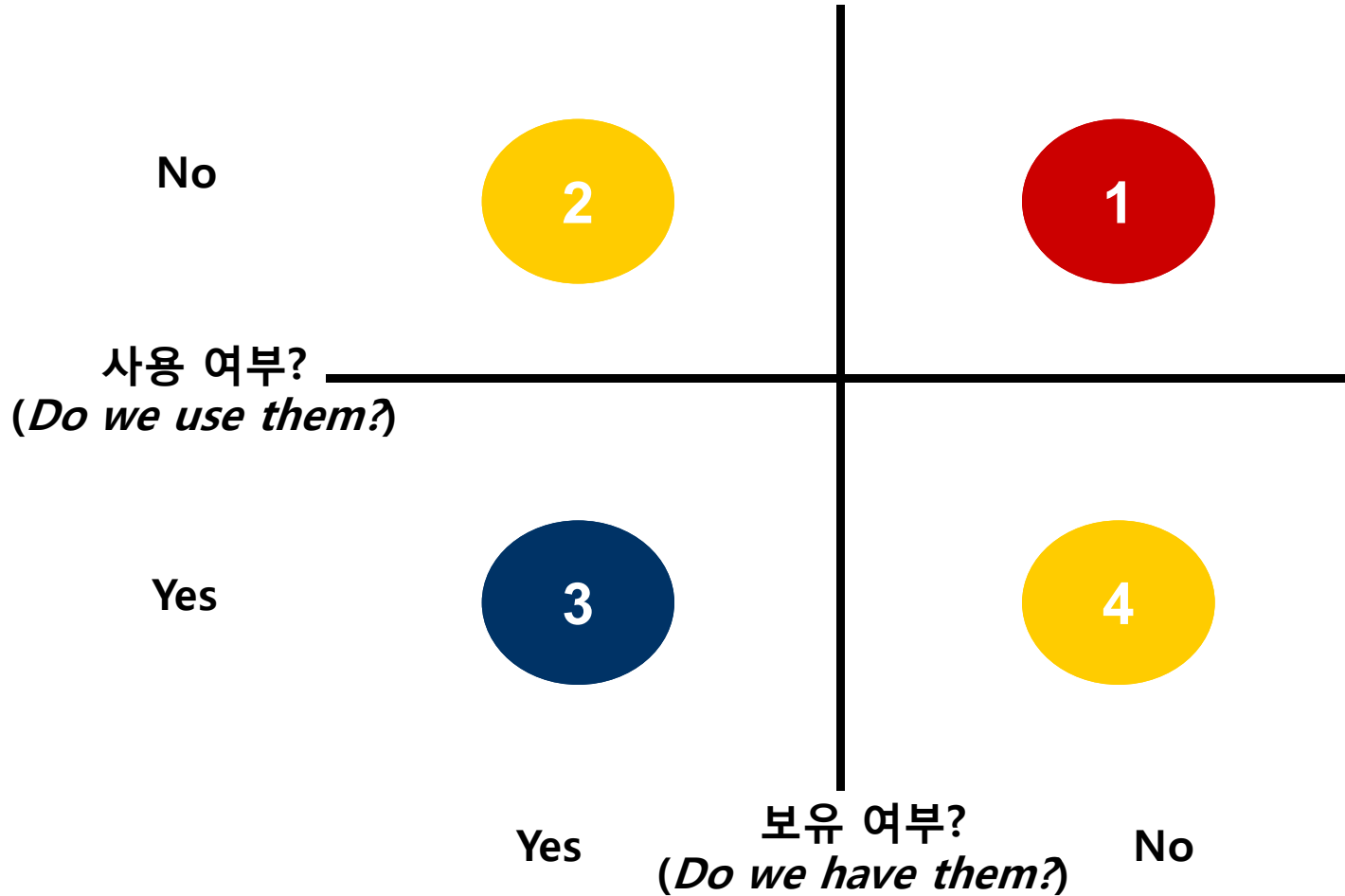
**Volume:** 테라바이트에서 제타바이트까지 확장

**Veracity(\*)** 일관성 없음, 불충분, 모호함, 대기시간, 착시, 근사치로 인한 불확실성

(\*) 진실성



# 빅 데이터, how to approach?



***Start with "questions", not with "data"***





# 빅 데이터 활용을 위한 다양한 플랫폼

**Analytic Applications**

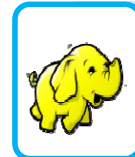
BI / Reporting	Exploration / Visualization	Functional App	Industry App	Predictive Analytics	Content Analytics
----------------	-----------------------------	----------------	--------------	----------------------	-------------------

**IBM Big Data Platform**

Visualization & Discovery	Application Development	Systems Management
<b>Accelerators</b>		
Hadoop System	Stream Computing	Data Warehouse
Information Integration & Governance		



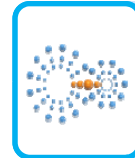
연계된(federated) 빅 데이터 소스의 이해와 탐색



엄청난 양의 데이터를 저장하고 관리



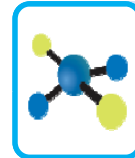
모든 데이터의 구조화와 관리



스트리밍 데이터의 관리



비 정형 데이터의 분석



모든 데이터 소스의 통합(integration)과 관리(governance)



