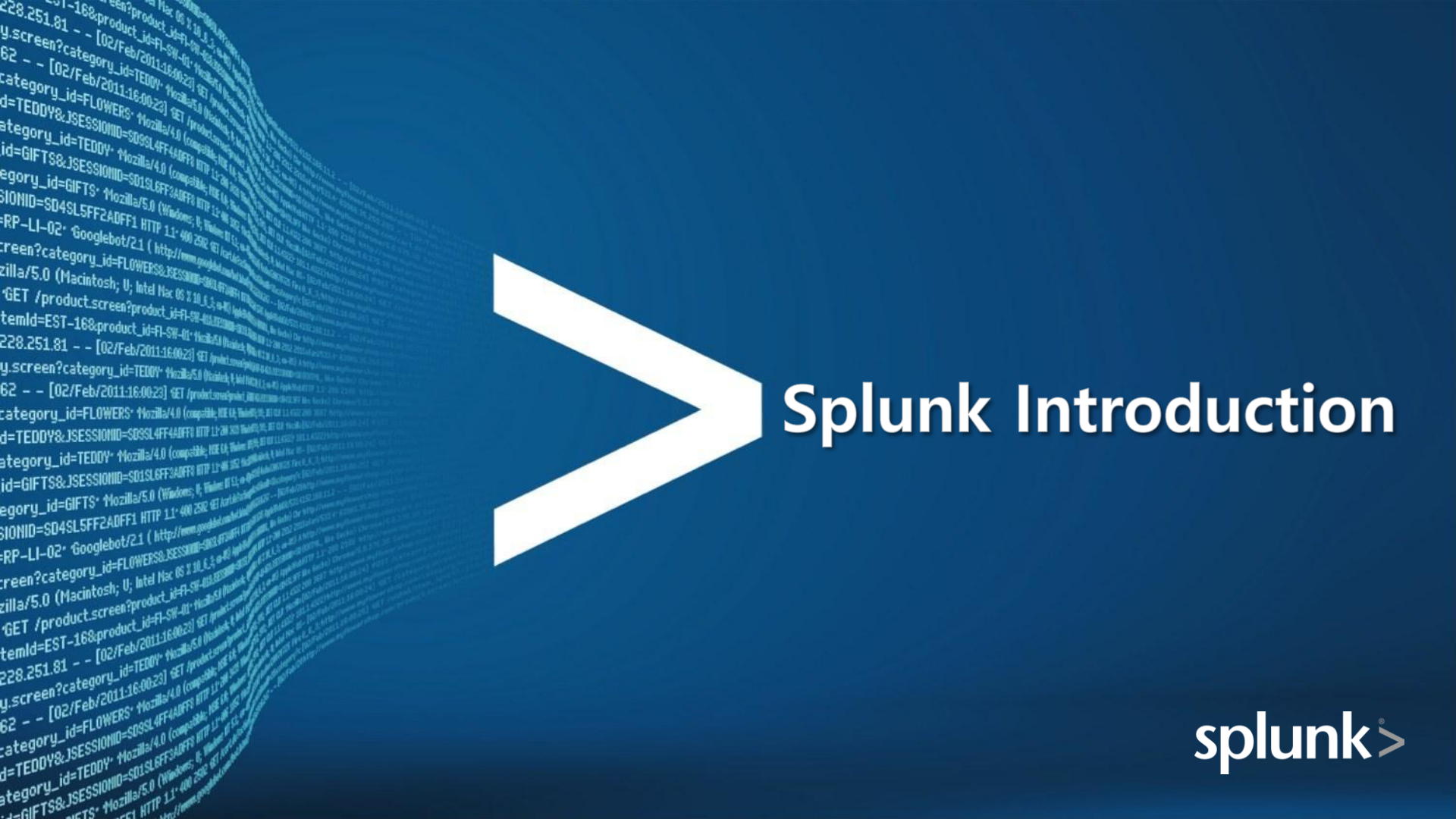

Splunk Introduction
splunk>

스플렁크 개요

회사

- 글로벌 HQs:
  - 샌프랜시스코(AMER)
  - 런던(EMEA)
  - 홍콩(APAC)
- 직원수 **1,800+**명
- 연매출 약 **8,000**억원 (YoY +49%)
- 나스닥 상장 : SPLK

제품

- 무료 버전에서 시작, 대규모 분산 확장
- 스플렁크 제품:
  - **Splunk Enterprise**
  - Splunk Cloud
  - Hunk
  - Splunk Light
  - Splunk MINT
  - Premium Apps

고객

- 고객사(전체): **11,000+**
- 고객사(한국): 350+
- 국가기준: 100개국+
- 중소기업, 대기업 그룹 계열사
- 포춘 100대 기업: 80+
- 최대 라이센스:
  - **1.3+ PB**/일

머신에서 발생하는 빅데이터...

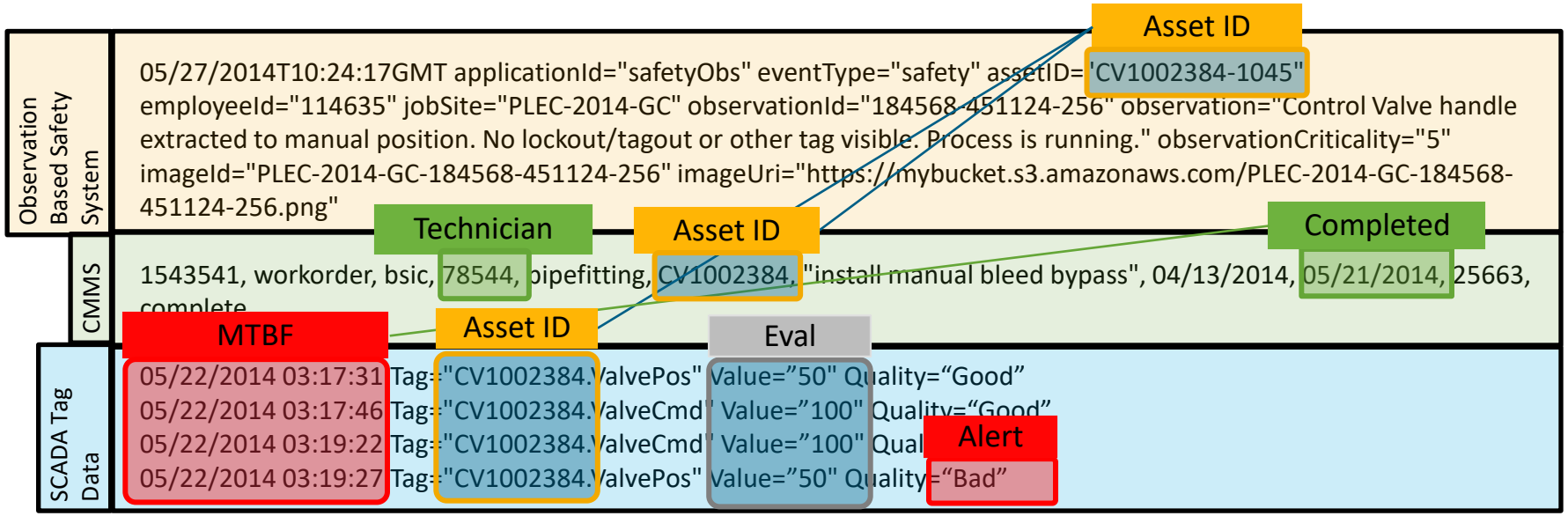
Volume | Velocity | Variety | Variability

GPS,
RFID,
웹 서버,
하이퍼바이저,
이메일, 메세징,
클릭스트림, 모바일,
전화, IVR, 데이터베이스,
서버, 보안기기, 데스크탑,
센서, 컴퓨터 통신, 스토리지,

