

HPE Data Service Innovation Day

빅데이터 분석 플랫폼의 성공적인 구축사례와 구축 고려사항

Hong JongHyun

VMware Data Solutions

VMware Korea

2023.06



Agenda

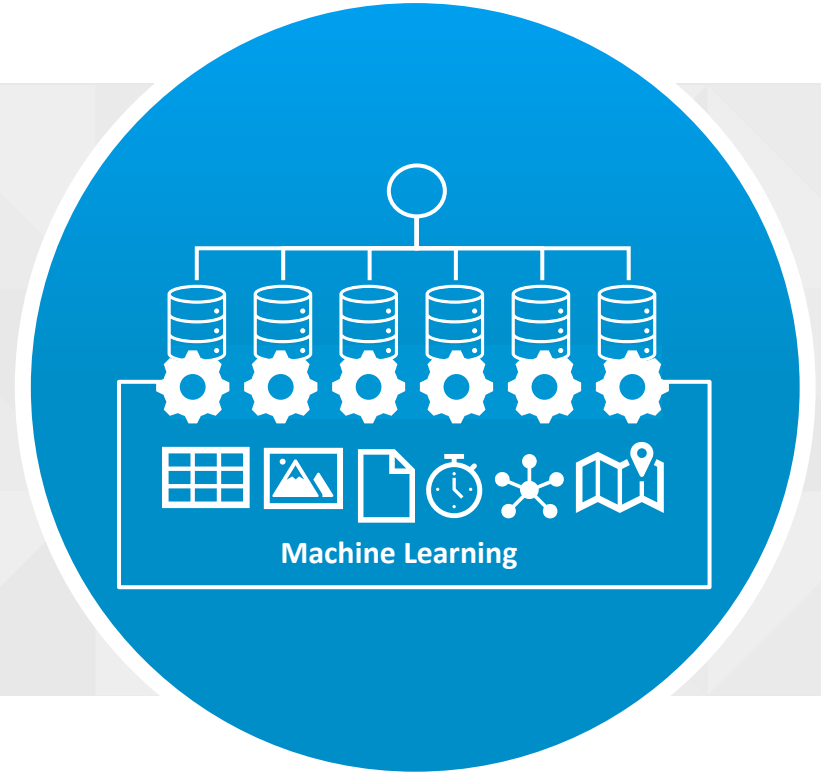
1. 성공적인 구축 사례

2. 핵심 적용 기술

3. 분석 시스템 구축 고려 사항



VMware Greenplum은 **초대용량**의 **정형 및 비정형** 데이터에 대한 분석력을 극대화한 데이터베이스로서 **AI/ML**을 위한 **데이터 분석 플랫폼**으로 계속 진화하고 있습니다