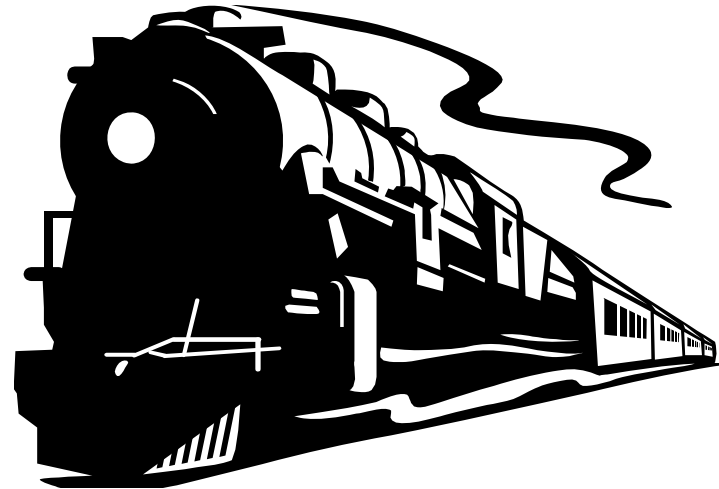
# IBM Predictive Analytics with Big Data



## 빅 데이터 VS. 예측분석

**“Big data is *the fuel* and predictive analytics is *the engine*”**

(Source : Forrester Wave, Big Data Predictive Analytics Solutions, Q1 2013 report)



## 예측분석이란....

**Predictive Analytics** helps connect **data** to **effective action**  
by drawing reliable conclusions  
about current conditions and future events.

— Gareth Herschel, Research Director, Gartner Group

“예측 분석”은  
Data를 기반으로 현재 상황과 미래에 발생할 일에 대해  
“신뢰할만한” 결론을 도출하여  
최적의 의사결정 및 실행을 하도록 “도와줍니다”.



# 예측 분석이란- How is it different from Business Intelligence?

Comparison	
<b>Business Intelligence</b>	<b>Predictive Analytics</b>
<i>What happened?</i>	<i>What's next?</i>
Insight	Foresight
Root Cause Analysis	Scenario Modeling

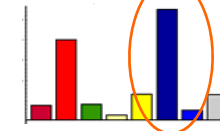
## Sheep B – Predictive Analytics

Reports showing “What will happen?”



*Conclusion: “You are about to walk off a cliff.”*

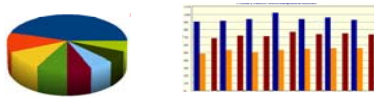
Reports showing “What should I do?”



*Conclusion: “Option F – Stop Walking is likely to produce the best result.”*

## Sheep A – BI Dashboard

Reports showing “What happened?”



*Conclusion: “You walked off a cliff.”*

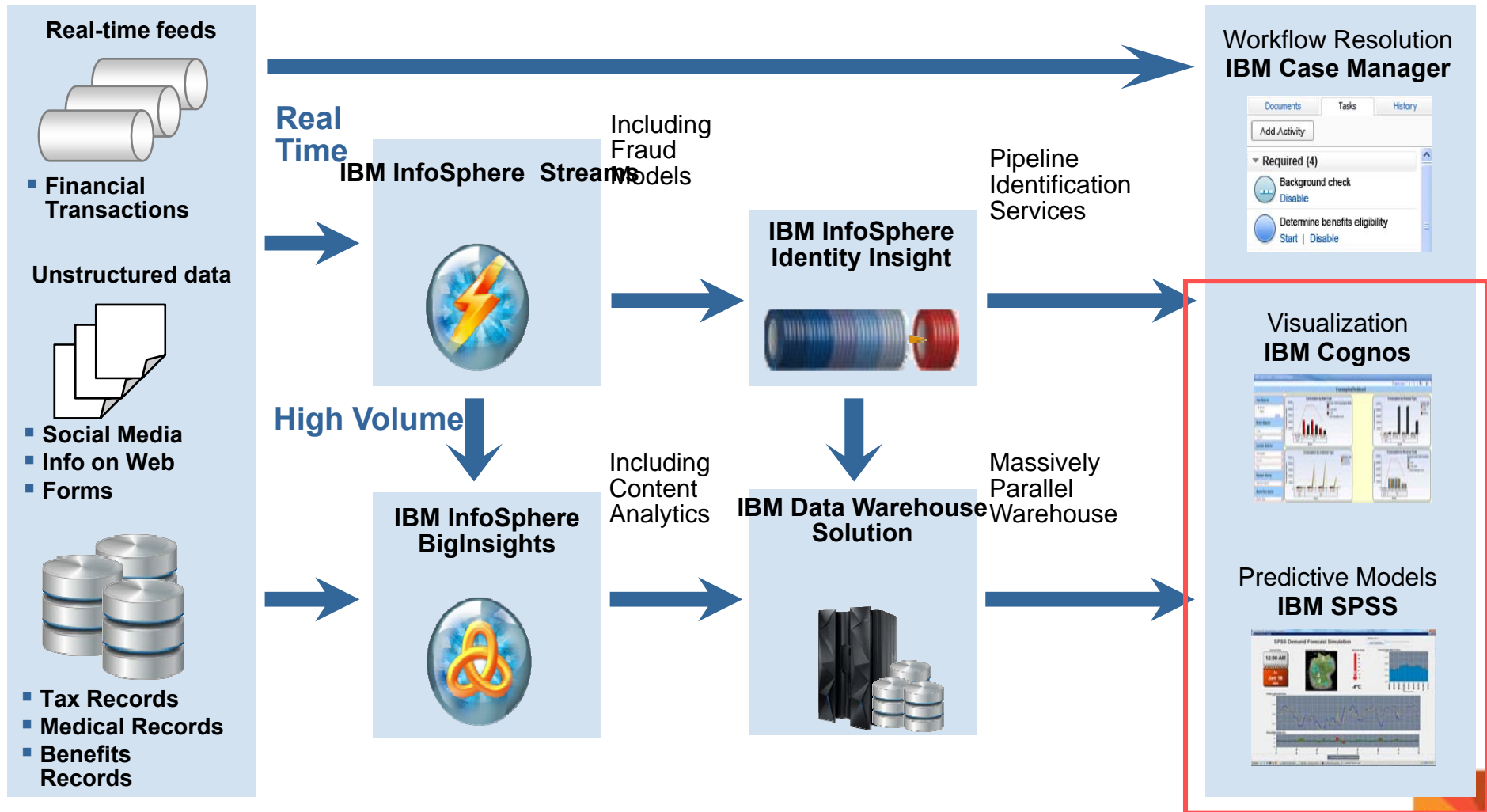
Reports showing “Why did it happen?”



