THALES

Vormetric Tokenization Server

개인정보 비식별화

탈레스코리아

비식별화, 가명화, 익명화

- 비식별화: 개인과 관련된 정보를 제공하는 과정에 대한 통칭. 익명화, K-익명화, 가명화.
- 익명화: 데이터 주체를 식별할 수 있는 모든 방법을 삭제
  - 예: 모든 이름을 “익명”으로 대체 또는 삭제
- 가명화: 추가적인 정보 없이는 데이터 주체를 재식별할 수 없도록 대체
  - 예: 각 이름에 대해 고유한 “가명들”이 존재
- K-익명화: 최소한 k 명의 사람이 동일한 준식별자 조합을 가지게 함으로써 식별할 수 없도록 함

개인정보 구성요소	익명화	가명화
Single Out	×	○
Linkable	×	×
Inferred	×	×

GDPR과 가명화 / 익명화

GDPR에서 가명화의 의미

- **강력한 가명화 기술이 적용될 경우**에 한함
- 가명화 데이터도 "개인정보"이므로 규제 대상

가명화 도입 시 면제 조항

- 데이터 처리를 위한 명시적 동의
- 삭제권 (사용자가 요청할 수 있는 잊혀질 권한)
- 정보주체권리 (열람권, 정정권, 이동권 등)
- 데이터 보안
- 개인정보 유출 시 유출 사실을 당사자에게 고지

익명화 도입 시 면제 조항

- 모든 조항 (위 항목에 수집/사용 시 고지, 국외 이전, 보유 기한 등 추가)

GDPR Obligation	Identified	Pseudonymized	Anonymized
1. Provide notice to data subject	Required	Required	Not Required
2. Obtain consent or have another legal basis	Required	Potentially helps	Not Required
3. Give right to erasure / right to be forgotten	Required	Depends on strength	Not Required
4. Other data subject rights (access, portability…)	Required	Depends on strength	Not Required
5. Basis for cross-border transfers	Required	Required	Not Required
6. Data protection by design	Required	Partially met	Not Required
7. Data security	Required	Partially met	Not Required
8. Data breach notification	Likely	Less likely	Not Required
9. Data retention limitations	Required	Required	Not Required
10. Documentation / recordkeeping obligations	Required	Required	Not Required
11. Vendor / sub-processor management	Required	Required	Not Required

GDPR 데이터 분석 예외 조항

- 정보주체의 위험을 감소시키고 컨트롤러와 프로세서의 의무를 충족시키는 보호조치 중 하나로 가명화를 제시
- 데이터 활용도를 높이기 위해 공익적 기록 보전, 과학적·역사적 연구 및 통계적 목적을 위한 연구
- 세부 조항

1) 목적 제한 예외(5조(1)(b), 전문 33)
2) 보관기간 제한 예외(5조(1)(e))
3) 민감정보 처리 제한 예외(9조(2)(j)): 제9조 '특별한 범주의 개인정보(special categories of personal data)' 로서, 인종·민족, 정치적 견해, 종교·철학적 신념, 노동조합 가입 여부, 유전자 또는 생체 정보, 건강, 성생활 또는 성적 취향의 정보 등이 포함됨
4) 제3자 활용시 개인정보의 고지의무 예외 (14조(5))
5) 삭제권 적용 예외(17조(3))
6) 거부권 적용 예외(21조(6))

대한민국 비식별화 규제

개인정보 비식별 조치 가이드라인 (2016)

- 데이터 이용 목적상 반드시 필요한 식별자는 비식별 조치
- 희귀병명, 희귀경력 등의 속성자는 구체적인 상황에 따라 개인 식별 가능성이 매우 높으므로 엄격한 비식별 조치 필요
- 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 여러 가지 기법을 단독 또는 복합적으로 활용
- 가명화만 활용된 경우는 충분한 비식별 조치로 보기 어려움
- 비식별조치 적정성 평가 항목으로 k-익명성, ℓ- 다양성, t-근접성 제시