Industrial Data Is Machine Data

Observation Based Safety System	05/27/2014T10:24:17GMT applicationId="safetyObs" eventType="safety" assetID="CV1002384-1045" employeeId="114635" jobSite="PLEC-2014-GC" observationId="184568-451124-256" observation="Control Valve handle extracted to manual position. No lockout/tagout or other tag visible. Process is running." observationCriticality="5" imageId="PLEC-2014-GC-184568-451124-256" imageUri="https://mybucket.s3.amazonaws.com/PLEC-2014-GC-184568-451124-256.png"
CMMS	1543541, workorder, bsic, 78544, pipefitting, CV1002384, "install manual bleed bypass", 04/13/2014, 05/21/2014, 25663, complete
SCADA Tag Data	05/22/2014 03:17:31 Tag="CV1002384.ValvePos" Value=”50" Quality=“Good” 05/22/2014 03:17:46 Tag="CV1002384.ValveCmd" Value=”100" Quality=“Good” 05/22/2014 03:19:22 Tag="CV1002384.ValveCmd" Value=”100" Quality=“Good” 05/22/2014 03:19:27 Tag="CV1002384.ValvePos" Value=”50" Quality=“Bad”

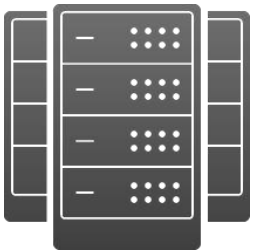
Connect the Dots – Make New Discoveries



산업용 데이터 분석의 어려움

Getting Insights Is Not Easy

Legacy systems limit high-velocity data collection



Analytics constrained by inflexible tools

Data silos prevent insights across the organization

Lack of platform capabilities restricts customization

스플렁크란?

splunk> = 머신데이터 플랫폼

"머신데이터를 아무런 제약 없이 수집>저장>분석>시각화 할 수 있는 실시간 분산 플랫폼"

머신데이터
(Machine Data)

- 서버/NW 로그
- 각종 설비 데이터
- 애플리케이션 로그
- 기타 모든 텍스트 형태의 데이터

제약 없음
(No Limits)

- 비정형/정형 데이터
- 데이터 포맷 무관
- 데이터 용량 무관
- 데이터 속도 무관
- 제약 없이 수용

엔드 투 엔드
(End-to-End)

- 별도의 외부 솔루션불필요
- 복잡한 코딩 및 SI 개발이 필요 없음
- 데이터의 생성부터 가치 획득까지 모두

실시간 및 분산
(Real-time)

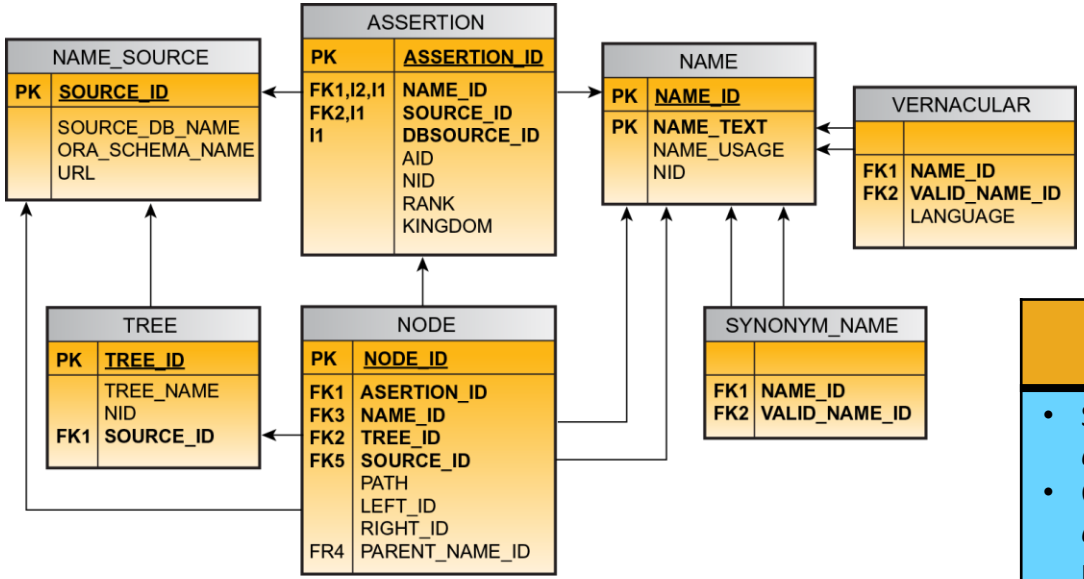
- 모든 데이터 실시간 처리, 즉시 결과 확인
- 분산 저장, 분산 검색
- 성능 및 용량의 선형적 확장

플랫폼
(Platform)

- 커스터마이징 용이
- 외부 시스템과 손쉬운 연동
- 앱을 통한 기능 확장
- 개발 프레임워크

Early vs. Late Structure Binding

Early Structure Binding – 전통적인 분석방법

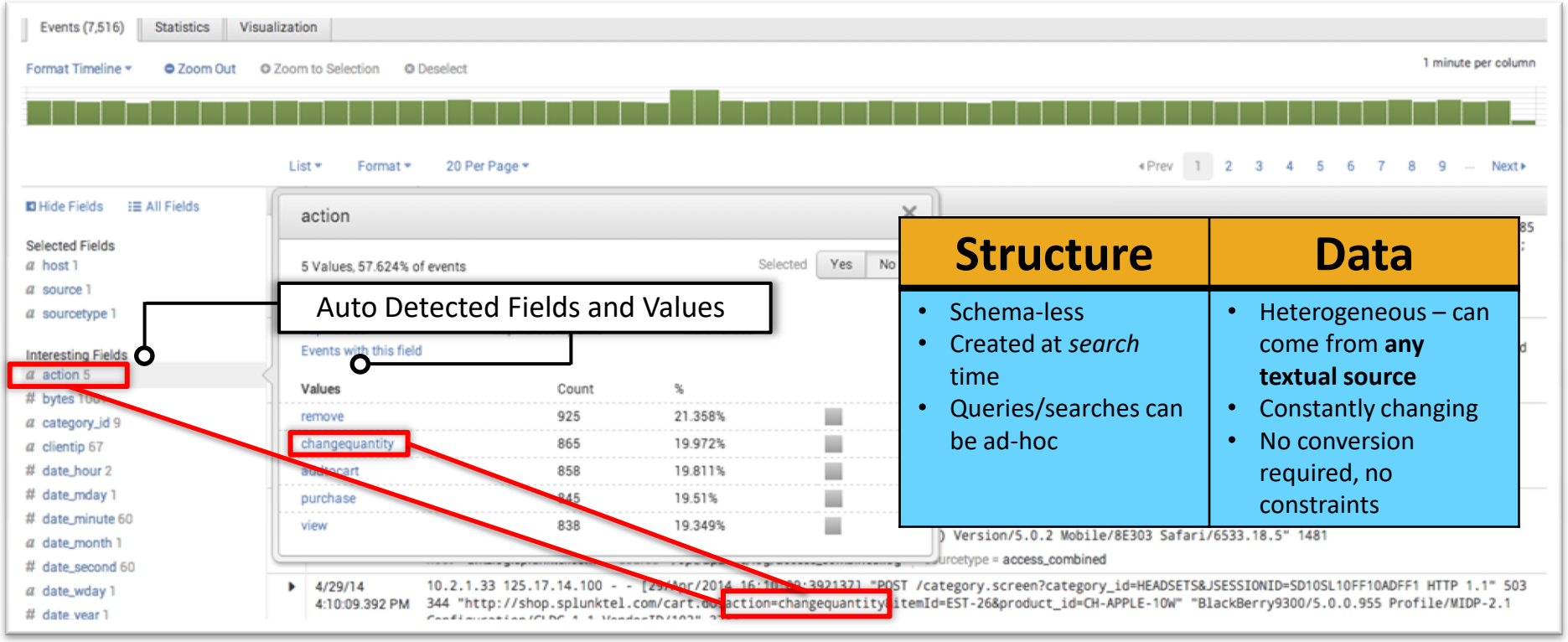


Structure	Data
• Schema – created at *design time* • Queries – understood at *design time* for maximum performance	• Homogeneous – must fit into tables or be converted to fit into tables • Must exactly match constraints

SELECT customers.* FROM customers WHERE customers.customer_id NOT IN(SELECT customer_id FROM orders WHERE year(orders.order_date) = 2004)