*Conclusion: “You walked too far.”*



# Predictive Analytics with Big Data 개념적 구성도



# IBM SPSS Predictive Analytics



## Predictive Customer Analytics

Acquire  
Grow  
Retain



- Up-sell/cross-sell
- Market basket analysis
- Churn prevention
- Customer segmentation

## Predictive Operational Analytics

Manage  
Maintain  
Maximize



- Predictive maintenance
- Condition monitoring
- Reverse logistics
- Allocation management

## Predictive Threat & Risk Analytics

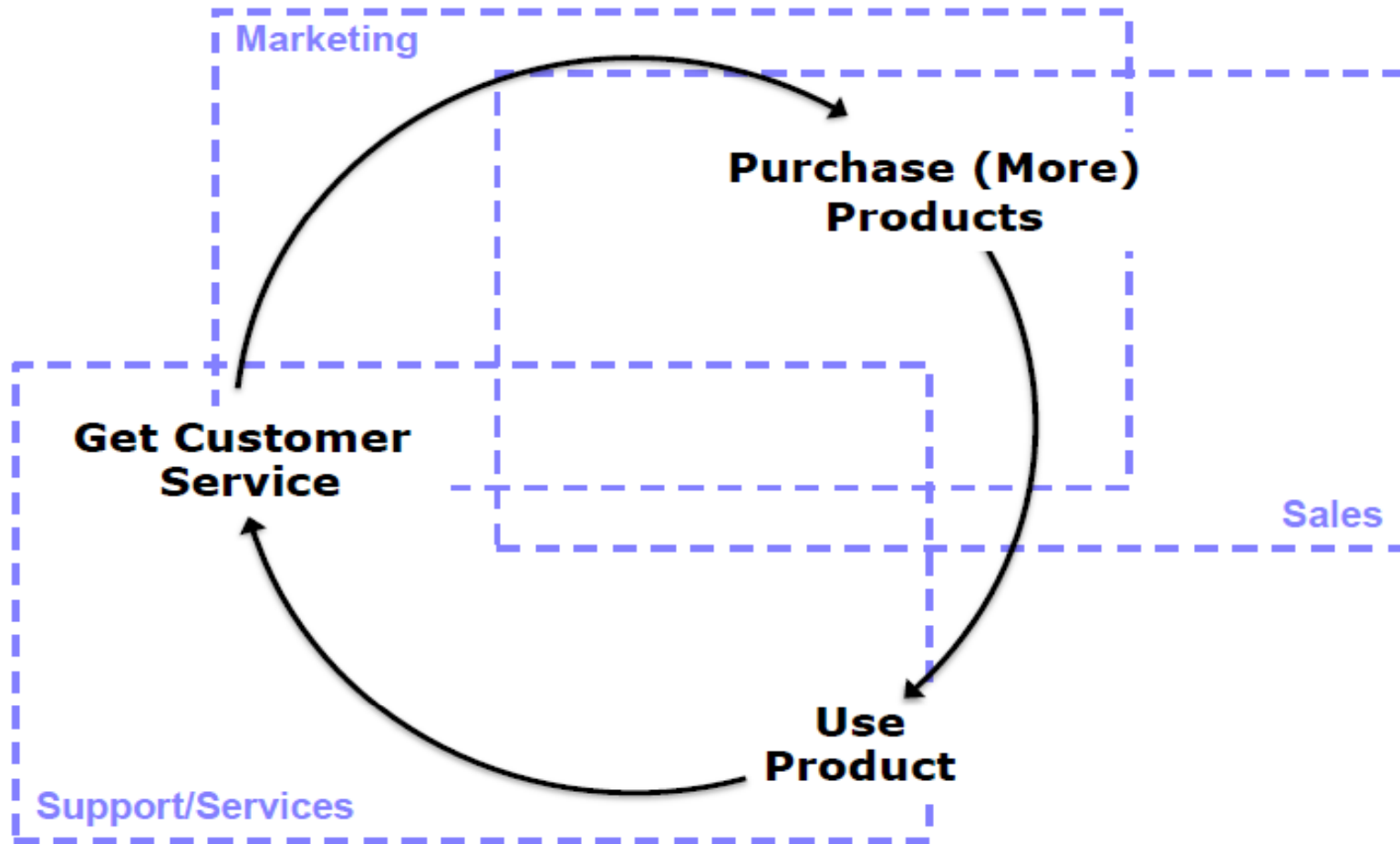
Monitor  
Detect  
Control



- Claims fraud
- Credit-card fraud
- Insider threat
