효성인포메이션시스템

AI 활용사례 중심의 DataOps 솔루션

Pentaho Data Integration & Pentaho Business Analytics

백요한 컨설턴트

Contents

1. **솔루션 소개**
   1) Pentaho소개
   2) Pentaho의 특장점

2. **활용 사례**
   - 정밀화학 제조사 (공장설비 예지보전)
   - 타이어 제조사 (타이어 성능 추론 및 최적화 시스템)
   - 테크노파크 (세라믹 공정최적화 TestBed)
   - 게임사 (real-time streaming)
   - 금융사 (IFRS17)
   - 공공기관 (DataLake를 위한 다수기관 Data 수집)

솔루션 소개

Pentaho 소개

HITACHI
Inspire the Next

Pentaho Data Integration & Pentaho Business Analytics

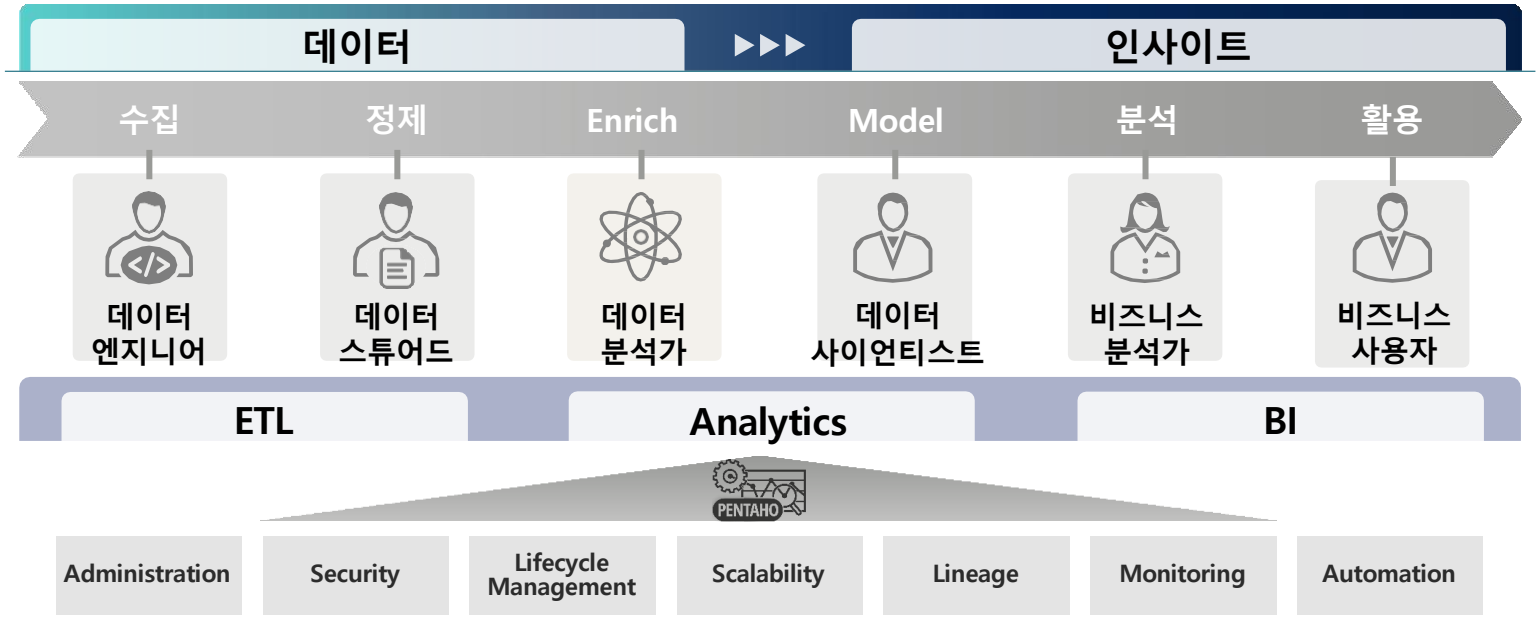
- Pentaho 솔루션은 단일 제품으로 ETL(데이터의 추출, 변환, 적재) 및 고급 데이터 분석(R/Python/weka)과 BI 시각화가 가능한 End-to-End 빅데이터 통합 분석 플랫폼

ETL
- ✓ 다양한 원천 데이터 수집, 강력한 빅데이터 통합 및 처리

분석
- ✓ R , Python 오픈소스 기반 분석
- ✓ 도구와의 유기적 연동 및 코딩 없는 GUI 분석 컴포넌트 제공

BI
- ✓ 웹 기반
- ✓ 손쉬운 데이터 시각화

운영
- ✓ 엔터프라이즈 레벨의 운영 및 보안 관리



© 2022 HYOSUNG INFORMATION SYSTEMS

Pentaho 소개

HITACHI
Inspire the Next

Pentaho Data Integration & Pentaho Business Analytics

통합

수집

Hadoop Cluster, snowflake, Amazon S3, Amazon Redshift, Cloud Dataproc, Google Cloud Storage, Google BigQuery, MINIO, LDC

⋮

처리

ETL metadata injection

- Custom Workflow Template
- Dynamic Input/Processing/Output

실시간

Messaging Protocol 지원

IBM MQ, AMQP, Apache ACTIVEMQ, kafka, MQTT

Producer → Broker → Consumer

분석

Module #1 Machine Intelligence

Module #2 Data Science Pack

python, R

Module #3 Knowledge Flow

WEKA The University of Waikato

자동화

Data Integration

Business Analysis

Monitoring

시각화

대시보드

리포팅

CSV, PDF

알림

PENTAHO

© 2022 HYOSUNG INFORMATION SYSTEMS

Pentaho 소개

HITACHI
Inspire the Next

Pentaho Data Integration & Pentaho Business Analytics

데이터 소스 → 데이터 정제(취합 및 전처리) → 분석 및 검증 → 시각화(대시보드/리포트)

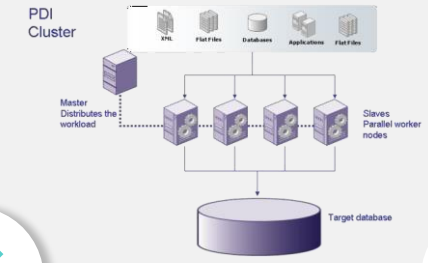
비정형/반정형 데이터

정형 데이터

실시간데이터

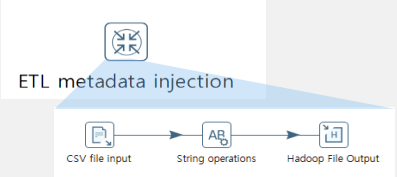
PDI

클러스터링 방식으로 처리속도 향상



데이터 탐색 → 데이터 변환 → 변환 결과

MDI(Meta Data Injection)



PDI

PMI (Machine Intelligence)

- 딥러닝
- Xgboost
- RandomForest
- SVM

⋮

DSP (Data Science Pack)