Late Structure Binding: Schema On The Fly

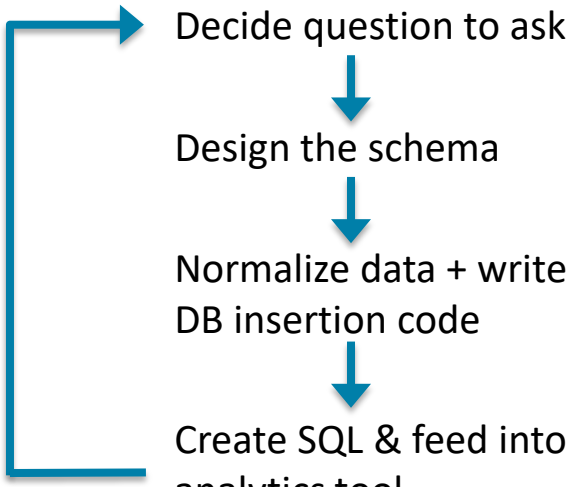
데이터 정규화 필요 없는 실시간 분석



Structure	Data
• Schema-less • Created at *search* time • Queries/searches can be ad-hoc	• Heterogeneous – can come from **any textual source** • Constantly changing • No conversion required, no constraints

빠른 데이터 인사이트 확보

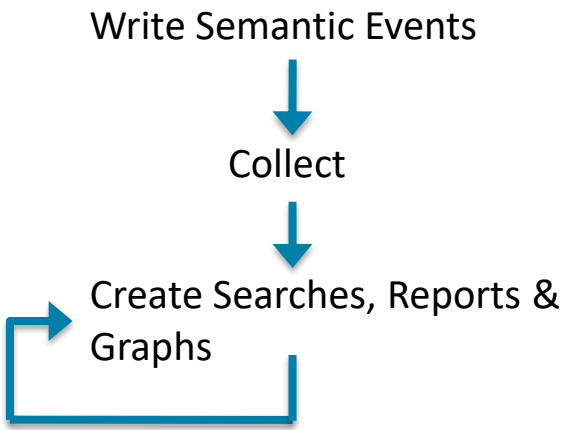
Early Structure Binding



- *Days – Weeks – Months*
- *Destructive*

Late Structure Binding

splunk>



- *Minutes*
- *Non-Destructive*

정형 데이터로 산업용 센서 데이터 연관분석

ICS Tag Data

Asset ID

Tag

9/8/15 4:41:48.055 PM 2015-09-08 23:41:48.055 +0000 Tag="Windfarm_10.Turbine_10.Wind_Direction" Value="132.959152"
AssetID="K23441gF4224" Quality="good" demo=Windfarm
host = 127.0.0.1 source = tcp:9997 sourcetype = opc 9/8/15 4:41:48.055 PM 2015-09-08 23:41:48.055 +0000
Tag="Windfarm_10.Turbine_10.Temperature" Value="19.3928394" Quality="good" demo=Windfarm host = 10.7.102.1 source = tcp:9997 sourcetype
= opc 9/8/15 4:41:48.055 PM 2015-09-08 23:41:48.055 +0000 Tag="Windfarm_10.Turbine_10.Stator_Oil_Temperature"
Value="85.4567337" Quality="good" demo=Windfarm
host = 127.0.0.1 source = tc
9/8/15

Host

Tag Value

Tag Quality

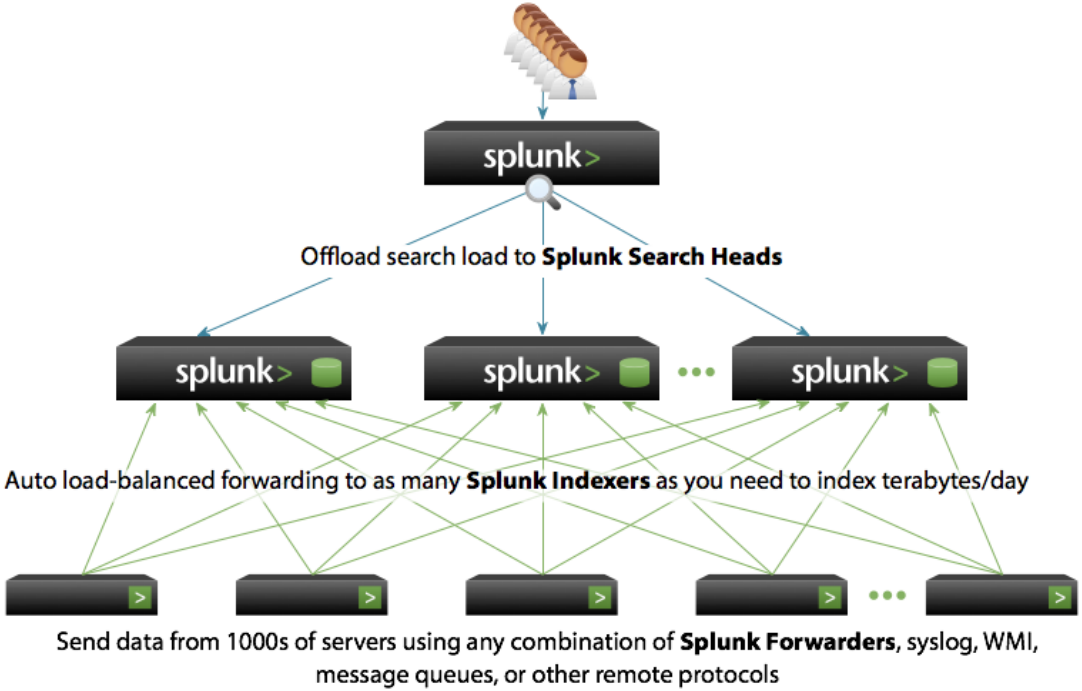
Workorder, Asset Databases

Asset ID	Technician	Date Serviced	Part Number	Lot Number
✓	50446	9/7/15	1224-56-A	B00747

Asset ID	Location
✓	Site 7

Location	Latitude	Longitude	Site ID	Address Line 1
Site 7	39.11515	84.45651	A345	409 Park St.

확장성 있는 분산 저장/처리구조

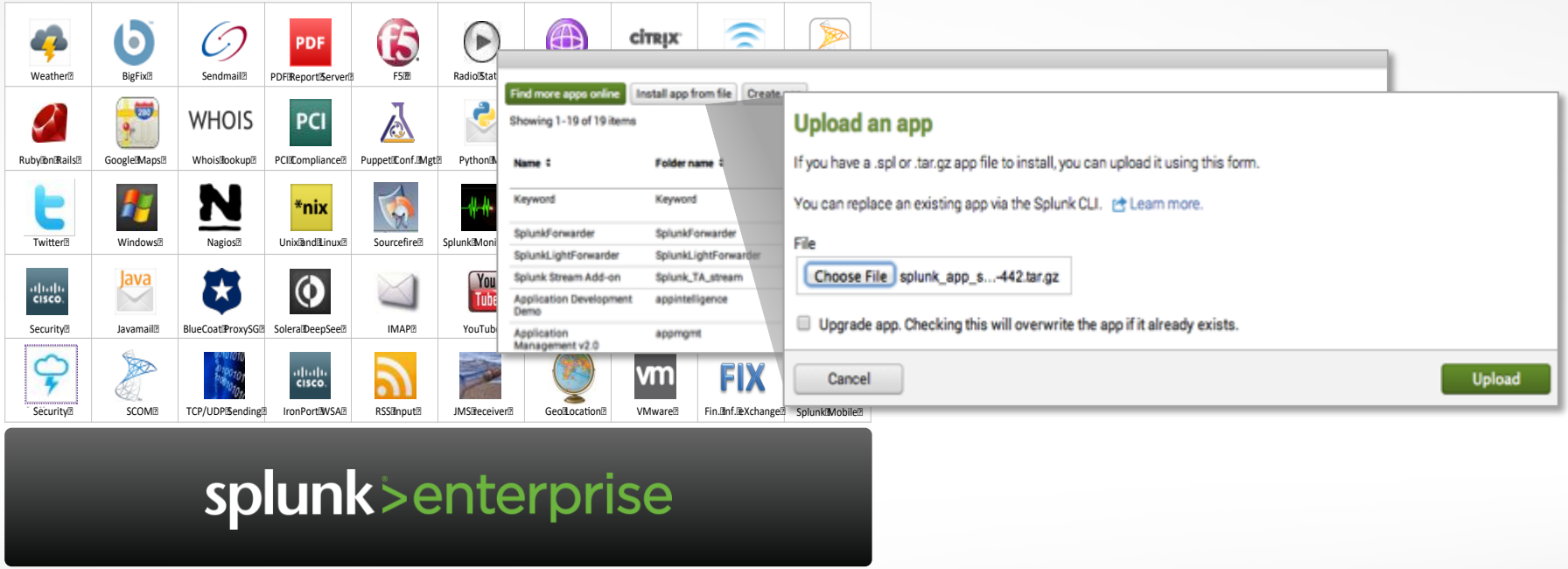


다양한 App 생태계를 사용한 빠른 구축

스플렁크 고객은 다양한 장비나 유즈케이스에 대한 기설정된 앱생태계를 활용하여 빠르게 쉽게 요구사항을 구현할 수 있습니다.

