AI 시대의 데이터 아키텍처
애자일 데이터 패브릭 (Agile Data Fabric)

TIBCO Software Korea

CONFIDENTIALITY & DISCLAIMER

The information in this document is confidential information of TIBCO Software Inc. and/or its affiliates. Use, duplication, transmission, or republication for any purpose without the prior written consent of TIBCO is expressly prohibited.

This document (including, without limitation, any product roadmap or statement of direction data) illustrates the planned testing, release and availability dates for TIBCO products and services. This document is provided for informational purposes only and its contents are subject to change without notice. TIBCO makes no warranties, express or implied, in or relating to this document or any information in it, including, without limitation, that this document, or any information in it, is error-free or meets any conditions of merchantability or fitness for a particular purpose.

The material provided is for informational purposes only, and should not be relied on in making a purchasing decision. The information is not a commitment, promise or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any features or functionality described for our products remains at our sole discretion.

During the course of this presentation, TIBCO or its representatives may make forward-looking statements regarding future events, TIBCO's future results or our future financial performance. These statements are based on management's current expectations. Although we believe that the expectations reflected in the forward-looking statements contained in this presentation are reasonable, these expectations or any such forward-looking statements could prove to be incorrect and actual results or financial performance could differ materially from those stated herein. TIBCO does not undertake to update any forward-looking statement that may be made from time to time or on its behalf.

 © Copyright 2000-2021 TIBCO Software Inc.

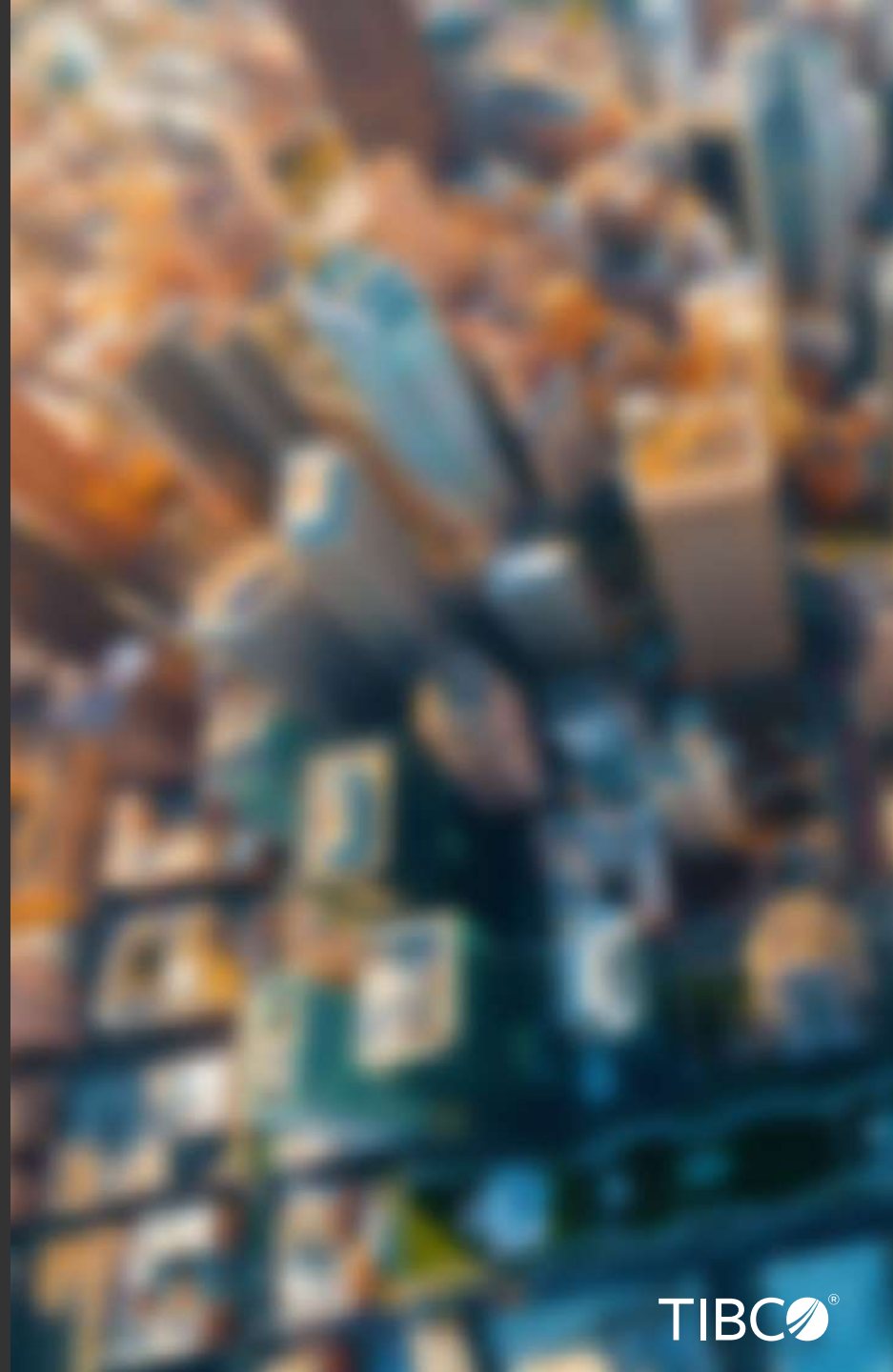
Agenda

Session 1

- **Becoming AI-driven**
- **Agile Data Fabric for AI Era**
  - 실시간 데이터 파이프라인
  - 데이터 서비스 레이어
  - 고급 데이터 분석

Session 2

- 국내 사례 소개
  - 국내 : 신한은행, 국내 반도체 H사



 © Copyright 2000-2021 TIBCO Software Inc.

TIBCO®

AI 시대를 위한 데이터 아키텍처

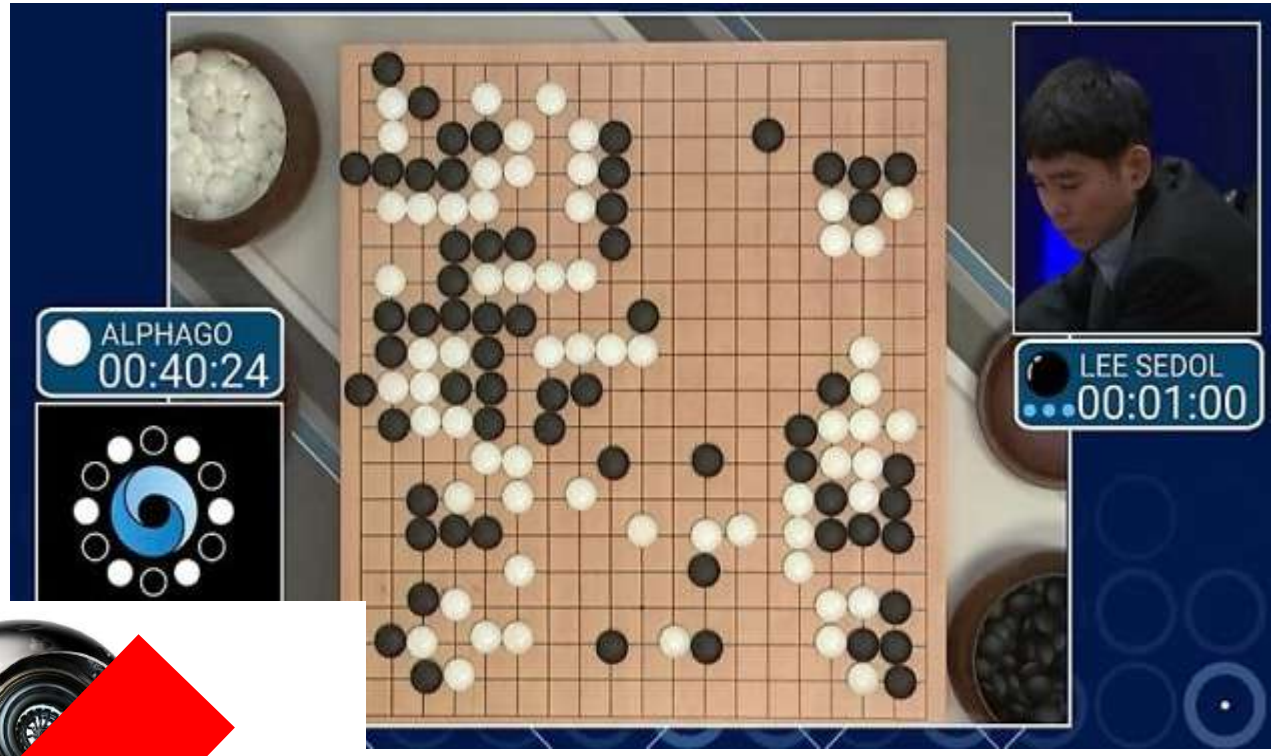
 © Copyright 2000-2021 TIBCO Software Inc.

Becoming AI-driven

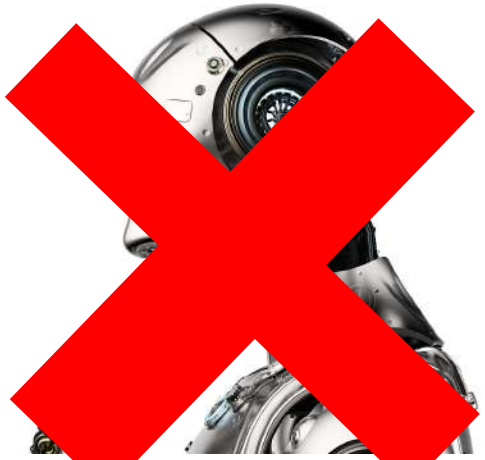
 © Copyright 2000-2021 TIBCO Software Inc.

AI = Super Intelligence?

IBM Watson Jeopardy 최 (2011년)



ogle Alpha Go 이세돌 대결 (2016년)



Artificial General Intelligence (AGI)

© Copyright 2000-2021 TIBCO Software Inc.

So What?

5년 전

해결하지 못했던 난제들을 해결해
세상을 바꾸겠습니다.

AGI

- 자연어 처리
- 자율 주행
- 질병 진단 및 치료
- 기타 고급 인지 활동

Out-sourcing
외부 조직에 대한 의존
업무와 직접 관련 적음

현재

Back to Reality

Five years ago, Watson, a nerdy, disembodied voice from the A.I. future, chatted and joked in advertisements with the tennis superstar Serena Williams. Today, the TV ads proclaim the technology's potential to save time and work in offices and on factory floors.

잘 할 수 있는 세부 영역에 집중하여
효율성과 생산성 증대에 기여합니다.

Practical AI

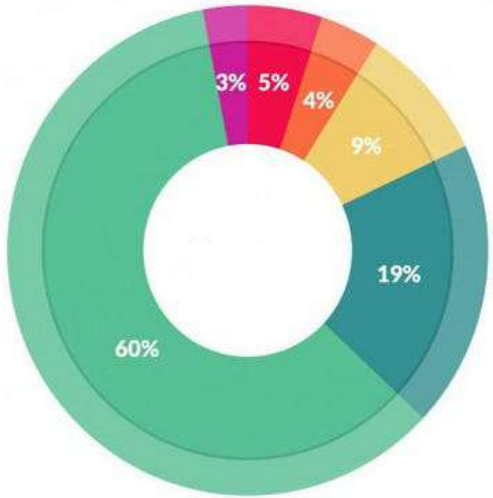
- 데이터 기반 의사결정
- 새로운 비즈니스
- 데이터 엔지니어링
- 예측 / 분류 모델

In-sourcing
조직 내 데이터 역량 향상
업무와 직접적 관련성

TIBCO can help

© Copyright 2000-2021 TIBCO Software Inc.

분석에 중요한 것은 Big Data? Good Data!



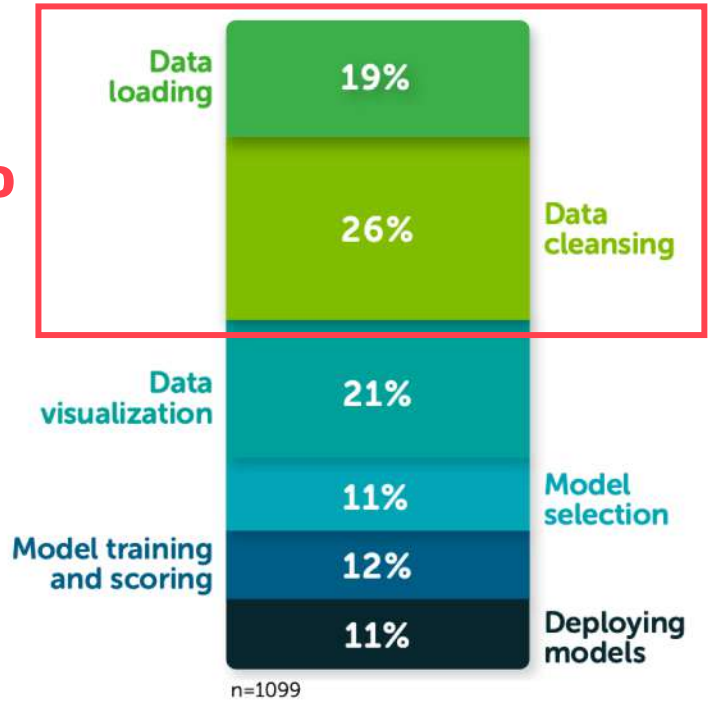
What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

79%

45%

THINKING ABOUT YOUR CURRENT JOB, HOW MUCH OF YOUR TIME IS SPENT IN EACH OF THE FOLLOWING TASKS? (PLEASE ASSIGN A PERCENTAGE; TOTAL MUST ADD UP TO 100%.)



For most respondents, data management tasks still consume a disproportionate portion of work time.

분석을 위한 Good Data 확보에 소요되는 시간과 자원

4년 간 34% (= 매년 8.5%) 감소

Andrew Ng

I feel like we need shift in mindset from **BIG** data to **GOOD** data.

- MIT Technical Review (2021.03.26)

Source : Forbes, Cleaning Big Data (2016)

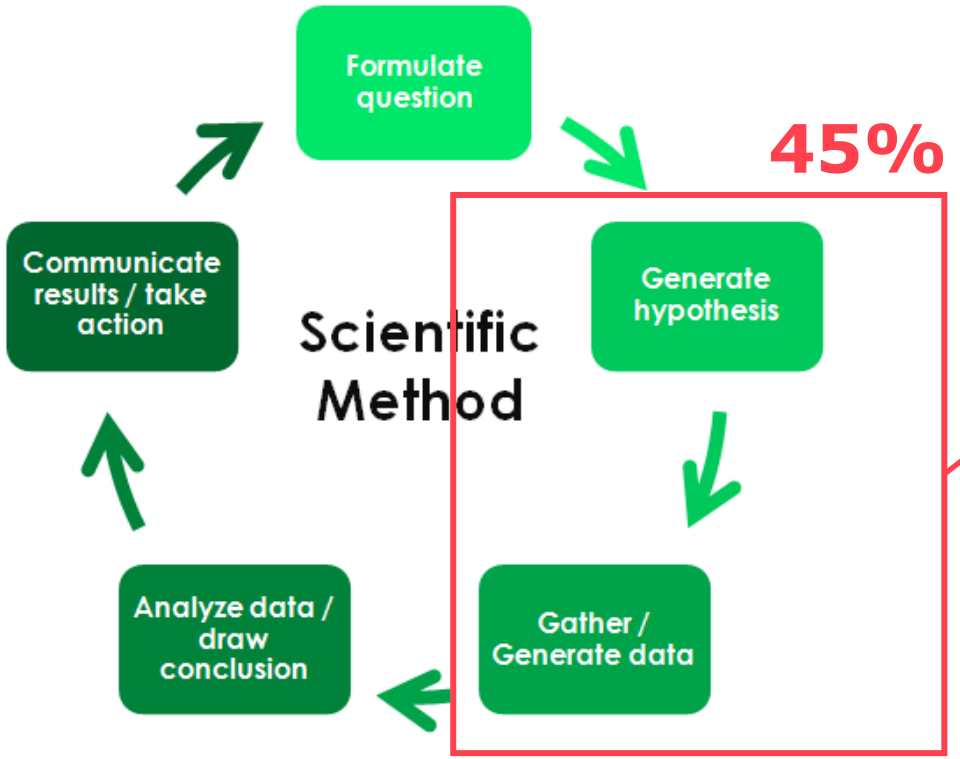
Source : Anaconda State of Data Science (2020)

© Copyright 2000-2021 TIBCO Software Inc.