R

Python

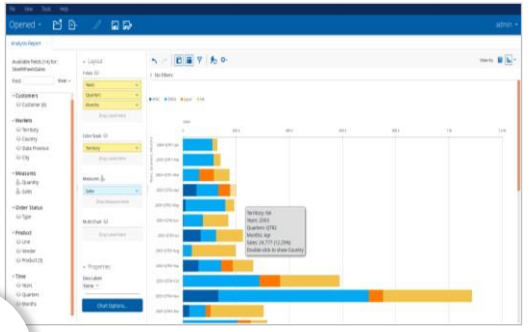
- 시계열회귀모형
- 오차원인분석
- 앙상블모형

⋮

Knowledge Flow

Weka

PBA



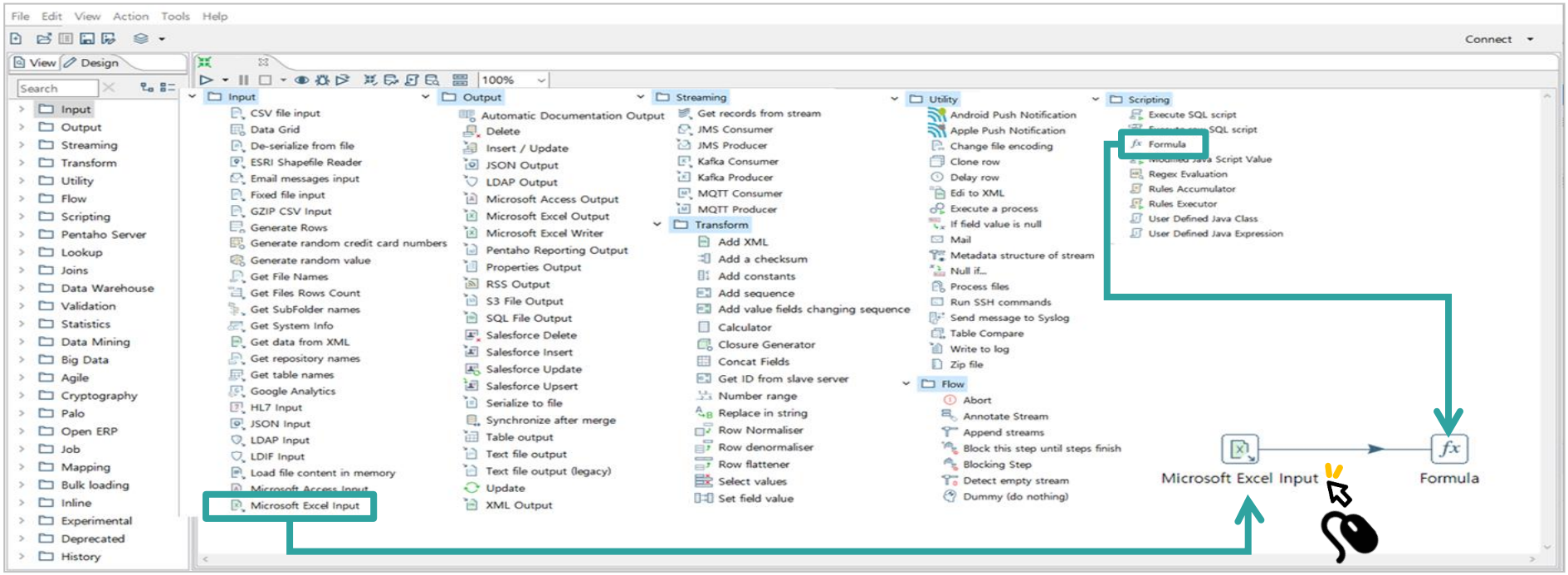
© 2022 HYOSUNG INFORMATION SYSTEMS

강력한 ETL을 위한 손쉬운 Step 제공

HITACHI
Inspire the Next

GUI기반으로 현업 커스터마이징 지원

- ETL을 위한 다양한 Step을 제공하여 정형/비정형/반정형 및 실시간 데이터 수집이 가능
- Script Base의 Step을 지원하기 때문에 사용자의 요구에 맞는 flow 구성 가능



드래그 앤 드랍을 통한 업무프로세스 커스터마이징

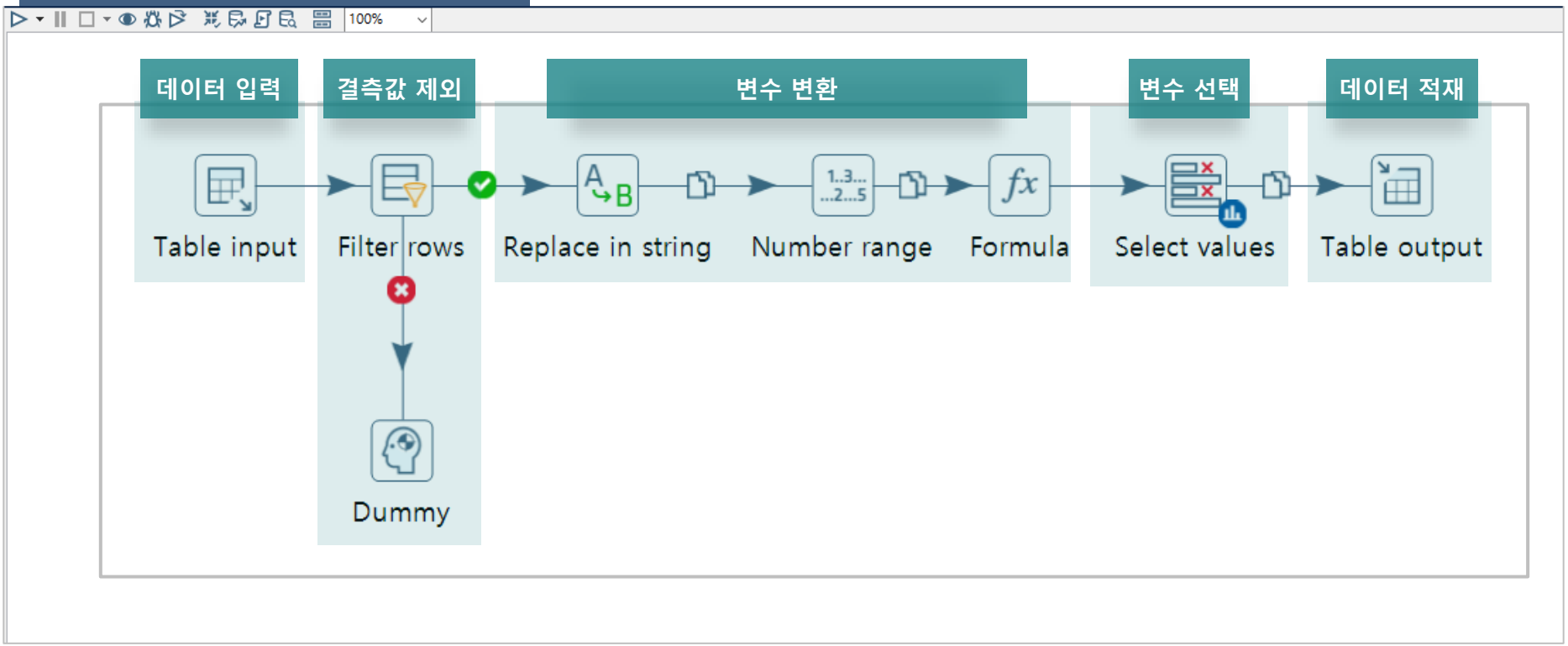
© 2022 HYOSUNG INFORMATION SYSTEMS

전처리 – 분석데이터 전처리

HITACHI
Inspire the Next

✓ Pentaho Enterprise Edition은 분석 데이터에 대한 전처리를 위한 다양한 전처리 Step제공
✓ Step을 활용하여 전처리 Flow 구현 가능

