

빅 데이터를 위한 데이터 거버넌스 전략

한국IBM 소프트웨어그룹, 정보관리사업부(Information Management)
조가원 차장(kwcho@kr.ibm.com)



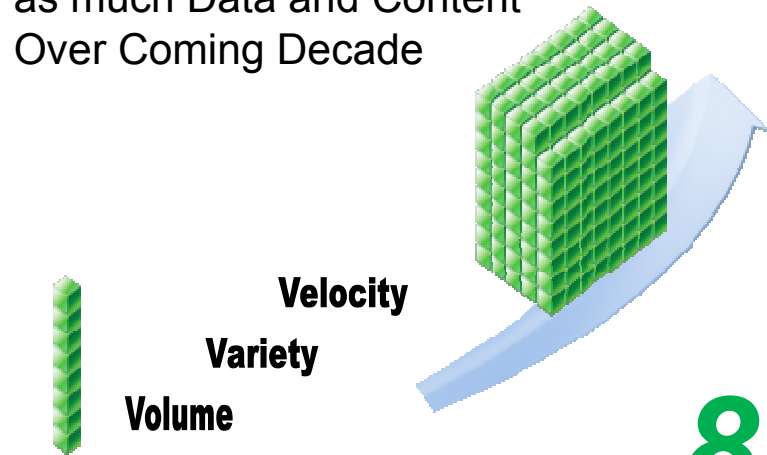
빅 데이터 시대의 도래

빅 데이터 시대, 폭증하는 데이터

44x

as much Data and Content
Over Coming Decade

2020
35 zettabytes



2009
800,000 petabytes

80%

Of world's data
is unstructured



1 in 3 업무 담당자 중 1/3은 의사 결정에 필요한 정보를 제대로 얻지 못하거나 신뢰할 수 없는 경우가 대부분이라고 말한다.

1 in 2 업무 담당자 중 1/2는 업무에 필요한 정보에 대한 권한이 없다고 말한다.

83% CIO 중 83%는 BI와 분석이 경쟁력 강화를 위한 기업의 전략 계획 중 하나라고 말한다.

60% CEO 중 60%는 경영 의사 결정을 위해 빠르게 데이터를 캡처/이해하는 향상된 기술이 필요하다.


빅 데이터를 활용한 새로운 통찰력이 필요합니다.




다채널 고객 성향 및 지식 분석



병원에서 즉각적으로 생명에 위협을 주는 양상 감지



풍력 발전 최적화 계획을 위한 기상 패턴 예측 및 풍력기 배치를 위한 설비 투자 최적화



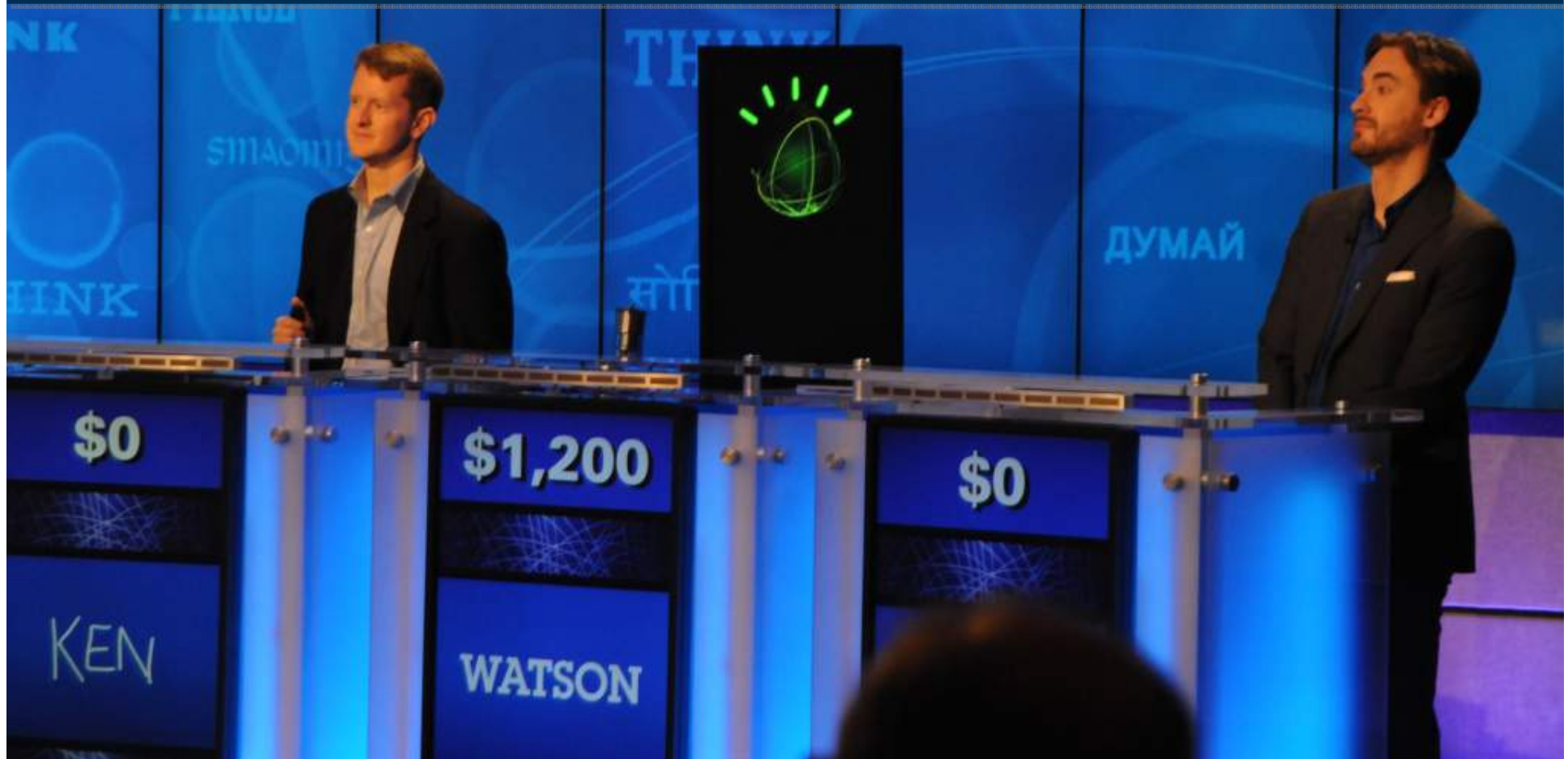
실시간 트랜잭션 데이터에 기반한 위험 요소 분별



서로 다른 비디오, 음성, 데이터 소스로부터 범죄 및 위험 요소 인식

Watson을 통해 살펴보는 새로운 도전

3초 이내 정답을 맞출 수 있는 높은 가능성을 가지고 있으며,
실시간으로 엄청난 양의 정보를 검색, 분석하고 해석하는,
인간의 능력과 경쟁할 수 있는 컴퓨팅 시스템을 고안할 수 있을까?



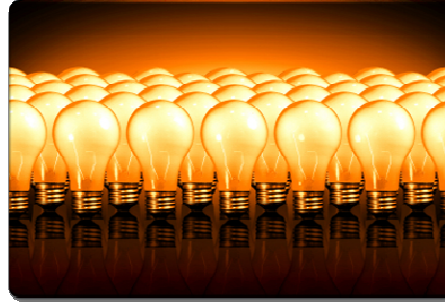
만약 이 모든 해답을 모두 가지고 있다면?



어떤 고객이
이탈을
고려하고
있는가?



어떤 트랜잭션이
사기로 청구된
것인가?



어떤 신규
상품이 성공
확률이 가장
높은가?