http://apps.splunk.com

1000+ 이상의 다양한 앱 생태계 구축



splunk>enterprise

파트너 에코시스템

Advanced Analytics and ML

IoT and ICS Security

Custom User Interfaces

splunk>

Ingest and Platforms

Services and Delivery

스플렁크 포트폴리오

스플렁크 프리미엄 솔루션

- Splunk IT Service Intelligence™
- Splunk Enterprise Security™
- Splunk User Behavior Analytics™

강력한 앱/애드온 생태계

Microsoft .net, Symantec, cisco, salesforce.com, Kepware, NetApp, Linux

splunk>enterprise | splunk>cloud | splunk>light | Hunk

splunk> 운영 인텔리전스(OI)를 위한 머신데이터 플랫폼

포워더 | Syslog/TCP | 모바일 | IoT 단말기 | 네트워크 패킷데이터 | 하둡 | 관계형 DB | 메인프레임 데이터

100여개국 의 11,000이상의 고객 보유

Fortune 100 회사중 80개 회사가 스플렁크를 사용

Cloud and Online Services

Education

Energy and Utilities

Financial Services and Insurance

Government

Healthcare

Manufacturing

Media

Retail

Technology

Telecommunications

Travel and Leisure

splunk>

기술과 사업 전반에 걸친 가치 전달

국내 사례 - 통합 관제 대시보드

S사

K사

H사

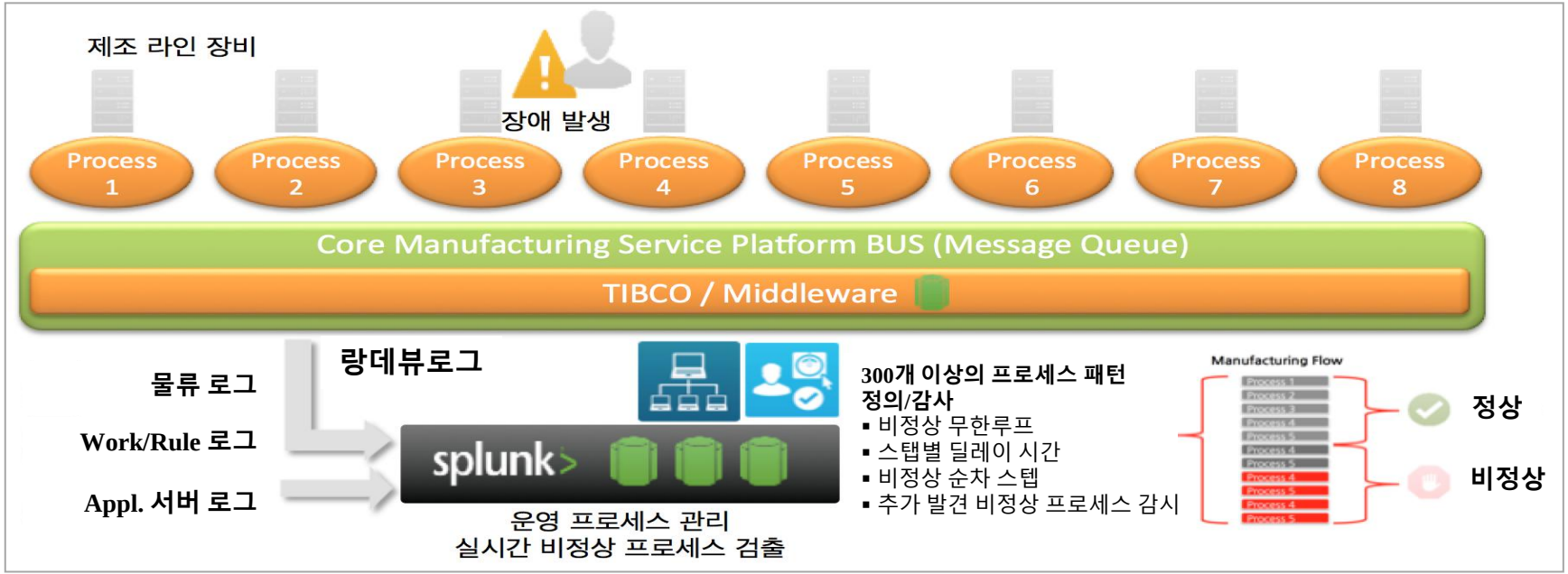
Measuring Results in Real Time

Printer Manufacturer

Splunk Use Cases

- Collect final printer configuration device data
- Monitor testing and configuration of printing devices
- Gain insights into product performance and production

국내사례: A 반도체 제조 운영 공정 실시간 관제/경보 시스템 구축 → 장애 대응 속도 향상으로 장비 가동율 증가 및 반도체 수율 향상



“ 장애 대응속도 향상 → 생산성 향상 ”

- 공정 프로세스들의 정상/비정상 구분화
- 기존에 파악하지 못한 비정상 프로세스들의 발견
  해당 프로세스에 대한 감지 룰 생성
  → 장애 선대응체제 구축
- 장애 탐지시간 감소 (10~30여분 → 1~2분)

“ 실시간 통합 품질 공정관리로의 확산 ”

- 적용 대상 확대 중 (라인 내 서버 → 전 서버 및 PC)
- 머신러닝 기반의 상관분석의 자동화
- 단위작업별 품질측정 모니터링
  → 실시간 통합 품질 공정관리 구현

A전자 사례 : 실시간 제조 공정 관제 확대

- **S사는 MOS 실시간 공정 모니터링의 성공을 바탕으로 실시간 통합 품질 공정관리를 목표로 스플렁크의 활용도를 확대하고 있음**

- 랑데뷰 로그
- 각종 설정 파일 변경
- 물류 로그
- Rule 로그
- Work 로그
- EES 어플리케이션 로그
- YMS 어플리케이션 로그
- 공정 PC 성능 로그

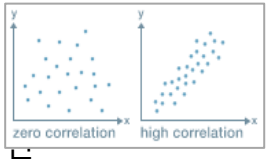
splunk>

반도체 전 공정의 **실시간 관제 및 알람**

→각 공정 세부 레시피 별 분석
→장애 대응 시간 감소 및 가동률 증가
→장애 선 대응 체제 구축

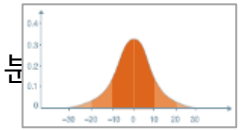
장애에 대한 인과 관계분석

장애에 영향을 미치는 설정 파일 변경, 센서 데이터 연관성 분석
→장애 재발 방지 및 모니터링 체계 개선



변수별 품질 영향도 분석 (TO-BE)

각종 센서 측정값들과 결과 품질의 상관 분

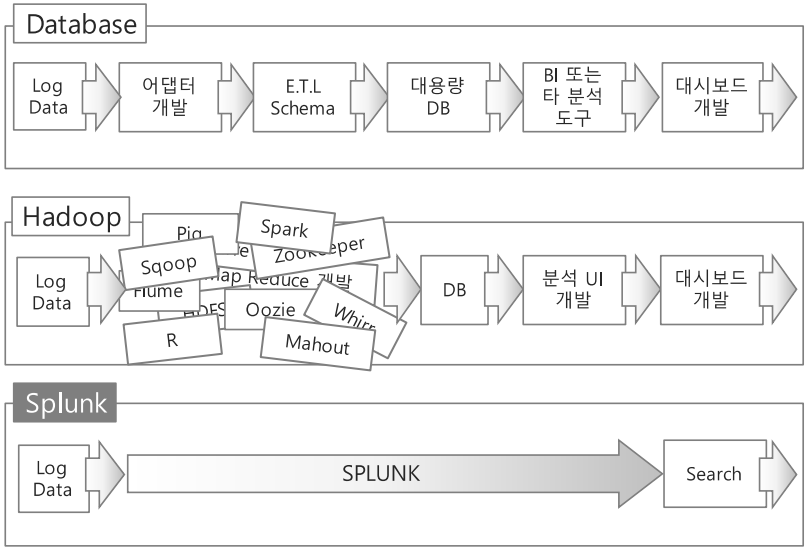


기존 기술 대비 차별점

- **원시 데이터 수집을 위한 <u>어댑터 개발이나 정제를 위한 별도의 추출 작업 없이 원시 데이터에서 즉각적인 분석</u>**
- **손쉬운 데이터 수집 및 무한한 확장성, 검색 기반의 손쉬운 분석 환경을 이용하여 효과적으로 개선**

원본 데이터

ex>
CDR
Network Packets
Database
Hypervisor
Telematics
Servers
Applications
Clickstreams
Custom log files
Manuf. Sensors
JMX
SNMP
Syslog
JMS
...



