

KRAFTON  
GAME UNION

PLAYERUNKNOWN'S  
BATTLEGROUNDS  
MOBILE

이제, 모든 곳이 배틀그라운드

# 물리학 전공자가 바라 본 게임 산업에서의 빅데이터 분석

모바일분석실  
실장 최영규



## + 개요

- 게임 산업에서의 Big Data 분석과 현황 그리고 미래 소개
- 물리학 전공자로서 Big Data 분석가/Data Scientist가 되기 위한 Skillset 소개
- 해당 분야에 대한 로드맵 제시

## + 약력

- 성균관대학교 물리학 박사
  - 입자 및 핵실험 전공
- CJ E&M
  - 게임 사업부문 인사이드팀 과장 입사
- PUBG Corporation
  - 전) 데이터 모델링팀 팀장
  - 전) PUBG 모바일인사이드팀 팀장
- KRAFTON
  - 현) 모바일분석실 실장

# 입자 실험 그리고 현업 분석

*Similarities / Differences*



# High Energy Physics & Big Data

## What is a LHC ?

Large Hadron Collider : the world's largest and highest-energy particle accelerator !

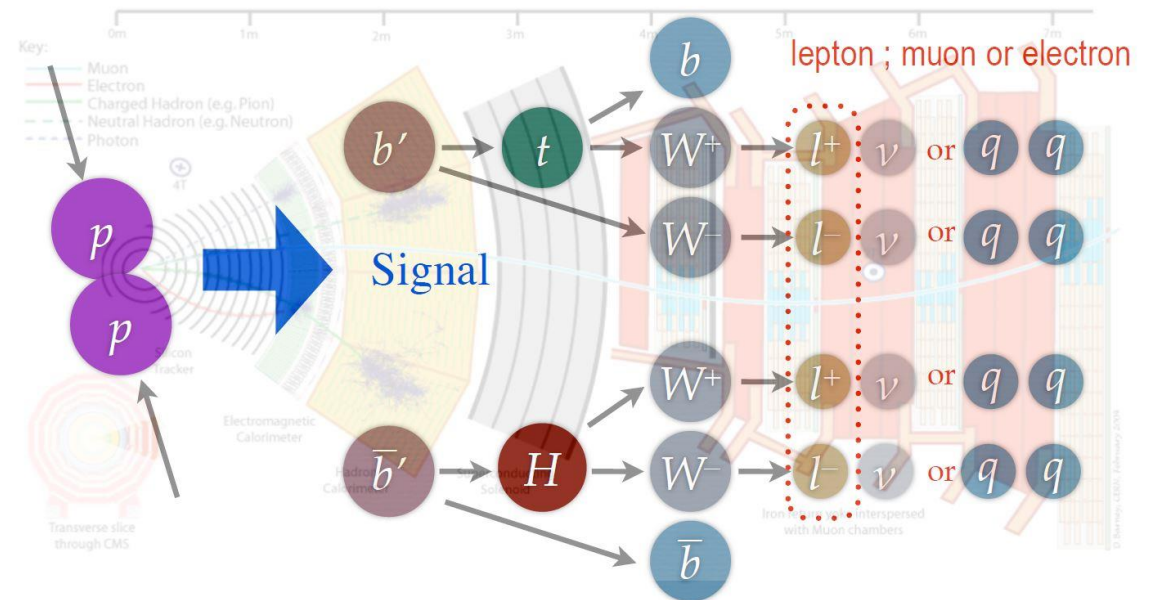


- ☞ Location in Swiss and France.
- ☞ 27km circumference with 50~175 meters depth.
- ☞ 40 MHz bunches of proton proton beam.
- ☞ 4TeV(Tera electron Volt) per beam(current)
- ☞ Hunting for Higgs, SUSY, extra dimension, dark matter, vector-like quark and so on.

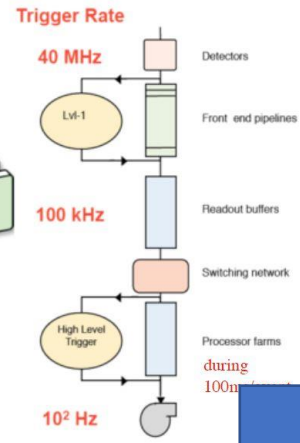
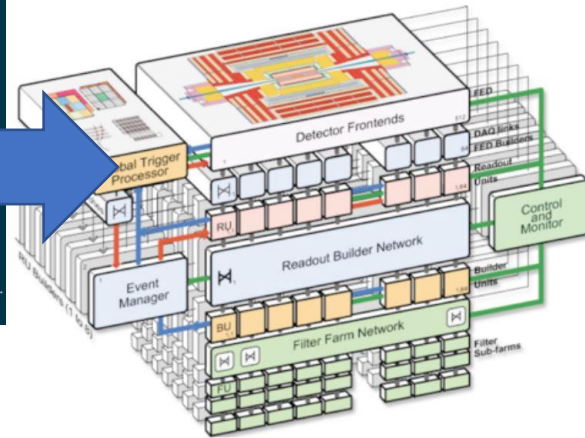
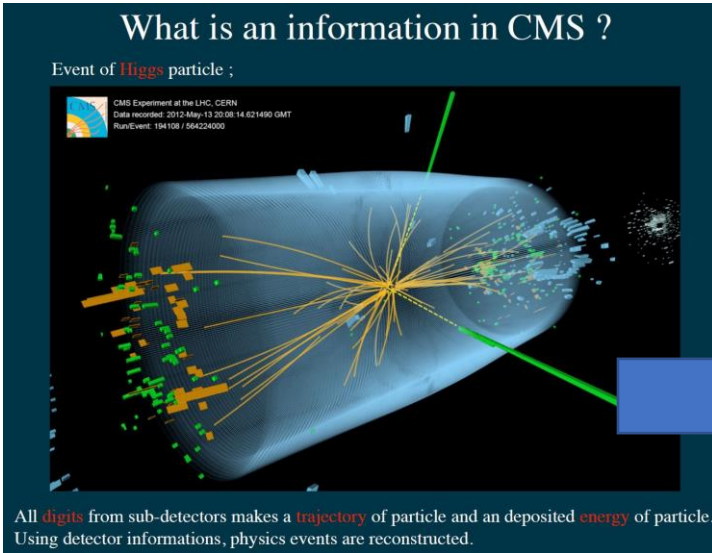
## What do we want to search for ?