어떻게 나의
모든
정보로부터
통찰력을
얻어낼
것인가?

빅 데이터 시대에 기업의 차별화 요소는

기업의 매출을 예측하고 창출시킬 수 있는 정보에 기반한 의사결정을 할 수 있는 기술을 가지고 있는가에 있다.

데이터의 관리 이슈

- 복잡성 증가**
- 품질 저하**
- 정보 보호**
- 비용 절감**

“데이터의 일관성이 부족해요.”
North American Multi-Line Insurer

“데이터의 품질 관리가 안 되고 있어요.”
United States Government Agency

“데이터를 안전하게 보호할 정책과 프로세스가 필요해요.”
Healthcare Insurer

“어쩔 수 없이 모든 데이터를 영구 저장하고 있어요.”
A large chemical manufacturer

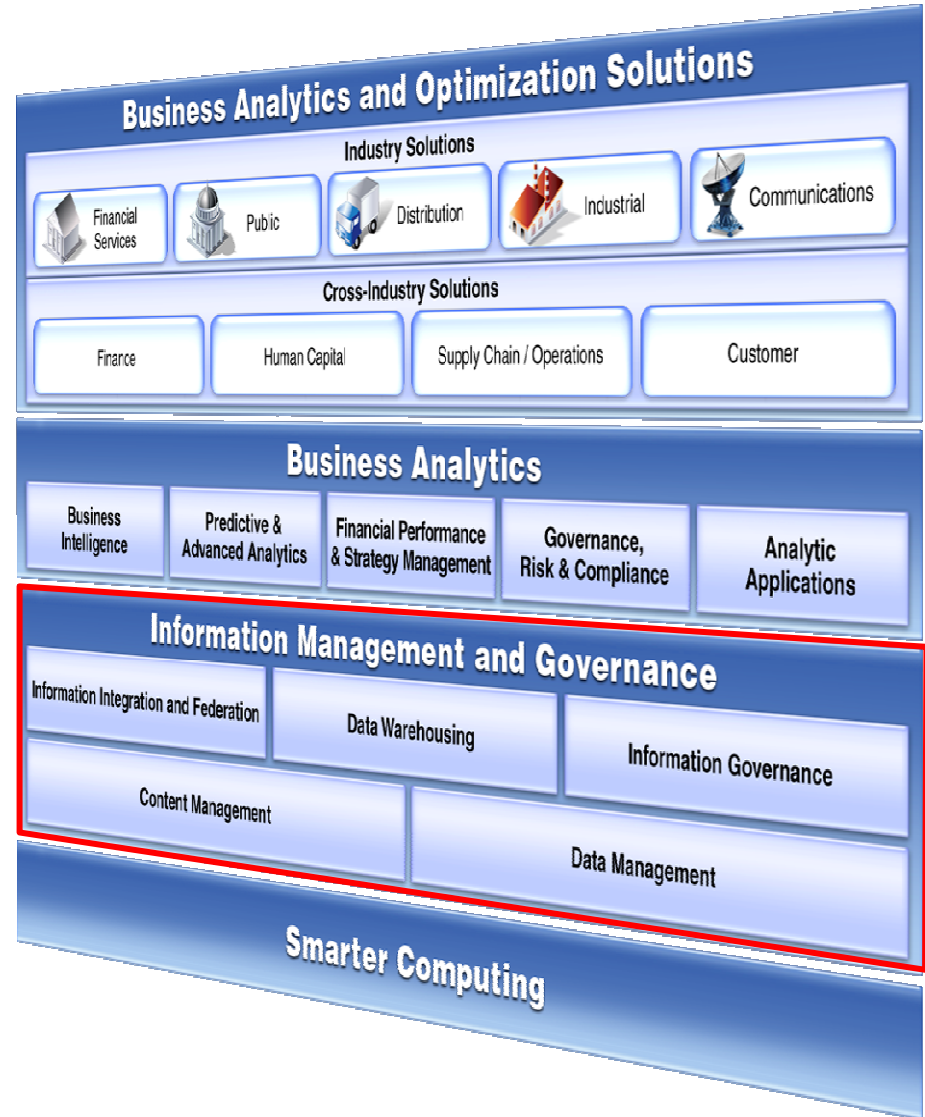
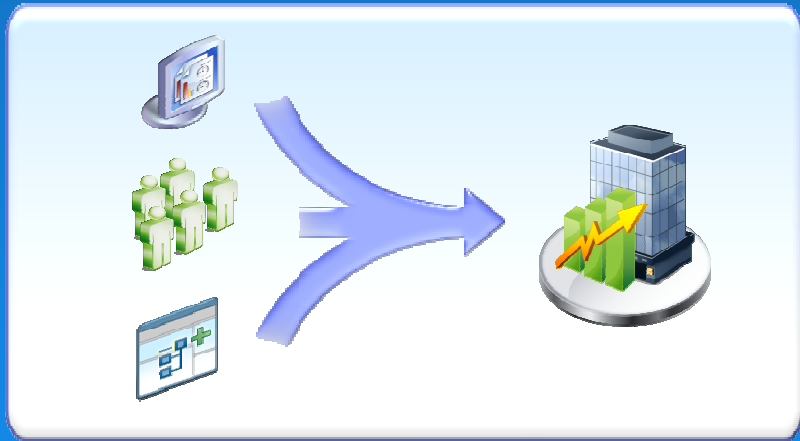
“데이터 증가를 관리할 수 있는 시스템적인 방법이 필요해요.”
CFO Survey