Business Value

ex>
데이터 시각화
로그 관리
이벤트 검색
실시간 운영 모니터링
장애 원인 분석
운영 인텔리전스
Capacity Planning
보안 관제
보안 사고 분석
APT 탐지
고객 분석
서비스 분석
마케팅 효과 분석
...

A 전자 : Why Splunk?

- **S사 구축 PoC 결과 경쟁사 대비 스플렁크는 구현 편의성, 가격, 구축 후 신규 요구사항 대응 및 유지 보수 등에서 우월함이 입증**

PoC 시나리오

운영 시스템의 예비 로그 서버에서 daily 300GB이상 발생하는 랑데뷰 로그를 실시간 수집, 필드 추출. 각 공정 레시피 단계별 소요 시간 분석 대시보드 및 알람 구성.

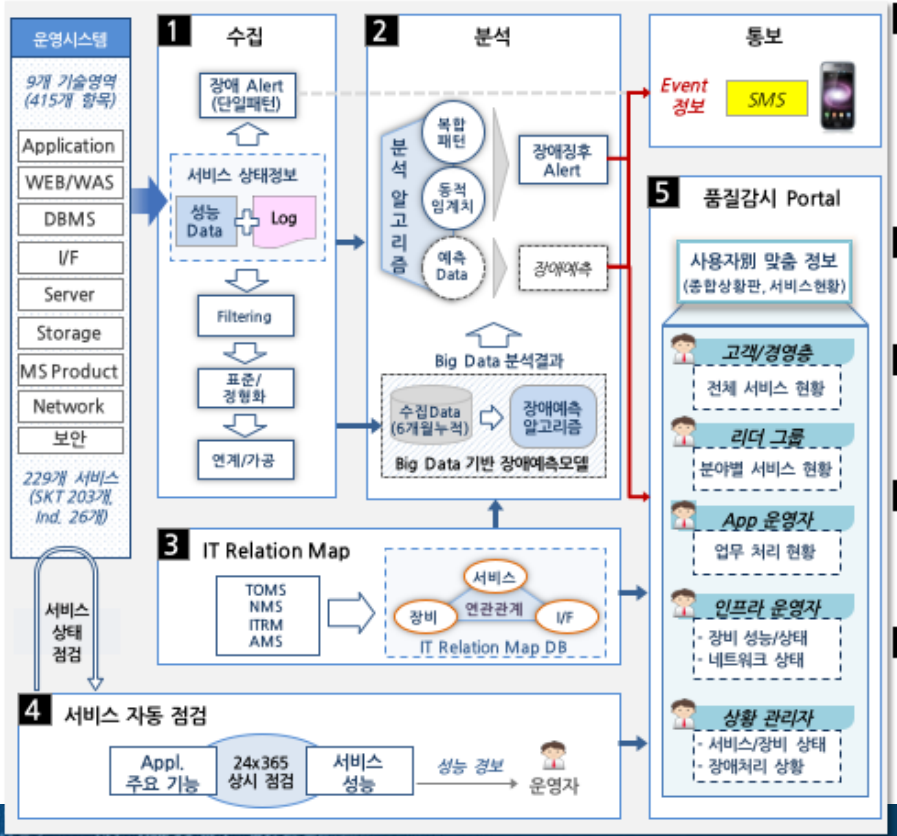
PoC 평가 기준

- 수집/저장/쿼리 성능
- 데이터 정합성
- 대용량 데이터 관리 기능 요건 충족
- 노드 장애 시 클러스터 안정성
- 분석 쿼리 및 대시보드 구현 유연성
- 요구 사항 별 예상 소요 M/M

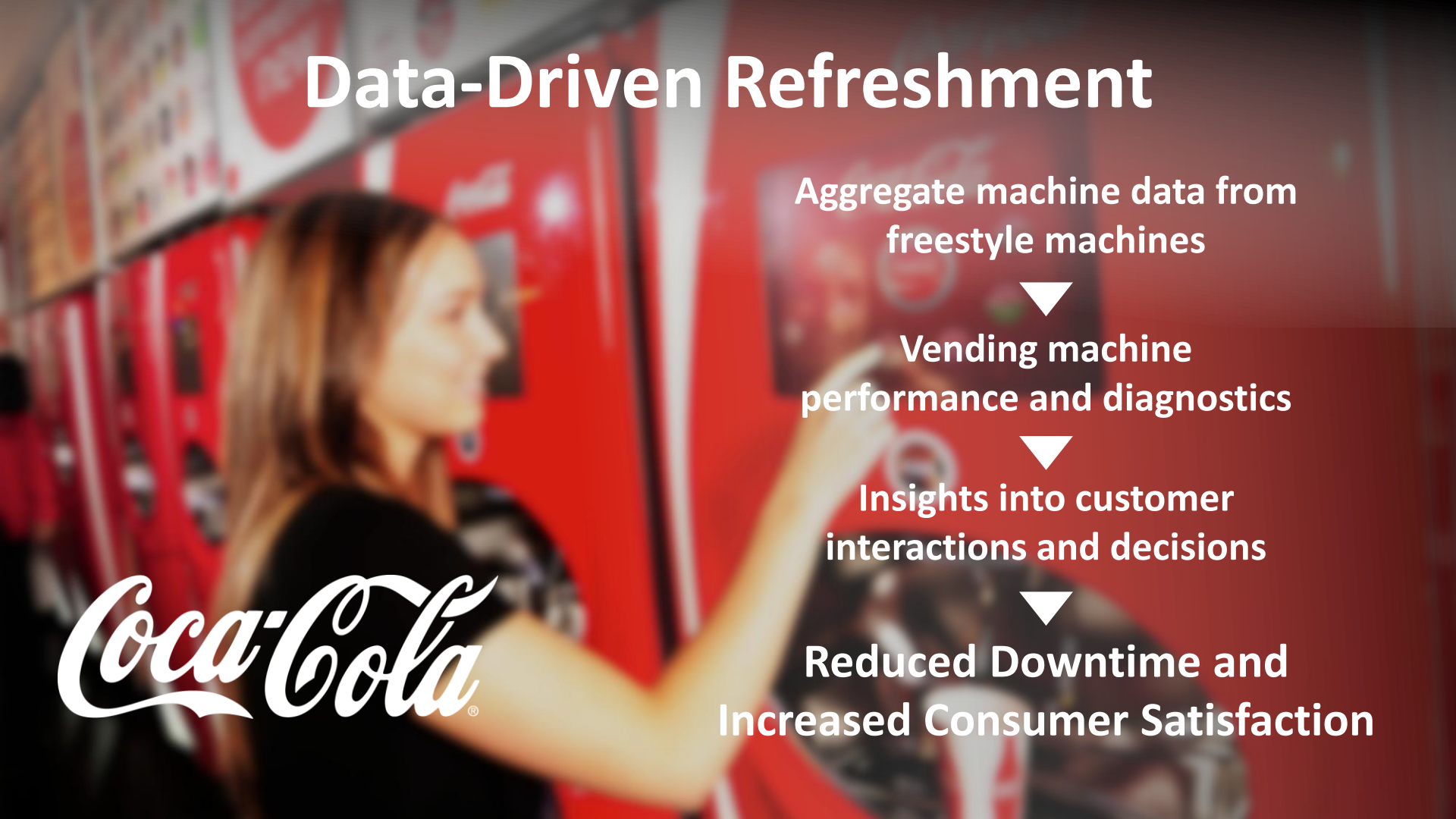
	경쟁사	*스플렁크*
시스템 구성	• 데이터 수집 시스템(개발) • CEP(Complex Event Processing) • 데이터베이스/하둡 • 시각화 툴	• 스플렁크(단일 플랫폼)
POC 결과	• 주어진 시간내에 전체 시나리오 구현 불가 판단	• 모든 요건 구현 완료
본 프로젝트 예상 M/M	• 컨설턴트 6명 • 개발자 12명	• 컨설턴트 2명 • 개발자 2명
본 프로젝트 예상 기간/TCO	• 12 달 • 5~10X	• 3 달 • 1X