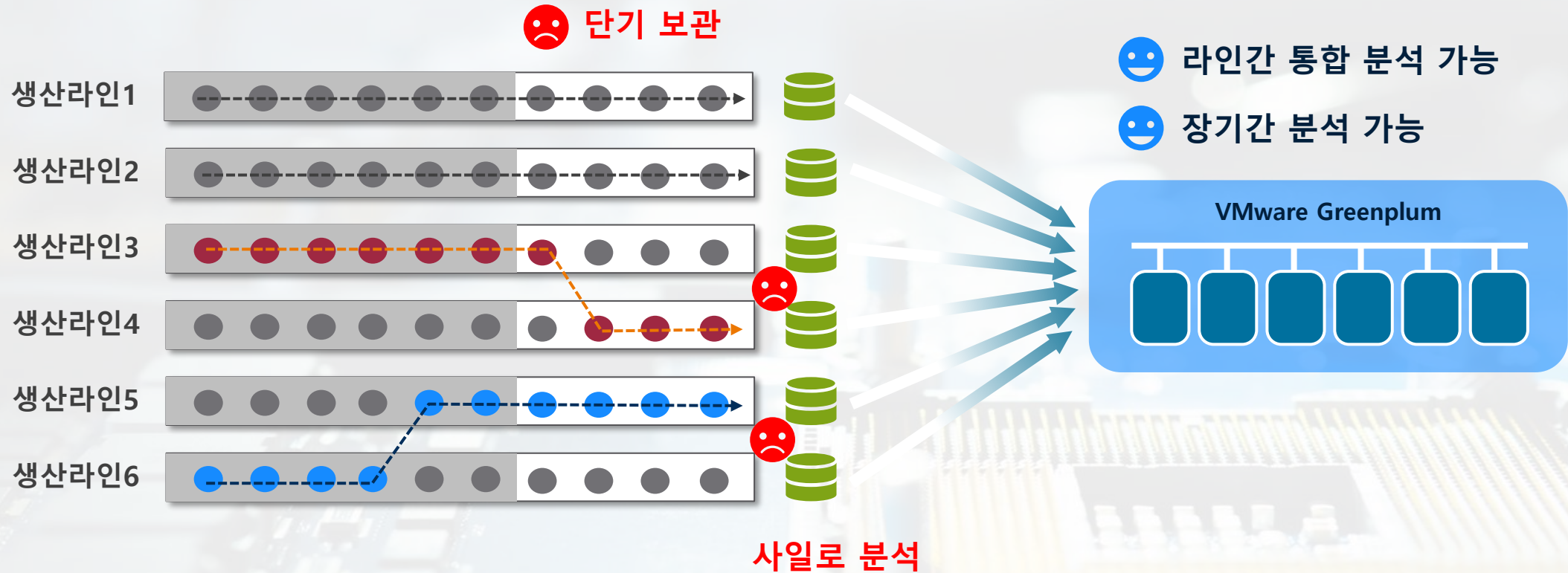


구축 사례

Big Data on Private Cloud

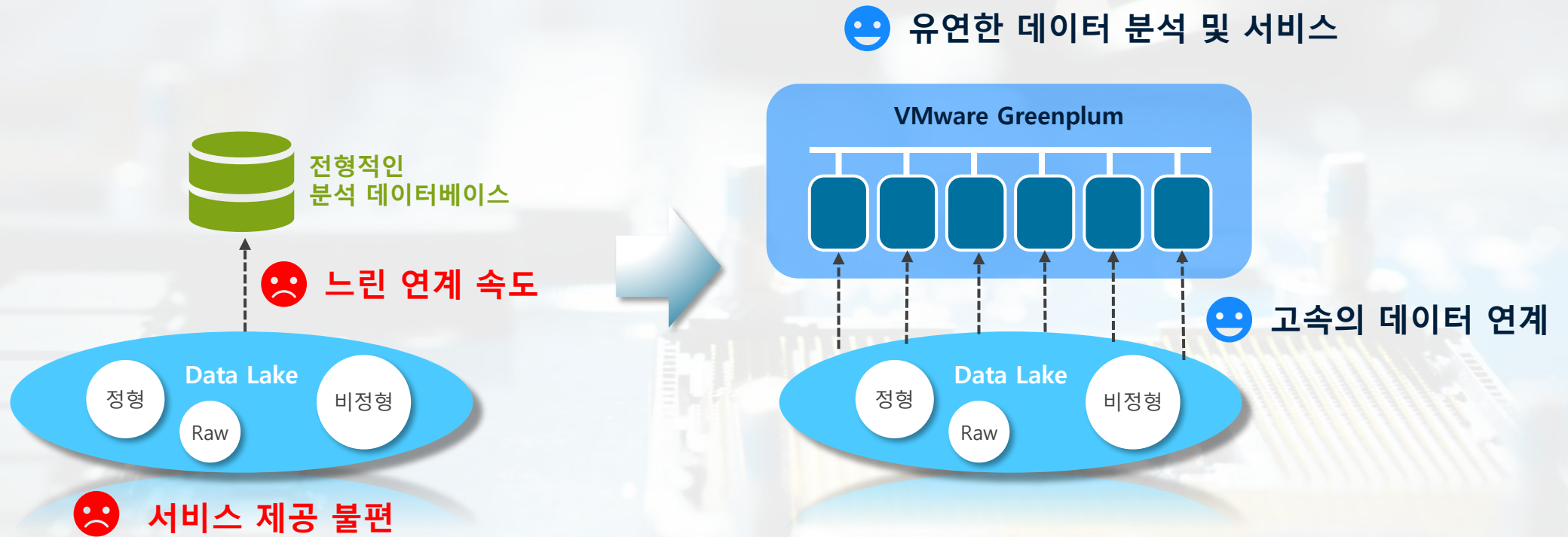
국내 하이테크 제조사의 10여년의 여정

분석 데이터베이스의 진화 - 단위 DW에서 통합 분석 플랫폼으로



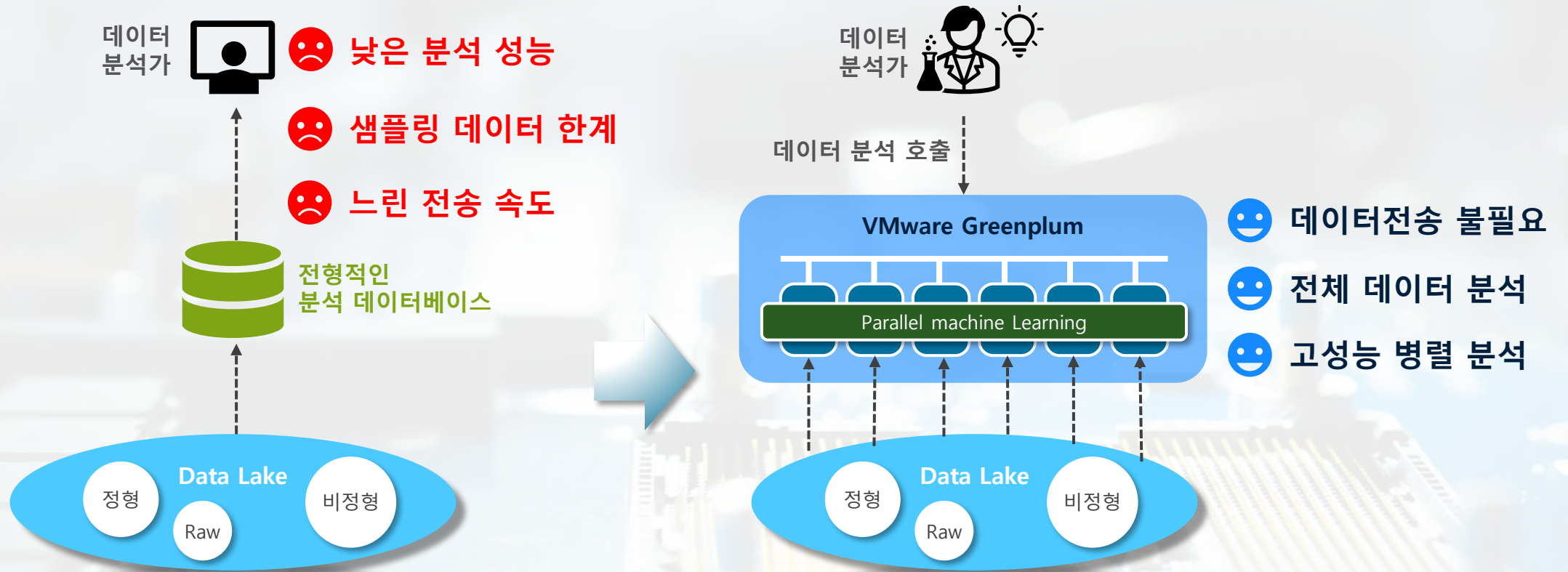