Searching same sign di-lepton channels.

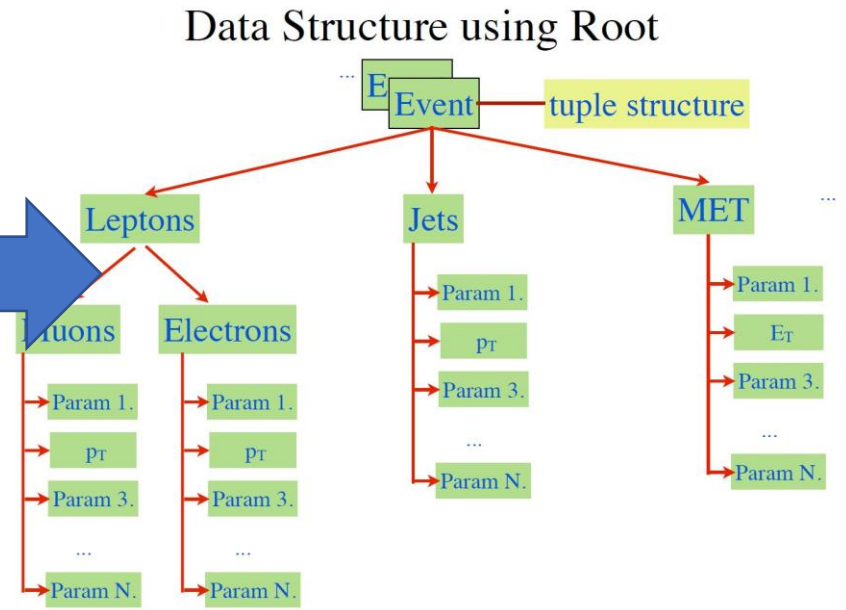
☞ Need two leptons and six jets(from q and b)



# 입자 실험에서의 데이터 적재 과정



Saving event data ;  
1MB/event



$$S_T = \text{lepton } p_T + \text{jet } p_T + \text{MET}$$

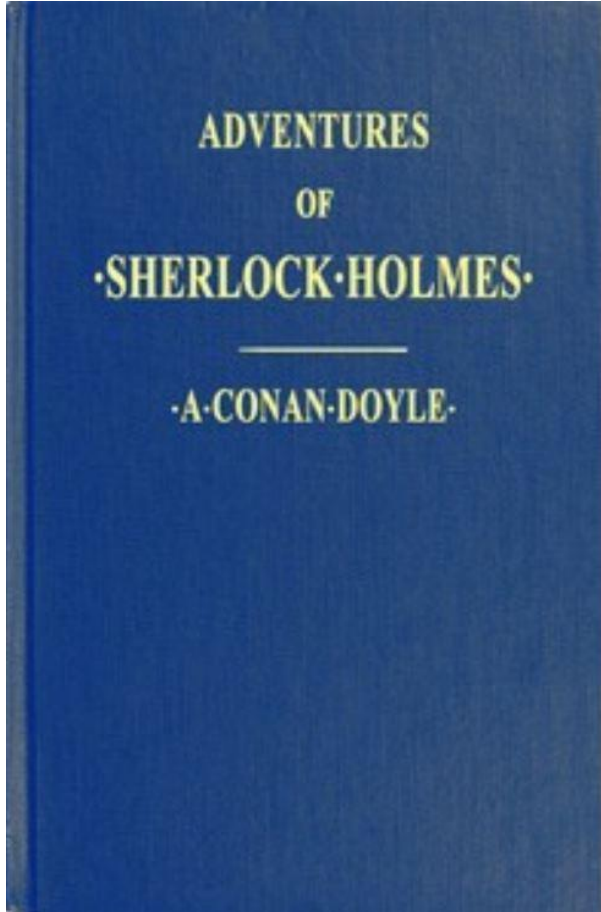
# BATTLEGROUNDS & Big Data

## IN MEDIA ↘

- 2022년까지 보도 자료 기준

- '배틀그라운드 모바일' 2018년에는 글로벌 일평균 이용자 수 1천만 명을 달성했고, 2020년 4월에는 국내 누적 가입자 수 2천만 명을 돌파
- '배틀그라운드 모바일 인도' 출시 1주 만에 3,400만 이용자 돌파
  - 일일 최대 이용자 1,600만, 최대 동시 접속자는 240만 기록
- 펍지유니버스, 단편영화 '방관자들' 100만뷰 돌파