● 〈 예시 〉 비식별 조치 방법 ●

처리기법	예시	세부기술
가명처리 (Pseudonymization)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대, 서울 거주, 국제대 재학	① 휴리스틱 가명화 ② 암호화 ③ 교환 방법
총계처리 (Aggregation)	• 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm	④ 총계처리 ⑤ 부분총계 ⑥ 라운딩 ⑦ 재배열
데이터 삭제 (Data Reduction)	• 주민등록번호 901206-1234567 → 90년대 생, 남자 • 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리	⑧ 식별자 삭제 ⑨ 식별자 부분삭제 ⑩ 레코드 삭제 ⑪ 식별요소 전부삭제
데이터 범주화 (Data Suppression)	• 홍길동, 35세 → 홍씨, 30~40세	⑫ 감추기 ⑬ 랜덤 라운딩 ⑭ 범위 방법 ⑮ 제어 라운딩
데이터 마스킹 (Data Masking)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학	⑯ 임의 잡음 추가 ⑰ 공백과 대체

대한민국 비식별화 대상

〈 예시 〉 식별자

- 고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호)
- 성명(한자 · 영문 성명, 필명 등 포함)
- 상세 주소(구 단위 미만까지 포함된 주소)
- 날짜정보 : 생일(양/음력), 기념일(결혼, 돌, 환갑 등), 자격증 취득일 등
- 전화번호(휴대전화번호, 집전화, 회사전화, 팩스번호)
- 의료기록번호, 건강보험번호, 복지 수급자 번호
- 통장계좌번호, 신용카드번호
- 각종 자격증 및 면허 번호
- 자동차 번호, 각종 기기의 등록번호 & 일련번호
- 사진(정지사진, 동영상, CCTV 영상 등)
- 신체 식별정보(지문, 음성, 홍채 등)
- 이메일 주소, IP 주소, Mac 주소, 홈페이지 URL 등
- 식별코드(아이디, 사원번호, 고객번호 등)
- 기타 유일 식별번호 : 군번, 개인사업자의 사업자 등록번호 등

〈 예시 〉 속성자

개인 특성	• 성별, 연령(나이), 국적, 고향, 시 · 군 · 구명, 우편번호 병역여부, 결혼여부, 종교, 취미, 동호회 · 클럽 등 • 흡연여부, 음주여부, 채식여부, 관심사항 등
신체 특성	• 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔 등 • 신체검사 결과, 장애유형, 장애등급 등 • 병명, 상병(傷病)코드, 투약코드, 진료내역 등
신용 특성	• 세금 납부액, 신용등급, 기부금 등 • 건강보험료 납부액, 소득분위, 의료 급여자 등
경력 특성	• 학교명, 학과명, 학년, 성적, 학력 등 • 경력, 직업, 직종, 직장명, 부서명, 직급, 전직장명 등
전자적 특성	• 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 접속로그 등 • 인터넷 접속기록, 휴대전화 사용기록, GPS 데이터 등
가족 특성	• 배우자 · 자녀 · 부모 · 형제 등 가족 정보, 법정대리인 정보 등

대한민국 규제에 따른 비식별화 예시

범주화+

삭제

마스킹+

식별자 삭제

구분	이름	연령	성별	지역	질병
1	*	< 40대	*	130**	전립선염
2	*	< 40대	*	130**	전립선염
3	*	< 40대	*	130**	고혈압
4	*	< 40대	*	130**	고혈압
5	*	>= 40대	*	148**	위암
6	*	>= 40대	*	148**	전립선염
7	*	>= 40대	*	148**	고혈압
8	*	>= 40대	*	148**	고혈압
9	*	< 40대	*	130**	위암
10	*	< 40대	*	130**	위암
11	*	< 40대	*	130**	위암
12	*	< 40대	*	130**	위암

k-익명성

l-다양성

t-근접성

THALES

GDPR 규제에 따른 비식별화 예시

식별자 가명화

가명화 (보고서에서 복원)