국내 하이테크 제조사의 10여년의 여정

분석 데이터베이스의 진화 – 데이터레이크와 고속의 연계



국내 하이테크 제조사의 10여년의 여정

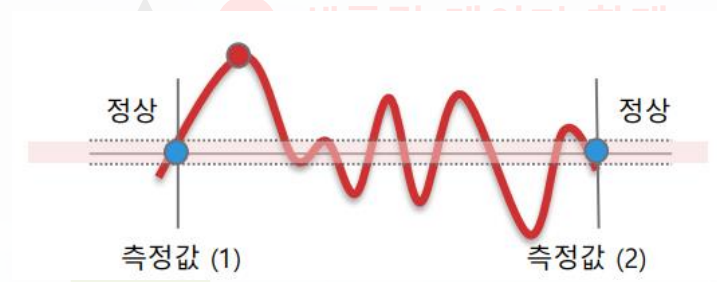
분석 데이터베이스의 진화 – in-Database 분석 활성화



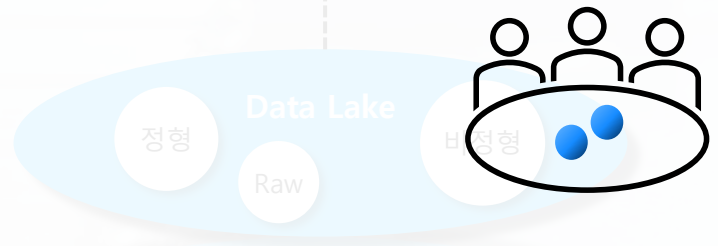
국내 하이테크 제조사의 10여년의 여정