분석에 중요한 것은 Big Data? Good Data!

Good Data란?

1. 기술적인 오류가 없는 데이터 (결측, 아웃라이어, 입력 및 연산 오류 등)
2. **분석가의 가설 (or 의도) 가 정확하게 반영된 데이터**



Source : Business Broadway 2015

- 다양한 데이터 활용 + 데이터 유통 효율 향상
  → ***배치성 정형 데이터 + 실시간 비정형 데이터***
- 반복되는 검증 과정 = 반복되는 가설의 변경
  = 반복적인 분석 데이터 셋의 변경
  → ***손쉽고 빠른 다양한 원천 데이터에 대한 접근성***
- 분석가의 가설을 가장 잘 아는 것은 분석가
  → ***Self-service analytics***

TIBCO® **Agile Data Fabric**

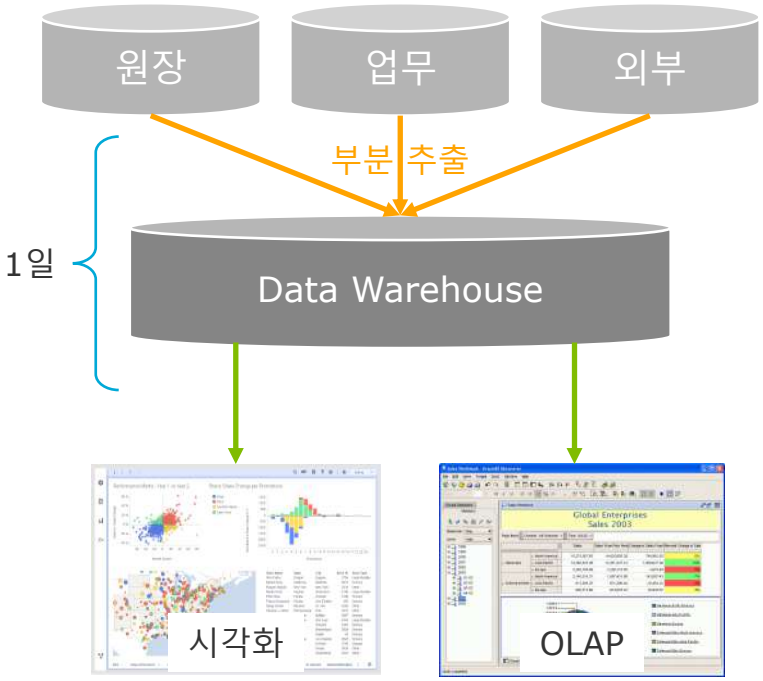
© Copyright 2000-2021 TIBCO Software Inc.
TIBCO®

Agile Data Fabric for AI Era

 © Copyright 2000-2021 TIBCO Software Inc.

TIBCO Agile Data Fabric

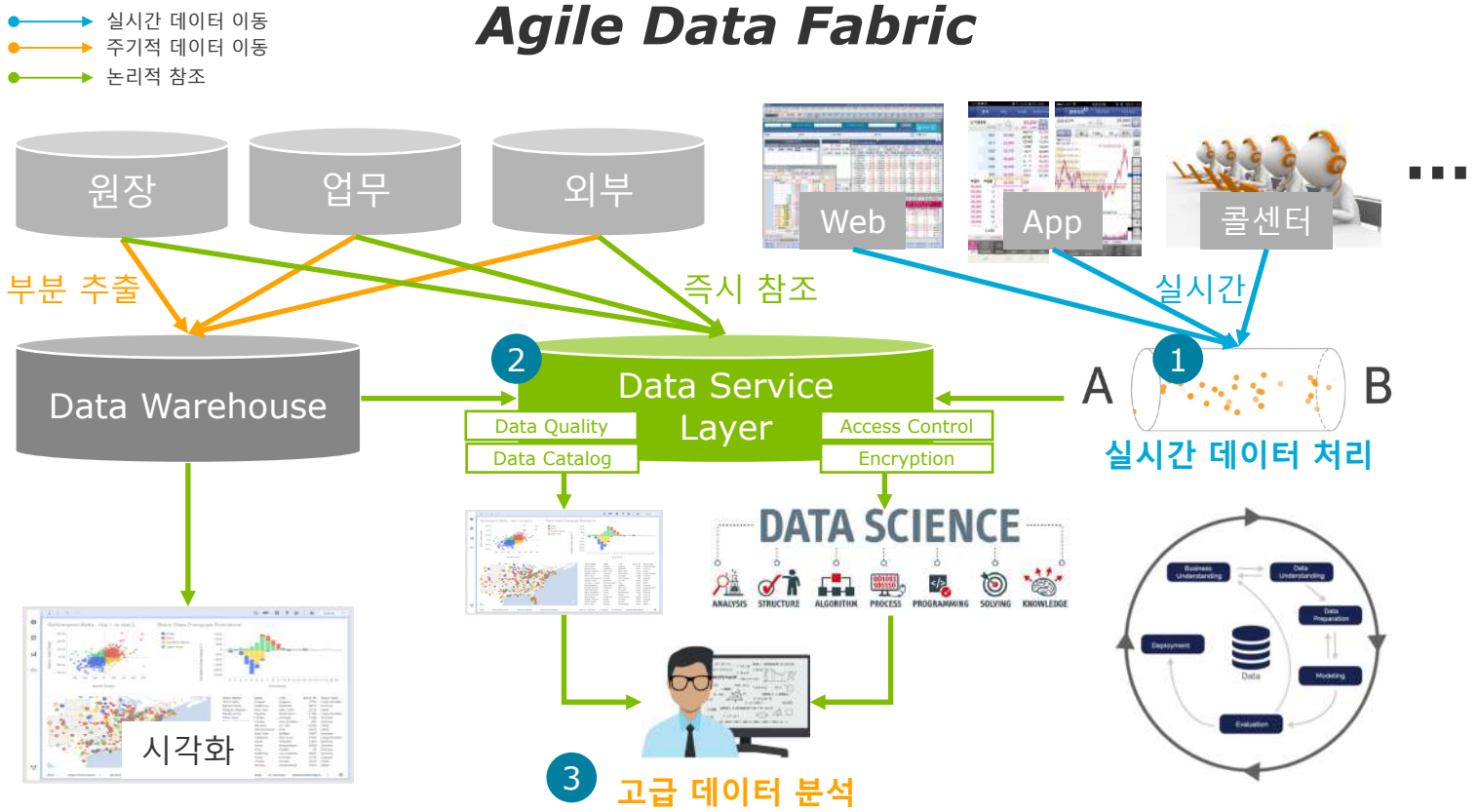
과거



- 직관적이고 아름다운 화면 구성
- 빠르게 더 빠르게!
- 추출된 과거 데이터만 활용/분석

데이터 분석 < 데이터 시각화

Agile Data Fabric



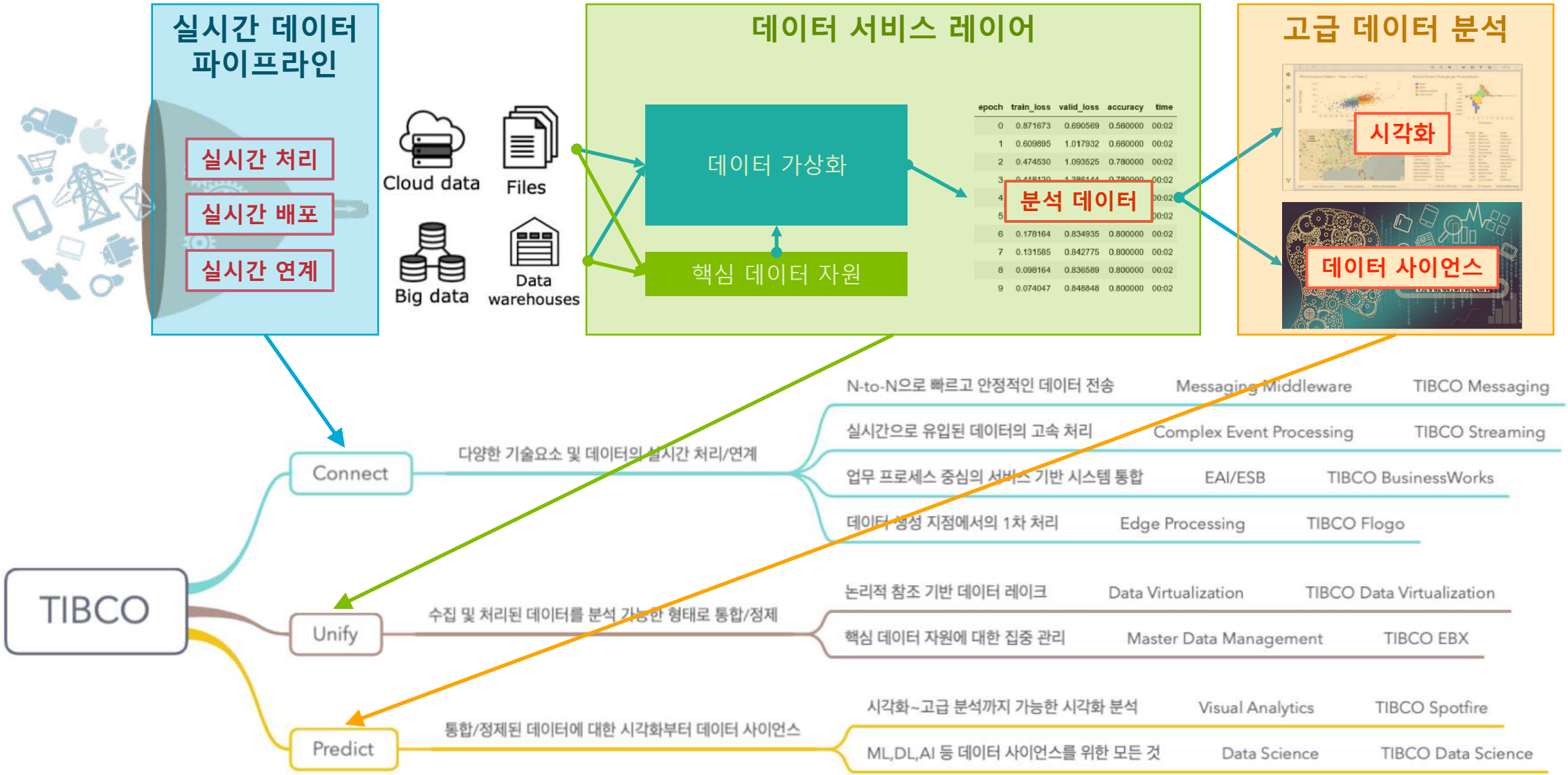
- 가용한 데이터 자원 최대 활용
- 데이터 지연 시간 최소화
- 빠르고, 자유로운 데이터 접근

- 데이터 사이언스에 적합 = AI에 적합
- 가설 → 데이터로 검증 → 해결책
- 빠른 반복으로 검증과 개발 용이

다양한 대용량 데이터에 대한 자유로운 접근 보장 → Built for AI

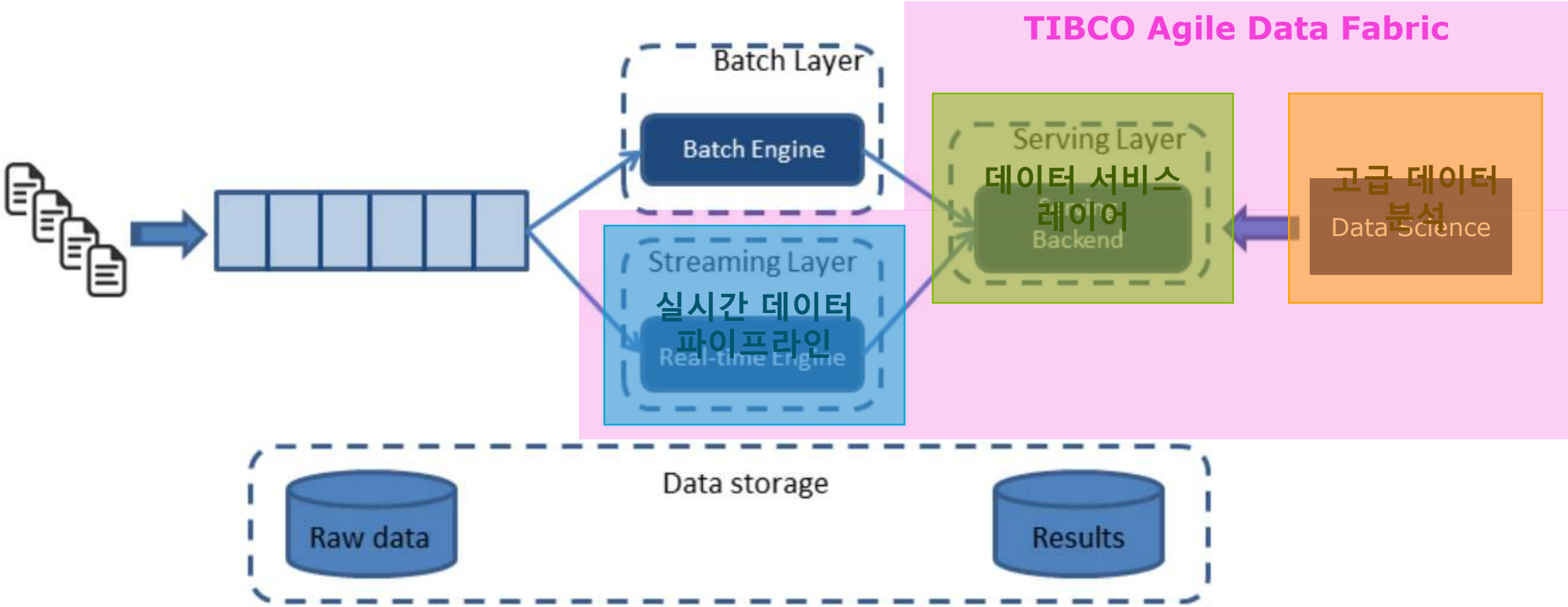
© Copyright 2000-2021 TIBCO Software Inc.

TIBCO Agile Data Fabric



 © Copyright 2000-2021 TIBCO Software Inc.

Lambda Architecture



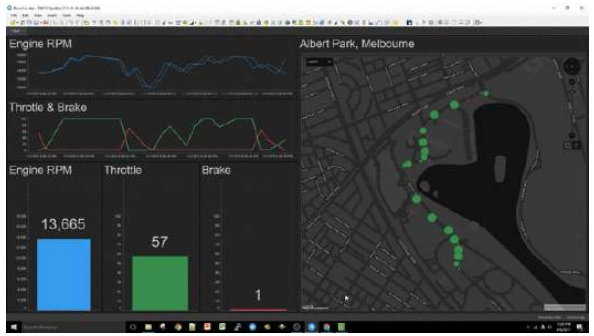
 © Copyright 2000-2021 TIBCO Software Inc.