# PUBG 모바일 유저의 하루 Log 적재 용량?



## CONTENTS

	PAGE
I.-A SCANDAL IN BOHEMIA	3
II.-THE RED-HEADED LEAGUE	29
III.-A CASE OF IDENTITY	56
IV.-THE BOSCOMBE VALLEY MYSTERY	76
V.-THE FIVE ORANGE PIPS	104
VI.-THE MAN WITH THE TWISTED LIP	126
VII.-THE ADVENTURE OF THE BLUE CARBUNCLE	153
VIII.-THE ADVENTURE OF THE SPECKLED BAND	176
IX.-THE ADVENTURE OF THE ENGINEER' S THUMB	205
X.-THE ADVENTURE OF THE NOBLE BACHELOR	229
XI.-THE ADVENTURE OF THE BERYL CORONET	253
XII.-THE ADVENTURE OF THE COPPER BEECHES	280

~



- 유저 한명이 하루에 발생시키는 log의 양은 단편 소설 한권 분량 정도
- 유저 클라이언트로부터 전송된 빅데이터를 처리 과정에 분석가의 참여가 늘고 있는 추세
  - 데이터 수집(Kafka, Cassandra)
  - 빅데이터 소프트웨어(Hadoop), 클라우드(Google BigQuery, Amazon AWS, MS Azure)





나누지 않으면 보이지 않는 것들

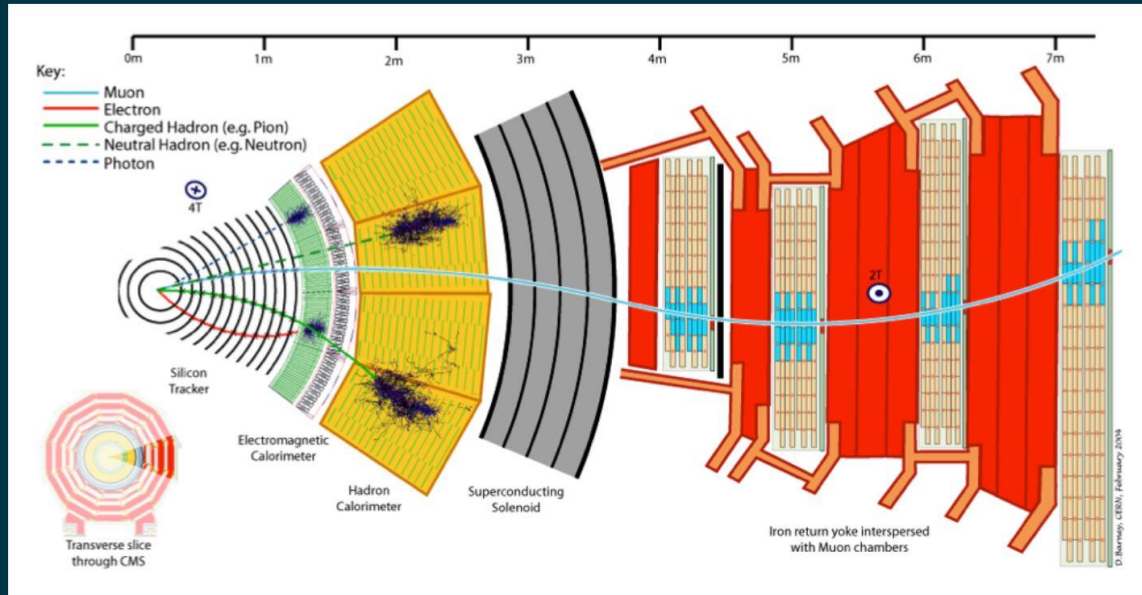
# 分析

나눌 분

쫓갠 석

# 나누지 않으면 보이지 않는 것들

## Particle reconstruction in offline process



Muons  
Electrons  
Hadrons  
Photons  
Jets

Reconstructed by **full detector informations** in offline process.

Digits → Segments → Tracks → Global Tracks

## In Physics

입자는 물질과 반응을 한다

- Muon, Electron, Hadron, Photon, Jet은 각기 다른 물질에 다르게 반응한다

## In Game Industry

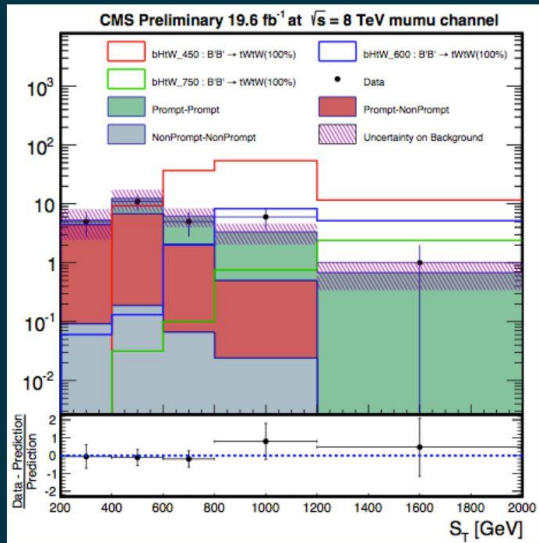
유저는 게임과 반응을 한다

- 유저는 성향에 따라 게임에 다른 반응을 한다
  - 가입하자마자 이탈한 유저
  - 특정 스테이지까지 플레이하고 이탈한 유저
  - 가입하자마자 100만원을 과금한 유저
  - 3년 동안 게임을 이용하면서 과금을 하지 않은 유저

# Analytics: Physics vs Business

## Systematic Uncertainties

Systematic uncertainties are smeared in estimated background.  
 Total 9 nuisance parameters affect  $S_T$ .