분석 데이터베이스의 진화 – in-Database 분석 활성화

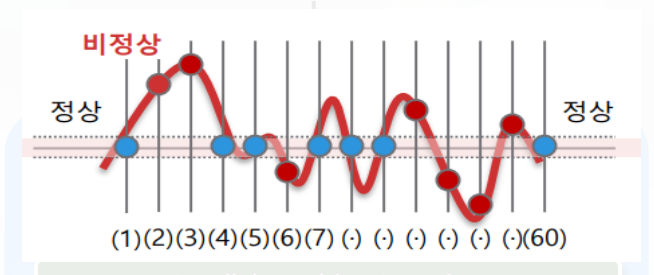
데이터 분석가  낮은 분석 성능




 샘플링 분석






데이터 분석가 



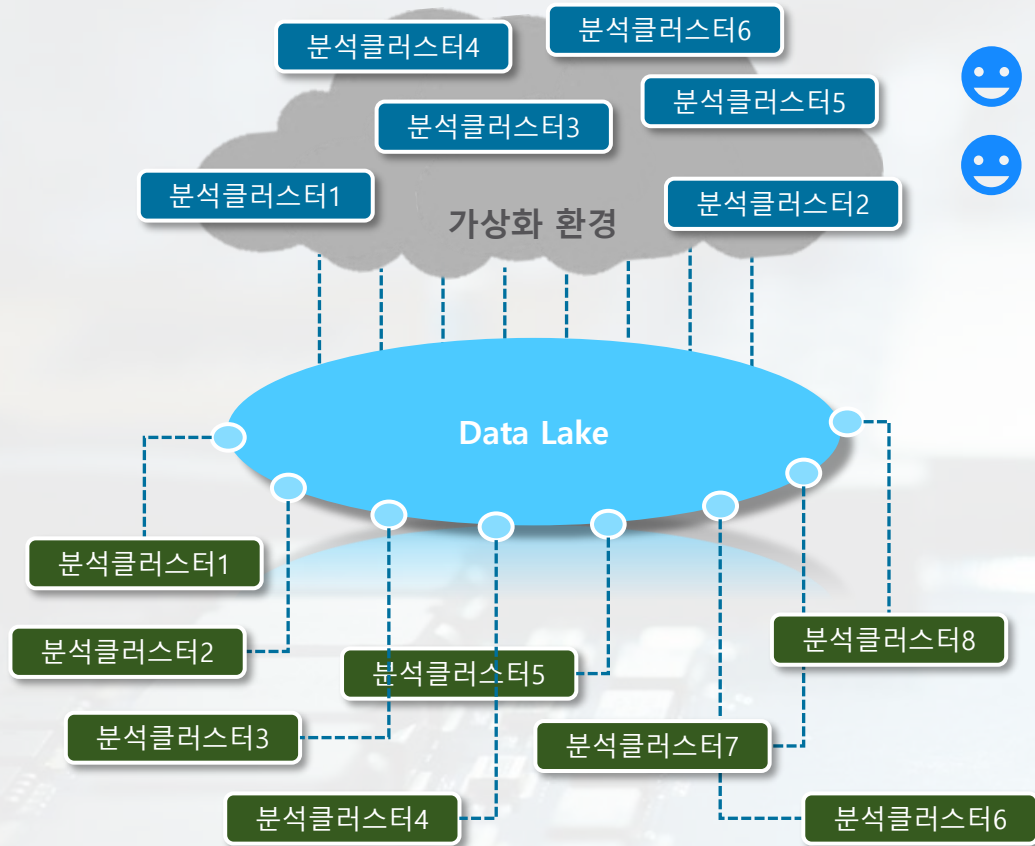
 전수 분석



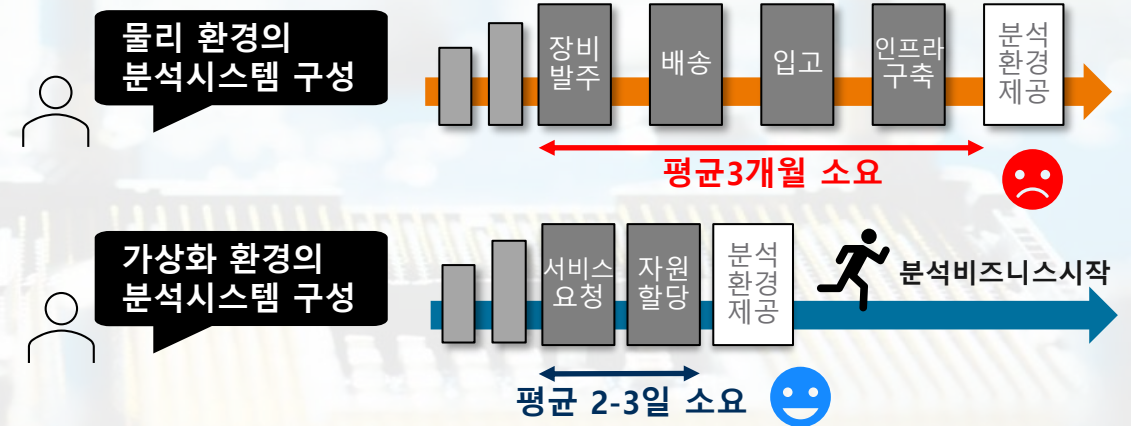