- Cyber security

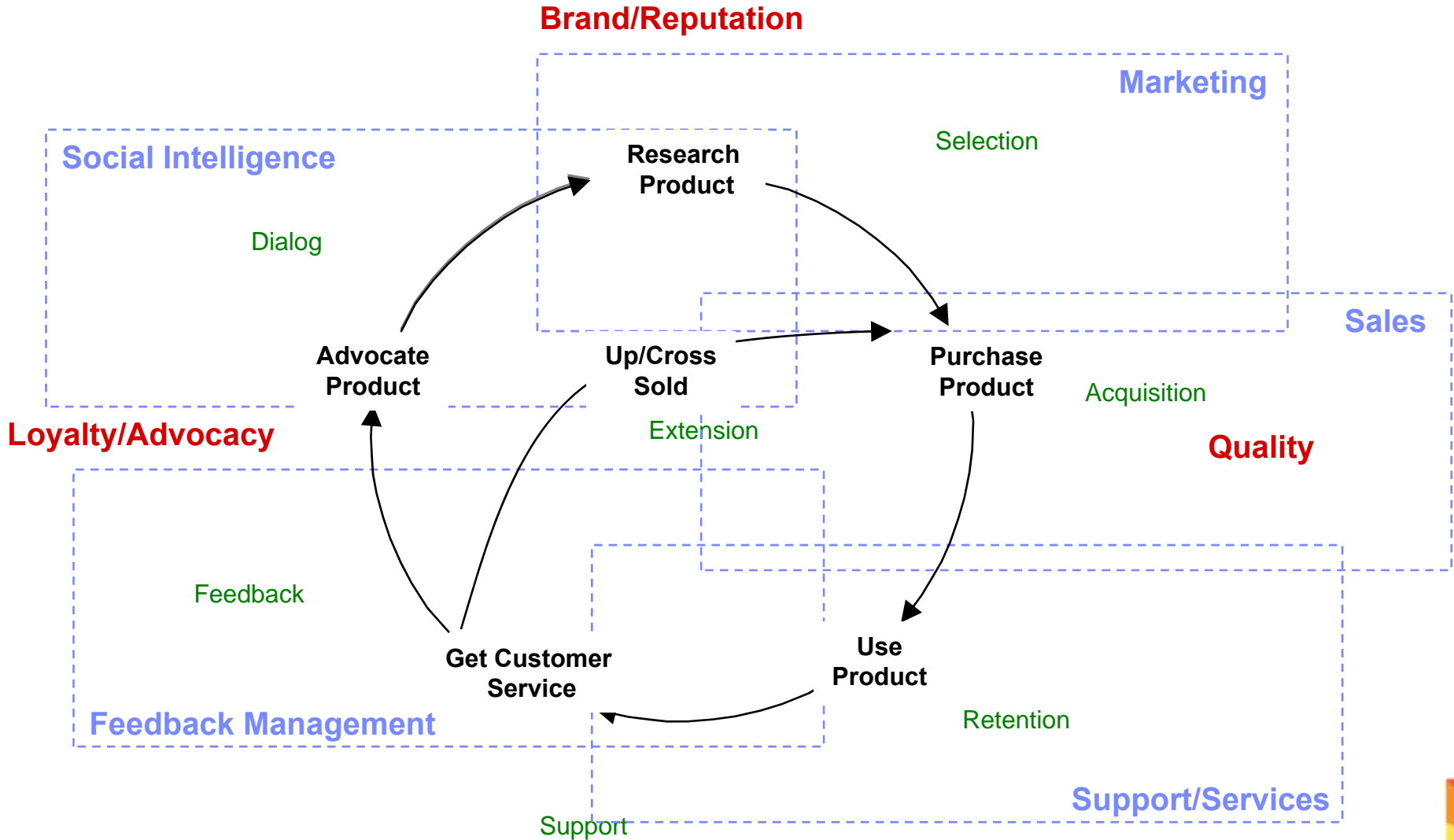


# 과거의 고객분석 접근 방향

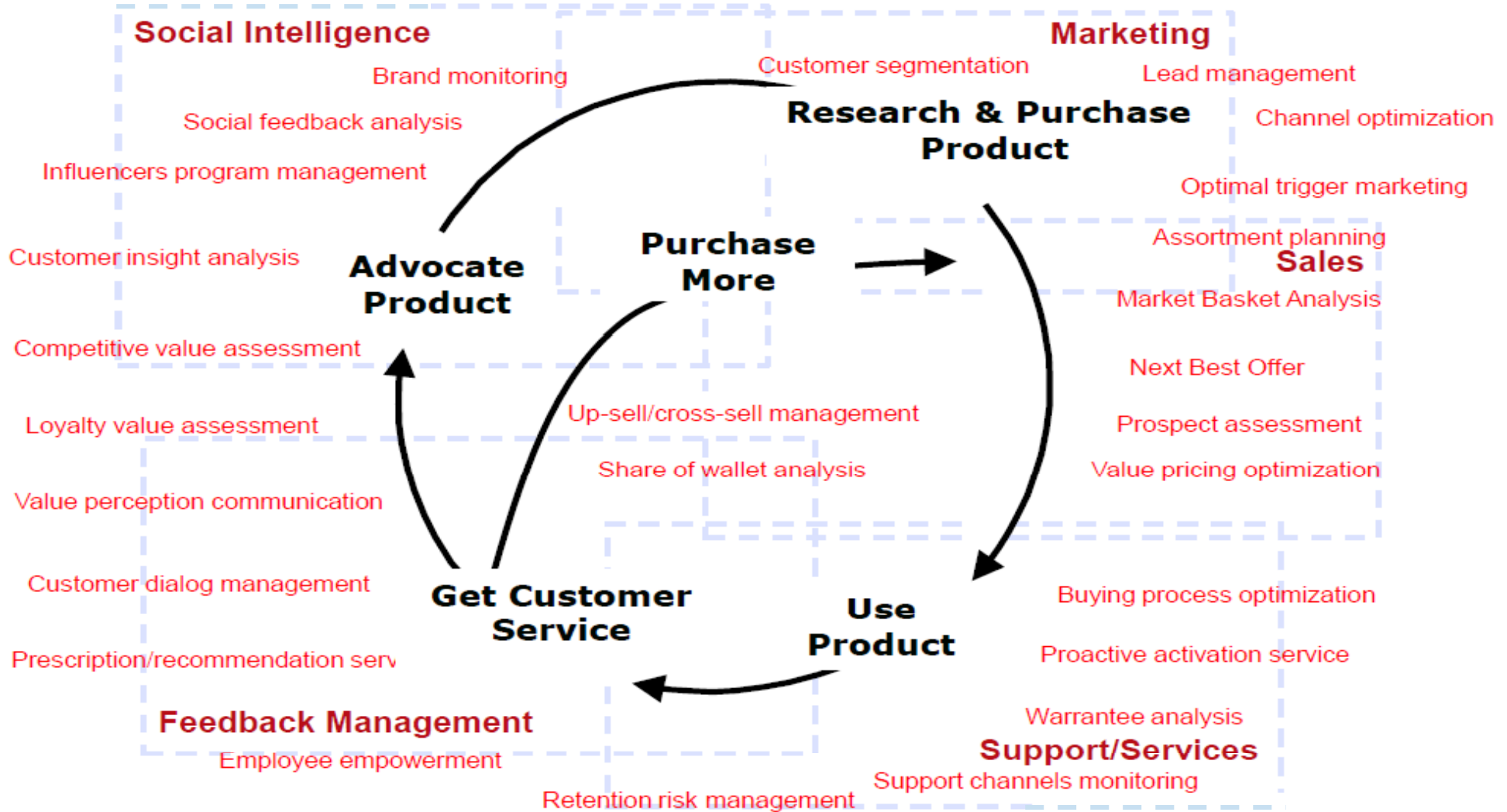




# 보다 진화된 고객 만족(Customer Intimacy) – 모든 접점에서 처리



# 빅 데이터 관점에서 접근 가능한 모든 고객 분석 영역



# Use Case



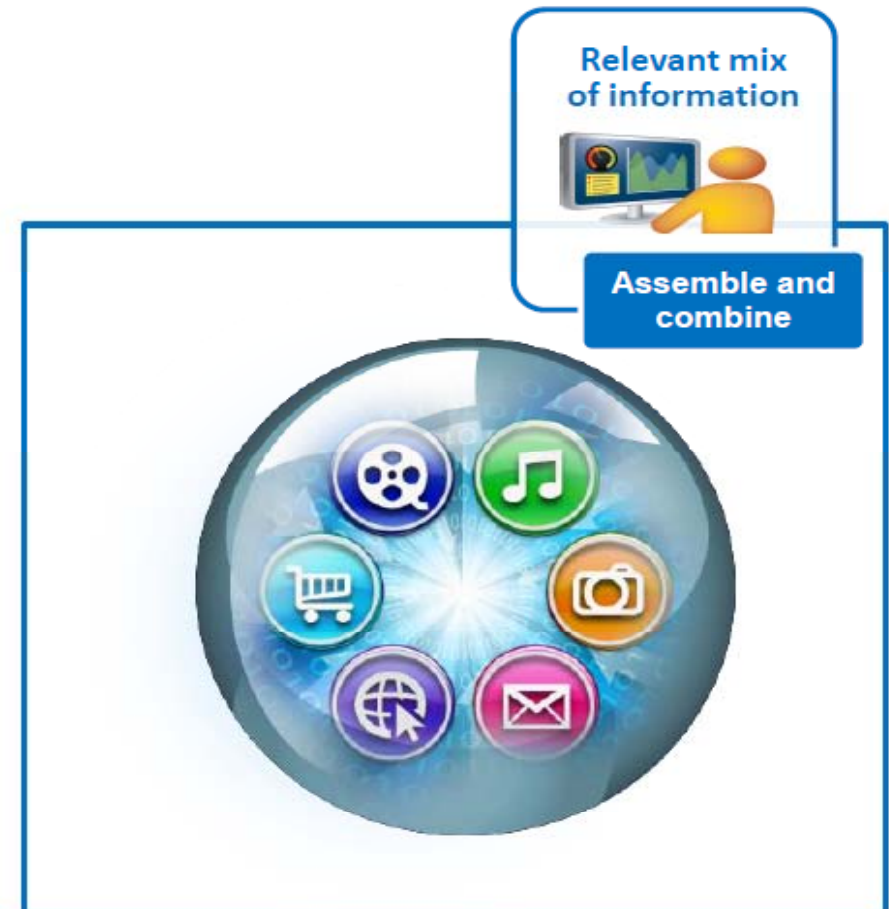
## Customer Analytics - 관련된 모든 정보를 연계 분석(통신)