구분	이름	연령	성별	지역	질병
1	로켁불	28	남	32659	전립선염
2	니앙굷	21	남	84562	전립선염
3	췡바이	29	여	84562	고혈압
4	뀋노기	23	남	32659	고혈압
5	쿄데습	50	여	63934	위암
6	됴약튜	47	남	03571	전립선염
7	똢뱤삸	55	여	63934	고혈압
8	역쥳꿞	49	남	03571	고혈압
9	숄컕폾	31	남	32659	위암
10	댹꽂벜	37	여	32659	위암
11	맣픫괮	36	남	84562	위암
12	훎젲뫈	35	여	84562	위암

가명화 규제 개선

국제 추세

- 전세계적으로 **GDPR**을 개정 방향으로 하고 있음 (미국, 일본, 중국, …)

향후 개인정보보호법 개정 방향

- 개인정보 개념 명확화
- 개인임을 알아볼 수 없도록 안전하게 조치된 가명화 도입
- 가명정보의 고의적 재식별시 엄격한 형사처벌과 과징금 부과

예외 조항을 통한 빅데이터 분석 유용성 향상

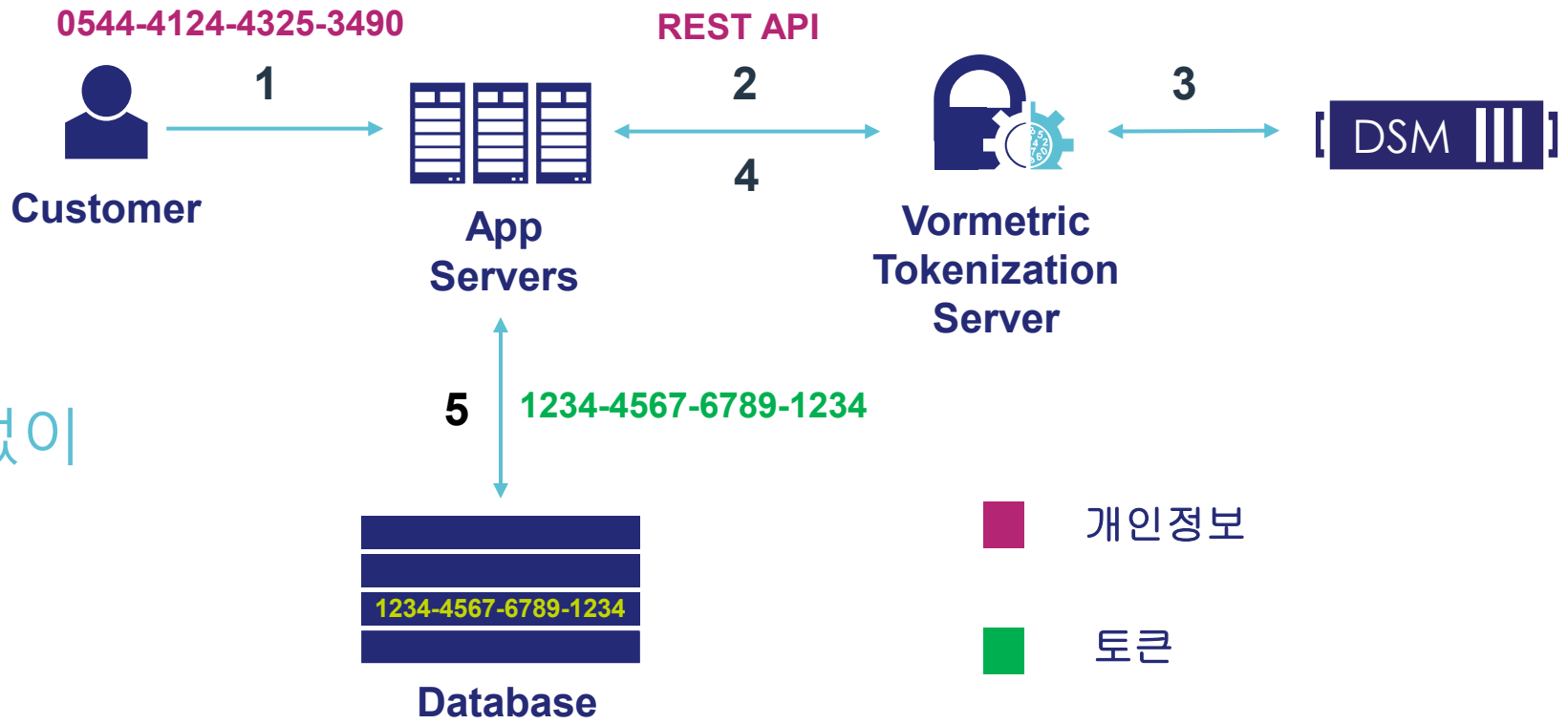
- 익명화 정보 만으로 분석된 빅데이터 분석 결과 효용성 저하
- 가명화 시 개인정보 유출 없이 정보 활용 수준 향상