-  데이터 전송 불필요
-  전체 데이터 분석
-  고성능 병렬 분석

국내 하이테크 제조사의 10여년의 여정

분석 데이터베이스의 진화 - 분석 플랫폼 확산



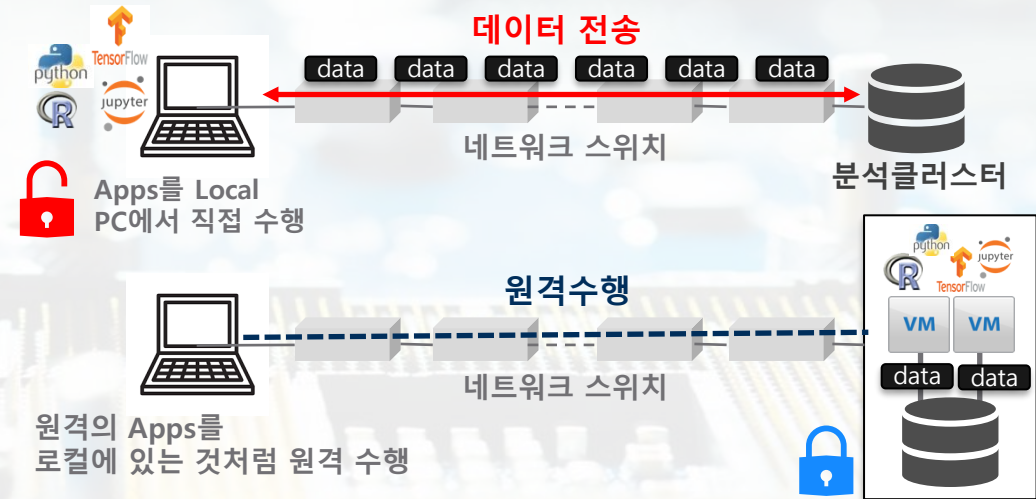
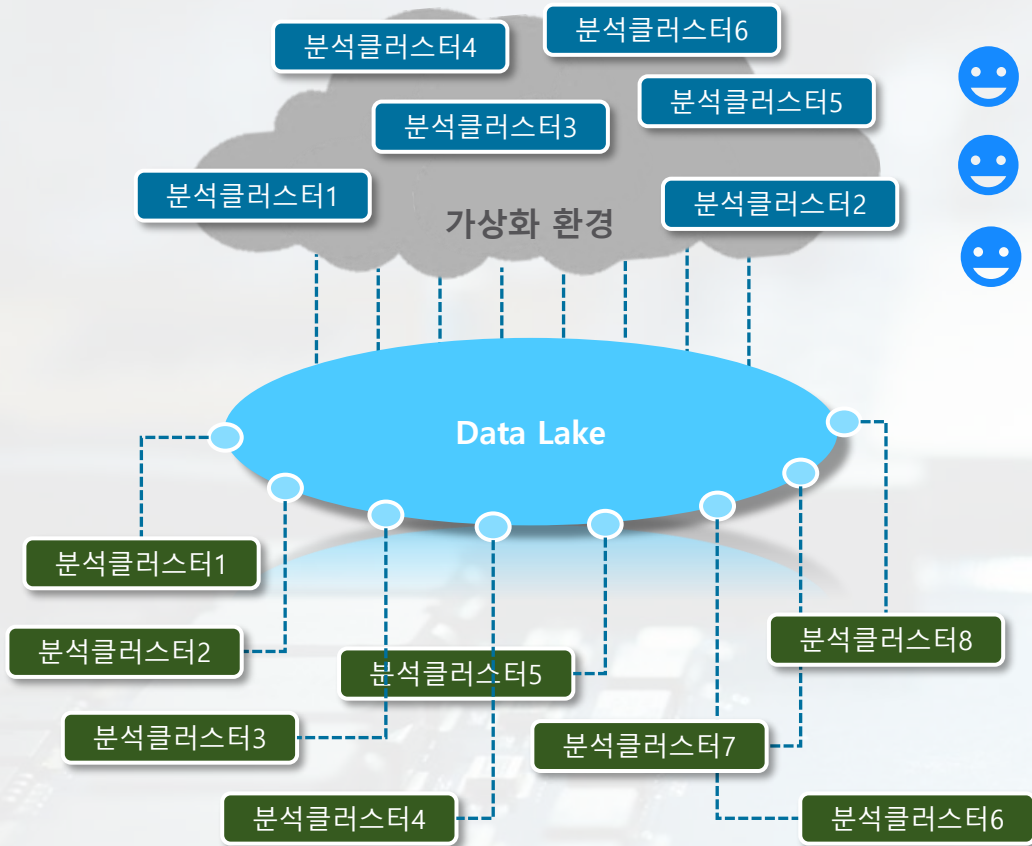
- 😊 유연한 분석 환경 즉시 제공
- 😊 시스템 자원 활용도 증가