Systematics uncertainties (%)	$\mu\mu$
Integrated luminosity	1.3
Trigger efficiency	1.7
Lepton selection	1.0
Pileup events	1.1
Jet energy scale	2.5
Jet energy resolution	3.3
Background Normalization	17.8
Accuracy of control-sample method Error	25.7
Parton distribution function reweighting	1.9
Total systematic uncertainty (%)	31.7
Statistical uncertainties (%)	5.9

예. 입자 실험에서 졸업하기 위해서 필요한 분석

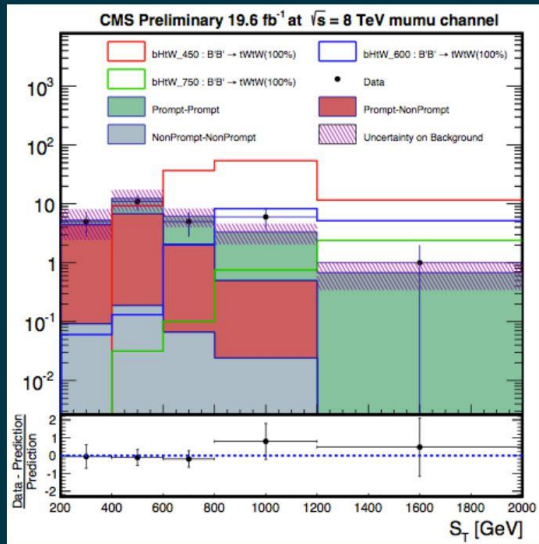
- Systematic Uncertainty를 분석하기 위해 1년 소요
- Data-Driven Method를 이용하기 위해 Fake rate 변수 분석에 6개월 소요



# Analytics: Physics vs Business

## Systematic Uncertainties

Systematic uncertainties are smeared in estimated background.  
 Total 9 nuisance parameters affect  $S_T$ .



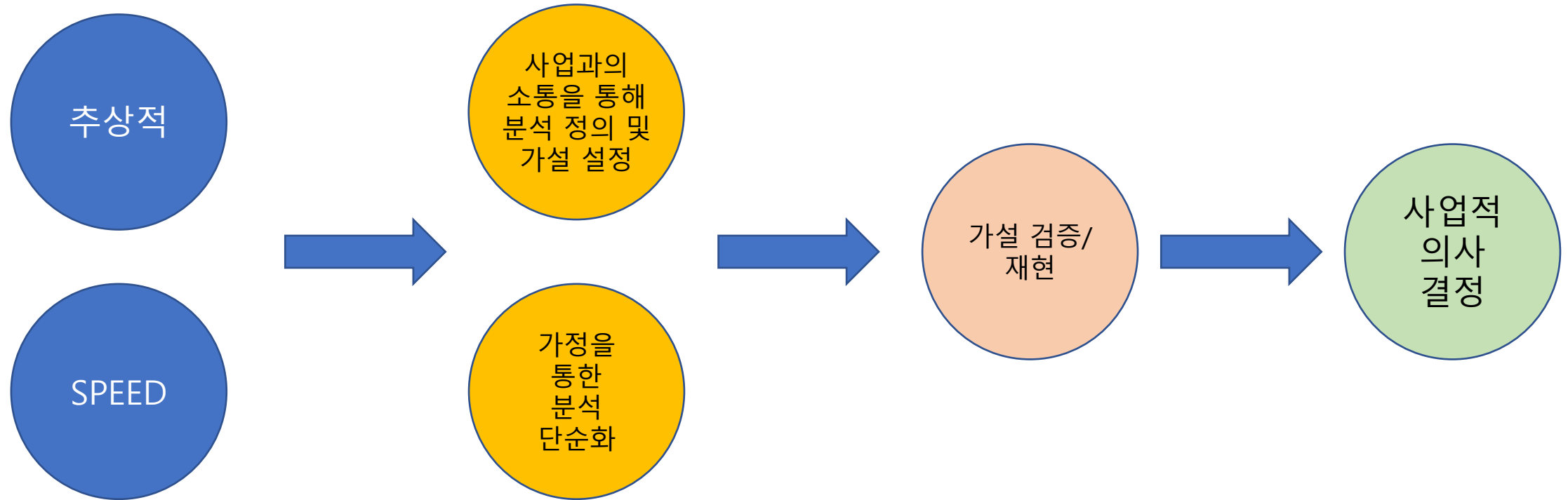
Systematics uncertainties (%)	$\mu\mu$
Integrated luminosity	1.3
Trigger efficiency	1.7
Lepton selection	1.0
Pileup events	1.1
Jet energy scale	2.5
Jet energy resolution	3.3
Background Normalization	17.8
Accuracy of control-sample method Error	25.7
Parton distribution function reweighting	1.9
Total systematic uncertainty (%)	31.7
Statistical uncertainties (%)	5.9

## 회사에서 요구하는 분석

- A: 유저들이 게임을 재미있게 하는 지 분석해주세요
- B: 유저들이 게임에서 왜 이탈을 하는 지 분석해주세요
- C: 이번 게임 상품은 왜 많이 안 팔렸나요?
- D: 어떻게 하면 매출을 올릴 수 있을까요?
- .
- .
- .
- E: **3일**안에 분석 결과를 공유해주세요.