분석 데이터 전처리 Flow - 예시



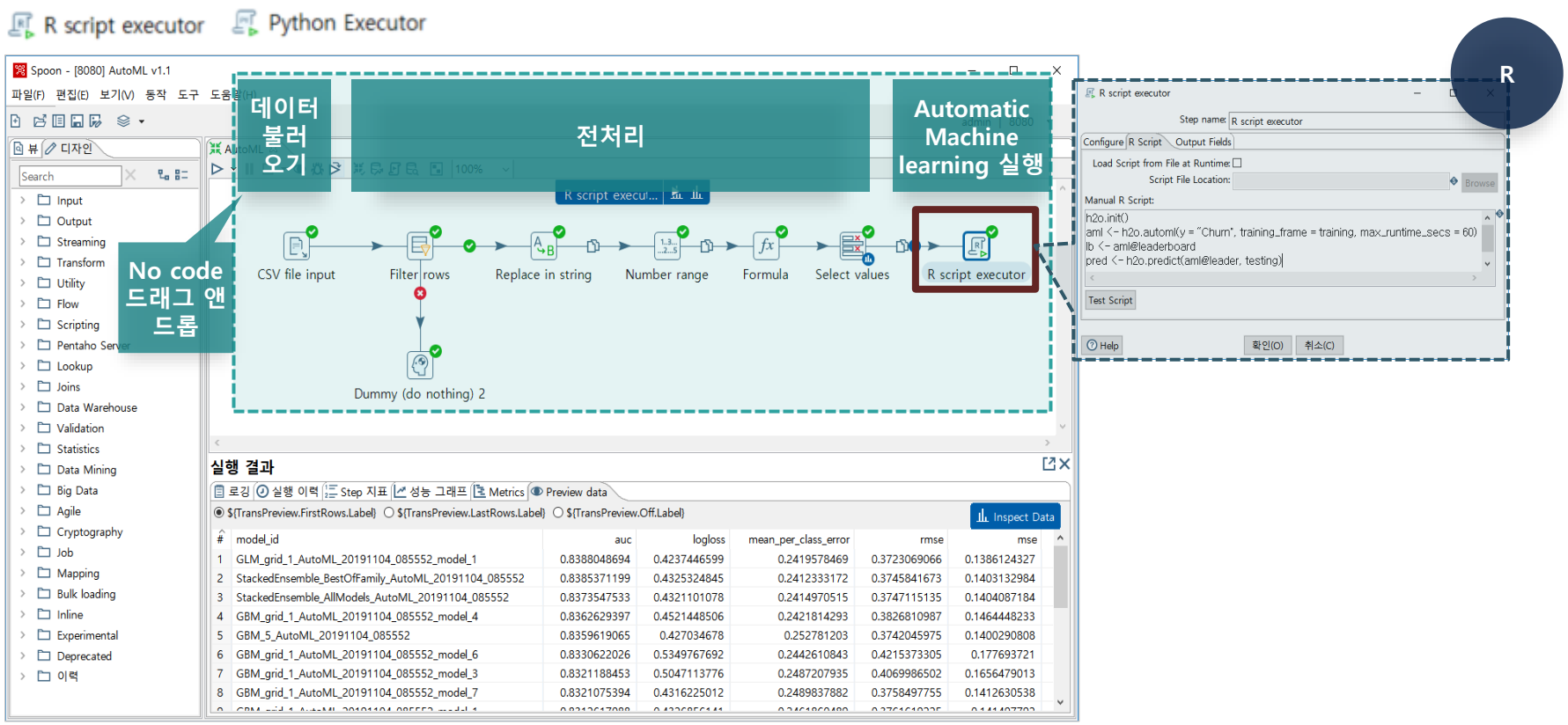
© 2022 HYOSUNG INFORMATION SYSTEMS

분석 – DataScience Pack(R)

HITACHI
Inspire the Next

- ✓ Pentaho Enterprise Edition은 R/Python Script를 수정없이 실행 할 수 있는 R/Python executor Step을 제공
- ✓ Script를 실행 할 수 있기 때문에 최신 R/Python library(package) 사용 및 분석 가능

R/Python 실행 Step

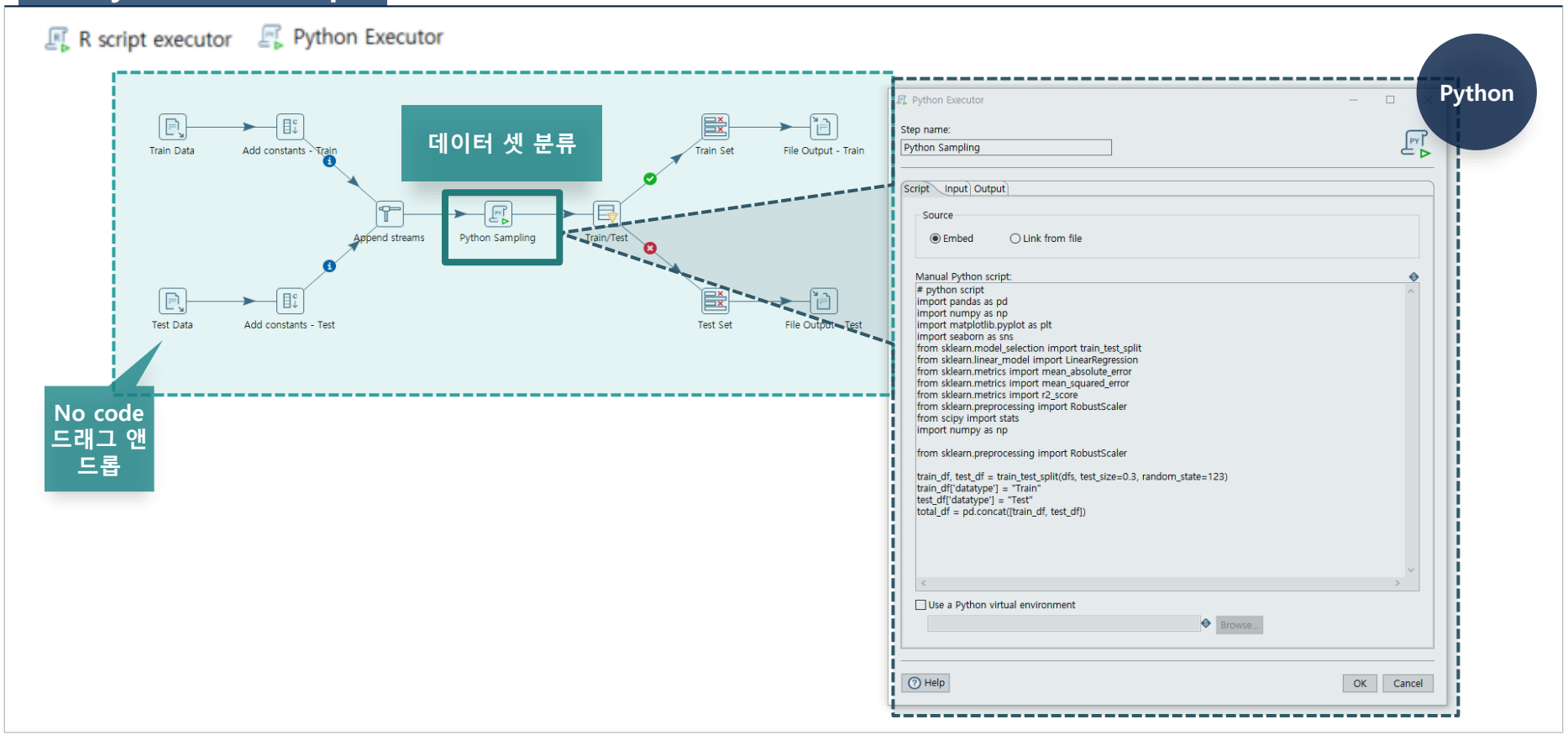


분석 – DataScience Pack(Python)

HITACHI
Inspire the Next

- ✓ Pentaho Enterprise Edition은 R/Python Script를 수정없이 실행 할 수 있는 R/Python executor Step을 제공
- ✓ Script를 실행 할 수 있기 때문에 최신 R/Python library(package) 사용 및 분석 가능