국내 하이테크 제조사의 10여년의 여정

분석 데이터베이스의 진화 - 분석 플랫폼 확산

- 😊 유연한 분석 환경 즉시 제공
- 😊 시스템 자원 활용도 증가
- 😊 보안과 성능 최적화



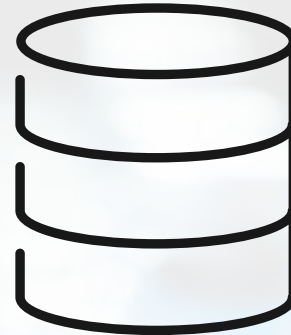
시스템 현황



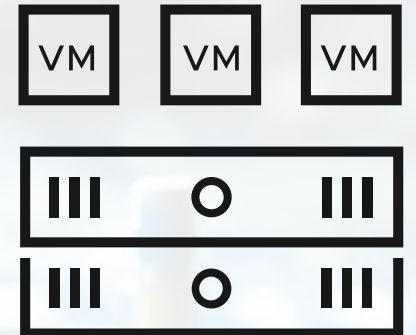
68 Clusters



681 VMs



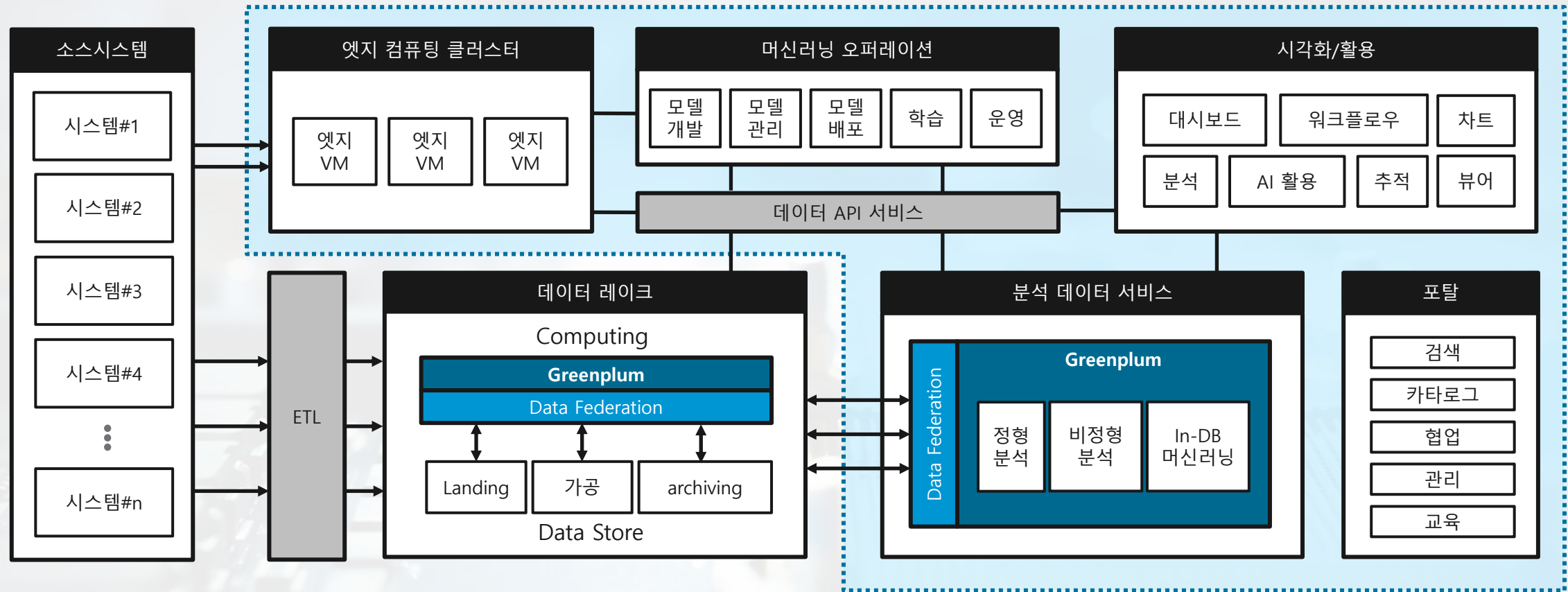
2.3 PB



100 VMs, 1.5PB

VMware Big Data Architecture Reference