Vormetric Tokenization with Dynamic Data Masking (VTS)

- GDPR 감사 범위 축소 및 보안 관련 규제 요건 충족
- 서비스 데이터베이스 보호
- 개발계/테스트 시스템, 클라우드환경 및 빅데이터 개인정보 비식별화
- 개인 정보 액세스 차단: 관리자, 해커, 및 업무와 관계 없는 모든 사용자

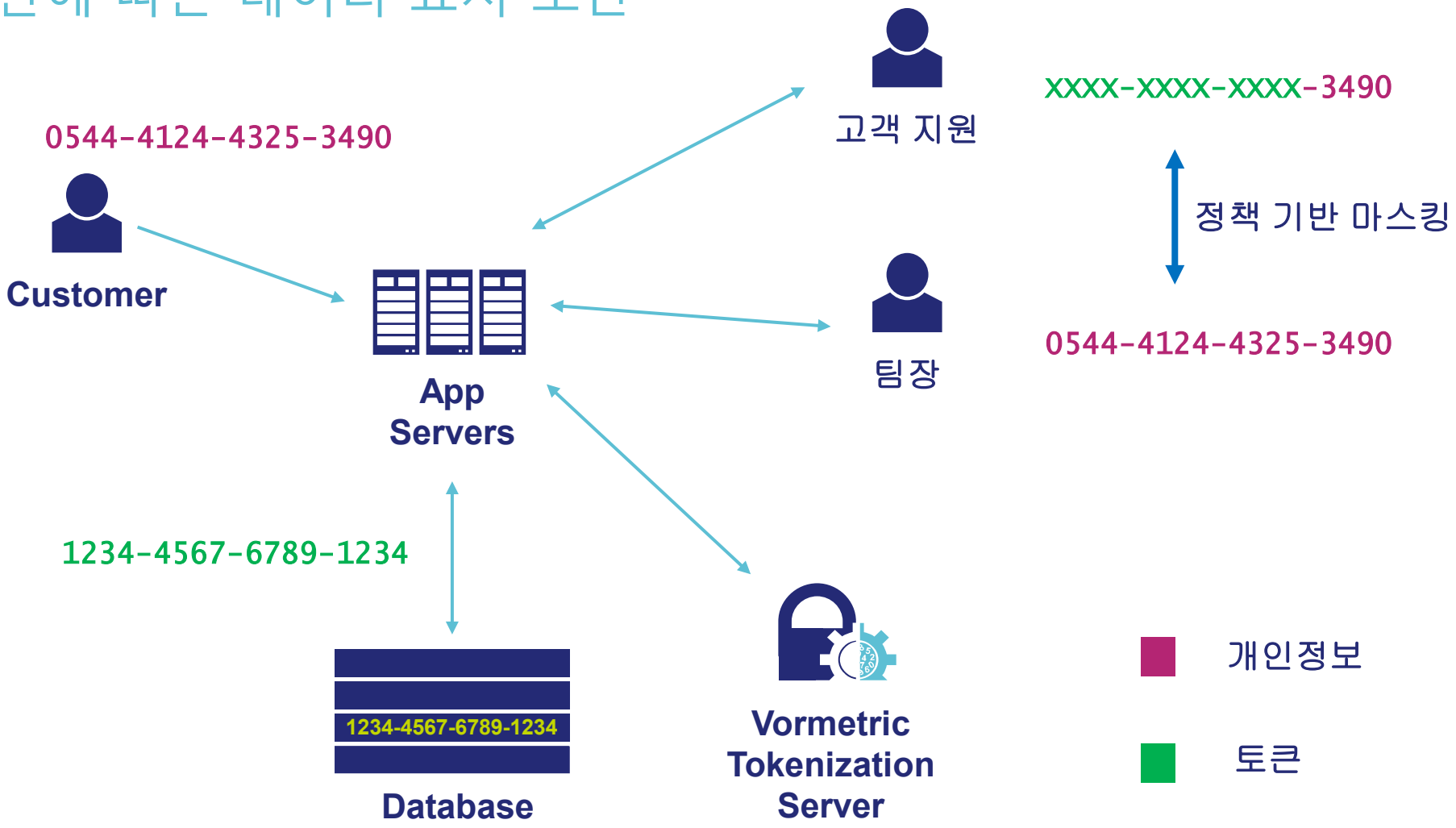
VTS 동작 구조

- 가상 환경 또는 AWS, Azure 등의 클라우드 환경에서 동작하는 가상 머신 형태
- 클러스터링 기능을 통한 다중 노드 성능 확장 방식
- 다양한 데이터 타입 지원
  - Number, Alphanumeric, Date
  - 2 Byte characters
- 동적 마스킹 지원
- 가명화된 데이터 복호화 없이 업무 처리 가능