R/Python 실행 Step



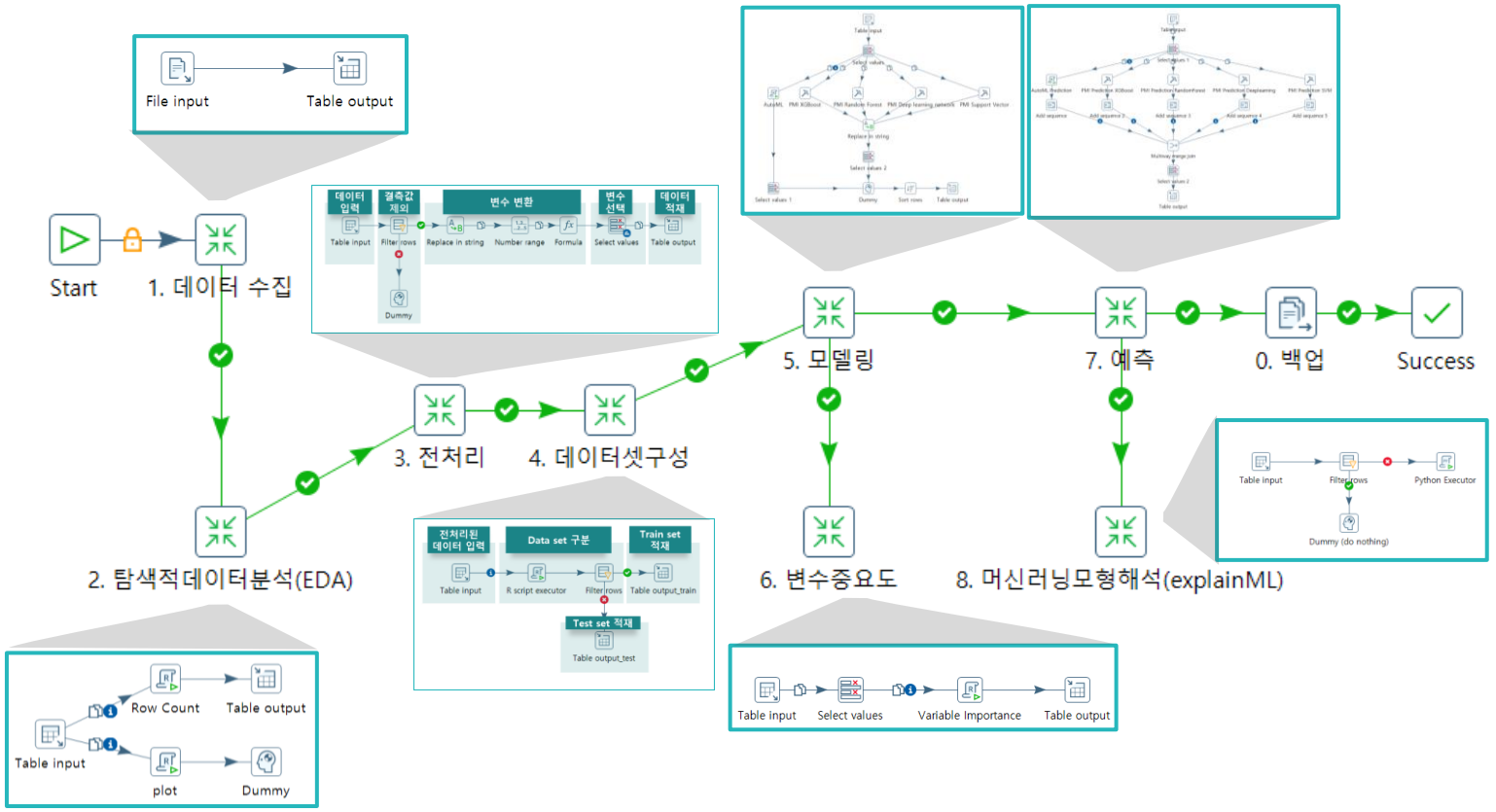
© 2022 HYOSUNG INFORMATION SYSTEMS

분석 – 실시간 모델 업데이트

HITACHI
Inspire the Next

✓ Pentaho Enterprise Edition의 데이터추출/처리/적재/머신러닝기반 예측모형을 하나의 Pentaho Workflow로 통합하여 Train, Tune, Test 및 업데이트 포함 End-to-End 자동화 가능

End-to-End Data-flow 예시



© 2022 HYOSUNG INFORMATION SYSTEMS

Pentaho 특장점

HITACHI
Inspire the Next

Pentaho Data Integration & Pentaho Business Analytics 특장점

강력한 빅데이터 수집

- Any Source Any Target
- Hadoop 환경에서의 빠른 처리를 위한 Shim driver 지원
- VFS(Virtual File System) Connection
- Bulk Load (Snowflake, Amazon Redshift, Google BigQuery, ···)

강력한 빅데이터 처리

- Scale-up, Partitioning, Clustering ➔ Performance enhancement
- MDI(Meta Data Injection)
- Carte

실시간 데이터 처리 및 다양한 처리엔진

- Real-time Data Processing (Kafka, MQTT, etc)
- Adaptive Execution Layer (Spark)

PENTAHO

PDI + PBA

분석 플랫폼간의 유기적 연동

- Data Science Pack (R/Python)
- PMI (Machine Intelligence)
- KnowledgeFlow (Weka)
- Data Service

보안 운영 강화

- Enterprise Security : LDAP, AD
- SSO

업무 자동화

- Orchestration
- 스케줄링, 모니터링

시각화

- Interactive Reporting
- Self-Dashboard
- 3rd Party BI 연계

© 2022 HYOSUNG INFORMATION SYSTEMS

시각화

시각화 - PUC – PBA(Pentaho Business Analytics) - Analysys Report

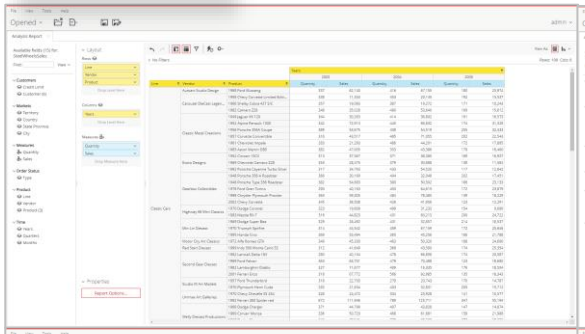
HITACHI
Inspire the Next

✓ PBA(Pentaho Business Analytics) -Analysys Report 기능은 Pentaho분석 데이터 소스에 포함된 비즈니스 정보를 필터링하고 드릴 다운하는 직관적 분석 시각화 도구

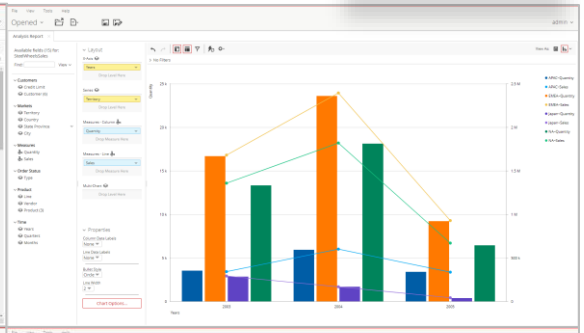
✓ 대화형 환경에서의 빠른 데이터 컴파일 및 데이터의 고급 정렬 및 필터링을 수행, 차트 하이라이트 시각화 가능

Analysys Report

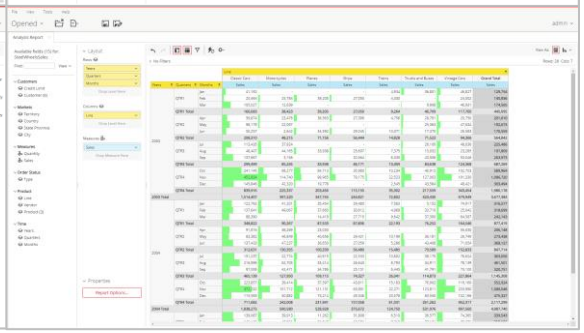
교차분석



막대&라인 차트



100%막대차트



교차분석 부분합

- Drag&Drop 으로 손쉬운 보고서 작성 가능
- 시간 /사용자 정의 계층에 따른 Drill-Up/Down 가능
- 측정값 Drill-through 를 통해 상세 Row Data 확인 가능
- Filter를 변수화 하여 다른 보고서와의 연계 가능
- Hyperlink 를 통해 다른 보고서와의 연계 가능
- Export 를 통해 PDF/CSV /Excel 내보내기 가능

© 2022 HYOSUNG INFORMATION SYSTEMS

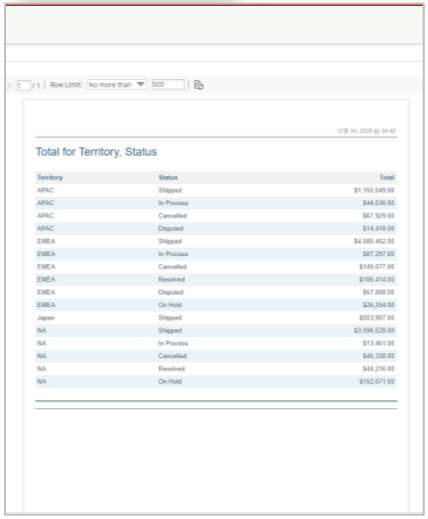
시각화 - PUC – PBA(Pentaho Business Analytics) - Interactive Reports