SDDC 플랫폼 소프트웨어, DW 및 App 가상화를 활용하여 Cloud향 Big Data 분석 환경 구현



핵심 적용 기술

신속한
인프라 구성과
유연한 운영

성능과 보안 강화

데이터레이크
고속 연계

최적화된 빅데이터
모델링 적용

핵심 적용 기술

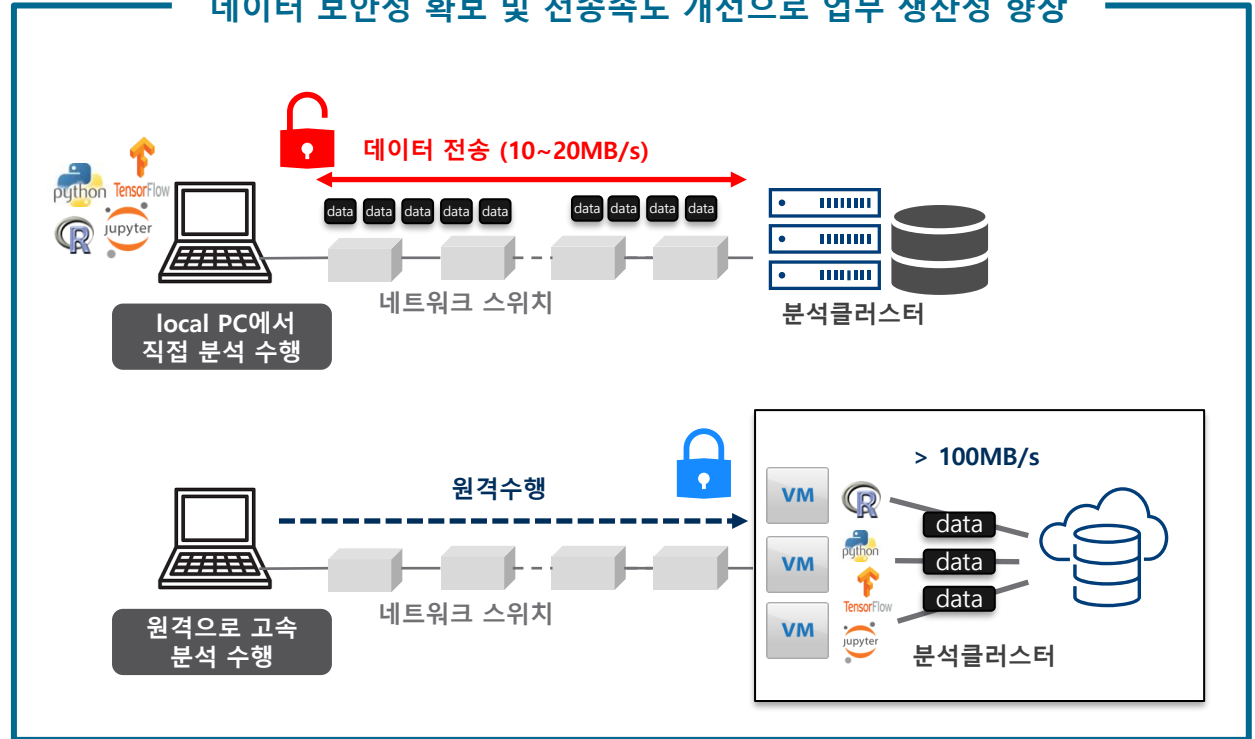
VMware 가상화 기반의 분석 플랫폼 구축



인프라 구성 시간 단축으로 비즈니스 민첩성 향상



데이터 보안성 확보 및 전송속도 개선으로 업무 생산성 향상



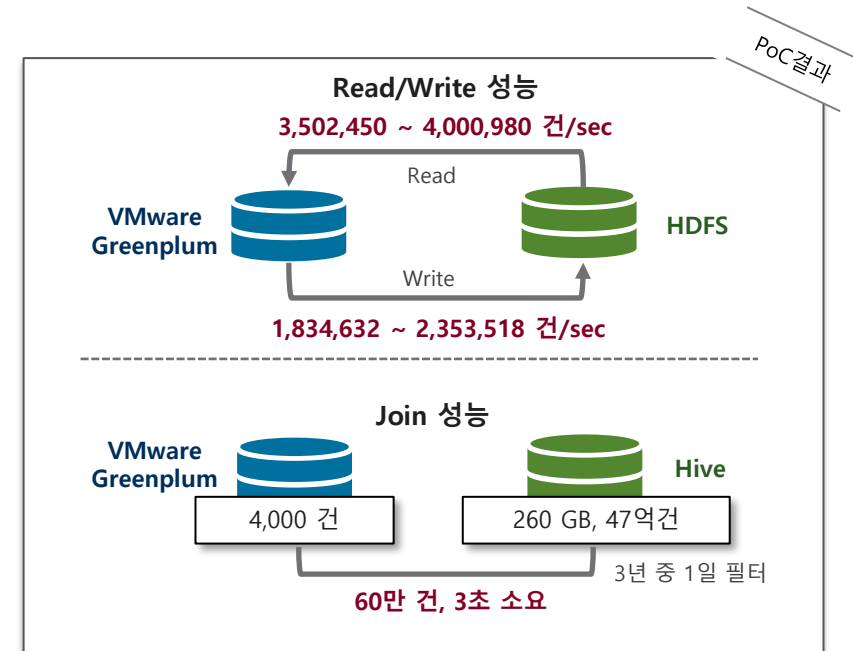
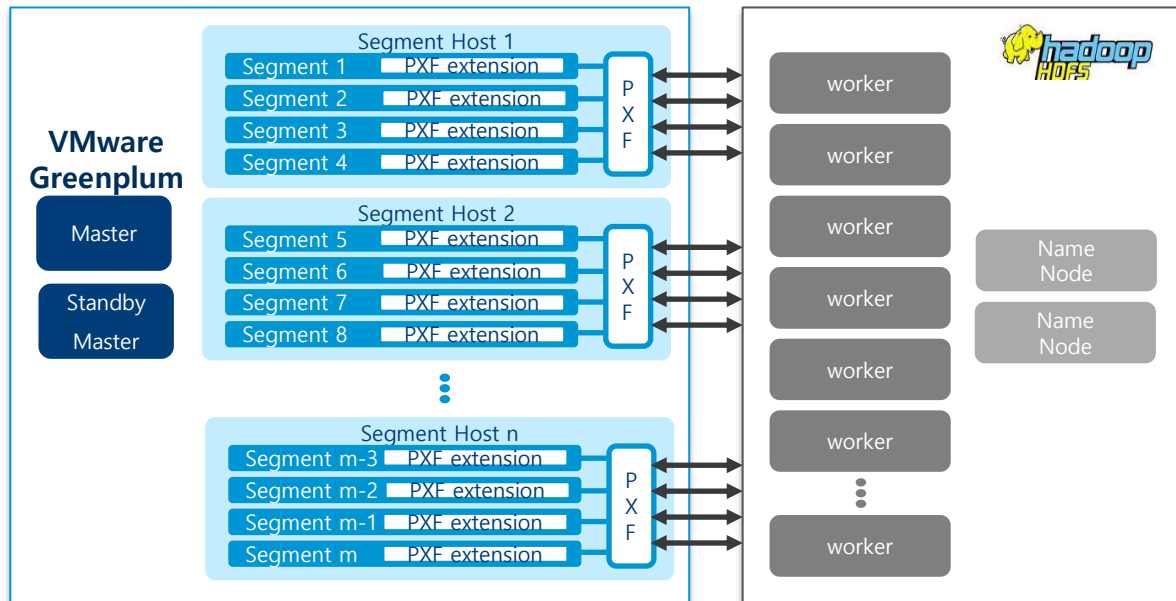
핵심 적용 기술