데이터 거버넌스

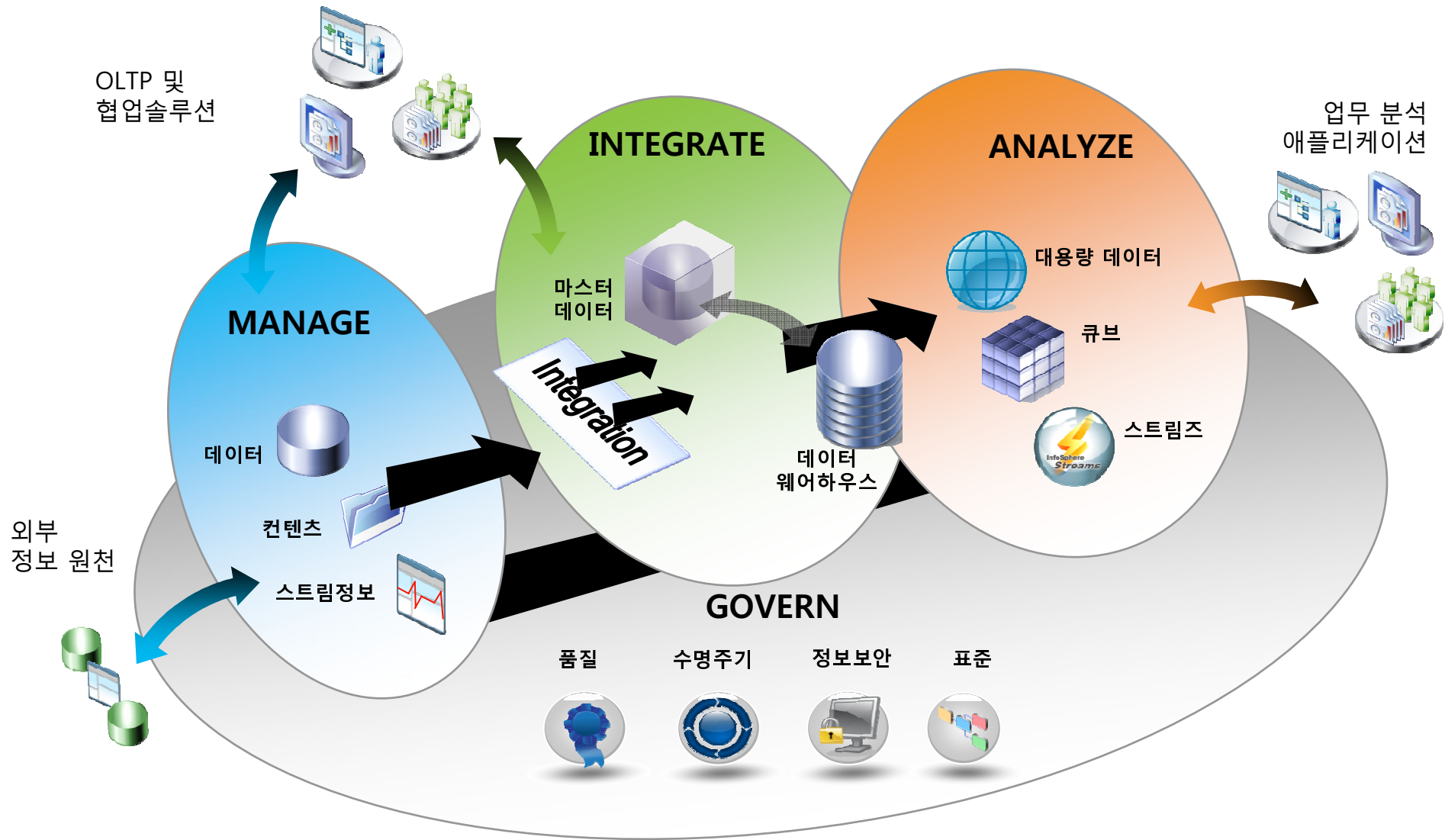
빅 데이터와 데이터 거버넌스

데이터 거버넌스는

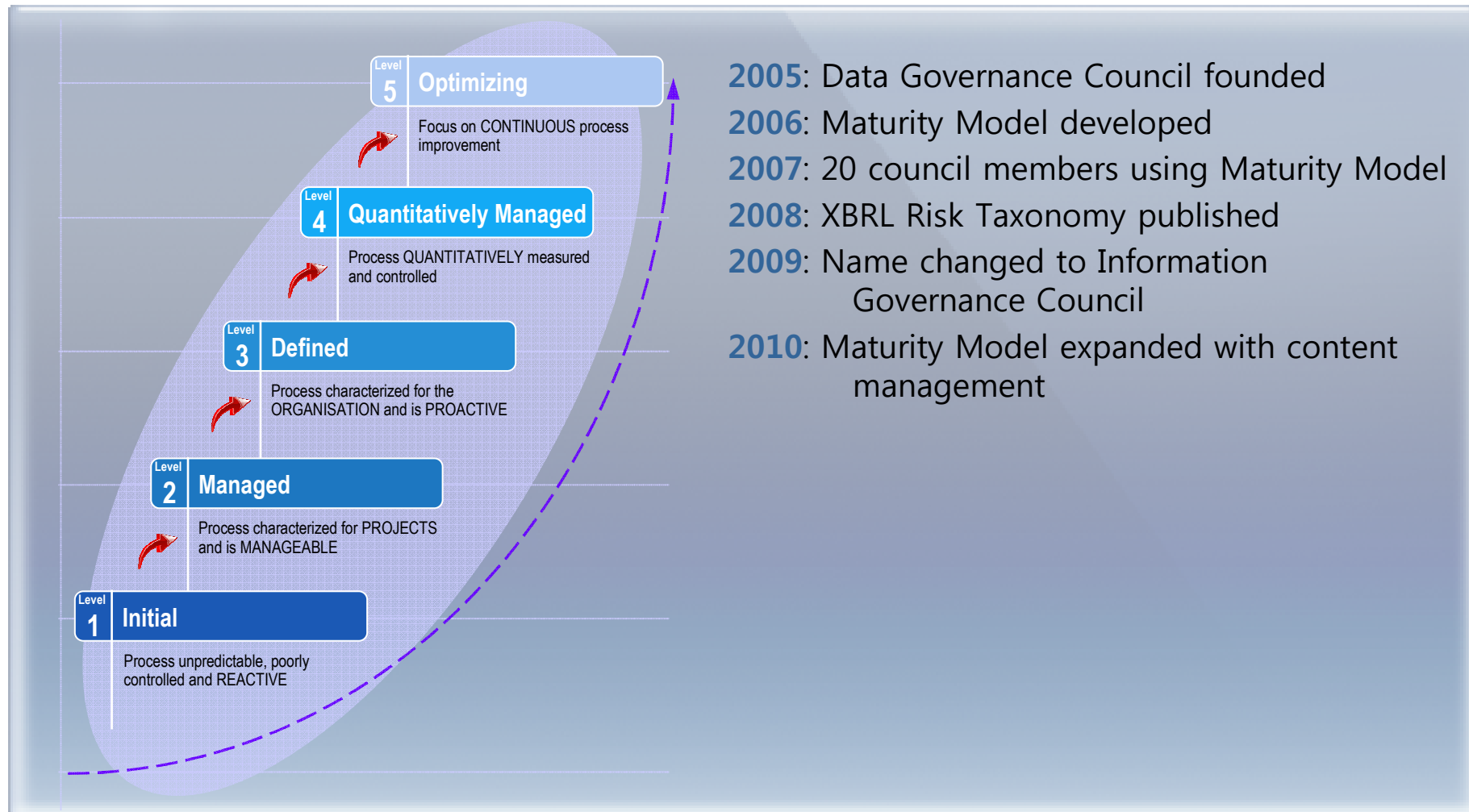
기업의 조직, 프로세스, 기반기술 등을 효과적으로 관리하여, 데이터를 기업의 비즈니스를 위한 핵심 자산으로 최적화하고, 보호하며, 활용하기 위한 방안



정보 공급망과 데이터 거버넌스



IBM의 데이터 거버넌스에 대한 준비



데이터 거버넌스 성숙도 모델



기업의 품질 관리 이슈

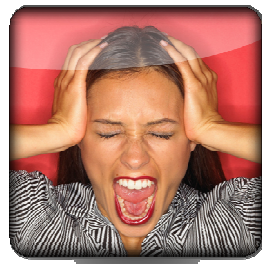
Data Quality Management



- 규제 준수 및 감사 투명성을 위한 데이터 품질 이슈 압박
 - 데이터의 품질 제고 방안 부재



- 신뢰할 수 없는 데이터에 대한 통찰력
 - 낮은 데이터 품질로 인해 업무 의사 결정에 있어서 적정 산출물이나 신뢰도 도출이 어려움
 - 품질 이슈의 원인 감지에 대한 불확실성
 - 신뢰할 수 없는 데이터로 산출된 업무 의사 결정이 다른 전략적 업무에 영향을 미침