HITACHI
Inspire the Next

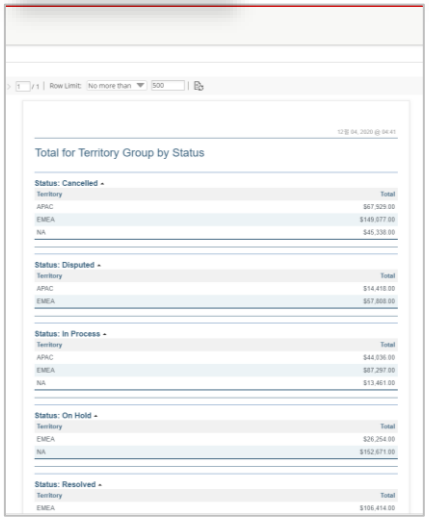
✓ PBA(Pentaho Business Analytics) - Interactive Reports는 웹 기반 디자인 인터페이스
✓ 빠른 보고서 작성 가능
✓ 보고서 서식 지정 가능

Interactive Reports

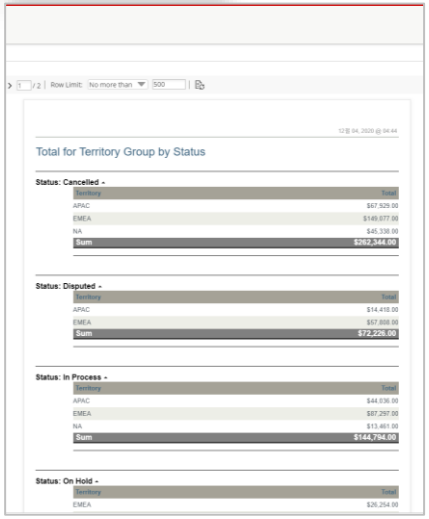
목록



그룹화



집계



- Drag&Drop 으로 손쉬운 보고서 작성 가능
- Group 별 집계 방식, 부분합 방식 설정 가능
- Filter를 변수화 하여 다른 보고서와의 연계 가능
- 각 Field, Title, Sub-title 등 항목들에 대한 스타일 편집
- Export 를 통해 PDF/CSV /Excel/HTML 내보내기 가능

© 2022 HYOSUNG INFORMATION SYSTEMS

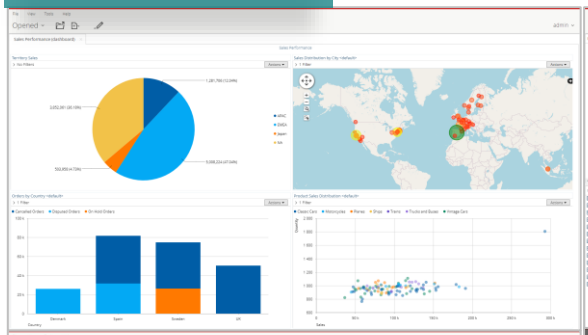
시각화 - PUC – PBA(Pentaho Business Analytics) - Dashboard

HITACHI
Inspire the Next

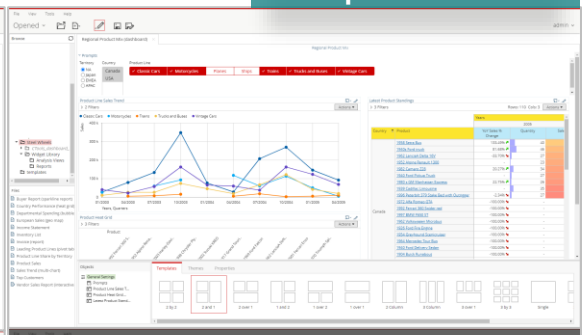
✓ Dashboard는 여러 보고서를 볼 수 있는 인터페이스를 생성
✓ 자주 방문하는 웹 페이지에 빠른 액세스 가능
✓ 동적 차트 및 그래프를 확인 가능

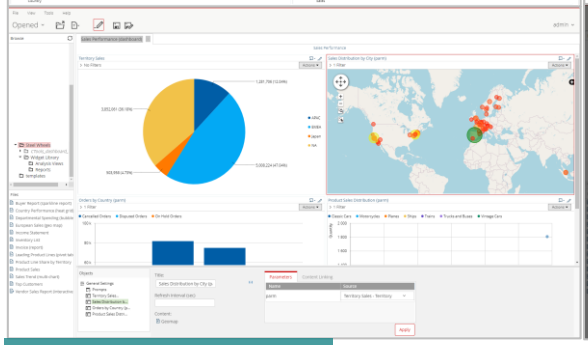
Dashboard

Dashboard

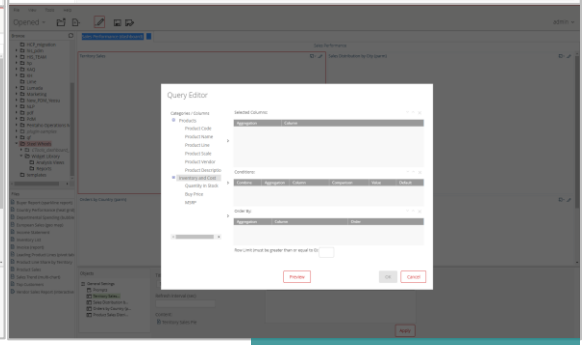


Template & Theme





Prompt & Content Link



Data Source Chart

- 생성된 Report를 Drag & Drop으로 손쉽게 추가
- 기본적으로 제공되는 Layout Template & Theme 제공
- 특정 시간 주기로 새로고침 기능 제공
- Prompt를 설정하여 사전에 정의된 보고서 Parameter로 사용 가능
- Content Link를 통해 보고서 간 Filter 동기화
- Data Source에 바로 접근하여 차트 생성 가능

© 2022 HYOSUNG INFORMATION SYSTEMS

활용 사례

도입사례_A 화학 제조사

HITACHI
Inspire the Next

예지정비 분석 프로젝트 (ETL + 데이터분석)