# Analytics: Physics vs Business

회사에서 요구하는 분석?





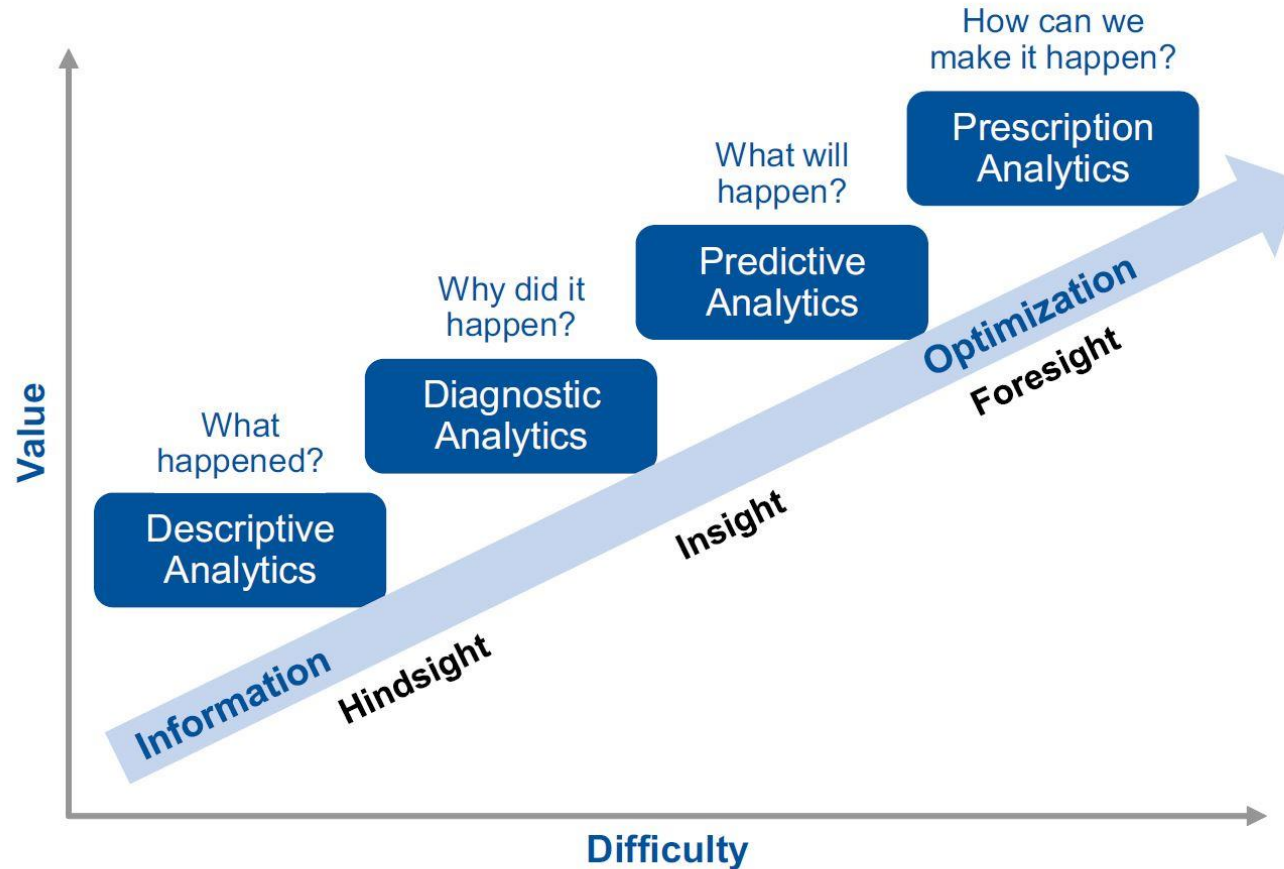
# 현업 분석의 현재와 미래

---



# 가트너의 데이터 분석 발전 모델

Figure 2. Gartner Analytic Ascendancy Model



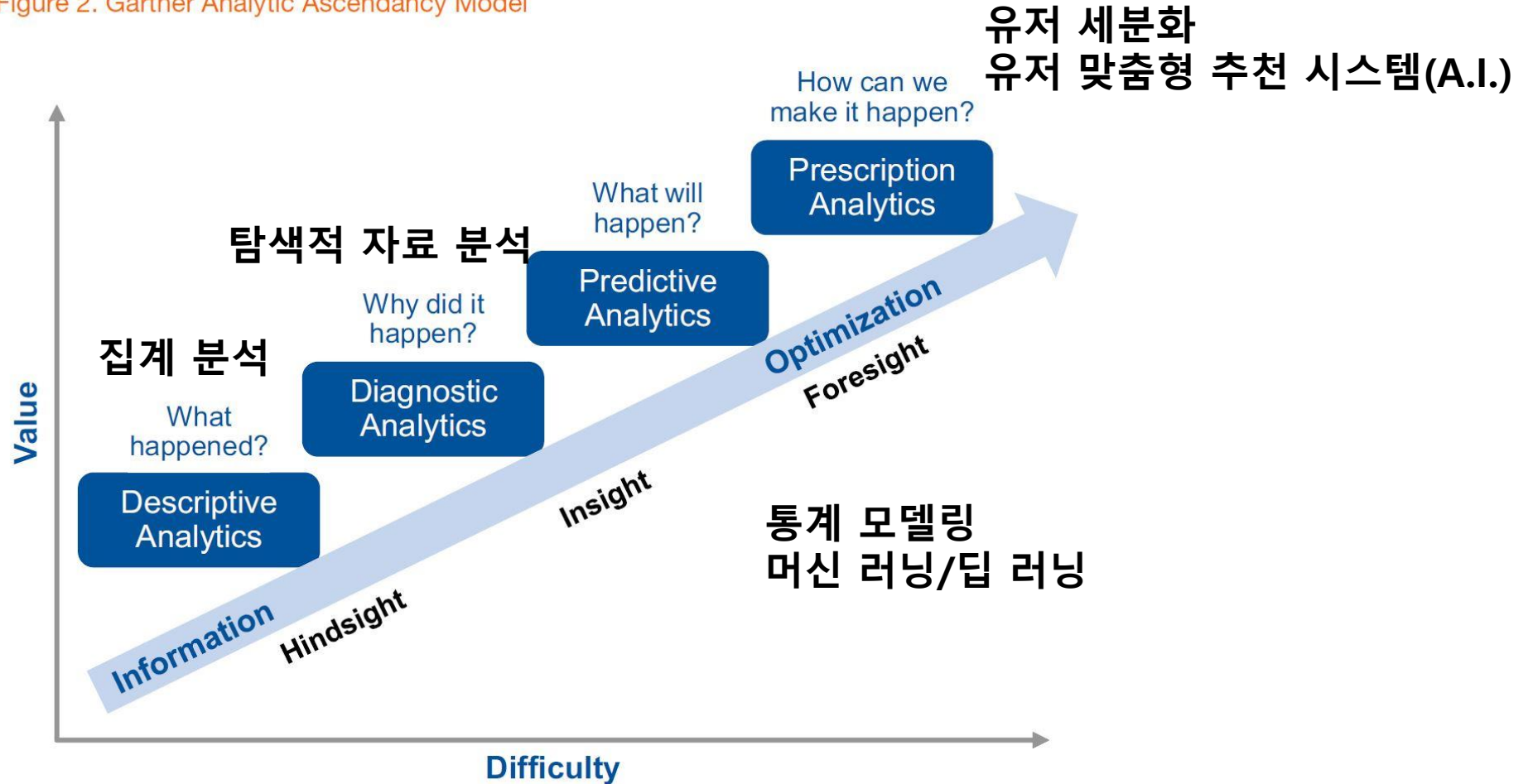
Source: Gartner (December 2012)



# 분석하는 단계

단순한 집계 분석에서 유저 맞춤형 추천을 하는 고도화 분석으로 진화 중

Figure 2. Gartner Analytic Ascendancy Model



Source: Gartner (December 2012)

# 분석 내용

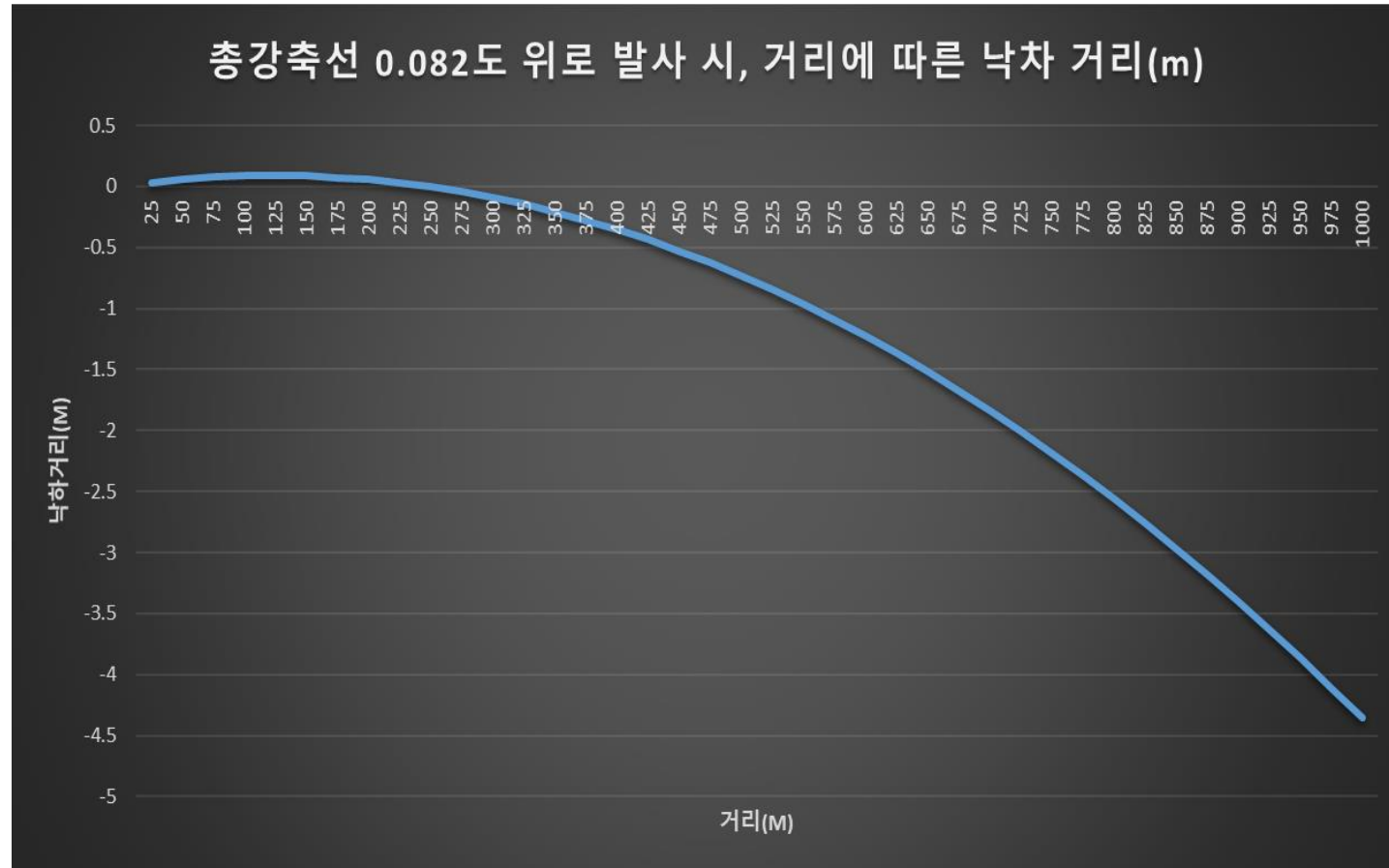
- 사업이 요구하는 집계 분석 및 인사이트 제공
- 게임 밸런스 분석
  - 유저 매칭 분석
  - 탄도 시뮬레이션
  - 게임 난이도 조정
  - 무기/탈것 밸런스 조정 등등
- 마케팅 최적화 분석
  - 광고 유입 매체 분석, Return on Ad Spend (ROAS), 광고 사기 등등
- 지표 예측 모형
  - 시계열 예측 모형을 활용한 예측
    - 일간 접속 유저수, 판매량, 동시 접속 유저수 등등
- 유저 과금 유무/액수 및 이탈 예측 모형
  - 머신 러닝을 활용한 Classification과 Regression
- 유저 상태에 따른 최적 상품 및 가이드 제공
- SNS 감성 분석
  - 일자 별 발생하는 단어 빈도
  - 유저들의 긍정/부정 단어 및 문장 분석
  - 정성적인 텍스트를 정량적인 분석으로 인사이트 제공
- *이외 회사가 급하게 요청하는 분석들 ππ*