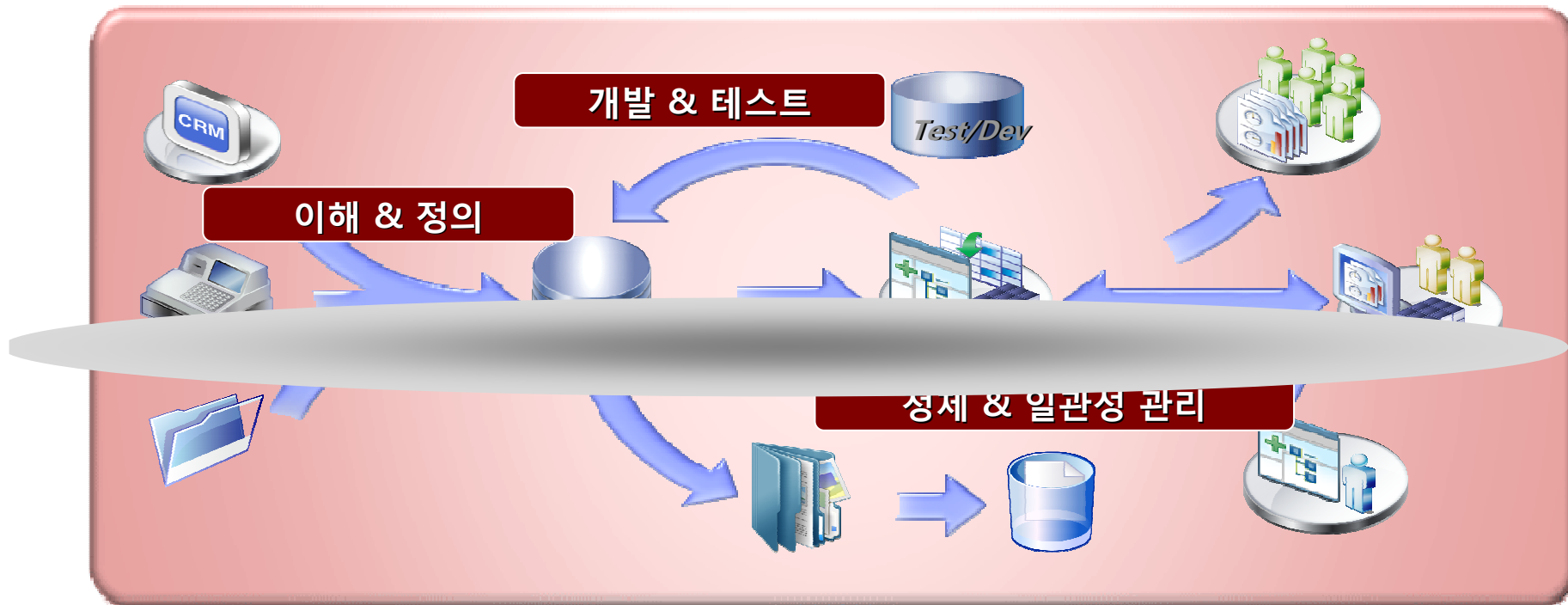


- 비용 증가 및 고객 불만족 증가
 - 매출 창출 기회 상실에 대한 직접 손실
 - 낮은 데이터 품질로 인한 늦은 대응으로 인한 고객 만족 저하

데이터 거버넌스 기반 품질 관리

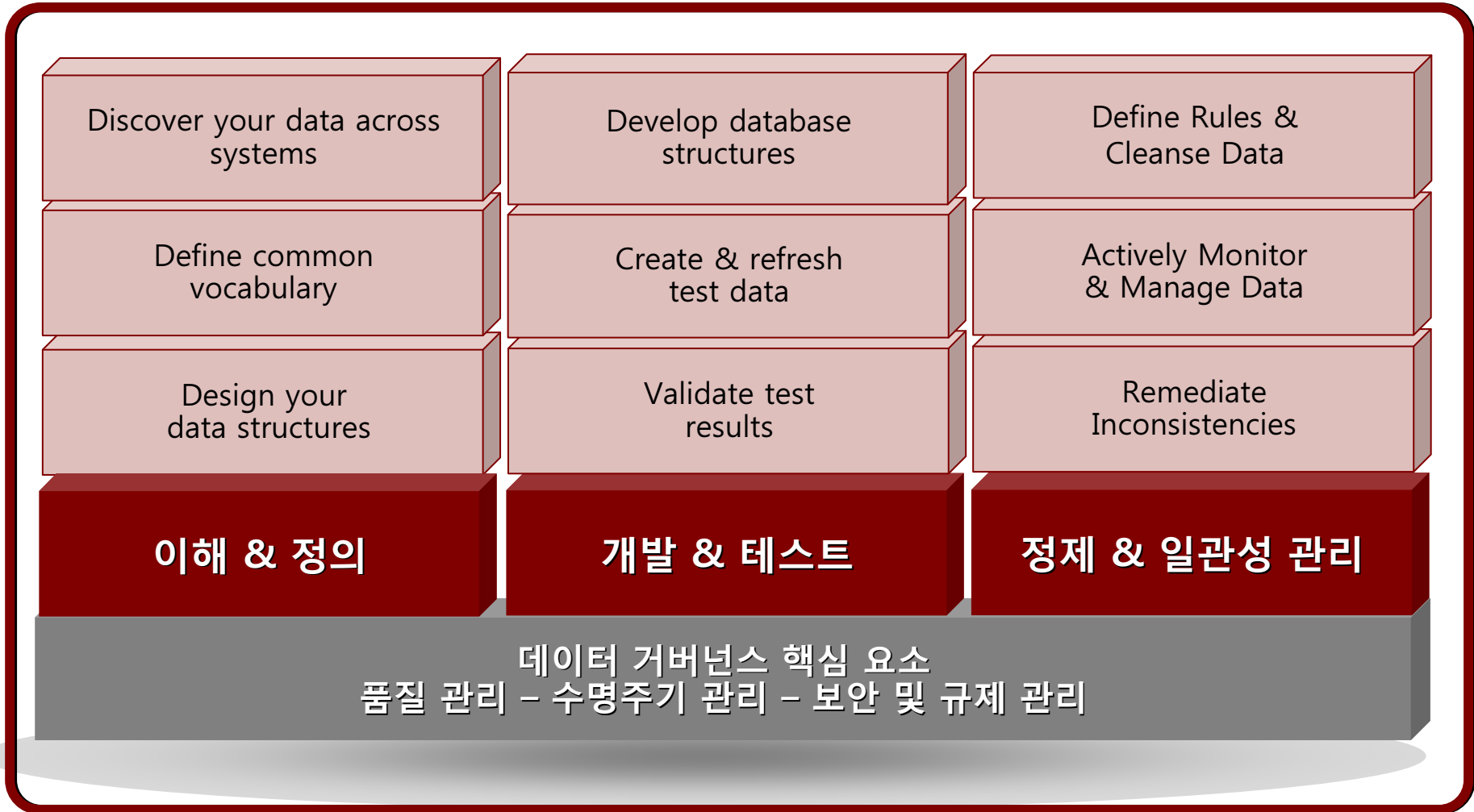
Data Quality Management

- 업무 내 "고품질" 데이터에 대한 정의
- 개인정보보호 정책 기반 실제적이고 재사용 가능한 테스트 데이터 생성
- 전략적 과제 수행을 위한 표준화 및 데이터 정제
- 지속적인 데이터 품질 모니터링 및 예외 처리 프로세스 이해
- 데이터의 이행 및 저장 경로 이해



데이터 품질 관리를 통한 통찰력의 실행

Data Quality Management



기업의 수명주기 관리 이슈

Information Life-cycle Management



- 신규 업무에 필요한 애플리케이션 개발 일정 지연
 - 반복되는 프로젝트 내 데이터 오브젝트 간 연관성 이해 부족
 - 폭증하는 데이터 규모로 인해 복제, 테스트, 검증에 걸리는 시간 증가



- 운영 및 인프라 비용에 대한 IT 예산 부담 증가
 - 운영 사본을 위해 DB 복제에 드는 하드웨어 비용 증가
 - 백업 등 유지보수에 필요한 운영 인력 생산성 저하
 - CPU 등 증설로 인한 추가 라이선스 비용 증가
- 이관 이후 애플리케이션 오류 발생
 - 개발 환경 대비 10-100 배 소요되는 운영 환경 오류 해결 비용