빠르고 효율적인 데이터 활용, Agile Data Fabric

Agile Data Fabric

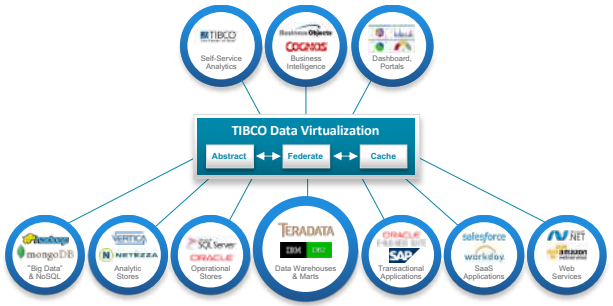
실시간 데이터 파이프라인

- 데이터가 생성되는 시점에 즉시 수집 / 정제 / 처리 / 배포
- Data-in-motion에 특화
- 실시간 모니터링, 실시간 예측



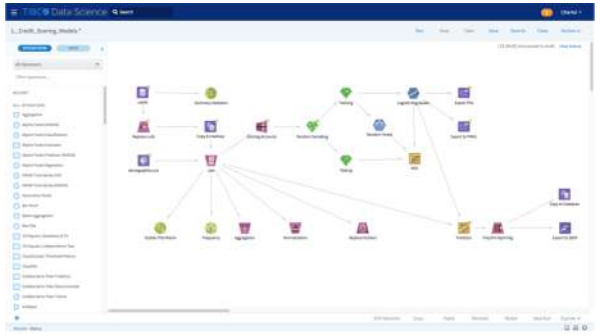
데이터 서비스 레이어

- 분산되고 고립된 데이터 자원에 대한 빠르고 안전한 접근
- 논리적 참조에 기반한 데이터 활용
- 전사 데이터 자원의 통합 창구



고급 데이터 분석

- Market leading 시각화 분석 플랫폼
- ML/AI를 위한 데이터 사이언스 end-to-end support



비즈니스 실시간 가시성 확보

손쉬운 데이터 연계 및 통합

AI를 활용한 예측 및 자동화

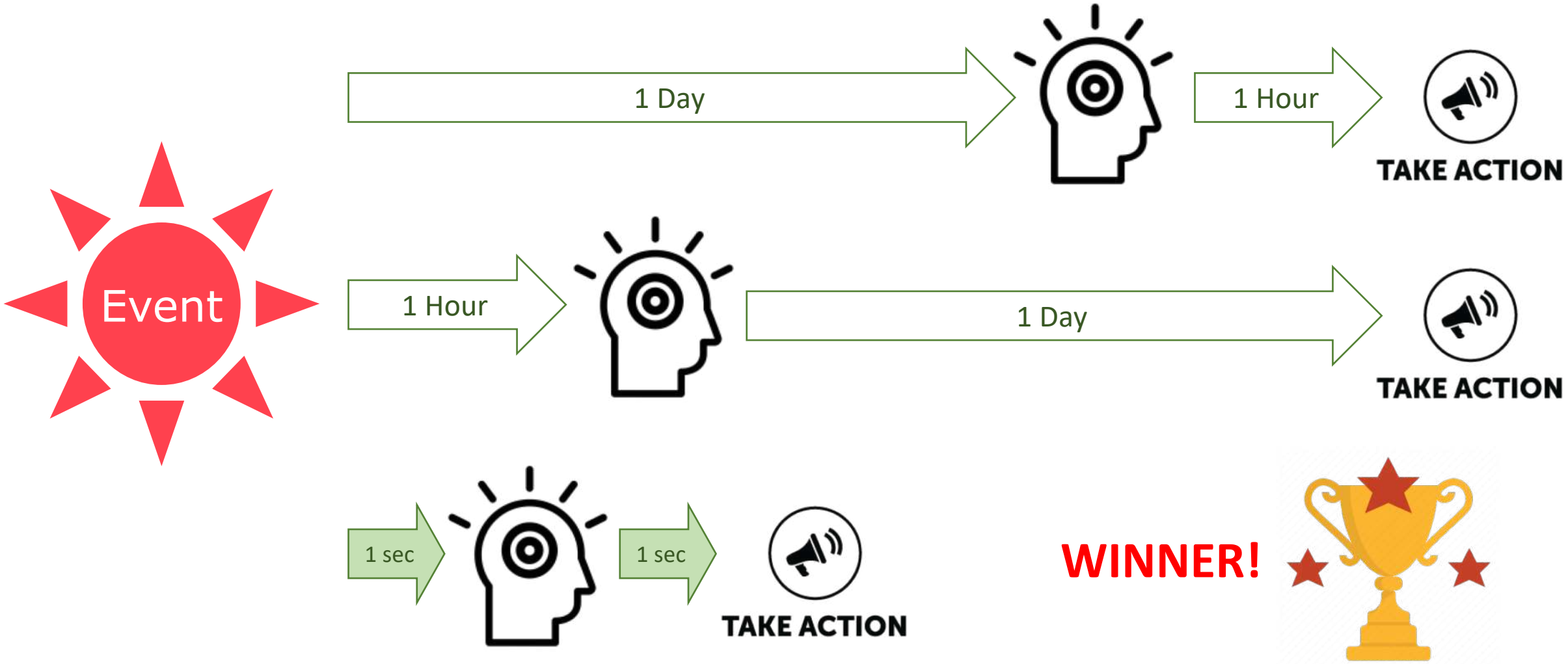
Real-time Smart Enterprise

© Copyright 2000-2021 TIBCO Software Inc.

실시간 데이터 파이프라인

 © Copyright 2000-2021 TIBCO Software Inc.

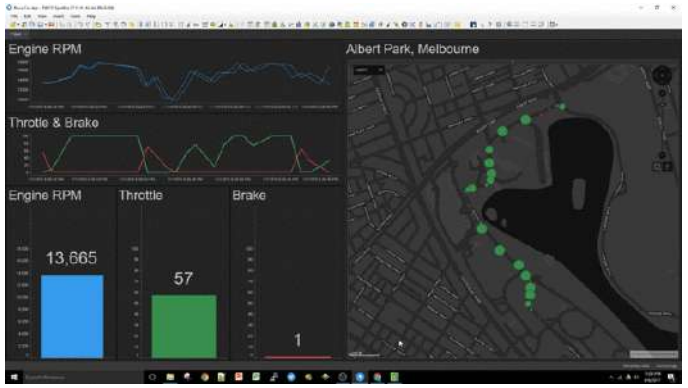
Being a Fast Fish

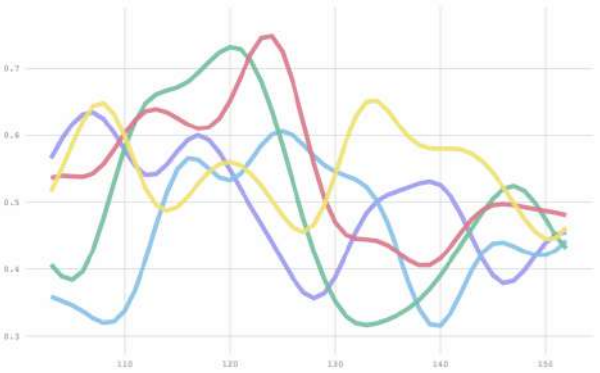


 © Copyright 2000-2021 TIBCO Software Inc.

실시간 데이터 아키텍처의 필수 요건

1 Event Processing





- *데이터는 멈추지 않고 계속 흘러야 한다.*
- *유입 데이터의 유실을 방지해야 한다.*
- *유입 데이터의 순서가 지켜져야 한다.*
- *데이터를 고속으로 처리해야 한다.*
- *실시간 데이터의 활용도를 높여야 한다.*

TIBCO® **Event Processing**

- Messaging
- Streaming
- Visual Analytics
- Data Science

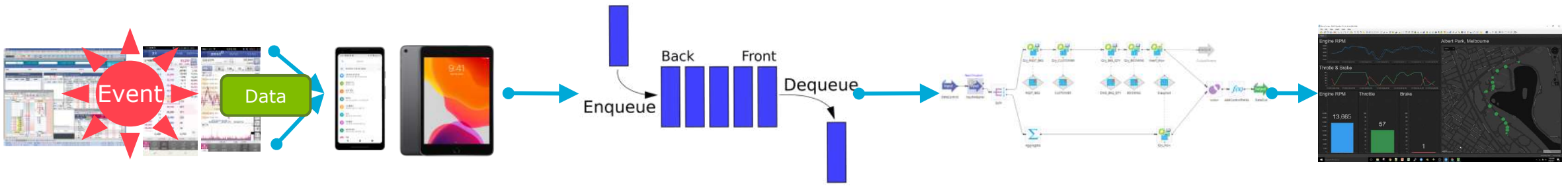
© Copyright 2000-2021 TIBCO Software Inc.

실시간 데이터 아키텍처 구성요소

1 Event Processing

실시간 업데이트

Source → **Edge** → **Messaging** → **Processing** → **Analytics**



- 이벤트가 발생하는 지점
- 센서 등으로 이벤트 감지
- 감지결과 전송(=데이터)

- 이벤트 데이터 수집
- 수집 데이터 1차 처리
- 다운스트림으로 전송

- 이벤트 데이터 수집 후 이벤트 처리 레이어로 전송
- Messaging queue 기반

- 유입된 대량의 이벤트 데이터를 고속으로 정제/통합/변환/적재
- 메모리 기반 병렬 확장성

- 실시간 데이터 활용
- 실시간 대시 보드
- ML 기반 실시간 예측 대응

TIBCO Event Processing

TIBCO **Flogo** | TIBCO **Messaging** | TIBCO **Streaming** | TIBCO **Spotfire**

kafka

RabbitMQ

PULSAR

Micro-Batch 기반 (준 실시간)

kafka Streams

Spark Streaming

+ableau

sas

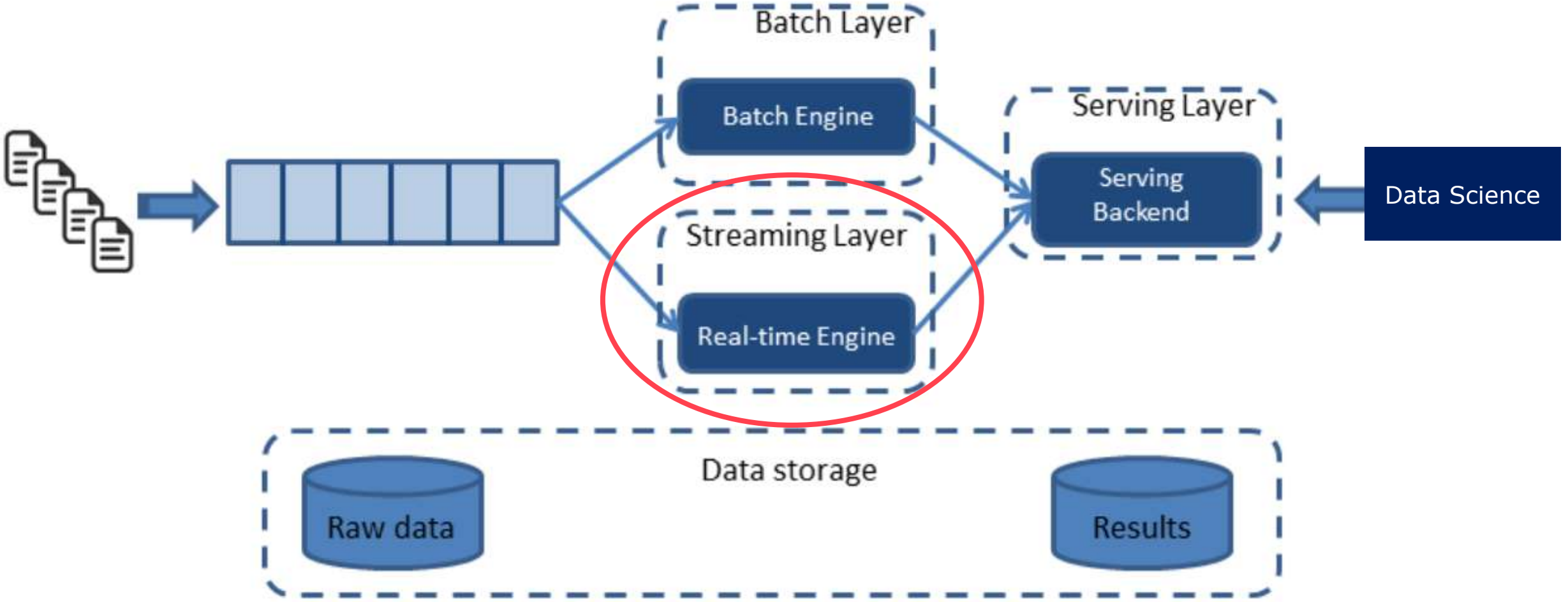
Power BI

© Copyright 2000-2021 TIBCO Software Inc.

TIBCO

제안 데이터 아키텍처 (= Lambda 아키텍처)

1 Event Processing



© Copyright 2000-2021 TIBCO Software Inc.

제안 데이터 아키텍처 (= Lambda 아키텍처)

1 Event Processing

실시간 업데이트

주기적 업데이트

PLC / OSI Pi

Batch Layer

DBMS

JDBC / ODBC

MES

PLC / OSI Pi

TIBCO Messaging

Streaming

ADAPTER

RULES

STREAM PROCESSING

JAVA / .NET / C++ / PYTHON

In-memory Fabric

Distributed Transactional Memory

ADAPTER

Streaming Layer

Any protocol

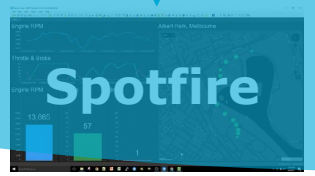
Any format

FDS

YMS

Model deploy

Real-time Dashboard
Real-time Prediction



TIBCO Data Science

Model development
Model management

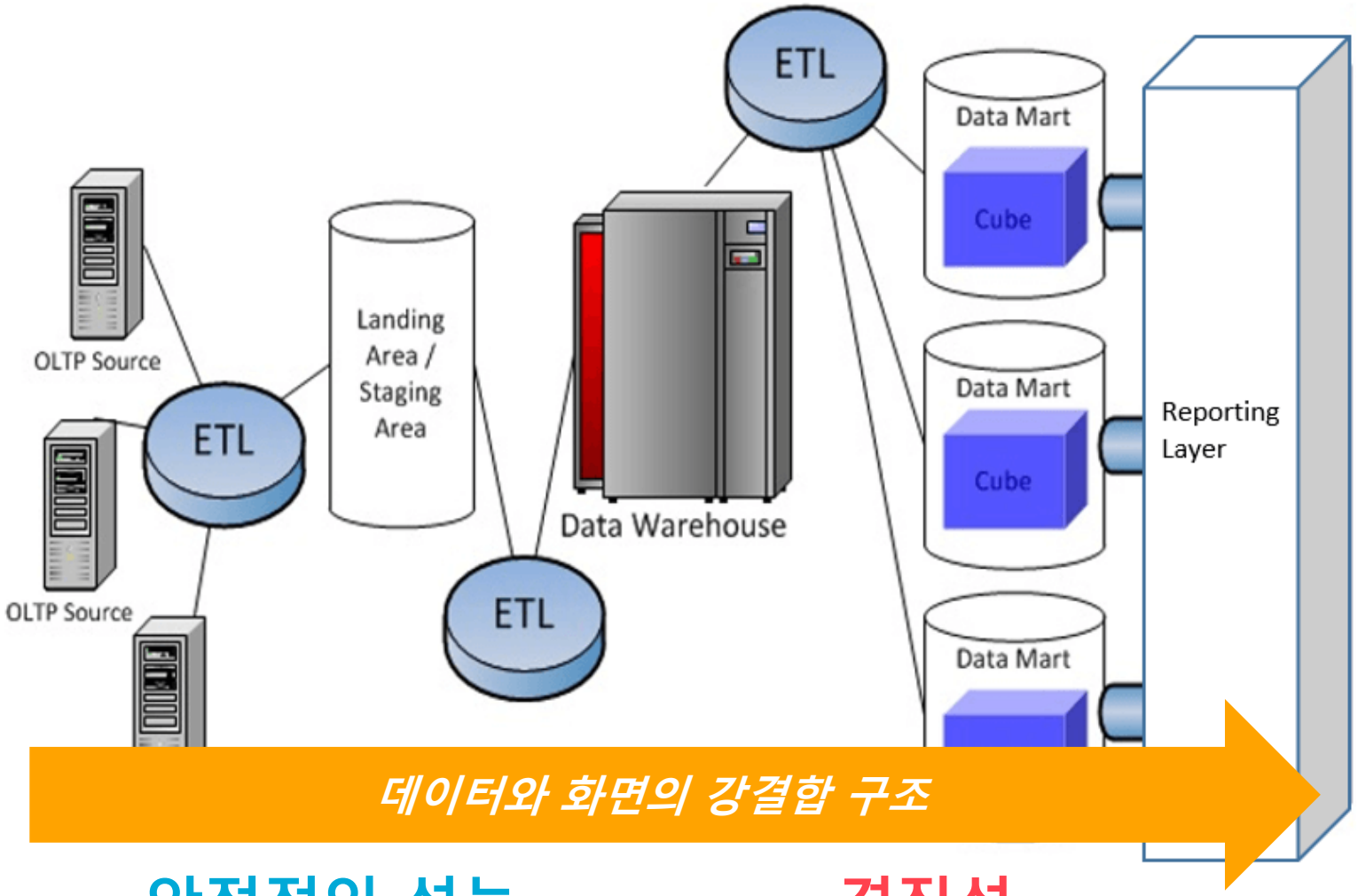
 © Copyright 2000-2021 TIBCO Software Inc.