# 탄도 시뮬레이션

K2 초속 920m/s 조준선과 총강축선 각도

공기 저항/탄알 모양 무시할 때

초속도(m/s)	각도(도)	목표거리(m)	Radian	목표도달시간(s)	낙차 거리(m)
920	0.082	25	0.00143117	0.027173941	0.032161001
920	0.082	250	0.00143117	0.271739409	-0.00403456



# 모바일 광고 매체

S.Korea Mobile Ad Technology Landscape.



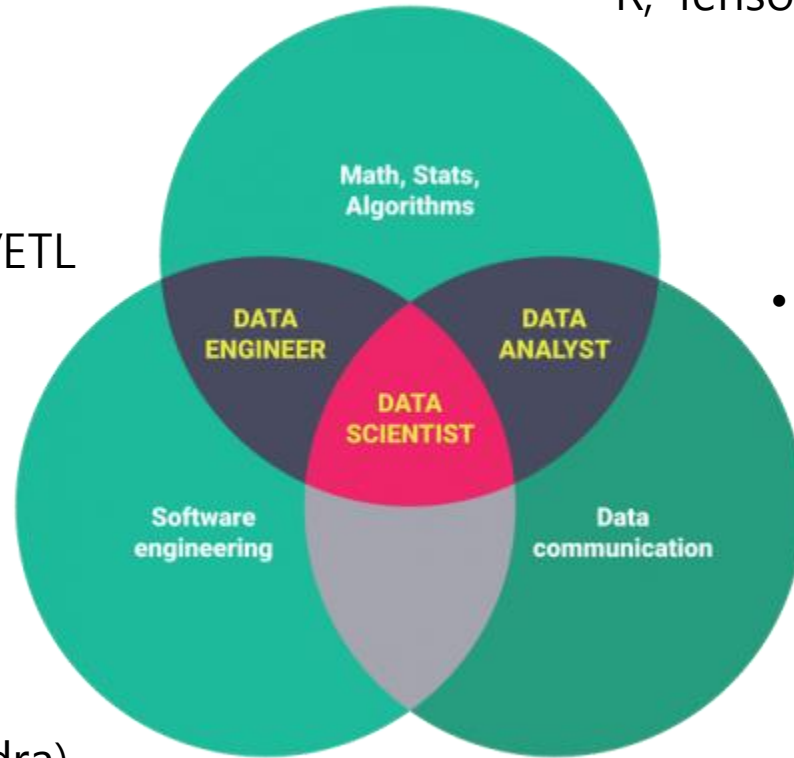
- 광고 유입 매체 분석 및 최적화
  - Cost per Install(CPI)
  - Cost per Action(CPA)
  - ROI(Return on investment)
  - Return on Ad Spend(ROAS)
    - LTV 예측 모형 활용
- 광고 사기 적발