- 개발/테스트 환경에서의 기밀 정보 노출

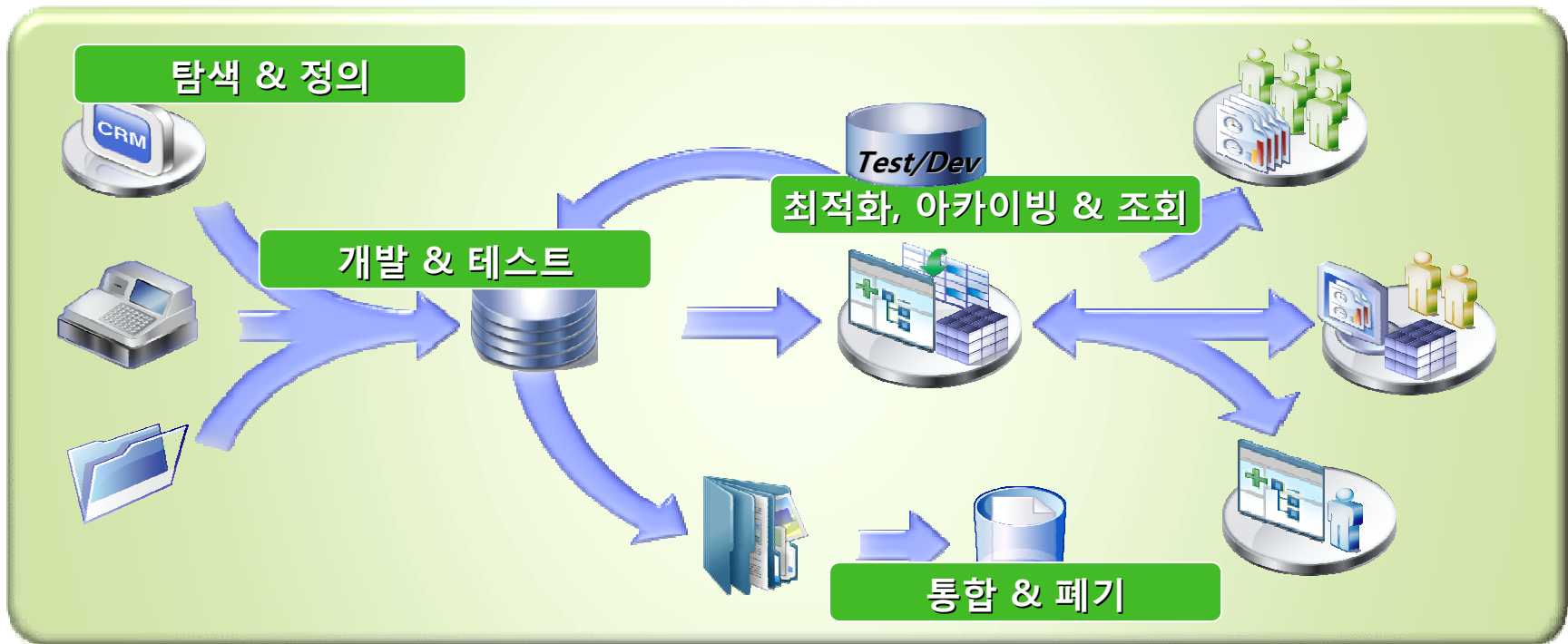
“ DB 내 저장된 85%의 데이터는 비활성 데이터 이다. Forrester

Source: Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, 8/13/07

데이터 거버넌스 기반 수명주기 관리

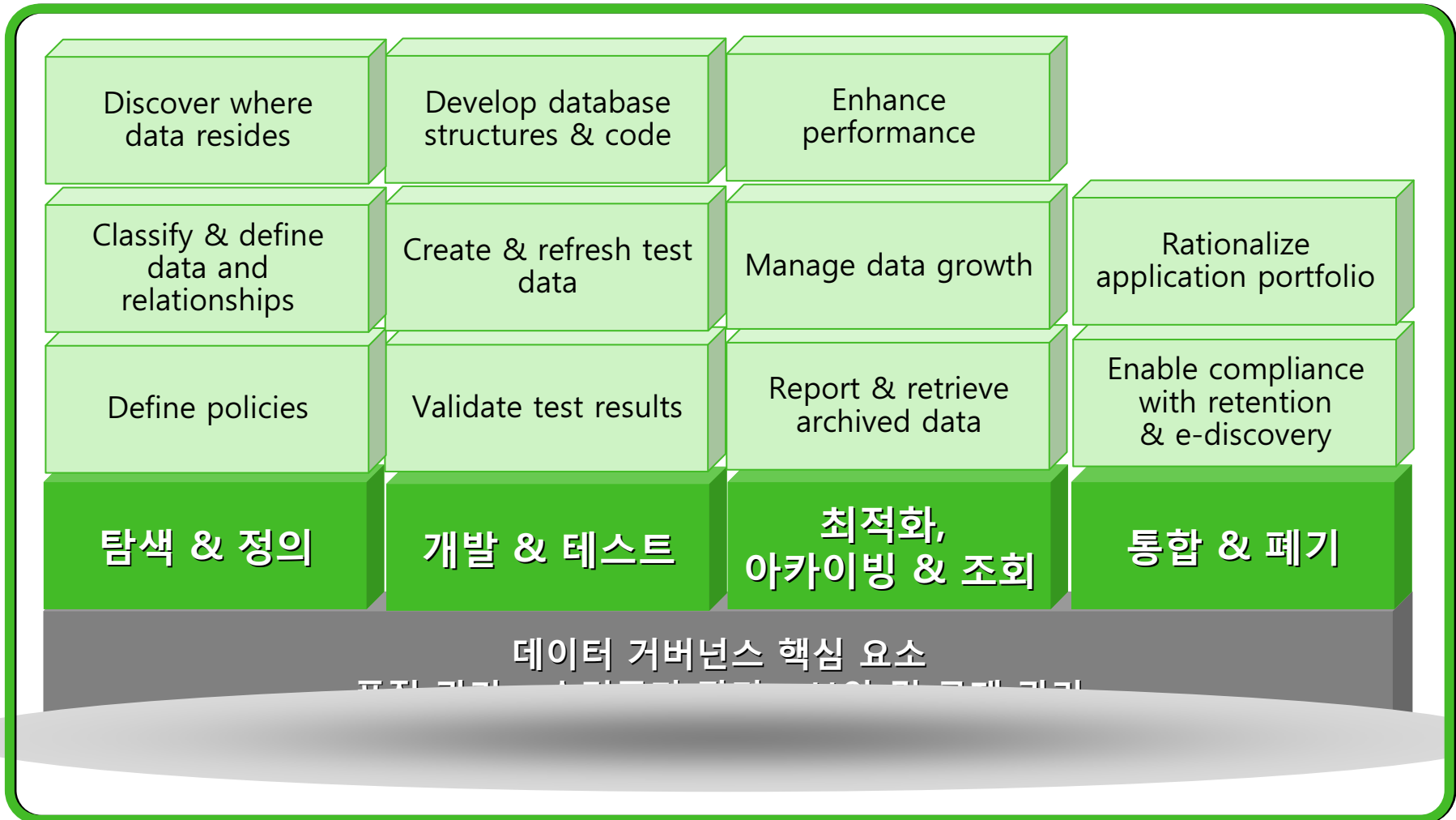
Information Life-cycle Management

- 전사 데이터 "용도" 및 "위치" 이해
- 테스트 환경 데이터 구성을 포함한 전사 데이터 저장과 조회를 위한 모델 개발
- 애플리케이션 내 병목 인식을 통한 성능 최적화
- 데이터 증가 규모 관리를 위한 적정 전략 수립
- 사용이 완료된 데이터에 대한 폐기/통합을 위한 일관성 있는 프로세스 구현