SKT 사례: 스플렁크를 이용한 통합품질관리 시스템

전사 IT 인프라에 대한 통합 실시간 모니터링 플랫폼 구축으로 IT 자산 품질 고도화



1. 유의미한 모든 Data 실시간 수집
2. 장애 감지, 예측의 정확도 향상
3. 서비스 中心 자원 연관 관계 Map 제공
4. 서비스 점검 시스템화
5. 사용자 계층별 맞춤정보 제공


Data-Driven Refreshment
Aggregate machine data from freestyle machines
Vending machine performance and diagnostics
Insights into customer interactions and decisions
Reduced Downtime and Increased Consumer Satisfaction
Coca-Cola

auto WOB-Q7047

SPEED

18

kmh

max 29

RPM

1288

rpm

max 2164

BATTERIE
Ladezustand

75%

REICHWEITE
auf bestehende Ladung

65km

UMGEBUNG
Außentemperatur

12.0°C

Scheinwerfer

TÜREN
Tür offen

0

TOP ZIELE

- Keine Treffer

ZEIT

All time

March 7 2014 3:5

March 7 2014 4:53 PM

10s

splunk> POWERED

Medical Devices Driving Better Patient Insights

Device

Manufactured

Prescribed
to patient

Shipped
to Physician

Returned
to iRhythm

iRhythm

Tracking Medical Device Supply Chain to Drive Critical Insights

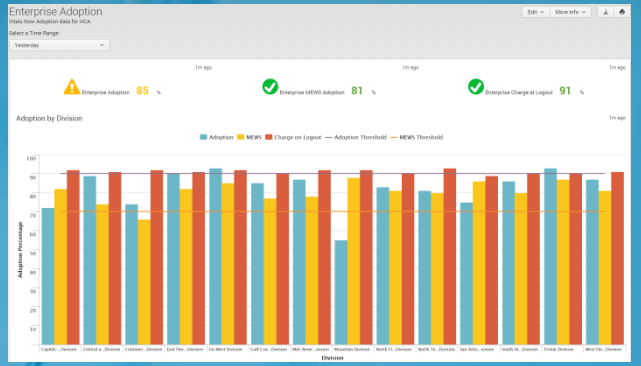
Patient Behavior

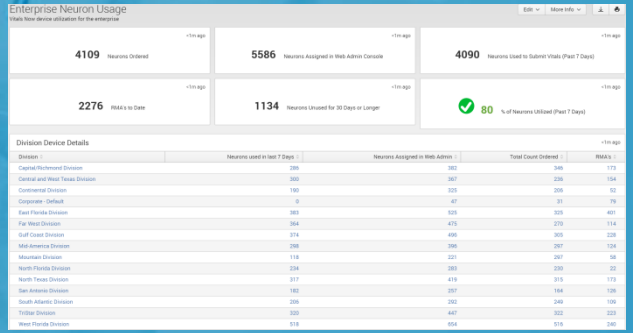
Prescription Patterns

Supply Chain Analytics

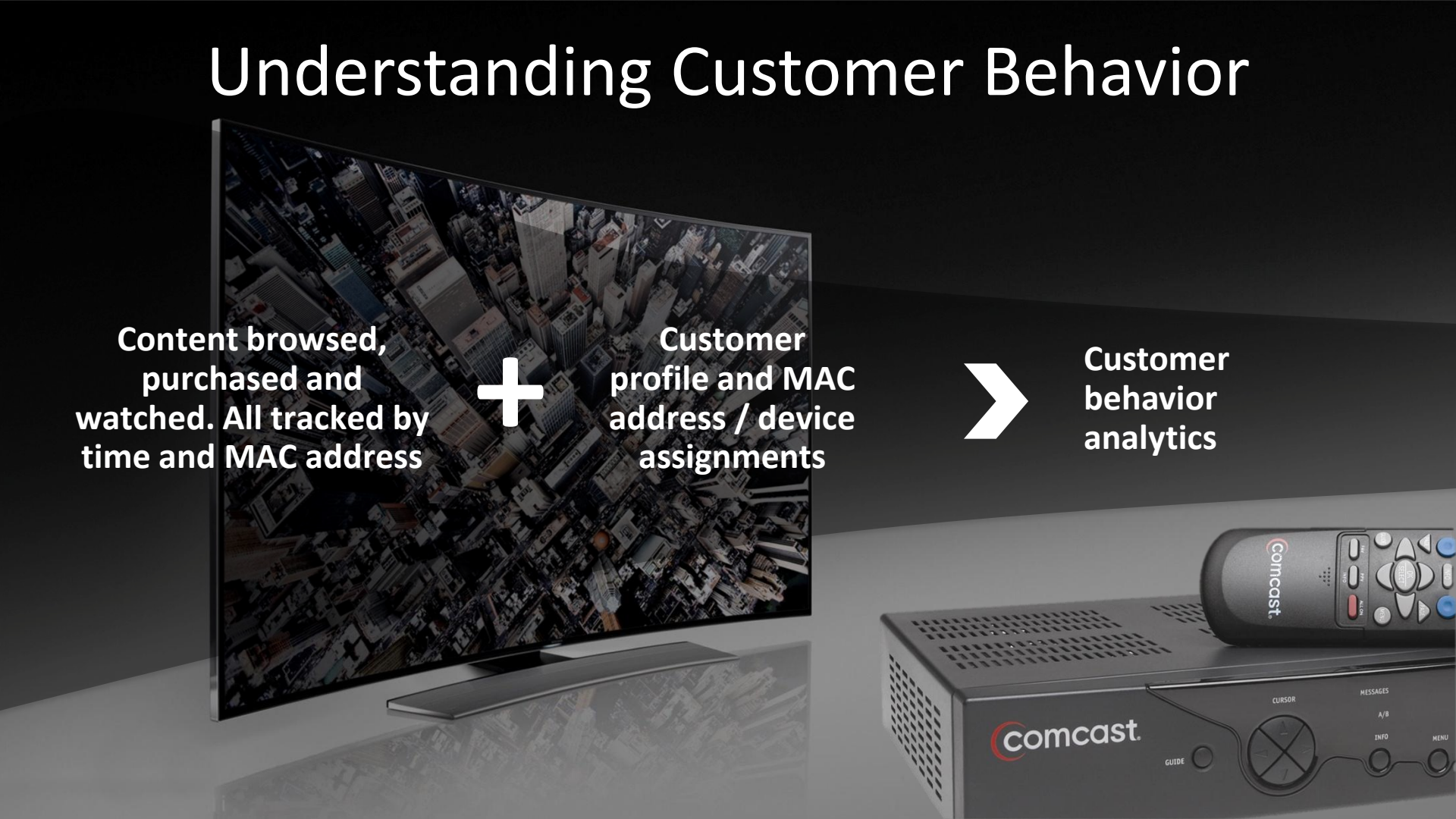
Operational Analytics for Connected Medical Devices

- Comply with Meaningful Use provisions of HITECH Act
- 5000+ devices across 165 hospitals
- 19,000 patients monitored daily
- Enabled change in nurse behavior: 27,000 nurses using the devices
- $3 million in cost savings for new batteries