CHALLENGE

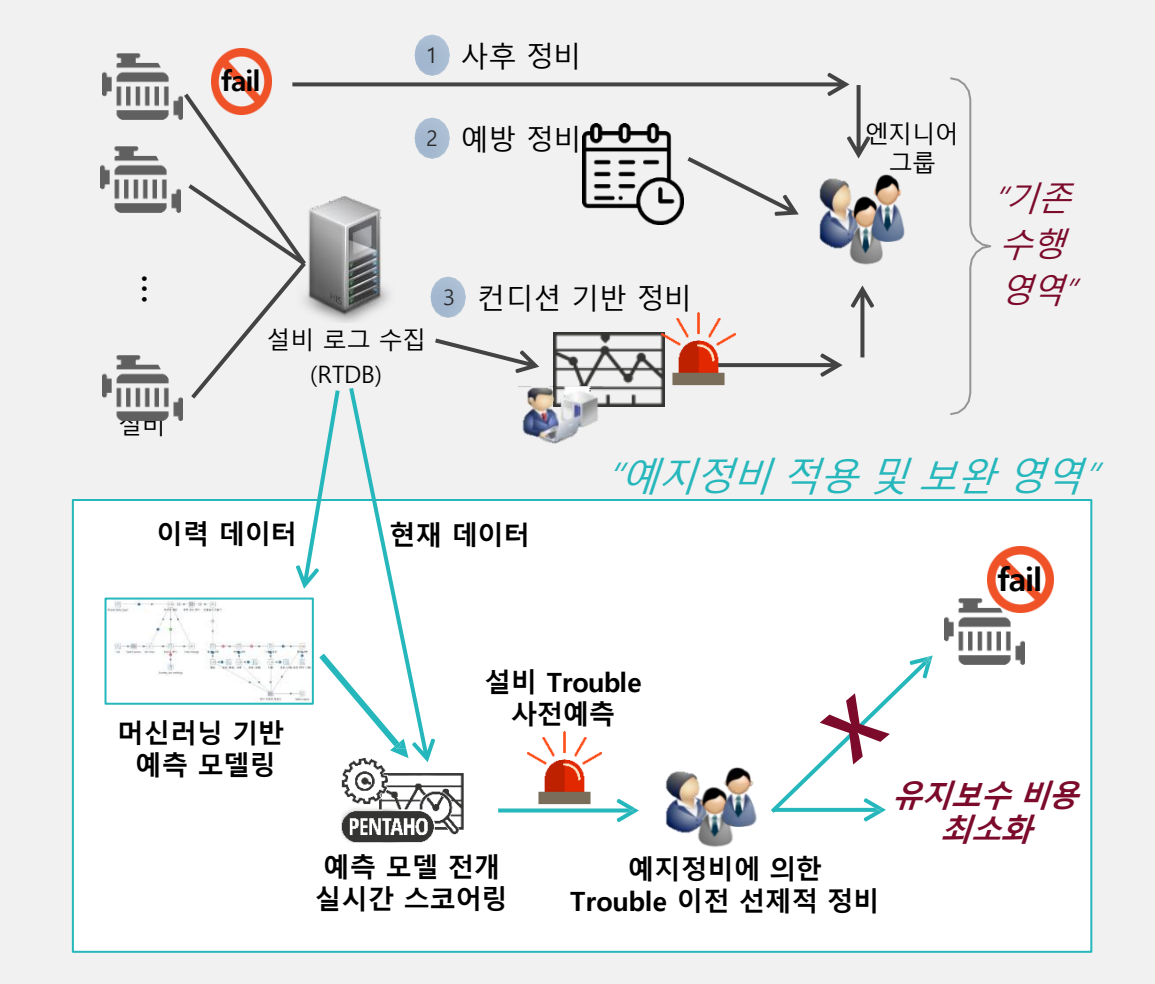
- 공장의 중요 설비의 이상 및 갑작스런 고장으로 비용 손실 발생
- 분석을 진행 하려고 해도 각 공장의 IoT 데이터 수집이 어려움
- 분석 모형을 통해 사전에 이상현상 예측하고 조치 필요

SOLUTION

- 예지정비 모델 생성
- IoT 데이터 실시간 수집 및 저장 처리
- 예지정비 모델의 장애 예측 일자 알람 제공

BENEFIT

- 다운타임 절감으로 인한 생산성 향상 (수십 억원의 손실 감소)
- 설비 관리의 신뢰성 향상
- 설비 관리 인력 효율화



© 2022 HYOSUNG INFORMATION SYSTEMS

도입사례_B 타이어 제조사

HITACHI
Inspire the Next

타이어 품질 예측 분석 프로젝트 (ETL + 데이터분석)

CHALLENGE

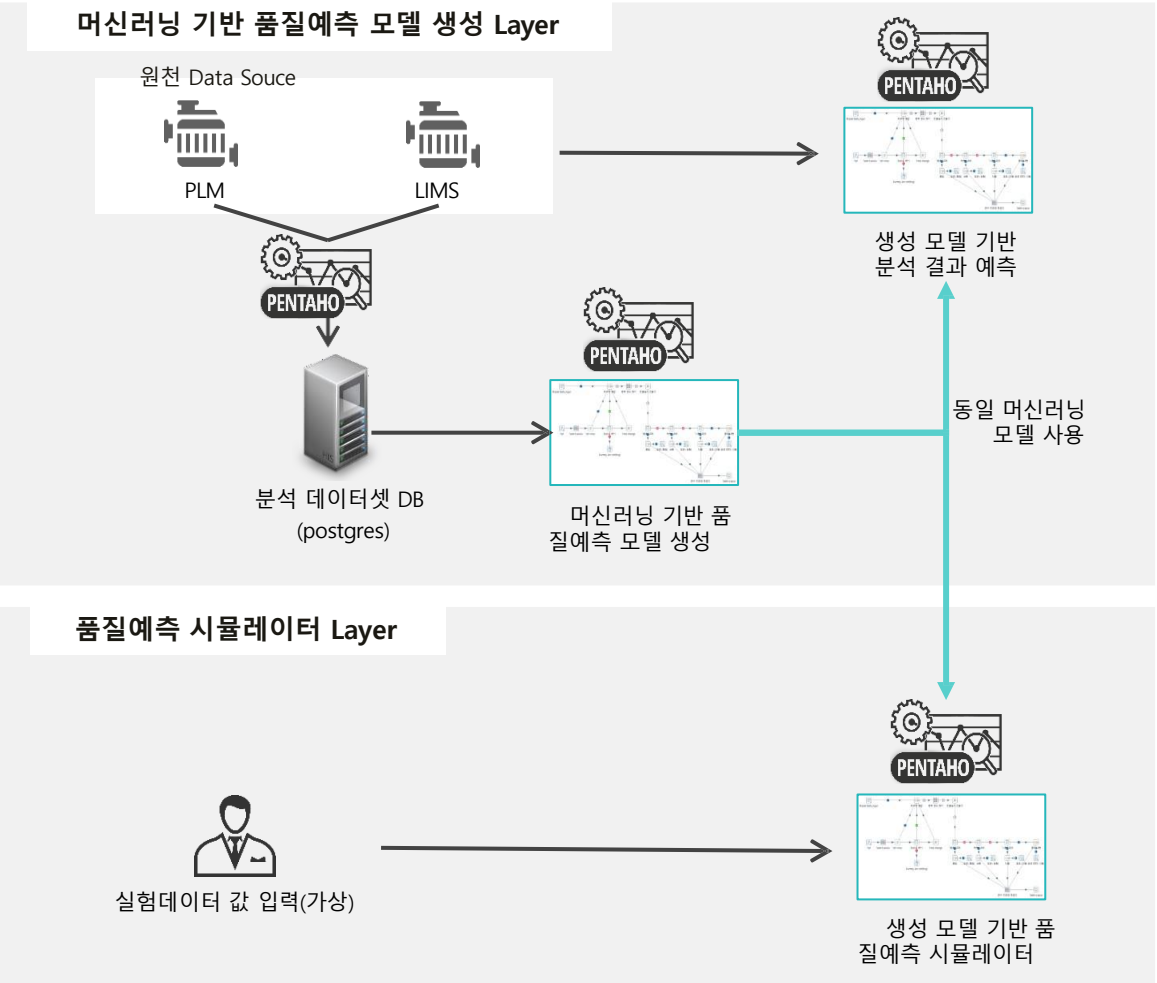
- 타이어 설계/재료/실험 데이터 활용 및 분석 범위 확대 필요성 증가
- 타이어 설계/재료 인자들의 관리 방안 및 타이어 성능 연구의 필요
- 타이어 개발 기간 단축, 개발 비용 절감 및 우수한 성능 타이어 개발 필요

SOLUTION

- Pentaho Data Integration 기능을 통한 실험 데이터 ETL 및 분석 알고리즘 적용
- 타이어 성능 및 재료 물성 예측 모델 시스템 자동화

BENEFIT

- 기존 수작업으로 1시간 이상 소요하던 데이터 조회를 데이터 마트 구축으로 바로 조회
- 기존 몇 일 소요하던 성능 예측 결과를 Pentaho 시스템 구축으로 예측 결과 바로 조회



© 2022 HYOSUNG INFORMATION SYSTEMS

도입사례_C 공공 제조사

HITACHI
Inspire the Next

AI기반 예지보전 & 공정분석 시스템 구축 (ETL + 데이터분석)