### 다른 소스의 정보를 기업 내부의 정보와 함께 혼합하여 분석

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

### Benefit

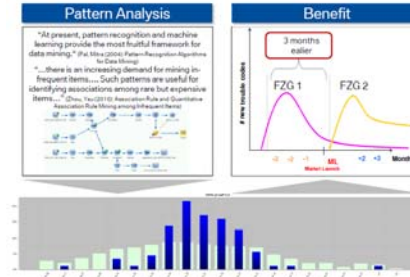
- 첫 해 이탈률 **1.9%에서 1.4%로 감소**
- 고객 유지율 **60% 개선**
- 콜센터 인력 감축에 의한 비용 감소



# Operational Analytics - 모든 프로세스에 예측 분석 적용(제조)

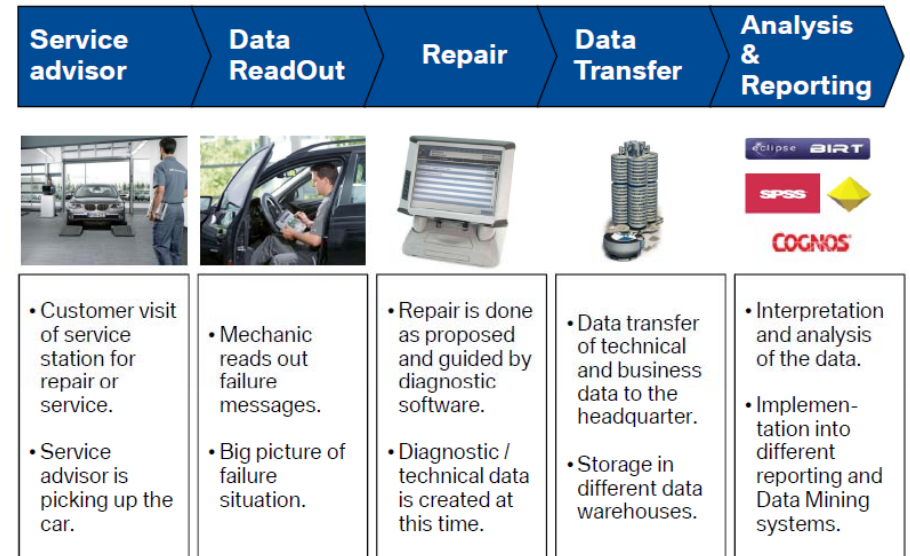
## 제품 생산에서 품질보증까지 전 영역에 예측 분석 적용

- Root-cause analysis 활용 품질 개선
- 품질 보증 청구 과다발생 차량의 옵션 조합 확인 및 개선
- 조기 부품 결함 내재 요소 선제적 대응