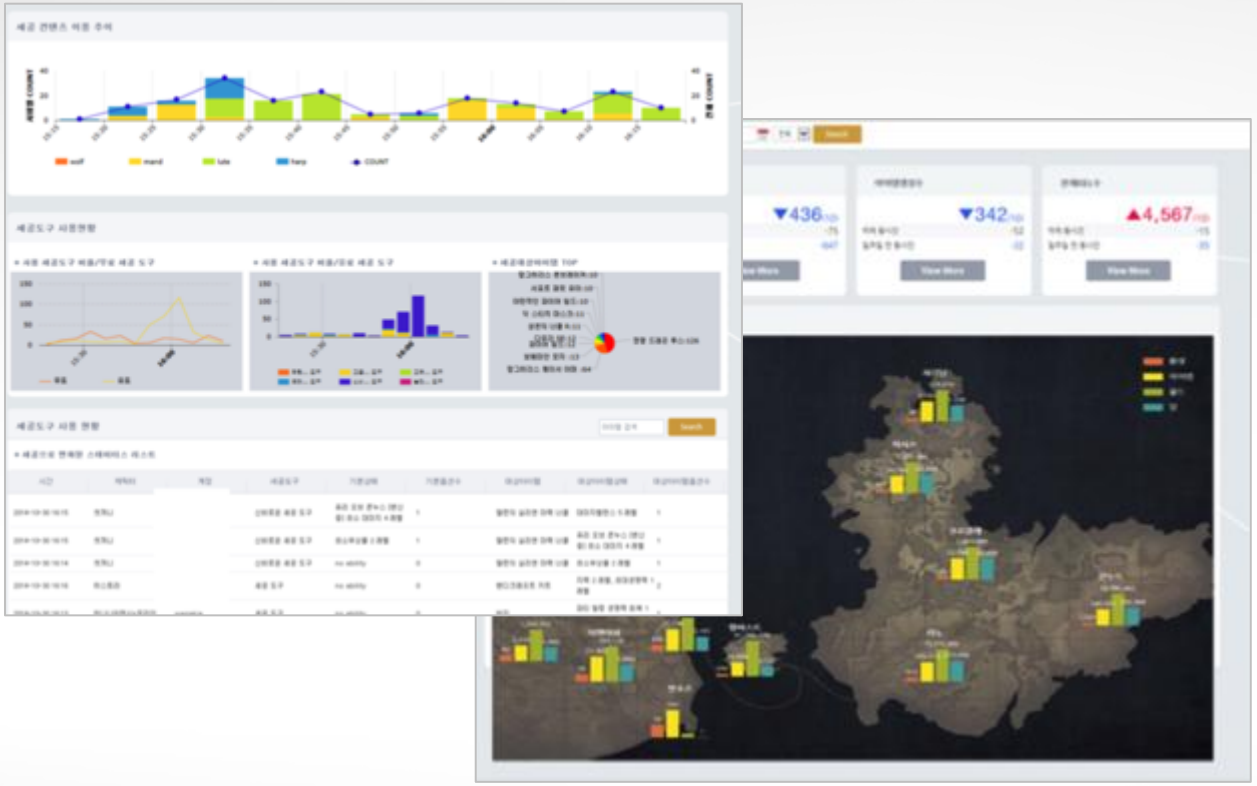




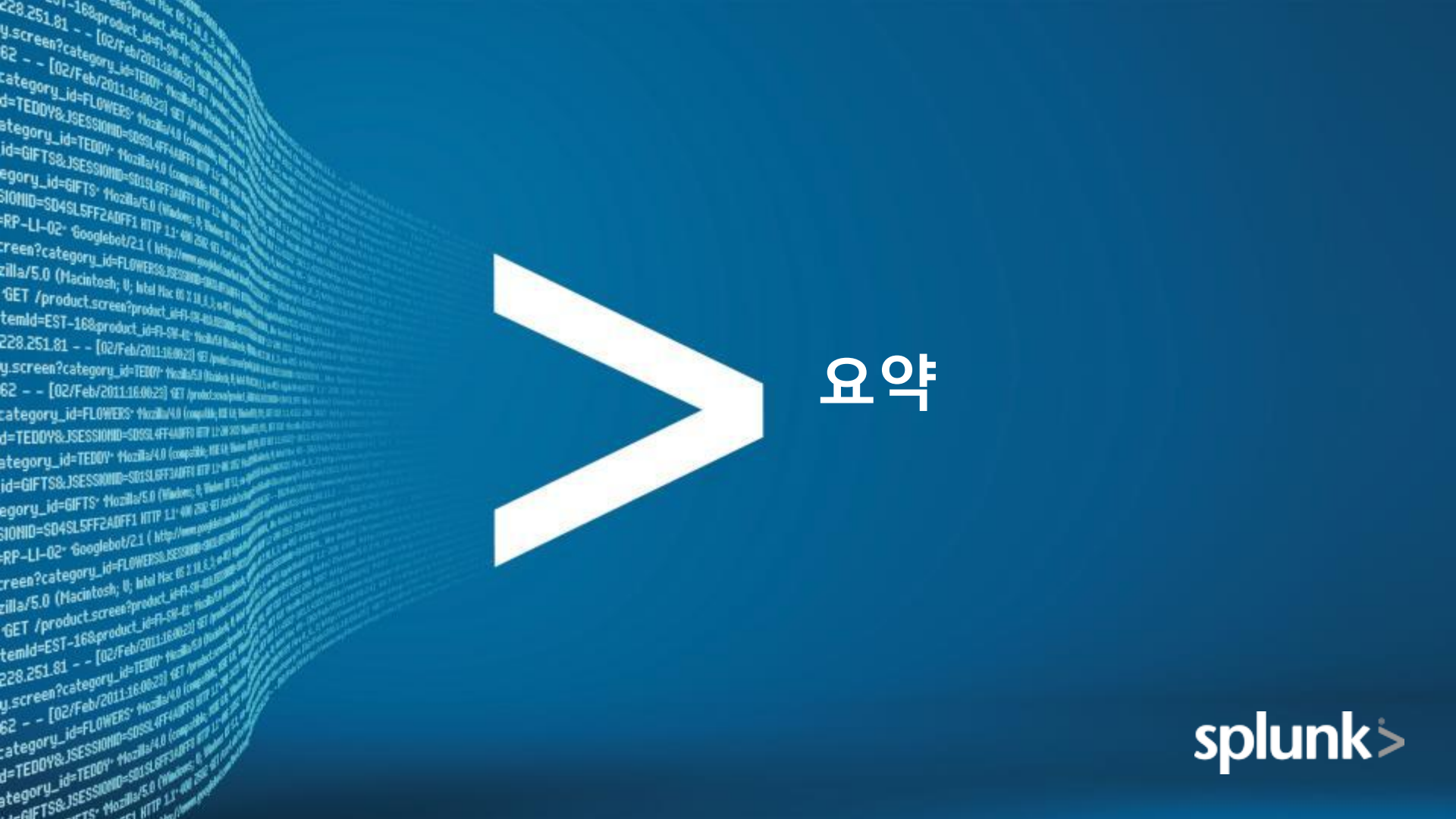
Understanding Customer Behavior



Nexon 게임 사용자 분석



- 실시간 사용자 분석
- 컨텐츠 이용 현황
- 실시간 재화/거래 분석
- 비정상 거래 탐지
- 어뷰징 예측 패턴 탐지



요약

splunk>

머신데이터를 운영 인텔리전스로 전환

실시간 Complex Event Processing

Collect, store and query data in real time

1

운영 인텔리전스 플랫폼

Store and summarize data in native distributed computing environment

2

Analysis /Integrations

Analyze, visualize and extend data to services

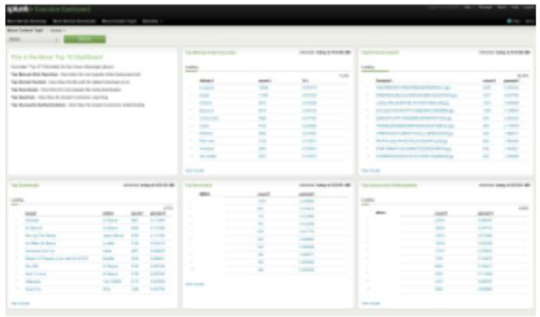
3

Use Cases for Manufacturing

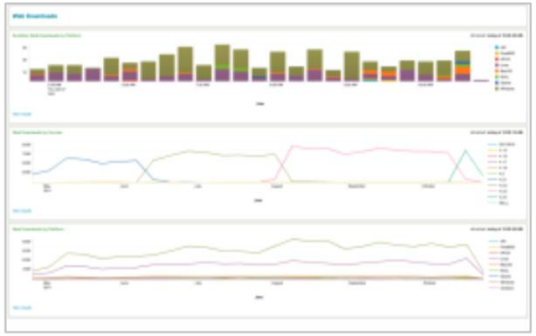
Heterogeneous Data: Sensors, Devices, SCADA/ICS/DCS Data, Application Logs, Wire Data, RDBMS-Backed Safety, Execution, Maintenance and Management Systems