CHALLENGE

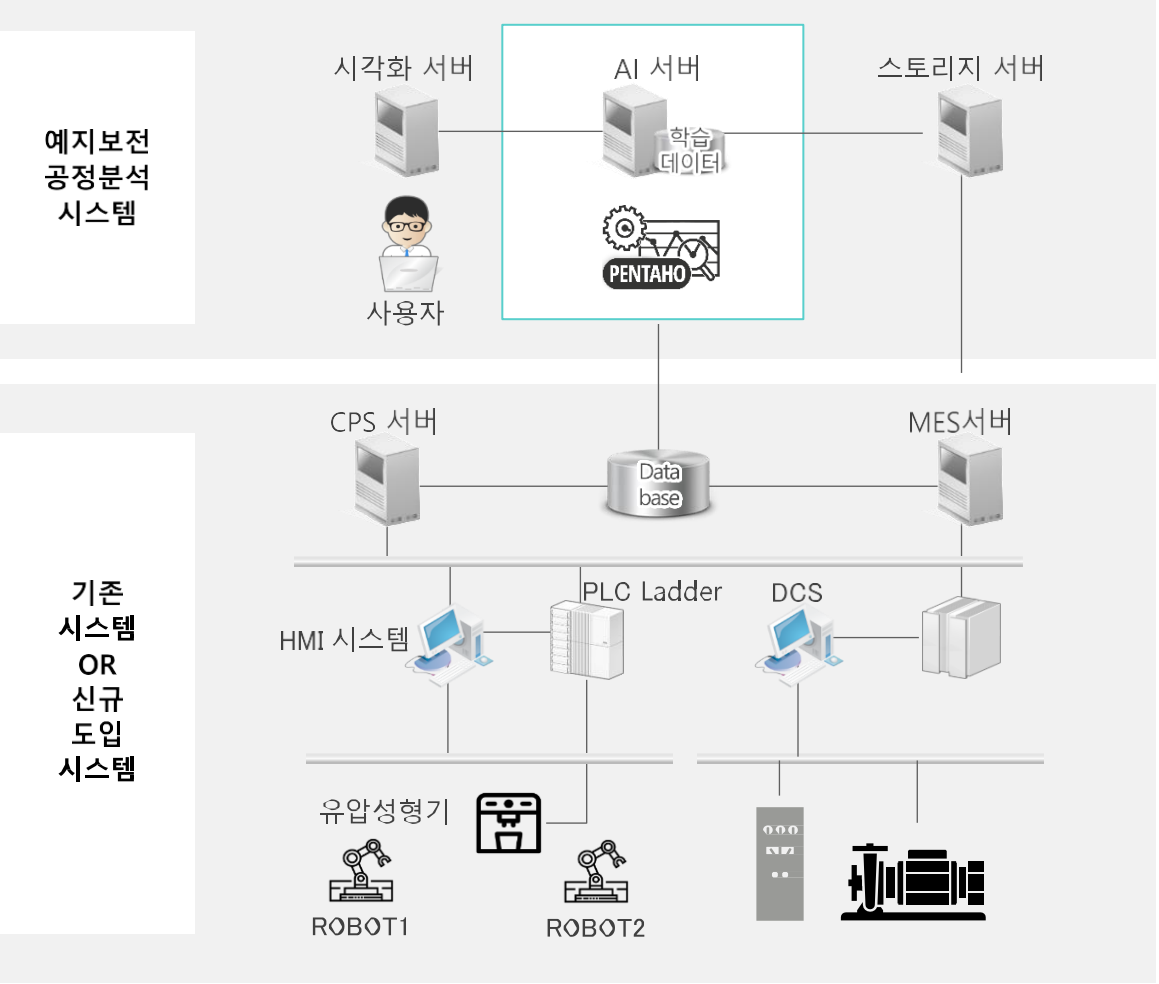
- 전라남도 지역 전략 산업인 세라믹 제조 산업의 활성화를 위한 방안 도출
- 기업 지원 시스템을 이해하고, 관련 설비에 대한 예지 보전과 원재료 공정분석, 성형공정 디지털 트윈의 필요성에 맞는 제안 방향 제시

SOLUTION

- 세라믹 특화 테스트 베드 구축
- 공급/수요 기업 기술 수요 대응
- 지역기업 밀착 지원

BENEFIT

- 대표 공정장비 스마트화 완료 및 실증 평가로 단위 공장 기술지원 체계 확립
- 스마트제조 통합운영 시스템 구축으로 전방위 스마트제조 기술지원 기반 확립



© 2022 HYOSUNG INFORMATION SYSTEMS

도입사례_D 게임사

HITACHI
Inspire the Next

안정적인 실시간 대용량 스트리밍 데이터 처리 (ETL)

CHALLENGE

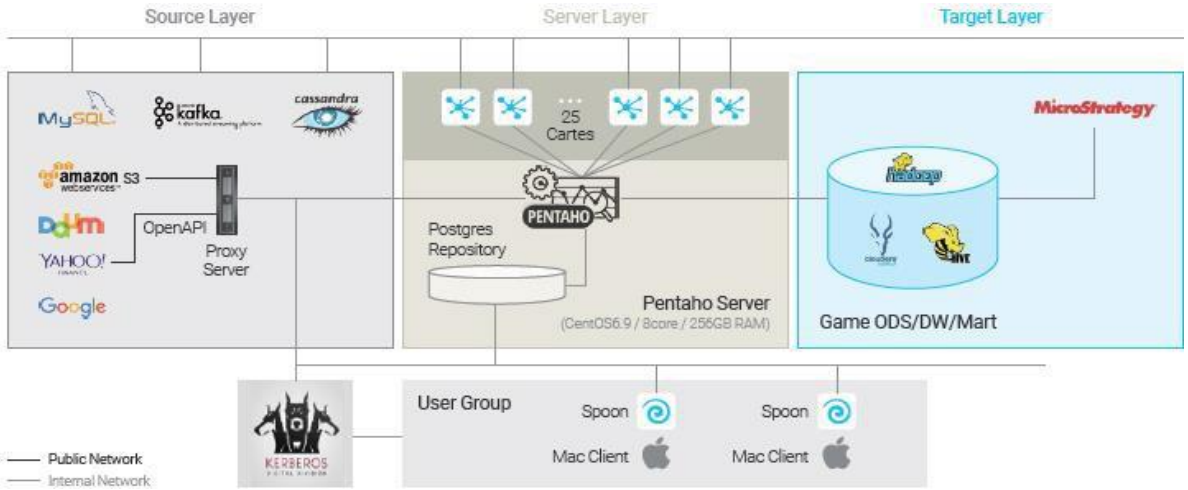
- ETL업무프로세스를 GUI기반으로 자동화
- 실시간 대용량 스트리밍 데이터 처리
- 데이터 업무프로세스에 대한 실시간 모니터링으로 생산성 증대

SOLUTION

- Pentaho 빅데이터 통합 기능을 통한 실시간 처리
- ETL 작업을 코딩 없이 GUI로 구현
- 작업 시간 단축으로 업무 생산성 증대

BENEFIT

- Kafka 데이터를 1시간에 약 2000만건 Hive에 저장하여 빅데이터 시스템을 ODS(Operating Data Store)로써 활용 가능
- 실시간 데이터를 활용한 집계정보 확인 시간을 1시간 이상에서 5분으로 단축



© 2022 HYOSUNG INFORMATION SYSTEMS

도입사례_E 금융보험사

HITACHI
Inspire the Next

IFRS17 work-flow (ETL)