## Benefit

- 품질문제의 원인을 신속하게 식별
- 차량 당 **1.1% 에서 0.85%**로 품질 보증 청구율 감소
- 연간 **3천만 유로 절감**



# Threat & Risk Analytics - 자동화된 예측 분석(금융)

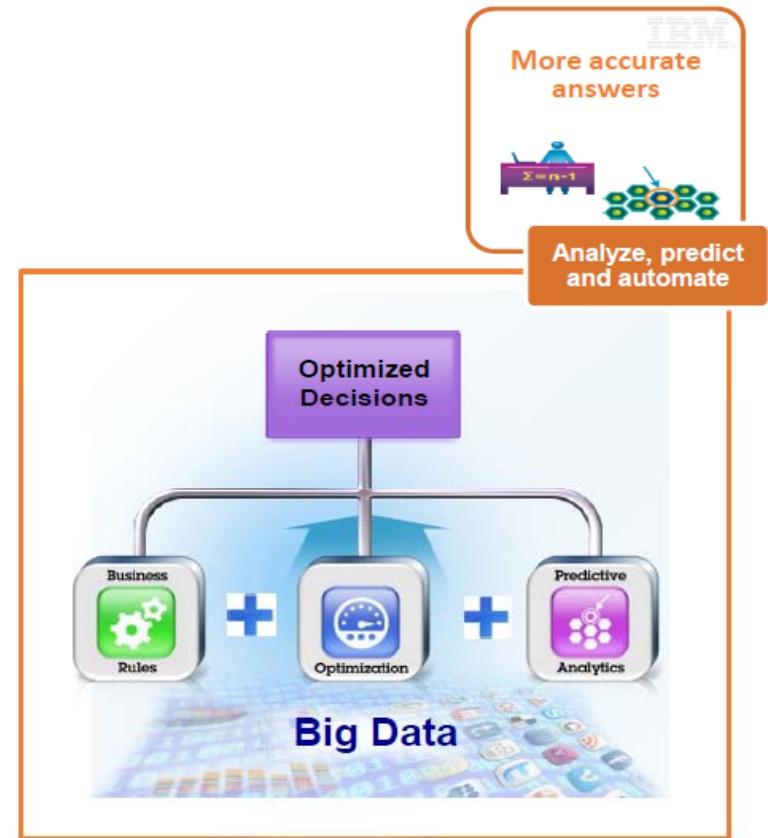
## 자동화되고 최적화된 의사 결정을 위한 분석

- 빅 데이터 활용 예측 모델 생성을 위한 알고리즘 사용
- Scoring과 Deployment를 위한 데이터 프로세싱



## Benefit

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



# Steel Industry Case - 비즈니스 이슈 및 개선목표

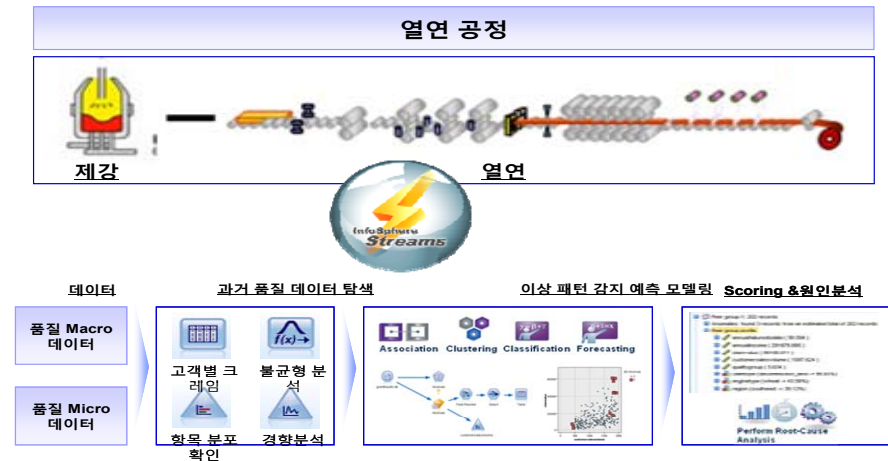
고객 Claim을 통해 결함을 인지하는 현 상황에서, 공정 중 결함을 예측하고 사전에 제거함으로써 품질 향상 및 고객 만족 제고를 도모하고자 합니다.

## As is



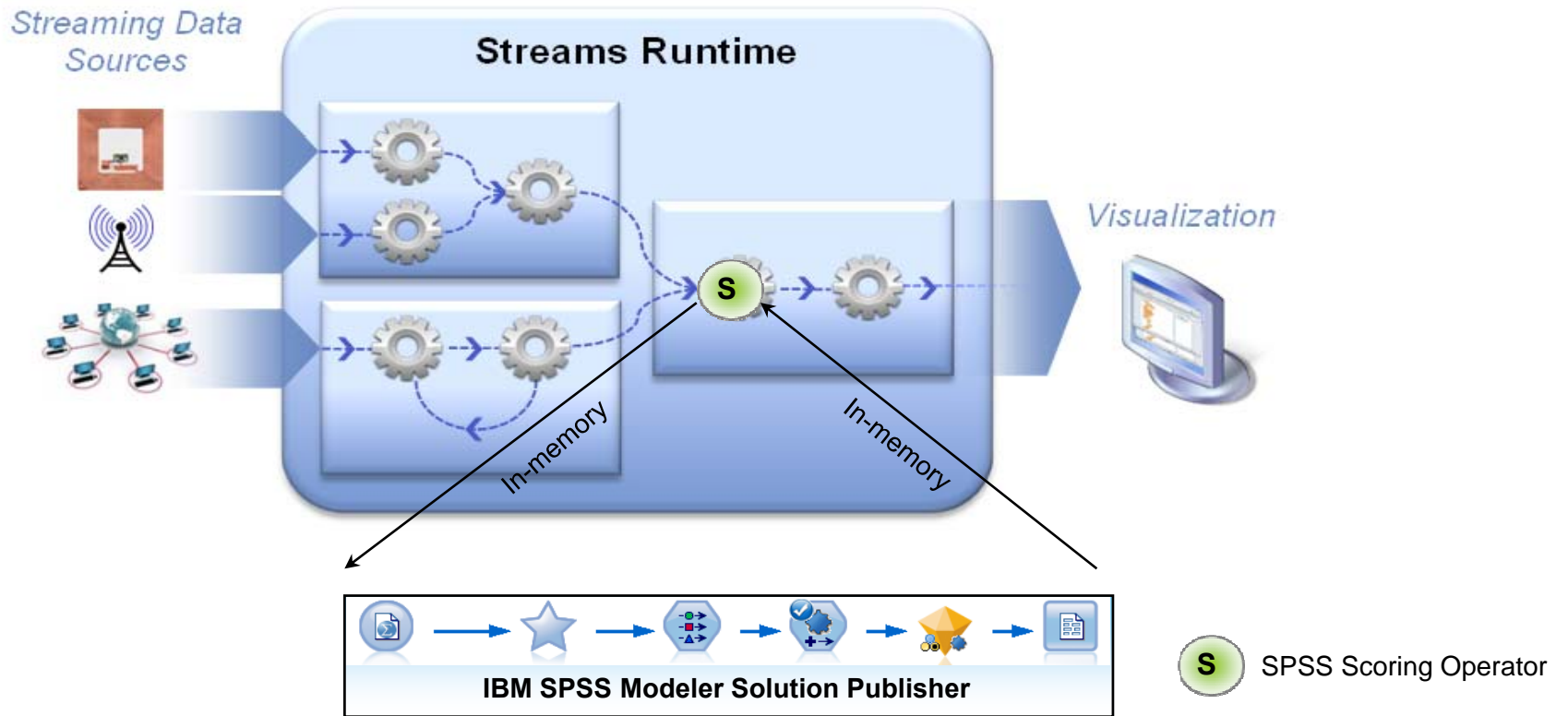
- 결함 발생 여부를 고객사 전달 전에 알 수 없음
- 결함 제거를 위하여 모든 코일에 결함 제거 작업을 적용할 수 없음 (비용 낭비 발생)
- 결함 발생 여부를 고객이 확인
- 고객 불만 발생