데이터 서비스 레이어

 © Copyright 2000-2021 TIBCO Software Inc.

시각화 및 리포팅 위주였던 시절

2 Data Service Layer



초기 요건에 맞춰
보기 좋게 설계된
대쉬보드

대쉬보드를 빠르고
안정적으로 실행

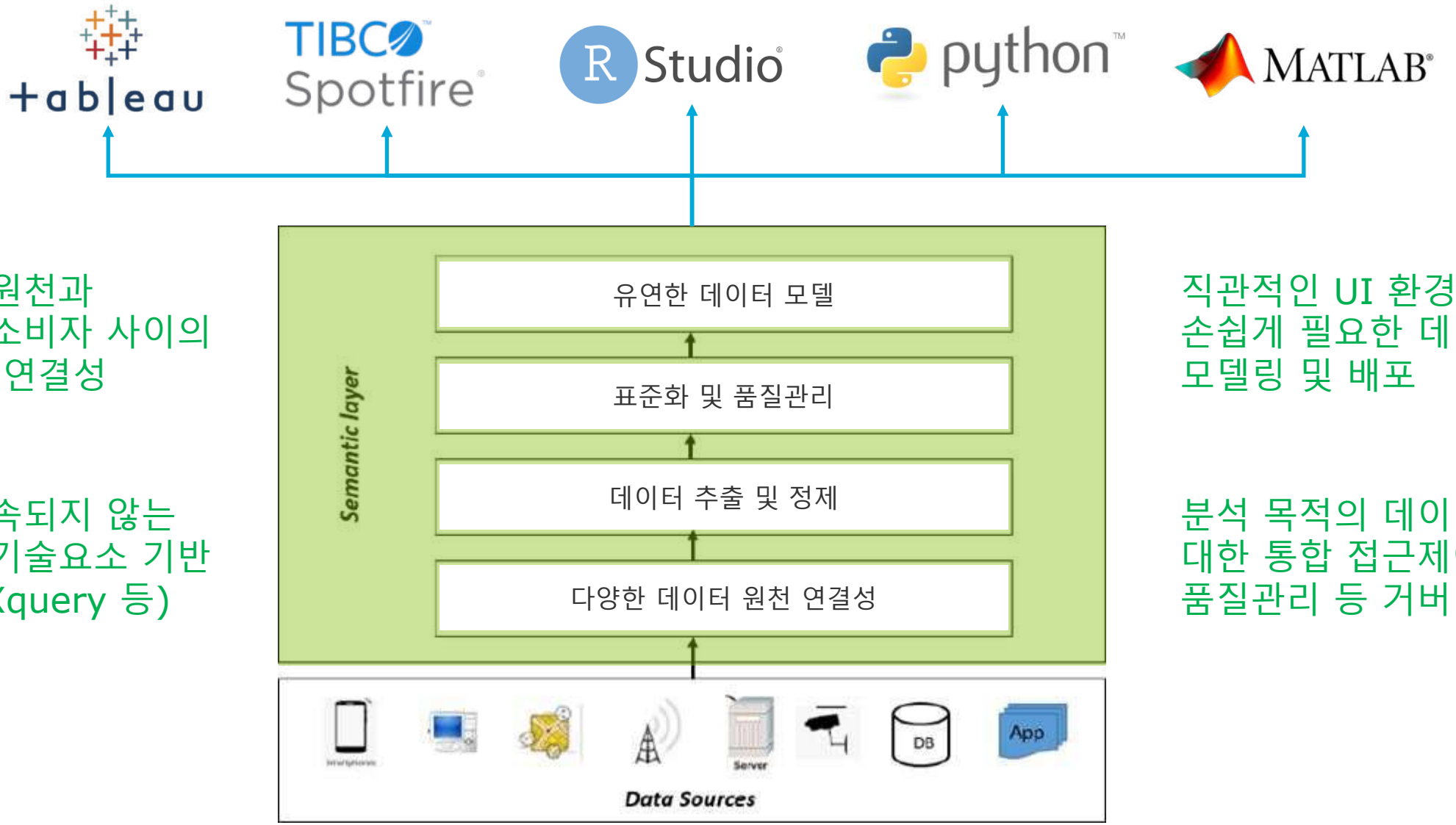
데이터 요건 수용성보다
화면의 성능과 안정성

데이터와 화면의 긴밀한 조합

© Copyright 2000-2021 TIBCO Software Inc.

데이터 사이언스의 도래

2 Data Service Layer



데이터 원천과
데이터 소비자 사이의
N-to-N 연결성

툴에 종속되지 않는
보편적 기술요소 기반
(SQL, Xquery 등)

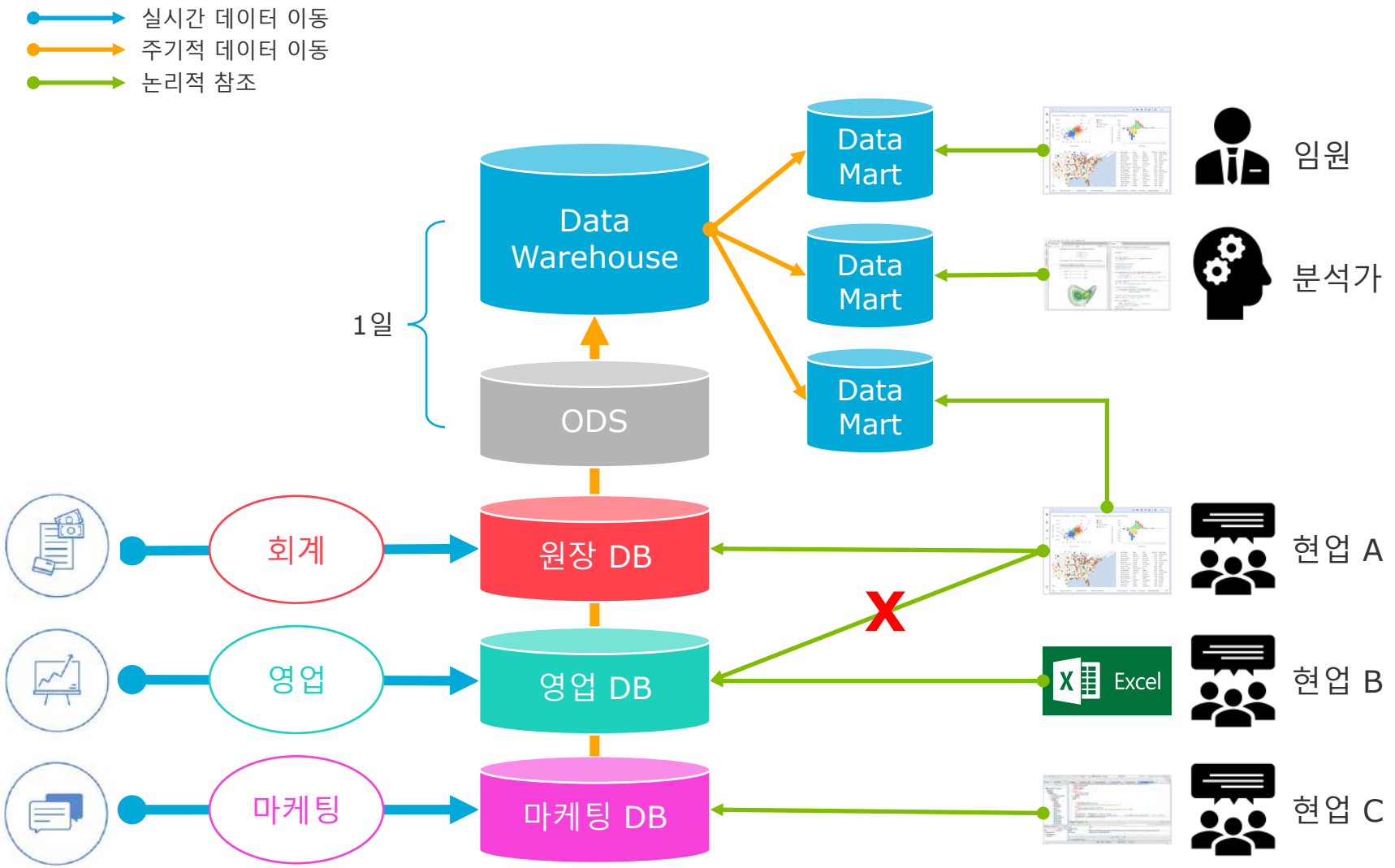
직관적인 UI 환경에서
손쉽게 필요한 데이터의
모델링 및 배포

분석 목적의 데이터에
대한 통합 접근제어 및
품질관리 등 거버넌스

© Copyright 2000-2021 TIBCO Software Inc.

일반적인 정보계 데이터 아키텍처

2 Data Service Layer



- 경직된 데이터 아키텍처
  ETL 변환 및 정제 작업,
  DW 데이터 모델은 한 번
  셋업되면 변경이 매우 어려움
  → 정형 리포트, 대쉬보드 적합
  → 고급 데이터 분석 부적합
- 변환 이력에 대한 가시성 부족
  (Data Lineage)
- 필요한 데이터 탐색 어려움
  (Data Catalog)

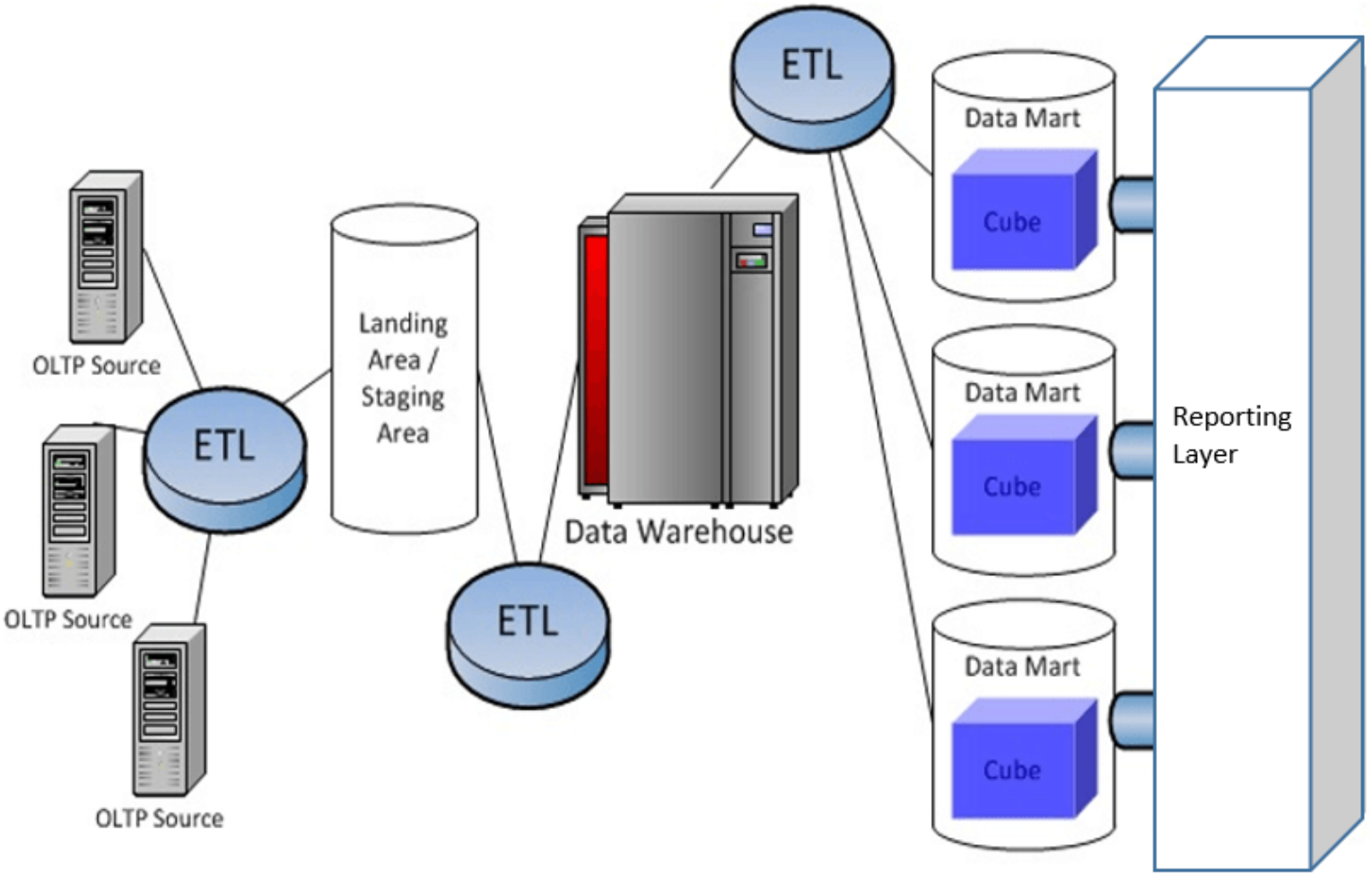
- 부서 별 원천 데이터 접근권한 제한 (데이터 이기주의)
- 물리적 데이터 고립
- 서로 다른 데이터 관리 주체
- 타 부서의 데이터 구조에 대한 이해 부재
- 전체 데이터 자원의 활용성 저하

© Copyright 2000-2021 TIBCO Software Inc.