# 분석에서 요구하는 데이터 사이언티스트 정의

- 통계모델링/머신러닝/딥러닝
  - R, TensorFlow, Scikit-learn 등등

데이터 분석 범위

- 탐색적 자료 분석 EDA
  - SQL, R



Visualization

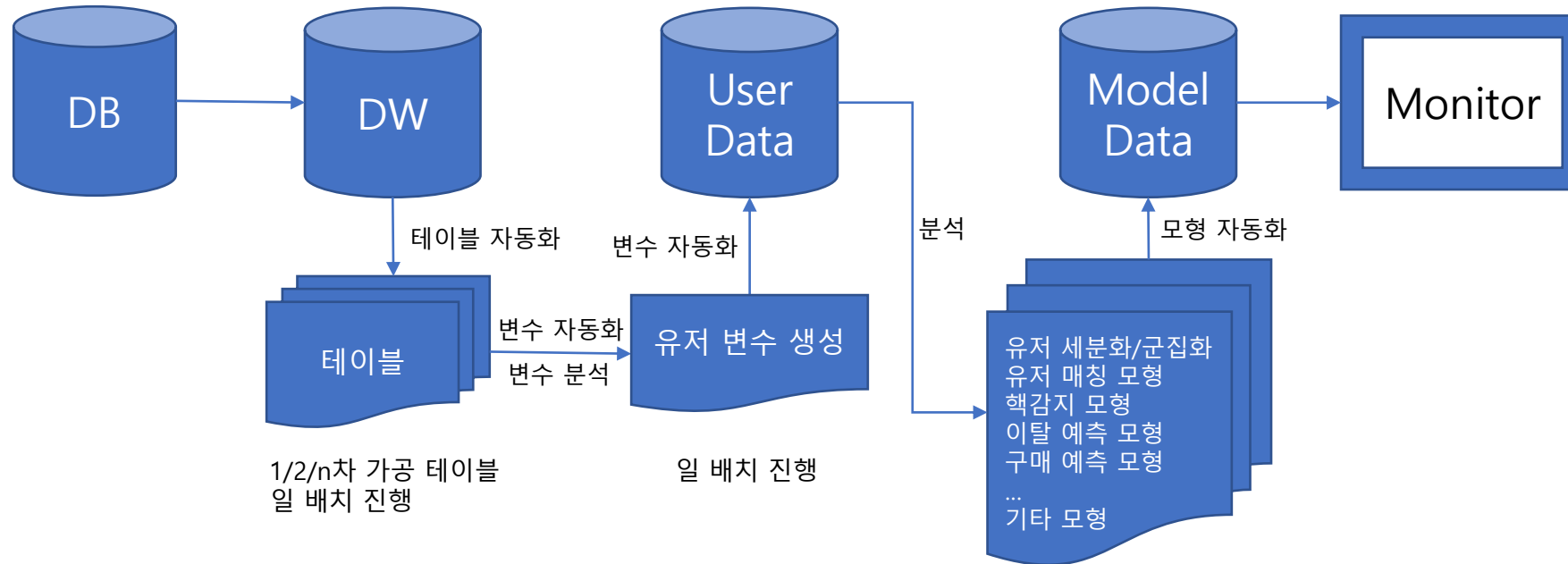
Text mining

- 자료추출(Acquire)/ETL
  - SQL, Python

데이터 수집 및 정제 스킬 이해

- 데이터 인프라
  - 데이터 수집(Kafka, Cassandra)
  - 빅데이터 소프트웨어(Hadoop), 클라우드(Google, AWS)

# 데이터 사이언티스트가 필요한 이유?



- 분석의 효율성을 위해서 분석 자동화를 진행
- 유저로부터 생성된 log를 자동으로 정제/집계하여 유저를 분류하고 예측

# 어렵지만 업무에 필요한 분석 기법

- 기초 통계 분석
  - 빈도분석, 기술통계분석, 교차분석, 상관관계 분석, 요인 분석, 회귀 분석, T-test 분석, 분산 분석 등등
- 시계열분석
- 분석 모델링 및 머신(딥) 러닝
  - Logistic, Lasso, Ridge
  - Association Rule
  - Random Forest, XGBoost, Conditional Inference Tree, Bayesian Network
  - SVM, NN, kNN, CNN, RNN, LSTM, GBM
  - Clustering(k-means), Anomaly Detection, Dimension reduction



# 꼭 알고 있으면 좋은 것들

- 업계에 대한 이해
- SQL
  - 배우기 정말 쉽지만 효과가 큰 언어
- Python
  - 머신 러닝에 많이 쓰이는 언어
- 기본 통계 개념
  - P value
  - Logistic regression 이해
  - 베이저안 추론 및 마르코프 체인 개념
- 그리고 가장 중요한 물리학





# 물리학을 통해서 바라본 빅데이터 분석

- 자연 현상을 설명하려는 물리학은 사회에서 어려운 문제를 정의/모델링/분석하는 데 큰 도움
- 실제로 머신 러닝에서 물리학 개념을 차용하고 있음
  - 볼츠만 머신, 이징 모형, 크로스 엔트로피 등등
- 현업에서는 데이터 분석가에서 데이터 사이언티스트를 필요로 하고 있는 추세
- 입자 물리 실험에서의 빅데이터 분석 경험은 빅데이터 분석 산업 영역에서 분석과 큰 차이가 없음
- 데이터 사이언티스트로 성장하기에 물리학 전공은 많은 이점이 있다고 판단



**THANK YOU!**



**PLAYERUNKNOWN'S  
BATTLEGROUNDS**