CHALLENGE

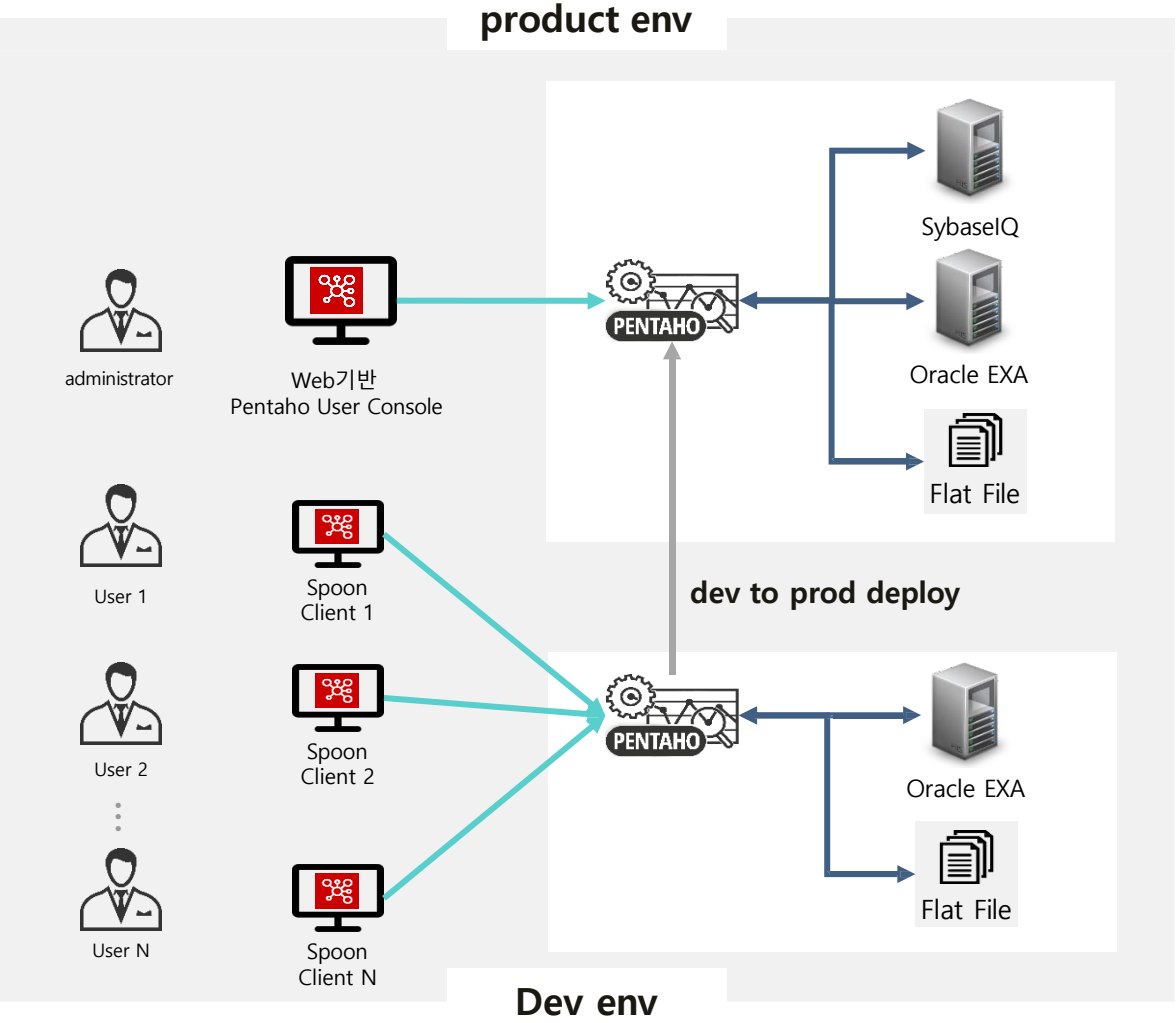
- ETL업무프로세스를 GUI기반으로 자동화
- Enterprise급 금융권 보안 감사 요건 충족 솔루션 필요
- 업무별 선-후행 작업간 스케쥴링 기능 필요
- 작업간 정합성을 위한 validation 체크 및 모니터링 필요
- 개발/운영 환경 분리

SOLUTION

- Pentaho Data Integration을 통한 업무프로세스 구축
- Enterprise급 금융권 보안 감사 요건 충족
- 선-후행 작업의 group job 스케쥴링
- 데이터정합성 validation 체크 로직 및 모니터링 요건 구축

BENEFIT

- 중앙집중식 repository를 통한 work-flow 단일화
- client tool 제공을 통한 개발 환경 제공
- 대용량 데이터 처리 프로세스의 스케쥴링을 통한 자동화
- 데이터 처리 로깅과 모니터링



© 2022 HYOSUNG INFORMATION SYSTEMS

도입사례_F 공공기관

HITACHI
Inspire the Next

다수 타 기관 Private Cloud 기반 데이터 중앙관리 저장소 수집(ETL)

CHALLENGE

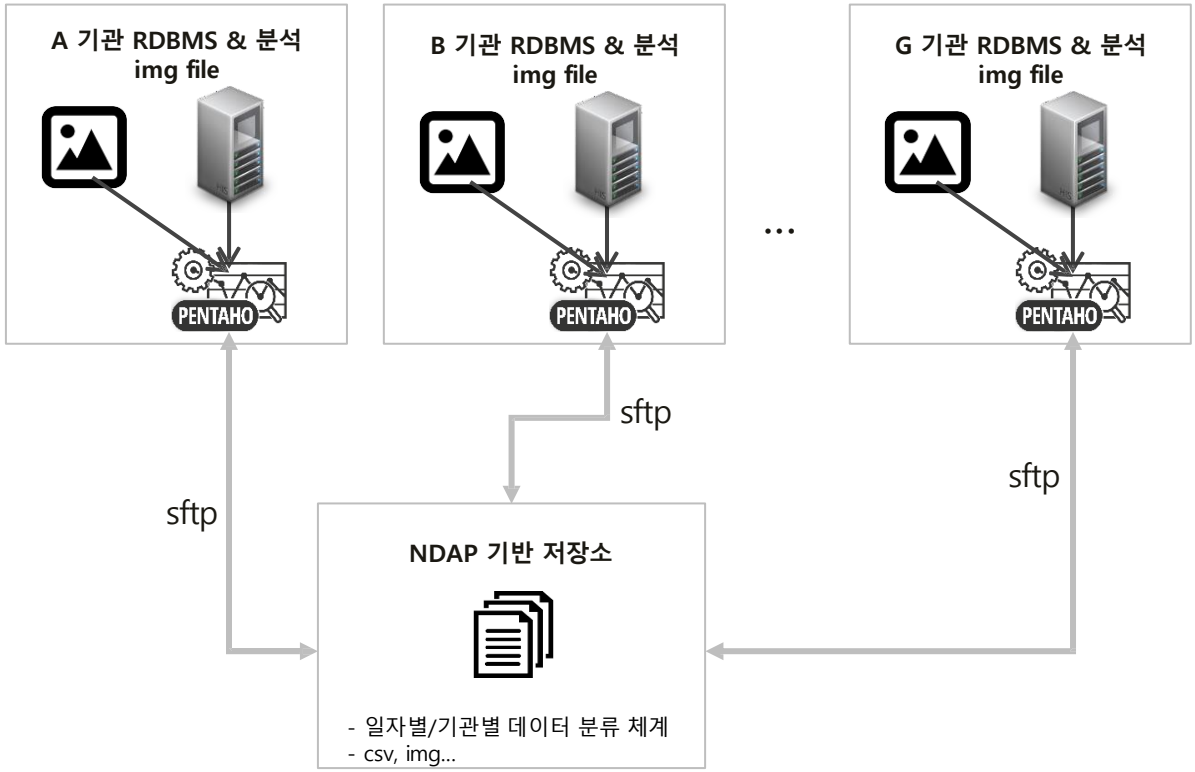
- 각 기관별 보유/생성 데이터 관리 필요
- 중앙집중식 관리 저장소 필요
- NDAP 기반 데이터 저장소로 수집에 대한 필요

SOLUTION

- NDAP기반 중앙 집중식 저장소 구현
- 기관별 다양한 RDBMS의 데이터 ETL 및 비정형 이미지 데이터 수집
- 수집에 대한 validation 및 수집 retry 로직 구현

BENEFIT

- 각 기관의 다양한 실험인증 데이터 수집을 GUI방식으로 손쉽게 처리
- 안정적인 배치작업(스케줄링)으로 각기관 업무 감소



© 2022 HYOSUNG INFORMATION SYSTEMS

Thank you

Q&A

Hyosung Information System
DataSolution Team

his-baek.yh@hyosung.com