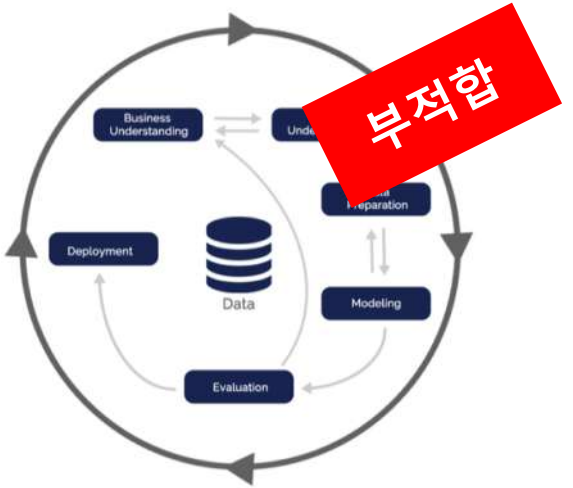
정형화된 시각화에 적합한 데이터 아키텍처

2 Data Service Layer

PIMS 담당자 ← ETL 담당자 ← DW 담당자 ← DM 담당자 ← 시각화 담당자 ← 변경 요건



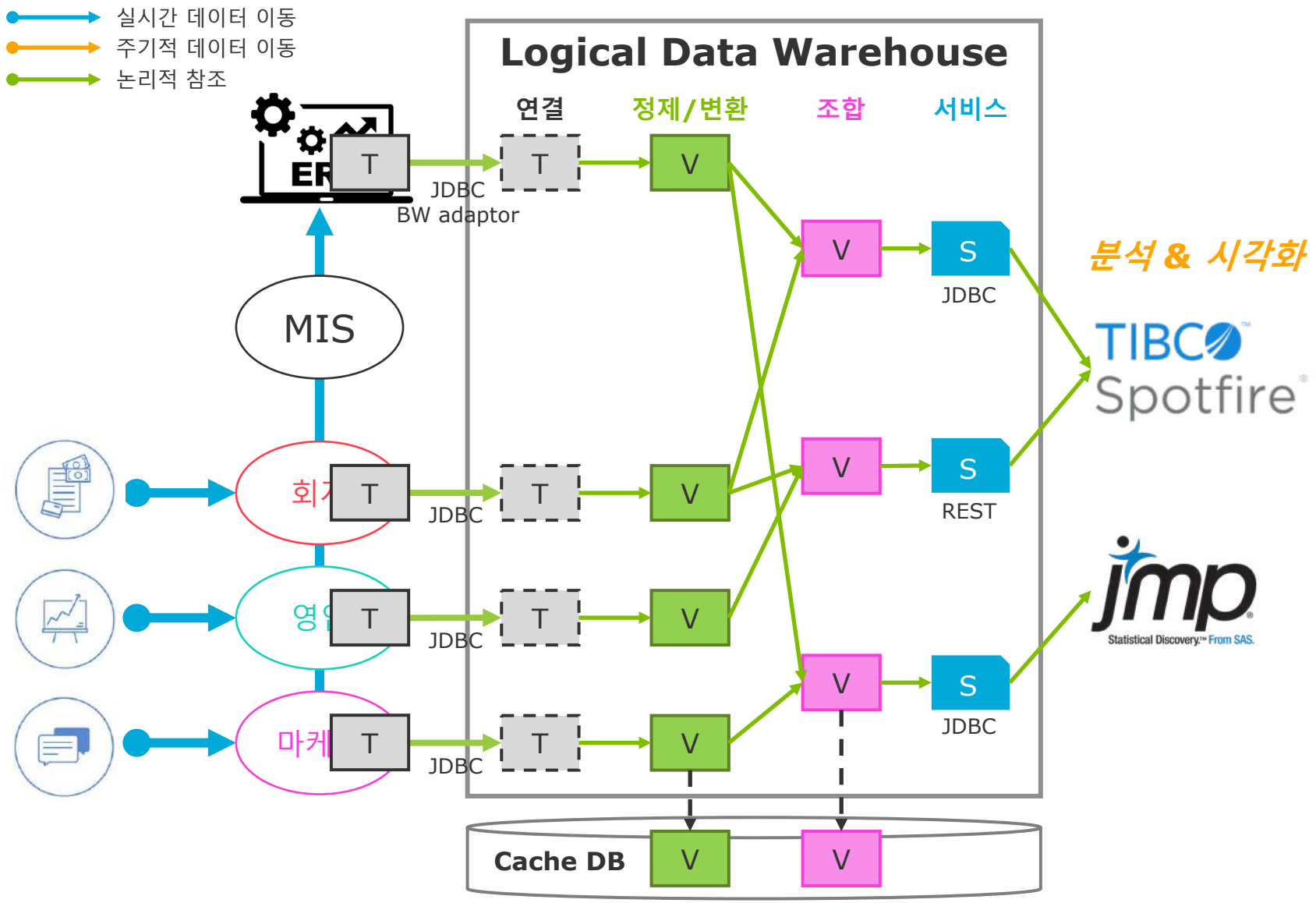
분기 1회



매일 수차례

 © Copyright 2000-2021 TIBCO Software Inc.

Data Service Layer



장점

- 원본 데이터의 **정제 및 변환 자동화** (정제/변환 view 생성)
- 논리적 참조 기반의 **유연한 데이터 아키텍처** 분석에 필요한 데이터 셋 (=뷰)을 필요한 시점에 쉽고 빠르게 개발 → **데이터 분석 효용성 ↑**
- 하나의 솔루션으로 구현해 **운영 관리 / 비용 효율 향상**
- 분석가들이 직접 원하는 데이터 셋을 개발 → **Self-service**
- 원천 DB의 부하 최소화 및 **쿼리 성능 최적화 메커니즘**
- 캐쉬를 통한 **원천 부하 해소**

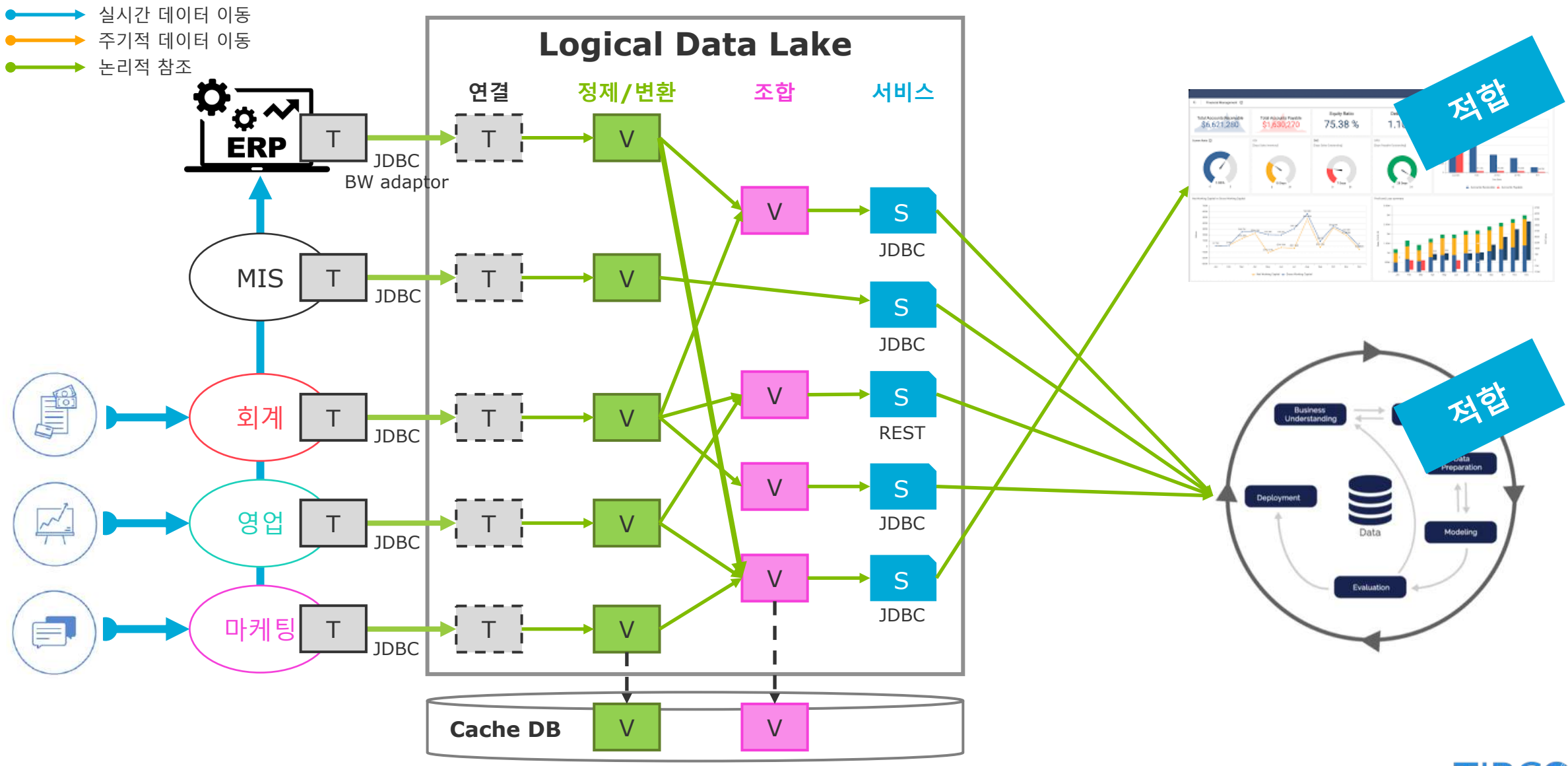
단점

- 대용량 데이터 조회 시 **성능 저하의 폭**이 더 클 수 있음
- 데이터 가상화라는 새로운 기술 요소에 대한 **경험 부족**

© Copyright 2000-2021 TIBCO Software Inc.

분석과 시각화 모두에 적합한 데이터 아키텍처

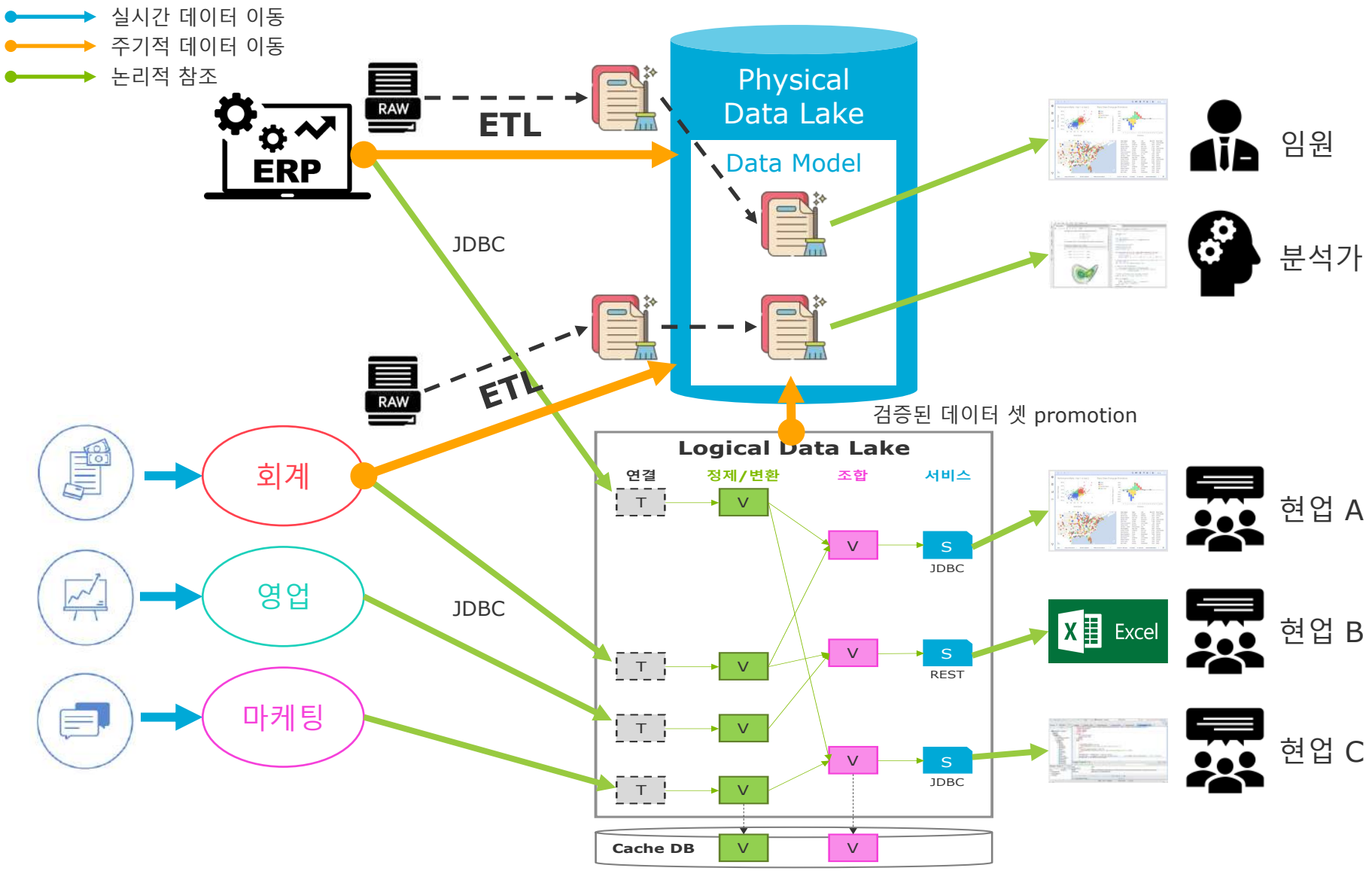
2 Data Service Layer



© Copyright 2000-2021 TIBCO Software Inc.

제안 정보계 아키텍처

2 Data Service Layer



장점

- 데이터의 정제 및 변환 자동화
- **Physical, Logical의 시너지**
- Physical : 반복적이고 변하지 않는 D-1 데이터 셋 처리 (시각화용)
- Logical : 분석을 위한 자유로운 D-0 데이터 셋 개발 (분석용)
- Logical에서 효용이 입증된 데이터 셋을 Physical에 반영
- 데이터 소비자의 요건의 성격에 따라 **Physical or Logical 유연 대응 가능**

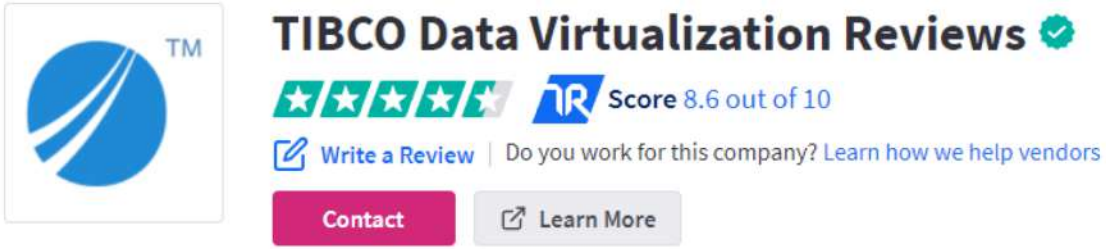
단점

- 시스템 복잡도 상승
- 구축 비용 증가 (ETL, DW, DV)
- D-1과 D-0의 차이로 **Physical과 Logical의 데이터 부정합** 가능성 존재

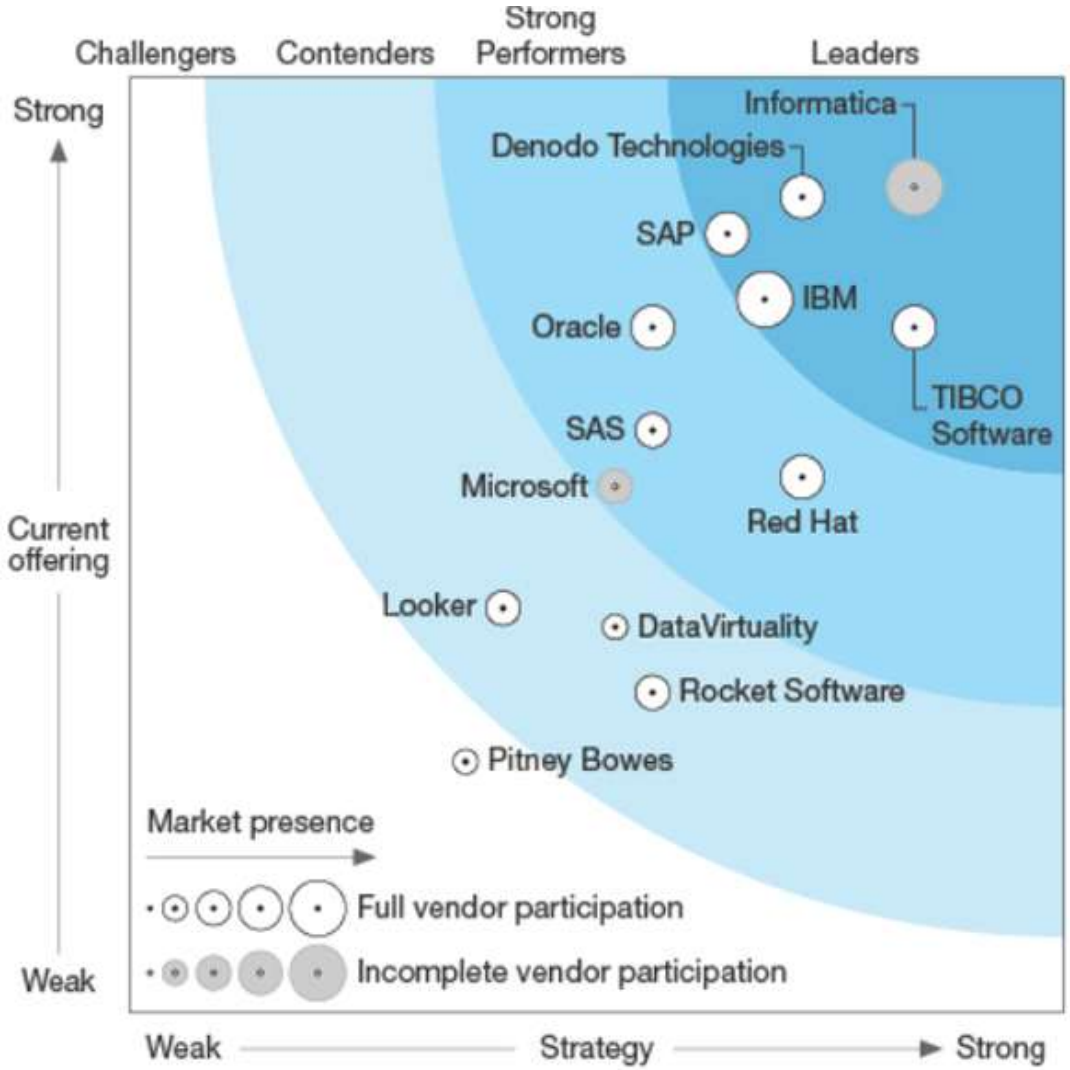
© Copyright 2000-2021 TIBCO Software Inc.