## To be



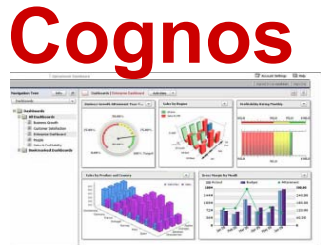
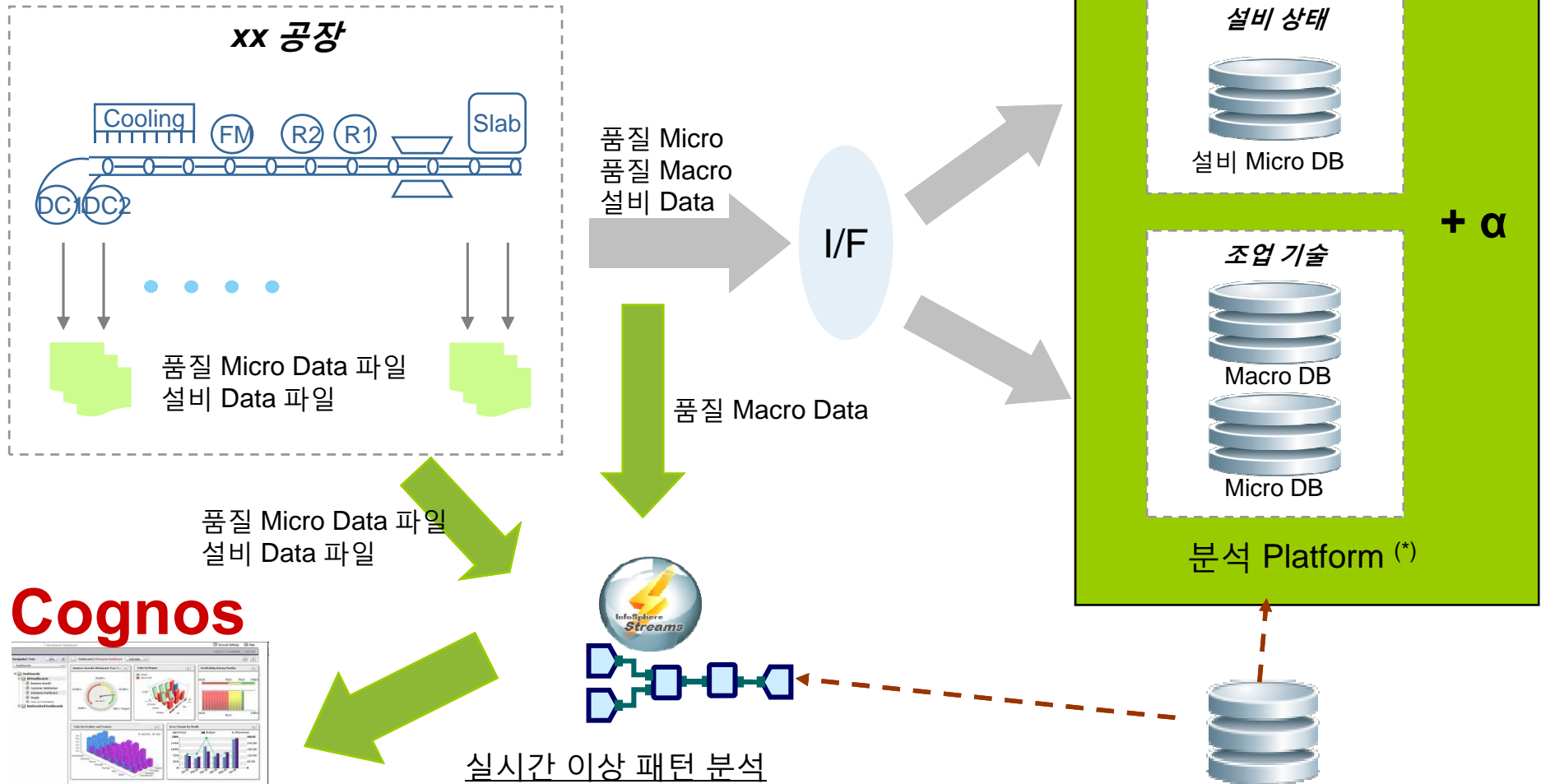
- 실시간 결함 발생 감지 솔루션 적용
- 결함 발생 여부를 조업자가 확인
- 결함 발생 예상되는 코일만 결함 제거 작업을 적용
- 고객 만족
- 고 품질이 요구되는 열연 시장에서 경쟁력 확보

# InfoSphere Streams & SPSS Modeler 연계





# Architecture(안)



- 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  
© 2013 IBM Corporation

**IBM®**