다양한 부서에서 활용할 수 있는 빅데이터 플렛폼

Executive Management



Product Managers



Marketing Managers

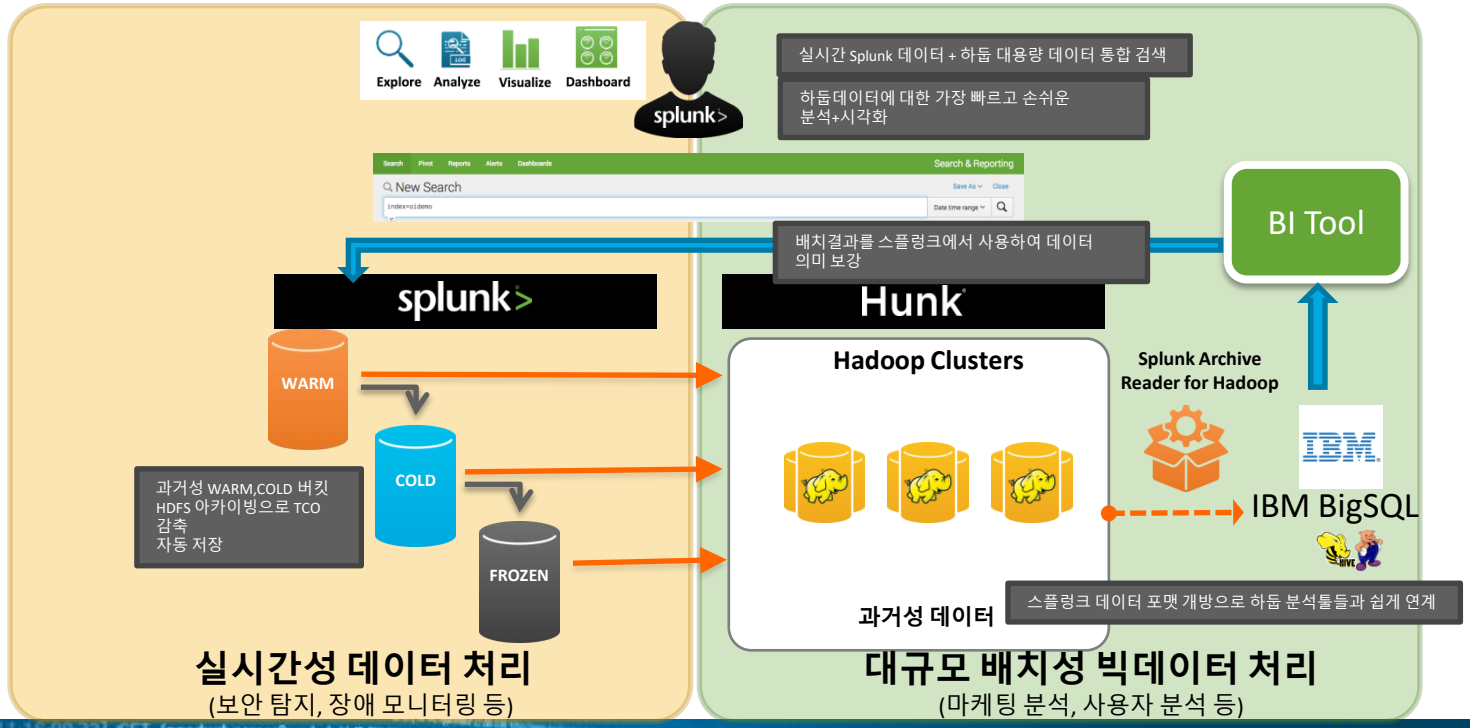
Sales Operations

IT Management & Operations

Customer Service & Support

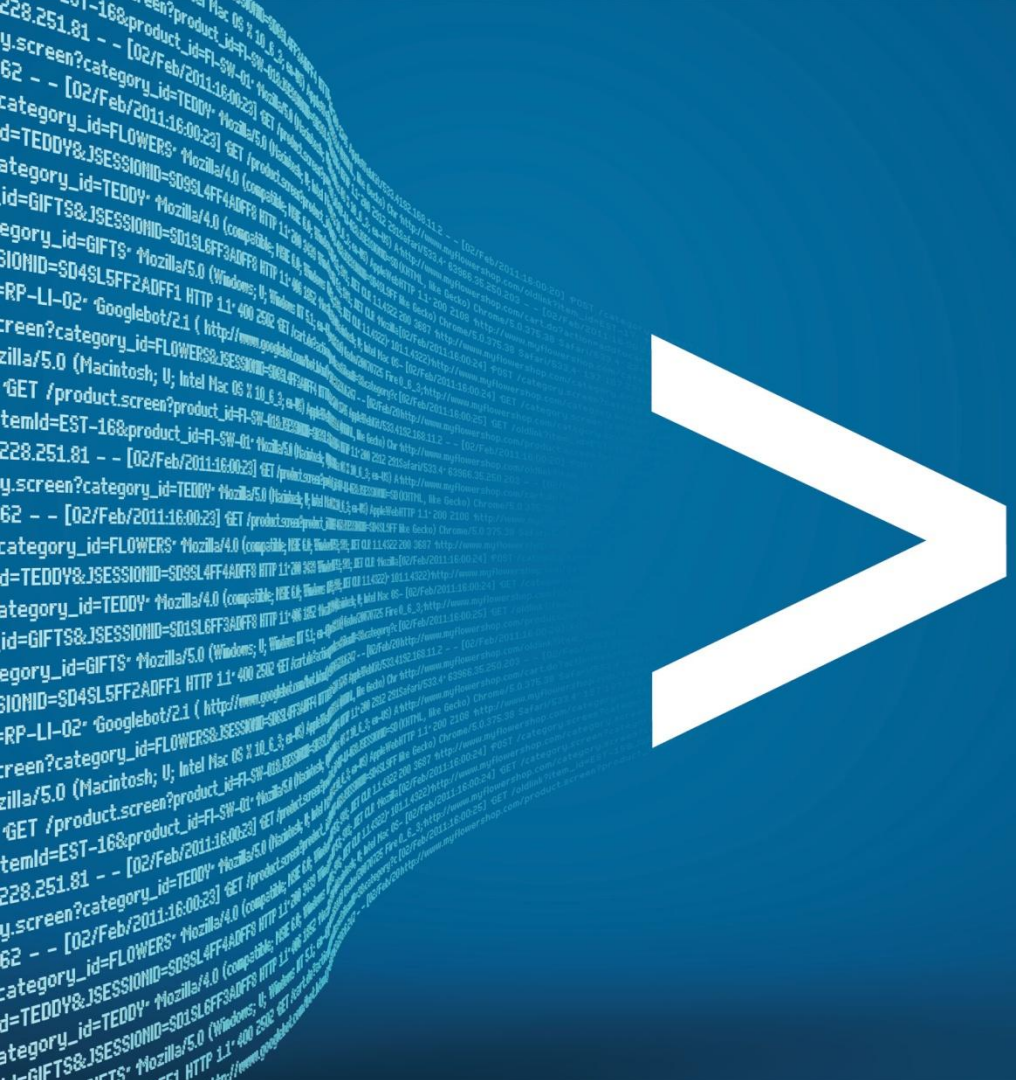
하둡연계성(고도화안)

1) 대량 데이터 이동 없는 분석 : 스플렁크 인덱스 데이터를 HDFS에 저장하고 데이터셋을 하둡의 분석툴로 직접 연계.
2) 하둡의 데이터 분석을 스플렁크의 검색문과 UI 동일하게 사용 : Map/Reduce 작업을 사용한 raw data 정제 작업 필요 없이 손쉽고 빠르게 하둡에 staging된 Raw Data에 대한 직접 분석 가능
3) Closed Loop Analysis: IBM BigSQL의 분석결과를 Splunk에서 참조하여 데이터 의미보강



splunk>

모두에게 머신 데이터에 대한
접근성과 **사용성** 그리고 **가치**를 제공



감사합니다.

Contact :
㈜한국밸런스
김 형덕 영업대표
Mobile : 010-7138-8889
Email : hdkim@valence.co.kr

splunk>