시장 평가

Gartner Peer Insights



Trust Radius



Enterprise Data Virtualization (Forrester Wave)

 © Copyright 2000-2021 TIBCO Software Inc.

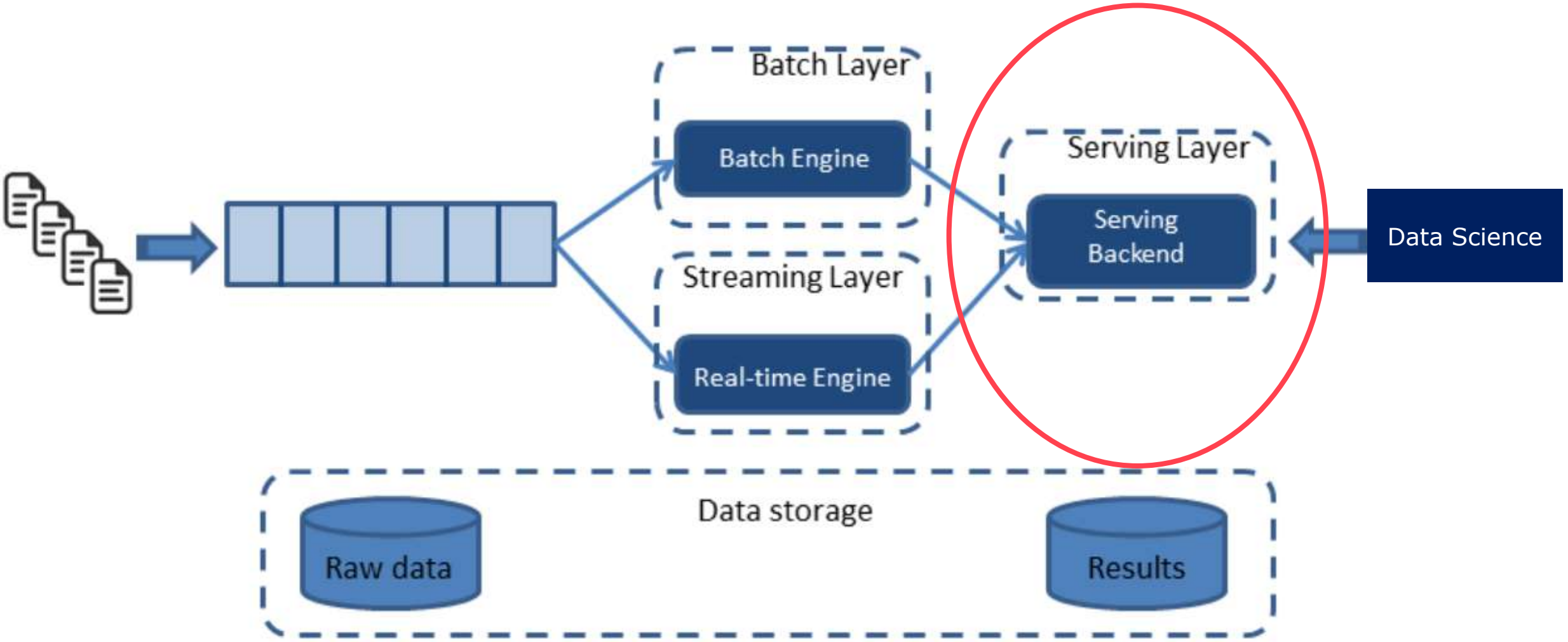
글로벌 고객사

Bank of America Merrill Lynch	comcast	Pfizer	BARCLAYS	Sony Entertainment Network
60% Reduction in Infrastructure Costs	**$21** Million Increase in Revenue	**60%** Reduction in R&D Project Delays	**$1.7** MM Reduction in Cost	**$9** Million in Revenue Acceleration

 © Copyright 2000-2021 TIBCO Software Inc.

제안 아키텍처 (= Lambda 아키텍처)

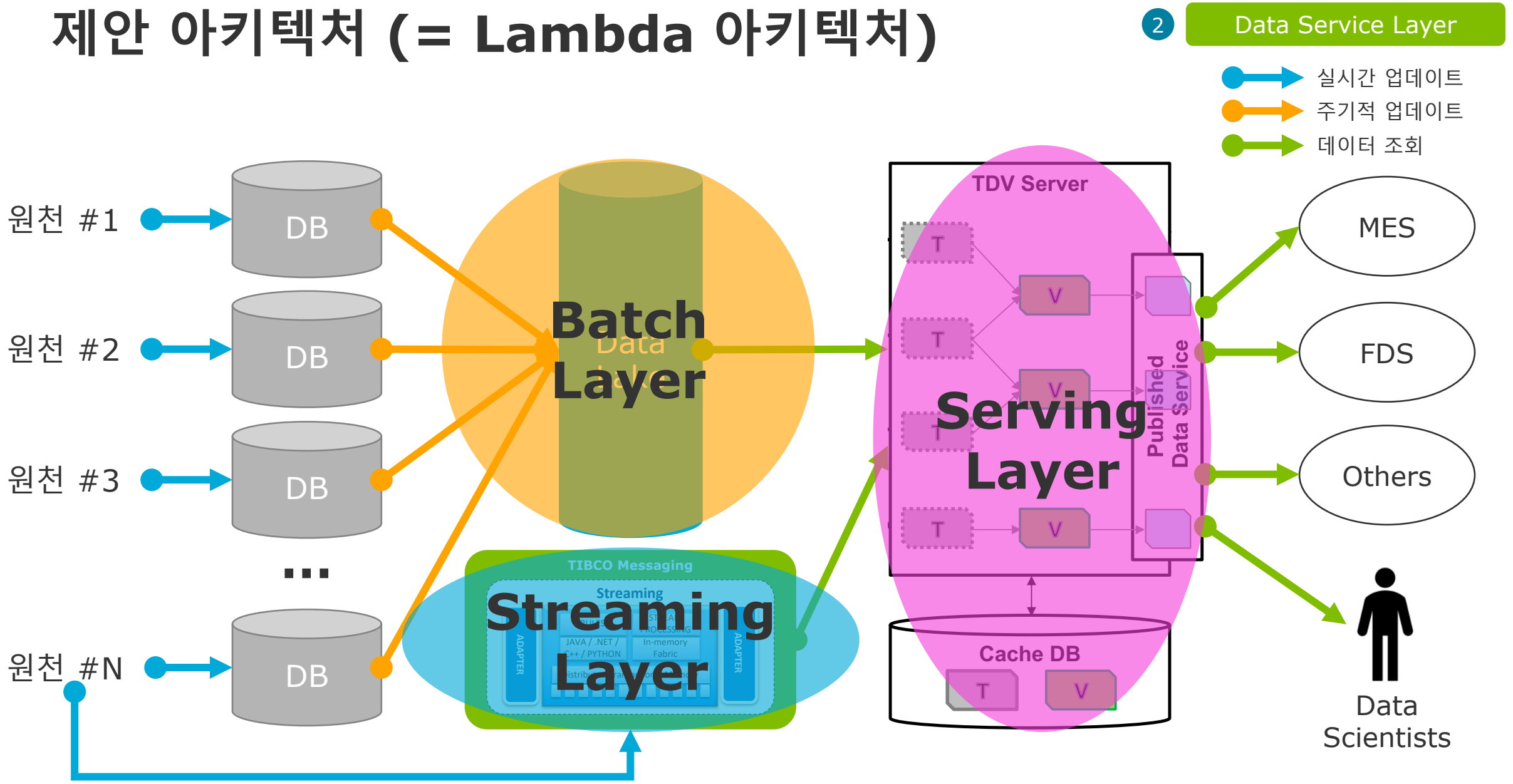
2 Data Service Layer



 © Copyright 2000-2021 TIBCO Software Inc.

TIBCO

제안 아키텍처 (= Lambda 아키텍처)



© Copyright 2000-2021 TIBCO Software Inc.

고급 데이터 분석

 © Copyright 2000-2021 TIBCO Software Inc.

Analytics Market Leader

S전자 전사 경영혁신 의사결정 지원 사례

[단독] 신한銀 국내 첫 '빅데이터 워룸'
경영정보 실시간 분석...금융권도 4차산업 광풍 매일경제 2017.01.02

신한은행이 국내 최초로 임원진이 경영 관련 빅데이터를 실시간으로 확인하면서 바로바로 의사결정을 내릴 수 있는 '빅데이터 워룸(war room)'을 만든다.

4차 산업혁명 시대를 맞아 가속화한 은행 간 무한 경쟁 속에서 살아남기 위해 870여 곳의 점포와 2400만명에 달하는 고객 정보를 총망라해 보여주는 종합상황실을 꾸리고 본격적인 빅데이터 기반 경영에 시동을 거는 것이다.

P&G
Business Sphere

P&G의 Business Sphere

전 세계 40개 지사에서
수집된 각종 시장상황의
빅데이터를 실시간 분석
→
임원들의 모바일기기 및
상황실 스크린에
일목요연하게 시각화

매주 전사 임원회의에서 빅데이터 분석을 통해 파악한
시장 동향을 공유하며 대응전략을 신속하게 결정

 © Copyright 2000-2021 TIBCO Software Inc.

TIBCO

TIBCO Hyper-Converged Analytics

TIBCO의 Predict 데이터 분석 플랫폼은 단순 시각화 및 리포팅에서부터 실시간 대쉬보드, 실시간 데이터 사이언스, 딥러닝, AutoML, ModelOps까지 데이터 분석의 모든 스펙트럼을 지원하며 이를 Hyper-Converged Analytics라고 합니다.

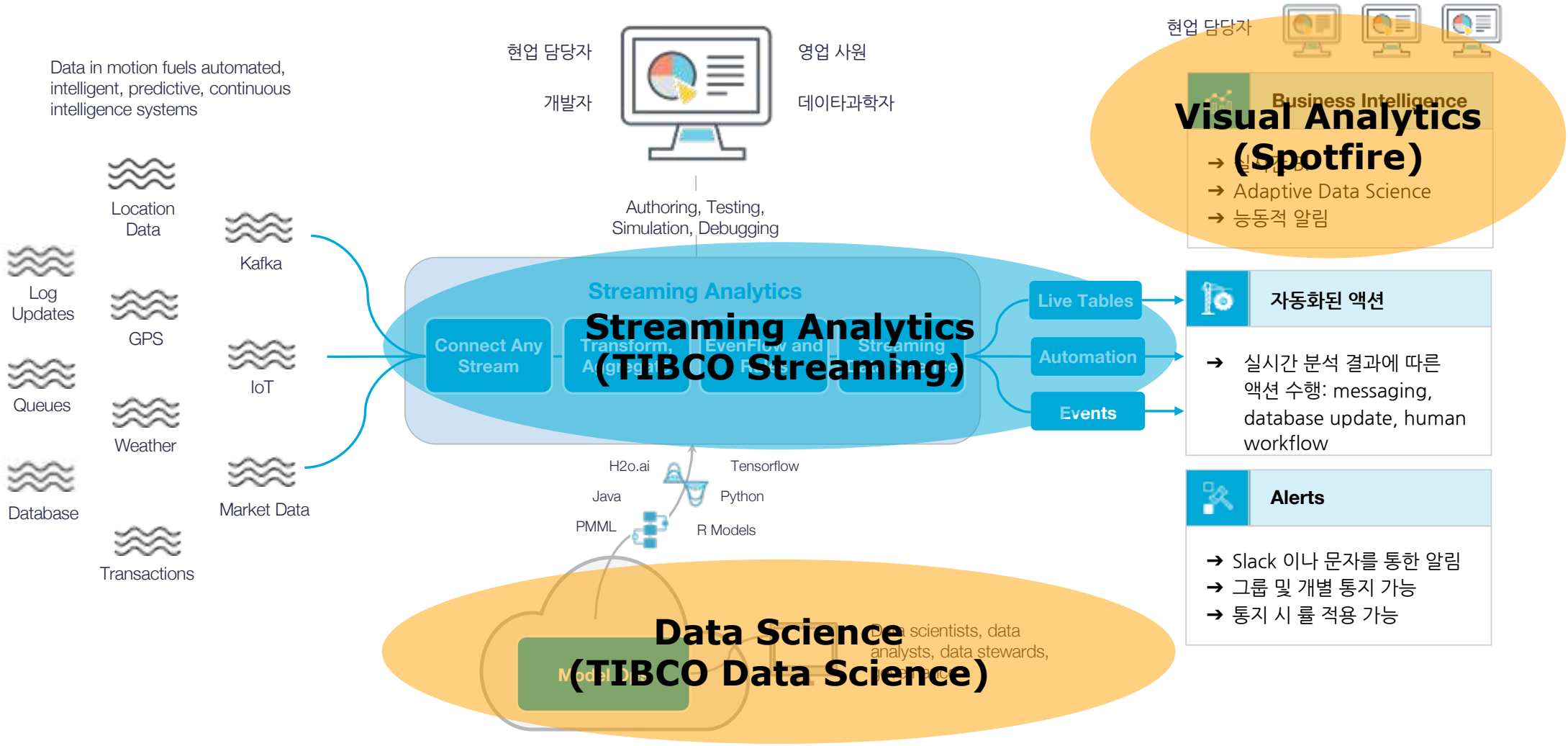
PREDICT

Visual Analytics	Data Science	Streaming Analytics
TIBCO Spotfire *Quickly generate high value business insights & predictions powered by AI & search* • Analytic Applications • Visual Augmented Analytics • Embedded Analytics	**TIBCO Data Science** *Operationalize and scale data science with AutoML and Model Ops* • AutoML & Explainability • Model Ops & Compliance • Collaboration	**TIBCO Streaming** *Run your business with real-time data-driven continuous intelligence* • Streaming Business Intelligence • Streaming Data Science • Continuous Intelligence Applications

 © Copyright 2000-2021 TIBCO Software Inc.