데이터 수명주기 관리를 통한 통찰력의 실행

Information Life-cycle Management



기업의 보안 및 규제 준수 이슈

Information Security & Privacy



- 시간 제약, 수많은 규제, 규제 준수를 위한 비용 증가
 - 업무 성장에 따른 다양한 법적 규제 저축
 - 적정 시간과 비용 안에서 개인정보보호 규제 준수



- 사용자 권한 기반 개인정보 보호에 대한 복잡성 증가
 - 민감한 개인정보 데이터에 대해 직군별 전사 데이터 조회 분리 필요
 - 권한 요건 충족을 위해 임시성 솔루션에서의 데이터 일부 추출/복제

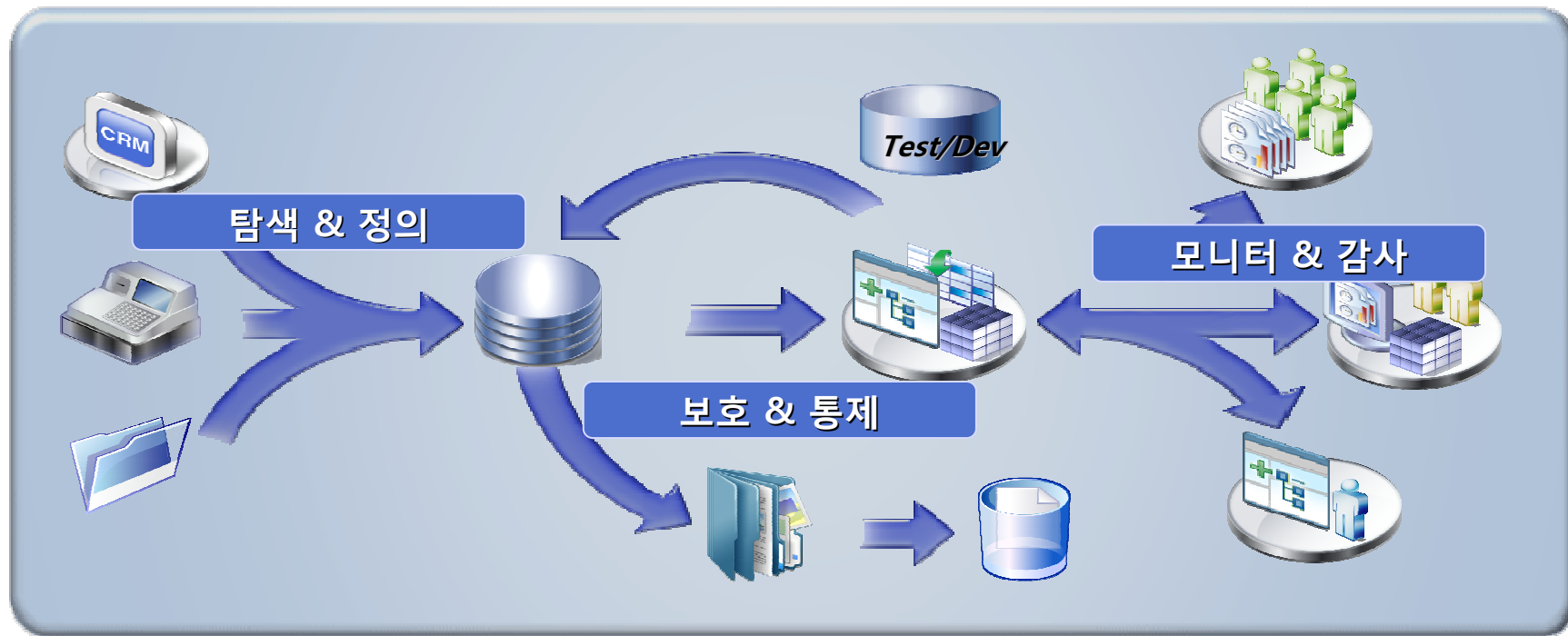


- 위험과 비효율성이 증가되는 수작업 현황
 - 정형/비정형 데이터 마스킹을 위한 비효율적인 자체 개발 프로그램
 - 민감한 데이터 탐색, 감사 수행 및 사용자 접근 모니터링을 위한 복잡한 수작업 프로세스

데이터 거버넌스 기반 보안 & 규제 준수

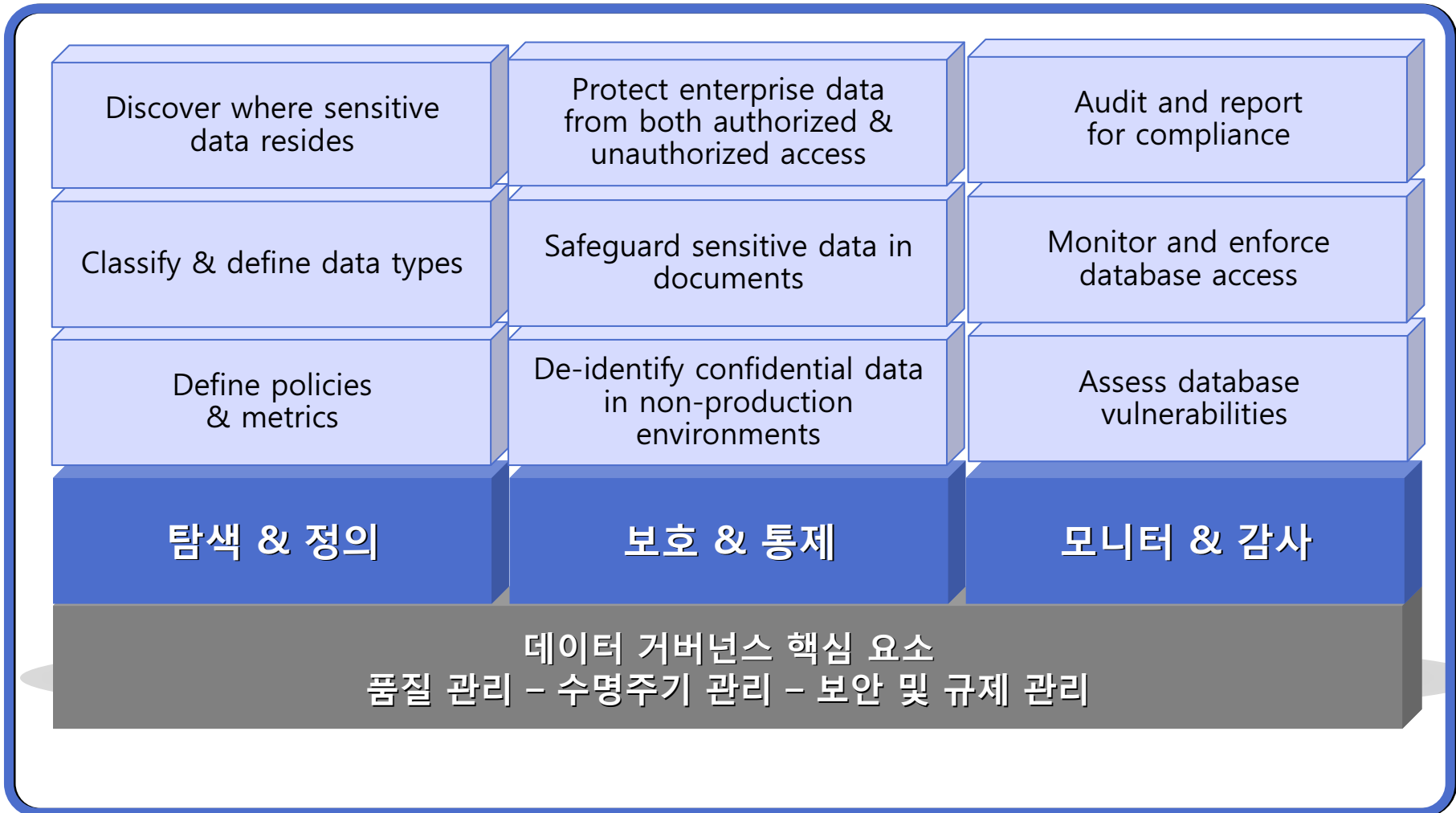
Information Security & Privacy

- 전사 데이터 "용도" 및 "위치" 이해
- 대내외 유출 방지를 위한 기업 내 전사 데이터의 보호
- 데이터 접근 기록: "누가", "언제", "어떻게"
- 감사 프로세스를 위한 DB 접근 내역 모니터링 및 리포팅