VMware Greenplum의 데이터 레이크 병렬 연계



데이터레이크
고속 연계

데이터레이크 데이터 고속 적재, 장기 보관 데이터 직접 조회



- 최적화된 빅데이터 모델링 적용
- 신속한 인프라 구성과 유연한 운영
- 성능과 보안 강화

핵심 적용 기술

빅데이터 처리를 위한 효과적인 데이터 모델링 적용



최적화된
빅데이터
모델링 적용

빅데이터 분석 및 추적에 최적화된 분석 전용 통합 모델링 필요

OLTP 정규화 모델링



목적

- 불필요한 중복 제거
- 정합성, 일관성 보장

물리설계

- 다수의 테이블과 인덱스
- 테이블별 최소한의 컬럼 구성

고려사항

- 다수의 복잡한 테이블 조인 필요
- Ad-hoc 쿼리의 복잡도 증가

신속한
인프라 구성과
유연한 운영

성능과 보안
강화

데이터레이크
고속 연계

핵심 적용 기술

빅데이터 처리를 위한 효과적인 데이터 모델링 적용



최적화된
빅데이터
모델링 적용

빅데이터 분석 및 추적에 최적화된 분석 전용 통합 모델링 필요

OLTP 정규화 모델링



DW 비정규화 모델링



목적

- 불필요한 중복 제거
- 정합성, 일관성 보장

물리설계

- 다수의 테이블과 인덱스
- 테이블별 최소한의 컬럼 구성

고려사항

- 다수의 복잡한 테이블 조인 필요
- Ad-hoc 쿼리의 복잡도 증가

- 테이블 조인 최소화
- 쿼리 복잡도 감소

- 테이블 통합
- 테이블별 컬럼 수 증가

- 중복 데이터 증가
- 인덱스, 저장 공간 등 유지 관리

신속한
인프라 구성과
유연한 운영

성능과 보안
강화

데이터레이크
고속 연계

핵심 적용 기술

빅데이터 처리를 위한 효과적인 데이터 모델링 적용



최적화된
빅데이터
모델링 적용

빅데이터 분석 및 추적에 최적화된 분석 전용 통합 모델링 필요

OLTP 정규화 모델링



DW 비정규화 모델링



빅데이터 직렬화 모델링



목적

- 불필요한 중복 제거
- 정합성, 일관성 보장

물리설계

- 다수의 테이블과 인덱스
- 테이블별 최소한의 컬럼 구성

고려사항

- 다수의 복잡한 테이블 조인 필요
- Ad-hoc 쿼리의 복잡도 증가

- 테이블 조인 최소화
- 쿼리 복잡도 감소

- 테이블 통합
- 테이블별 컬럼 수 증가

- 중복 데이터 증가
- 인덱스, 저장 공간 등 유지 관리

- 특정 레코드 검색 성능 향상
- 머신러닝 분석 활용

- 배열구조로 전환
- 그룹핑 컬럼 선정

- 비즈니스 도메인 지식
- 물리모델 관리

신속한
인프라 구성과
유연한 운영

성능과 보안
강화

데이터레이크
고속 연계

분석 플랫폼 구축 고려사항

검증된 최신
트렌드를 반영하는
기술요소 접목

- 멀티 클라우드
- 앱 가상화
- 빅데이터 모델링
- 병렬처리
- 머신 러닝
- 각종 기술의 장기적 로드맵

데이터 분석
흐름을 고려한
전체 아키텍처

- 데이터 수집,적재
- 데이터레이크
- 데이터분석
- 시각화 및 활용
- 데이터 사이언스
- 데이터 API 서비스
- 각 시스템간 효율적 연동

유연하면서 확장
가능한 인프라
구성

- 데이터 병렬처리 활용 구조
- 병목없는 대용량 데이터 처리
- 유연한 운영 관리 구성

경험있는 전문
인력을 통한
시행착오 최소화

- 전문 인력의 빠른 테스트와 템플릿 구성 및 확산
- 각 분야별 전문 인력의 협업을 통한 시너지

Thank you