Hyper Converged Analytics Concept

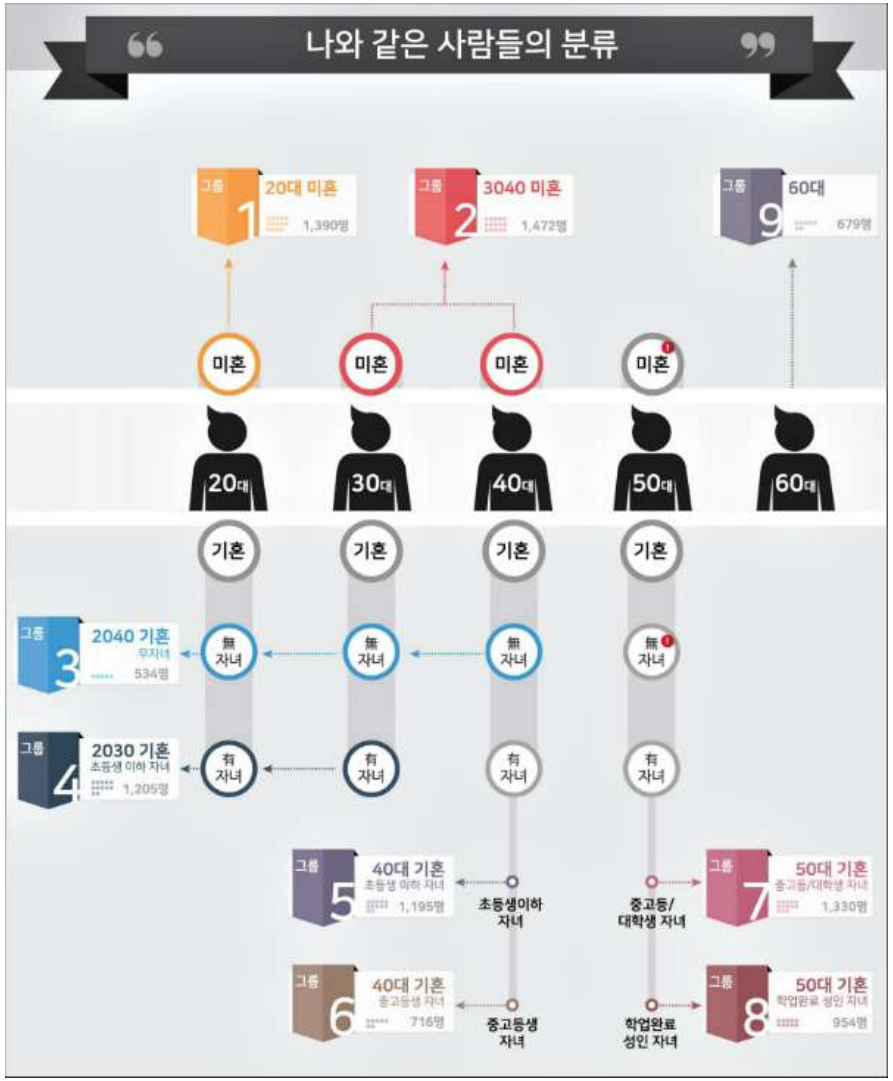
유입되는 데이터의 전처리를 위한 변환 / 필터링 / 통합 등 일반적인 이벤트 프로세싱을 위한 모든 기능은 물론, 데이터가 유입되는 시점에 즉각 반응하는 실시간 모니터링 및 분석, 머신러닝 모델을 적용한 실시간 예측 분석까지 가능합니다.



 © Copyright 2000-2021 TIBCO Software Inc.

VA System (since 2016)

신한은행



신한은행은 지난해 5월 빅데이터센터를 신설하고 '시각화 분석 시스템(VA, Visual Analytics System)'을 도입, 은행 내 있는 다양한 데이터를 분석해 영업현장에서 활용하고 있다.

지난 3월에도 고객 데이터 분석자료를 기반으로 '보통사람 금융생활 보고서'를 발간했으며 데이터 수집 및 분석을 고도화하는 등 단순 상품 추천에서 벗어나 고객 금융거래 분석자료를 기반으로 상담이 진행될 수 있도록 상담지원시스템을 준비하고 있다.

© Copyright 2000-2021 TIBCO Software Inc.

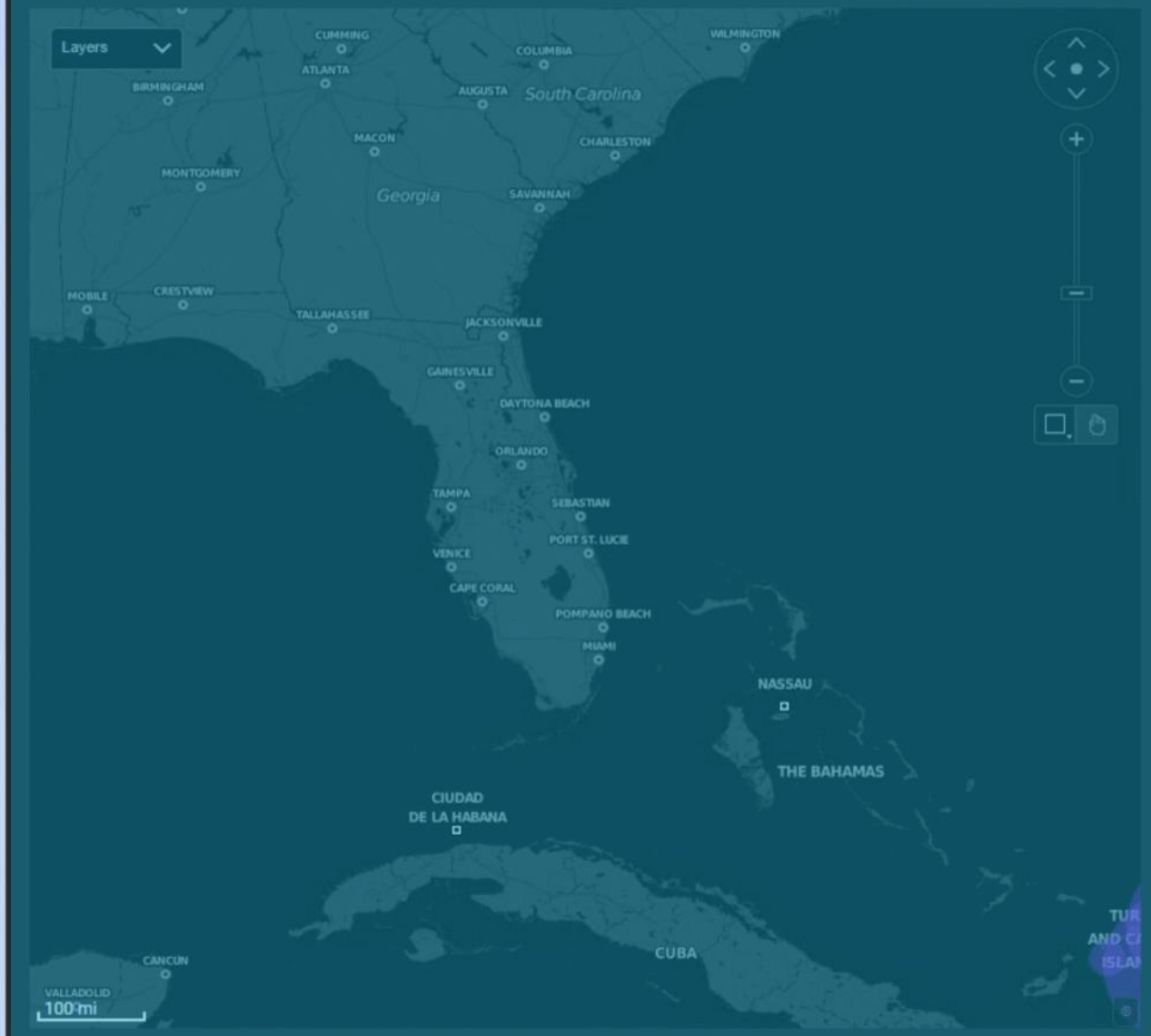
TIBCO

Property Loss Claims - Hurricane Irma

Refresh Data Table

Display: ClaimCount

Number of Claims	Claims Amount	Anomalies
0	(Empty)	0



Distribution by Location

Claims Received

Claims Details

Anomaly	ClaimNo	Claimstype	LossType	DateOfLoss	DateReceived

© Copyright 2000-2020 TIBCO Software Inc.

국내 관련 사례

 © Copyright 2000-2021 TIBCO Software Inc.

VA System (since 2016)

신한은행

2019 IDC Information Visionary Award

 © Copyright 2000-2021 TIBCO Software Inc.

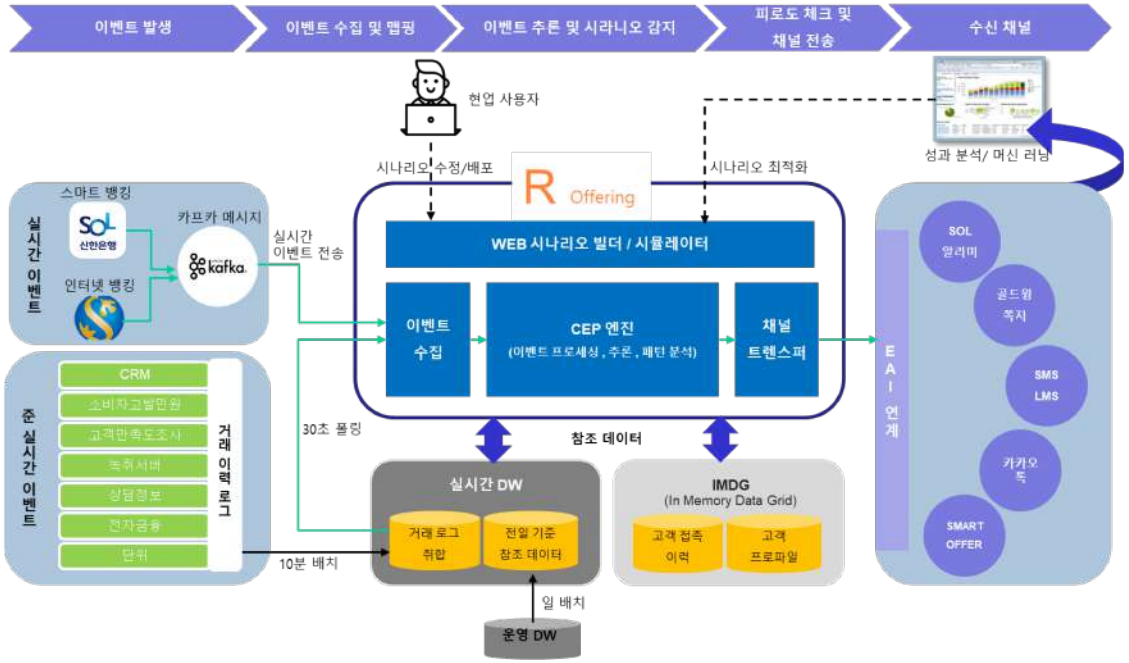
TIBCO

신한은행 이벤트 기반 추천/오퍼링

◇ 신한은행의 R-Offering 플랫폼 구축 프로젝트...옴니 경험 혁신(Omni Experience Innovator) 부문 수상

신한은행의 '실시간 제안 플랫폼(R-Offering)'은 2,500만 고객이 영업점, 콜센터, 모바일 앱, 인터넷 등 20개 이상의 대면/비대면 채널을 통해 발생시키는 접촉·거래 행태정보를 실시간으로 감지하여 고객니즈별 맞춤 상품·서비스 추천, 혜택정보 등을 최적의 채널로 실시간 전달해주는 시스템이다.

하루 평균 1억2천만건의 디지털 로그를 프로세스 마이닝 기법을 활용하여 분석하고, 고객행동패턴과 상황을 실시간 감지조건으로 자동화하여 고객관리부터 마케팅, 고객만족도 조사까지 은행업 전반의 서비스들을 '실시간', '디지털'로 전환하는 역할을 담당하고 있다. 현재 34개의 과제 수행결과, 기존 시스템 대비 고객 반응률 평균 4배, 상품/서비스 가입률 평균 14배 이상의 성과를 보이고 있다.



관심상품 등록율	자동해지고객 재가입율	대출상품 유도율
3배	2배	31배

평균 오퍼 시간	열람 / 응답률	고객 만족도 조사 비용절감
7분	3배	7억원

© Copyright 2000-2021 TIBCO Software Inc.

BD 1000 (since 2021)

신한은행

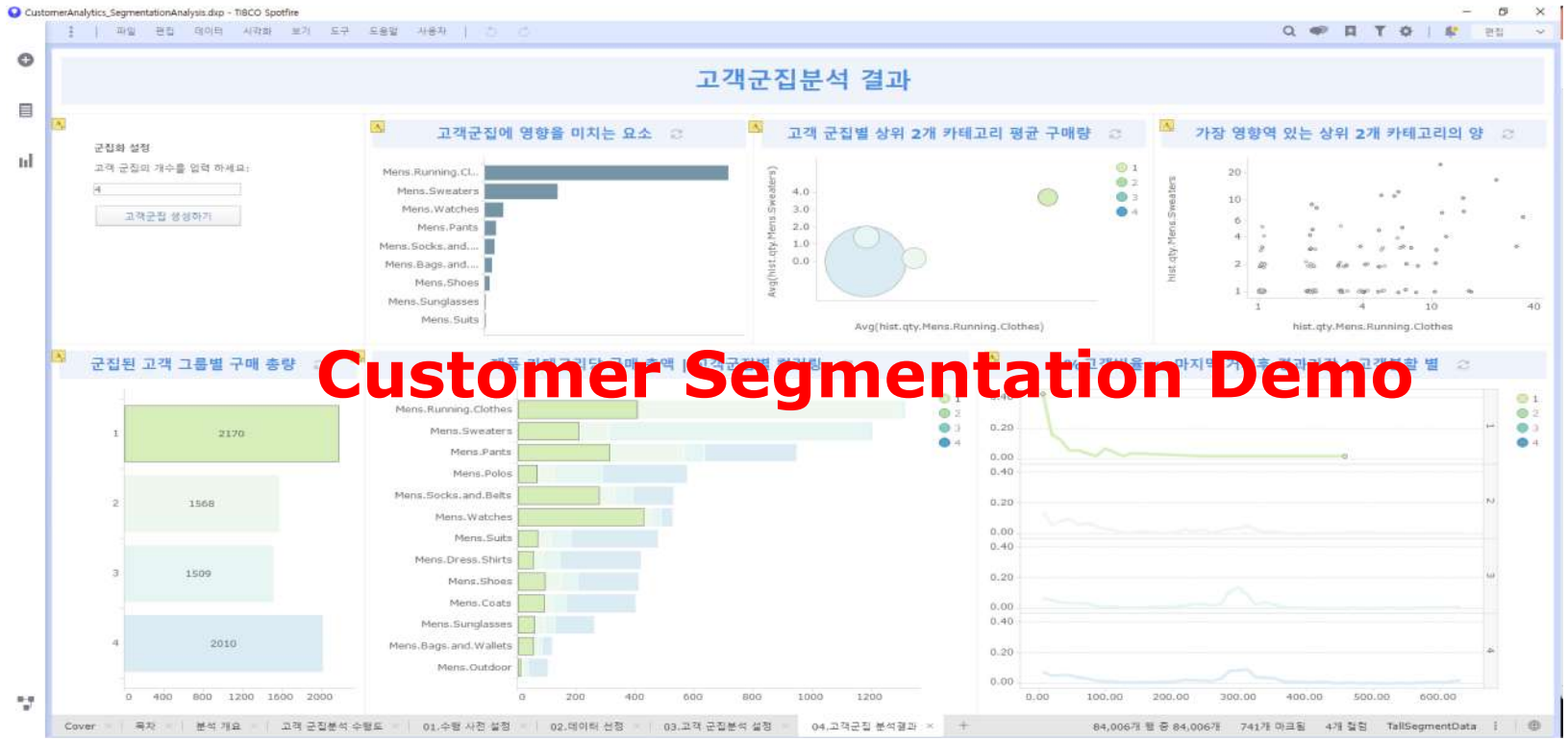
신한銀 빅데이터 전문가 1천명 키운다

신한 빅데이터 1000 프로젝트

현재 50명 수준 빅데이터 인력
20배로 키워 언택트금융 사활

본사 내 그룹·영업점마다 배치
고객관리·영업전략 수립 활용

- 각 지사 및 다양한 부서의 Self-service analytics
- CoE 기반의 중앙 집중형 분석 조직 구성
- Spotfire
  → 신한은행의 핵심 분석 플랫폼
- 현재 전 직원 대상 Spotfire 교육 진행 중



© Copyright 2000-2021 TIBCO Software Inc.