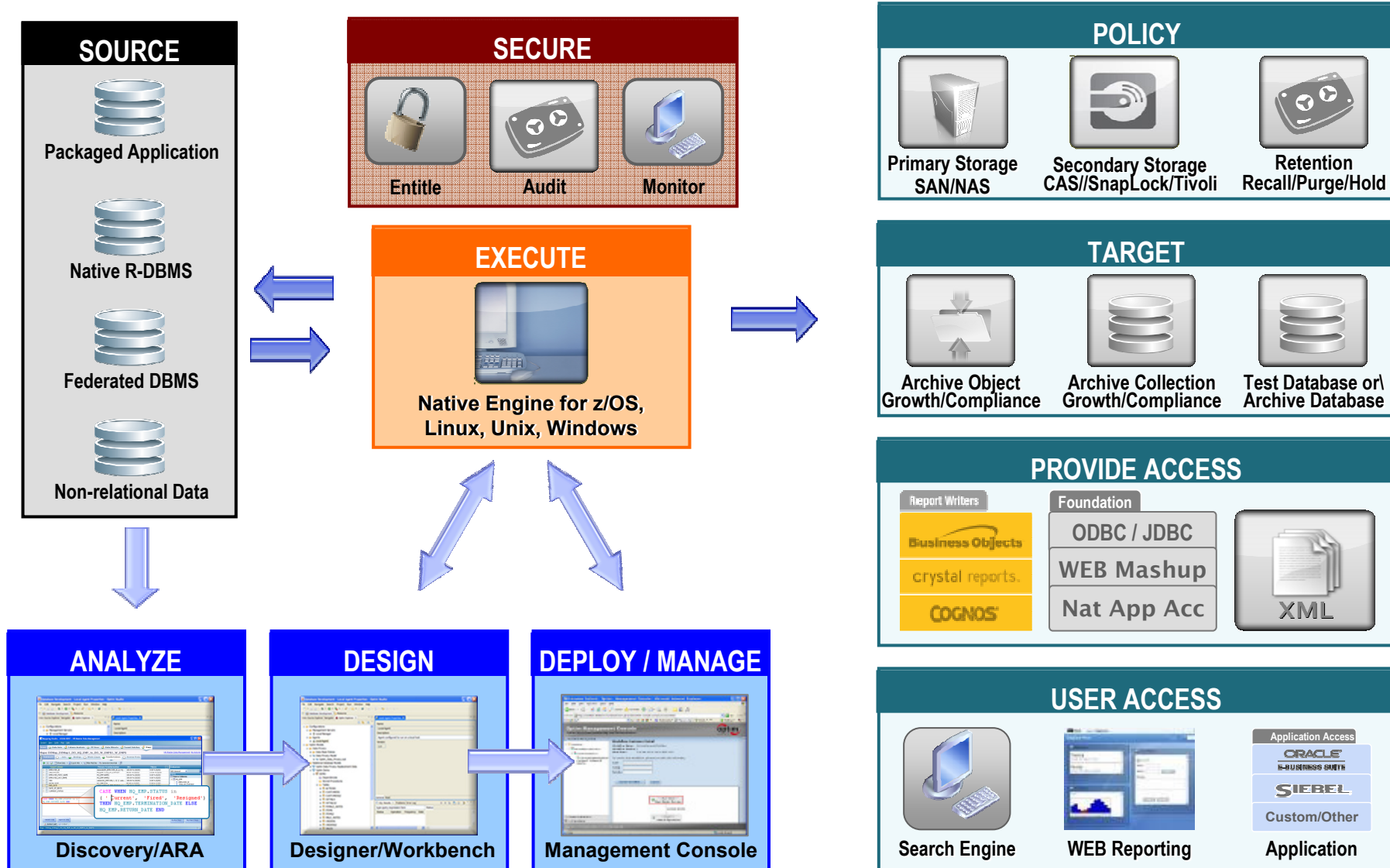


보안 & 규제 준수를 통한 통찰력의 실행

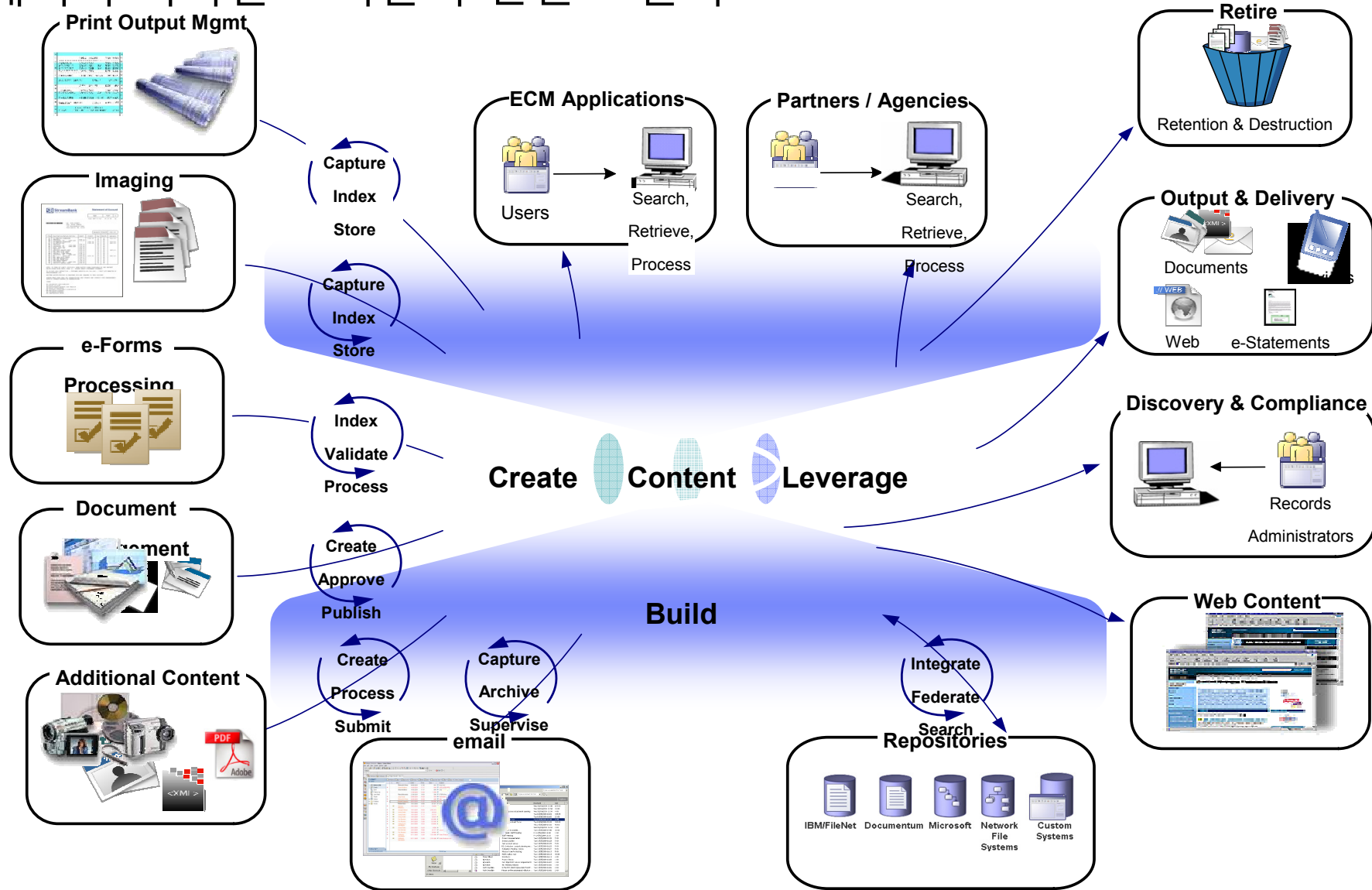
Information Security & Privacy



데이터 거버넌스 기반의 데이터 관리



데이터 거버넌스 기반의 콘텐츠 관리



데이터 거버넌스 구축을 통한 고객 가치 실현

위험 감소	관리 비용 절감	비즈니스 민첩성 확보
<ul style="list-style-type: none">❖ 내부 유출 방지❖ 외부 침입 방지❖ 비즈니스 연속성 보장❖ 개인정보보호법 등 컴플라이언스 및 사규, 감사 기준 준수	<ul style="list-style-type: none">❖ CPU, 소프트웨어, 스토리지 등의 IT 인프라 자원/운영 비용 절감❖ 최종사용자 및 IT 운영자 생산성 향상❖ 테스트라이프사이클 지원을 통한 어플리케이션 디버그 비용 감소	<ul style="list-style-type: none">❖ 테스트라이프사이클 개선을 통한 시장진입 가속화❖ 어플리케이션 성능 향상❖ 데이터 품질 향상



여러분의 현재는 어디입니까?

진단 워크샵 – 데이터 관리, 데이터 보안

데이터 관리 ROI 진단 워크샵 BVA

- 정보수명주기를 비롯한 데이터 관리를 통해 얻어지는 비용 절감과 추가적인 매출 창출에 대한 비즈니스 이윤을 정량화할 수 있도록 분석하는 워크샵
- 전략적인 파트너십 고객을 위해 무료로 제공되는 프로그램

진단 솔루션

- Alinean ROI Analyst 분석툴 사용
- 2008년 WW 80개 이상의 워크샵을 수행하여 100만 달러 이상의 비용 절감 결과를 도출

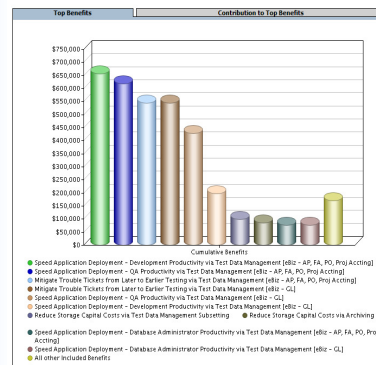


데이터 보안 ROI 진단 워크샵 BVA

- 데이터 보안 강화를 통해 얻어지는 비용 절감과 추가적인 매출 창출에 대한 비즈니스 이윤을 정량화할 수 있도록 분석하는 워크샵

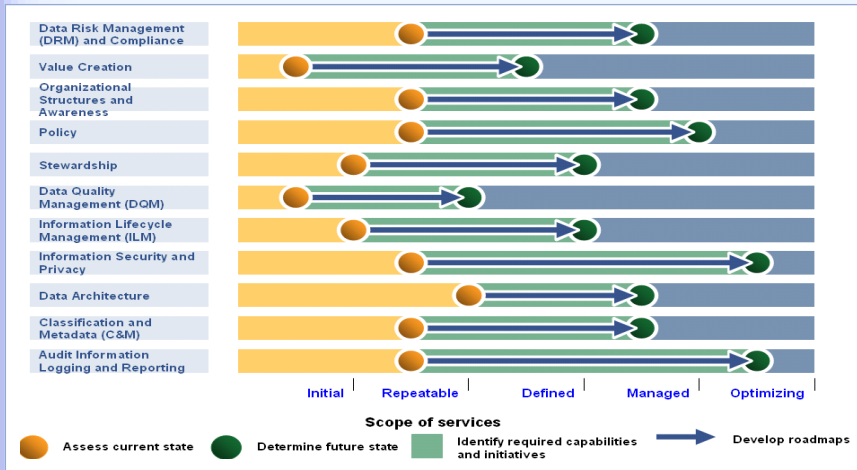
3 Year Financial Analysis for All Projects Examined

ROI Analysis	Initial (Q1 2009)	Year 1 (Q1 2009 - Q4 2009)	Year 2 (Q1 2010 - Q4 2010)	Year 3 (Q1 2011 - Q4 2011)
Benefits (Pre-tax)	\$0	\$1,101,865	\$1,151,202	\$1,239,827