TIBCO

BD 1000 (since 2021)

신한은행

Analytics Specialist

Business User

Develop & Update

Analytics
Center of Excellence

Distribute

Feedback

Sales Strategy

Field Office

Customer Care

Marketing

 © Copyright 2000-2021 TIBCO Software Inc.

TIBCO

Takeaway : Spotfire = Analytics

신한은행

Land & Expand

VA	BD1000
• Center-piece of analytic activities	• Company-wide analytic culture
• Internalized analytic capability	• CoE-driven analytic organization
• Familiarized user base	• Grassroot data-driven decision making
• Tangible outcome and performance	• Company-wide enablement on-going
• Heightened positive perception	• Data science at CoE level

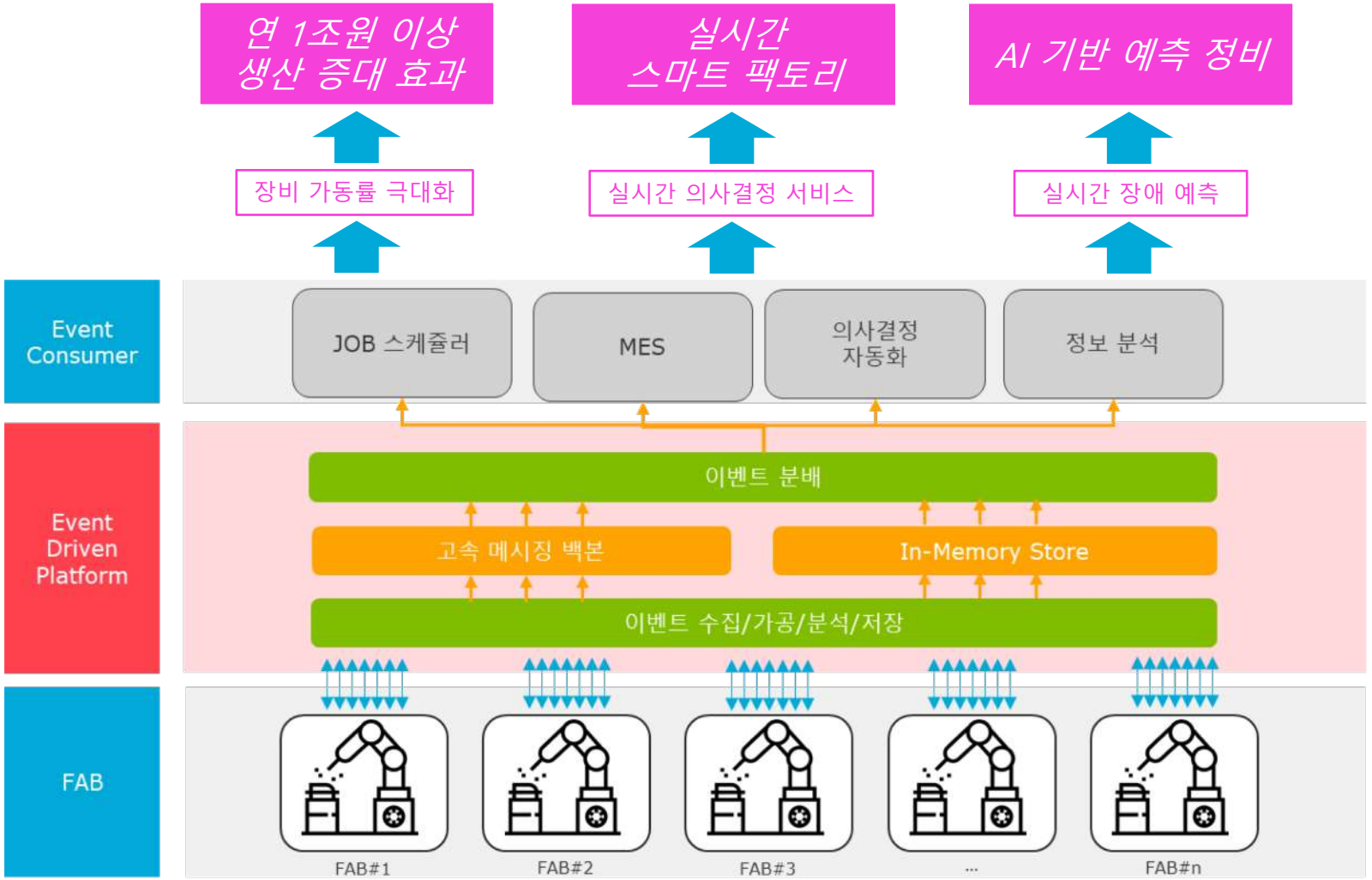
Shinhan and TIBCO Go Together!!

© Copyright 2000-2021 TIBCO Software Inc.

TIBCO

Event-Driven Platform

수많은 장비에서 발생하는 Event를 실시간으로 수집/가공/집계/분석/저장하는 Event 기반 플랫폼을 구축하였고 그 정보를 필요로 하는 시스템에 실시간으로 제공하여 Factory Full Automation 기반으로 활용되고 있습니다.



실시간 이벤트 데이터 활용

이벤트 데이터 수집 / 처리 / 분배 (Event-Driven Platform)

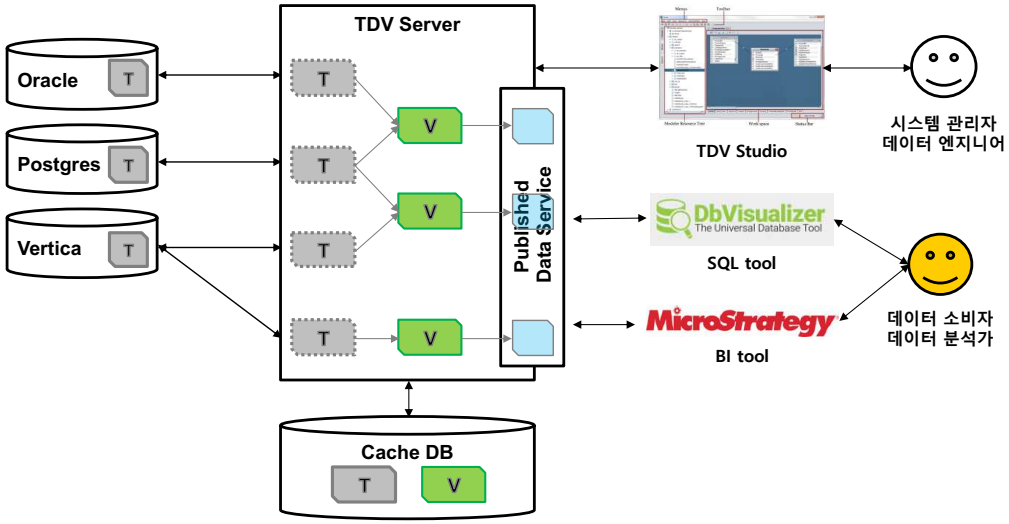
이벤트 데이터 생성

© Copyright 2000-2021 TIBCO Software Inc.

전사 데이터 자원에 대한 효율적 활용 및 거버넌스를 위한 DG&V 플랫폼

K사는 전사적인 데이터 사이언스 역량을 한 단계 향상시키기 위해 대대적인 데이터 아키텍쳐 업그레이드를 진행했습니다.
10가지 이상의 이기종 데이터 소스들로부터 빠르고 손쉬우면서도 암호화된 안전한 데이터 엑세스를 위한 DG&V (Data Governance & Virtualization) 플랫폼을 TIBCO Data Virtualization으로 구축하기로 하였습니다.

➔ **10개 이상, 500TB 이상의 이기종 데이터 소스 통합**



국내 통신사

K*

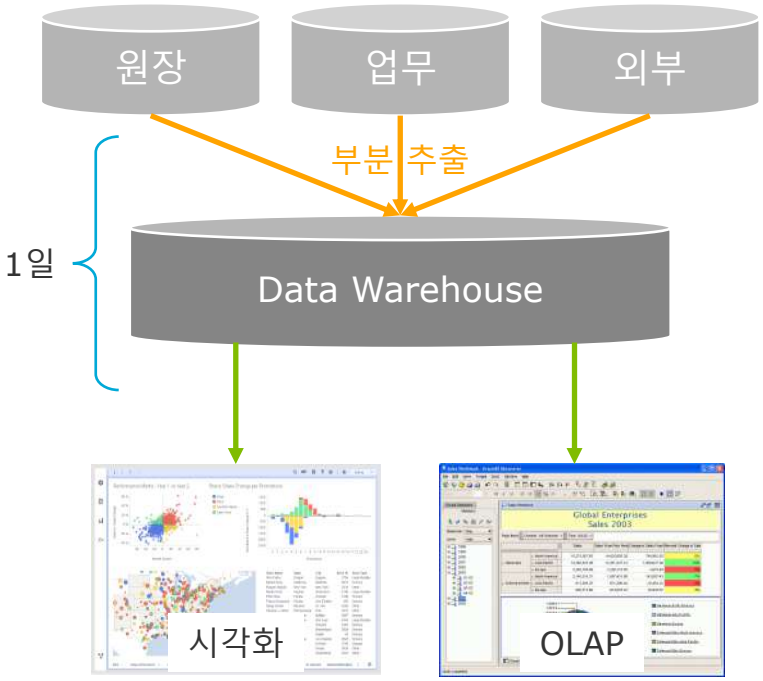
Benefits

- ➔ 빠른 iteration으로 데이터 분석 효용성 향상
- ➔ 데이터 카탈로그 확보
- ➔ 데이터 통합 비용 및 자원 절감
- ➔ 관리자 업무 생산성 향상
- ➔ Data as a Service 구현
- ➔ 데이터 자원에 대한 메타 데이터 관리, 접근제어, 민감정보 통제, 암호화 등 거버넌스 체계 확보

© Copyright 2000-2021 TIBCO Software Inc.

Wrap-up : TIBCO Agile Data Fabric

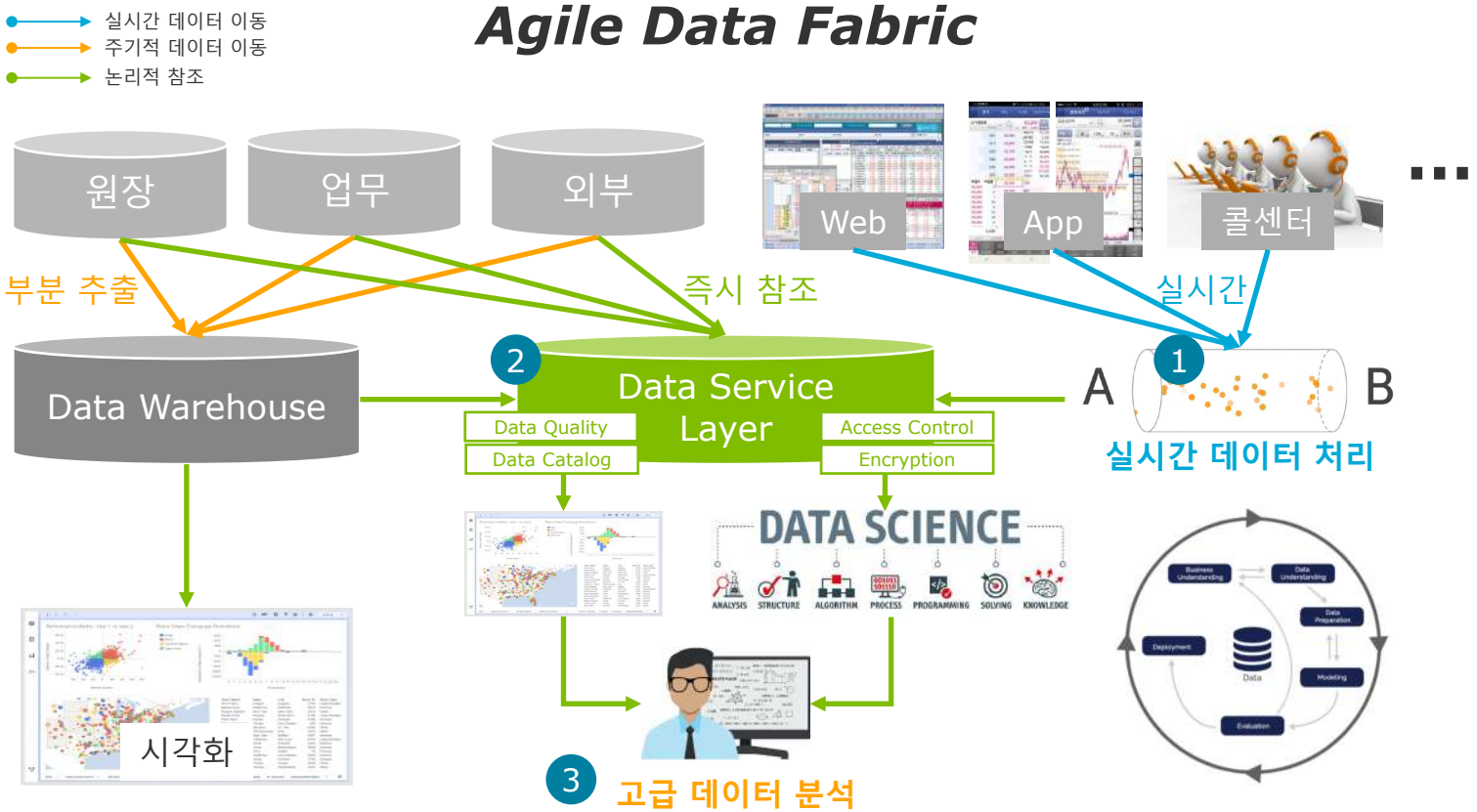
과거



- 직관적이고 아름다운 화면 구성
- 빠르게 더 빠르게!
- 추출된 과거 데이터만 활용/분석

데이터 분석 < 데이터 시각화

Agile Data Fabric



- 가용한 데이터 자원 최대 활용
- 데이터 지연 시간 최소화
- 빠르고, 자유로운 데이터 접근

- 데이터 사이언스에 적합 = AI에 적합
- 가설 → 데이터로 검증 → 해결책
- 빠른 반복으로 검증과 개발 용이

다양한 대용량 데이터에 대한 자유로운 접근 보장 → Built for AI

© Copyright 2000-2021 TIBCO Software Inc.

Wrap-up

Real-time Smart Enterprise

- **TIBCO : Agile Data Fabric for AI**
- **실시간 데이터 파이프라인**
  - Data-in-motion 활용 = Fast fish
  - 이벤트 발생 시점에서 최종 소비까지 소요되는 지연 시간 최소화
  - 실시간 데이터 처리를 위한 end-to-end 솔루션
  - Event processing을 위한 total solution

비즈니스 실시간 가시성 확보

- **데이터 서비스 레이어**
  - 분산 고립된 데이터 원천에 대한 쉽고 빠르고 안전한 접근
  - 솔루션 : 물리적 이동이 아닌 논리적 참조에 기반한 데이터 아키텍처
  - 핵심 데이터 자원에 대한 정합성 보장

손쉬운 데이터 연계 및 통합

- **고급 데이터 분석**
  - 단순 시각화, 실시간 분석, 데이터 사이언스까지 융합한 분석 플랫폼
  - 데이터 분석에 필요한 모든 활동 지원

AI를 활용한 예측 및 자동화

TIBCO는 데이터 기반 혁신에 필요한 핵심 솔루션들을 보유하고 있으며, 다수의 국내외 사례를 통해 구축 경험과 역량까지 갖춘 최고의 파트너입니다.

© Copyright 2000-2021 TIBCO Software Inc.

© Copyright 2000-2021 TIBCO Software Inc.