Cumulative Benefits	\$1,101,865	\$2,253,066	\$3,492,893	\$4,732,720
Investment	\$467,124	\$0	\$97,161	\$101,047
Cumulative Investment	\$467,124	\$467,124	\$564,285	\$665,332
Cash Flow	(\$467,124)	\$1,101,865	\$1,054,041	\$1,138,780
Cumulative Cash Flow	(\$467,124)	\$634,741	\$1,688,781	\$2,827,561
ROI		425%		
Risk-Adjusted ROI		366%		
NPV Savings		\$2,285,582		
IRR		227%		
Payback Period		5 month(s)		
Risk Adjusted Discount Rate		9.5%		



데이터 거버넌스 워크샵

- IBM은 Information Agenda Data Governance Workshop을 통해 고객이 데이터 거버넌스에 대한 이해도를 증대시키고, 중장기적인 데이터 거버넌스 전략/실행프로젝트를 정의하여 단계적으로 적용할 수 있도록 지원



[워크샵 수행]

- 수행 인력
 - Information Agenda 산업 전문가 (컨설턴트)
 - Information Agenda Architect
 - Business Value Assessment Architect
 - 수행 범위
 - 조직/프로세스 컨설팅을 제외한 IT 중심 거버넌스 전략/과제 도출
 - 조직/프로세스 측면 : 수립 가이드 , Blueprint 및 Best Practice제공
- **IBM GBS 거버넌스 컨설팅 서비스: 고객을 위한 거버넌스 전반 조직/프로세스 전략 수립을 목표로 하며 1~2개월 소요**

목적 및 기대효과

- 전사적 데이터 거버넌스 이해도 증대
- 비즈니스 이슈와 IT 연관도 리뷰 및 정리
- 데이터 거버넌스와 비즈니스 연관성 정의
- 비즈니스 가치 전달을 위한 장기/단기적 과제 정의
- 비즈니스 목표과 IT 과제의 연결 및 효과 도출
- 비즈니스 우선순위에 따른 거버넌스 로드맵 개발로 장기적인 목표 달성 가능

주요 activity

- 거버넌스 성숙도 진단
- 비즈니스 이슈 및 도전과제 정의
- 이슈해결을 위한 개선 과제 도출
- 필요 기술 요소 , 역량 및 프로젝트 정의
- 단계별 적용을 위한 로드맵 개발
- 비즈니스 효과 및 가치 정의

주요 Output

- 데이터 거버넌스 성숙도 진단 결과
- IBM 데이터 거버넌스 Blueprint
- 과제 정의 및 로드맵
- ROI/TCO 분석