VTS 동적 마스킹 – 가명화 정보에 대한 비식별 조치

정책 및 권한에 따른 데이터 표시 보안



VTS Random tokenization: Alphanumeric and Chaining

기능

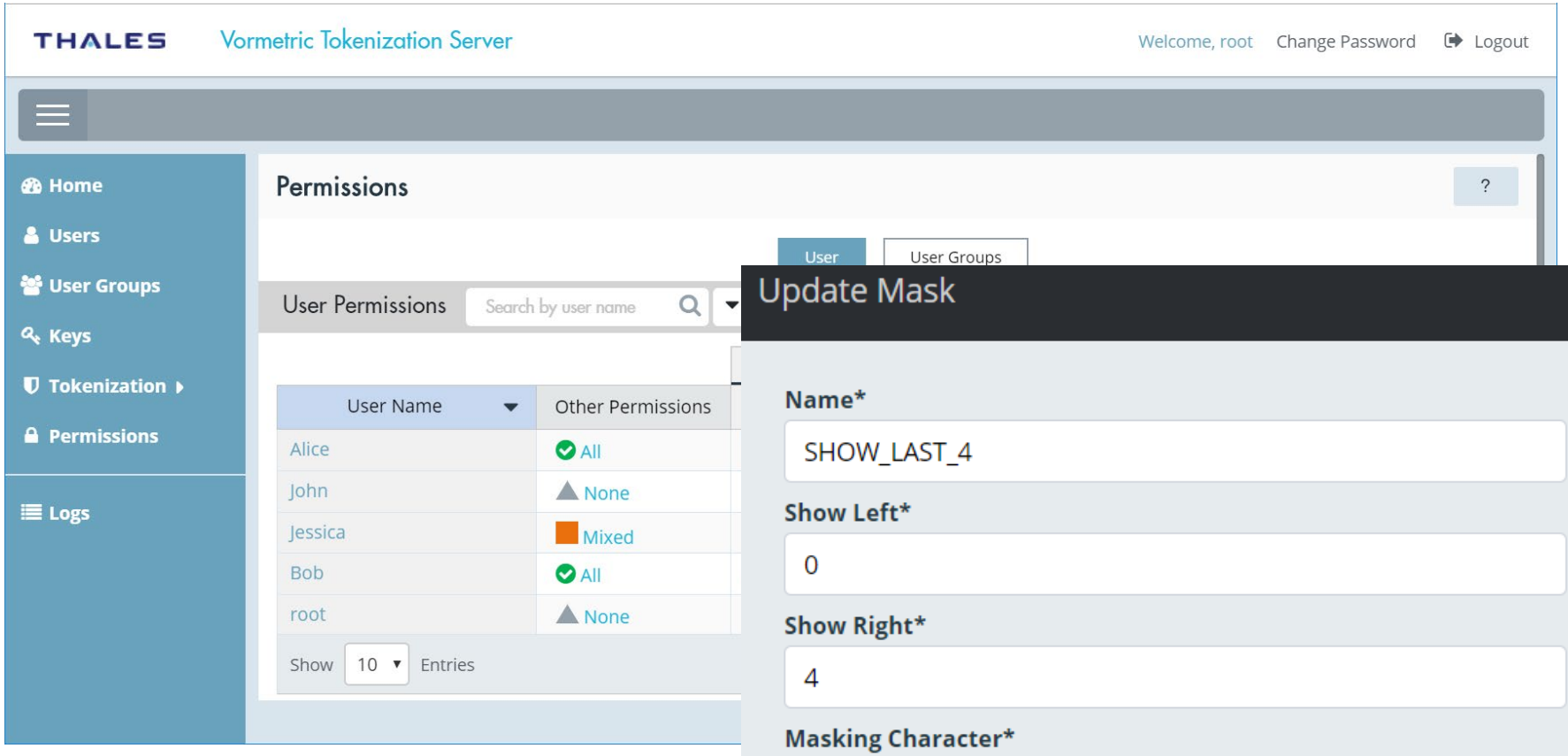
- 최소 5자리 숫자/문자 혼합 형태 지원
- Chaining을 통한 길이 확대: 5MB 이상일 경우 스크립트를 통해 지원
- 1,000개 항목까지 복수 처리 지원

적용 분야

- 빅데이터 및 GDPR 비식별화

웹 기반 관리 기능

- 사용자 관리
- 권한 관리
  - 마스킹 권한 포함
- 토큰 그룹 관리
- 토큰 템플릿 관리
- 마스킹
  - 마스크 위치
  - 마스크 자리 수
  - 마스킹 문자



Name	Show First	Show Last	Masking Character
SHOW_ALL	10	10	X
SHOW_LAST_4	0	4	X

Update Cancel

REST API 예제

Action	Rest API	Example	Return
Tokenize 민감정보에 대한 토큰 생성	**POST**	curl -k -X POST -u vtsUser1:Ssl123! -d'{"tokengroup" : “t-group", "data" : "0111-0222-0333-0444", "tokentemplate" : “testTemplate"}‘ https://192.168.10.41/vts/rest/v2.0/tokenize	{“**token**” : “6029-5413-1453-7206”, "status":"Succeed"}
De-tokenize 토큰에 해당하는 원래 민감정보 조회 (마스킹 적용)	**GET**	curl -k -X POST -u vtsUser1:Ssl123! -d'{"tokengroup" : “t-gruop", "token" : "6029-5413-1453-7206", "tokentemplate" : “testTemplate"}' https://192.168.10.41/vts/rest/v2.0/detokenize	{ “**data**” : “XXXX-XXXX-XXXX-0444”, "status":"Succeed"}

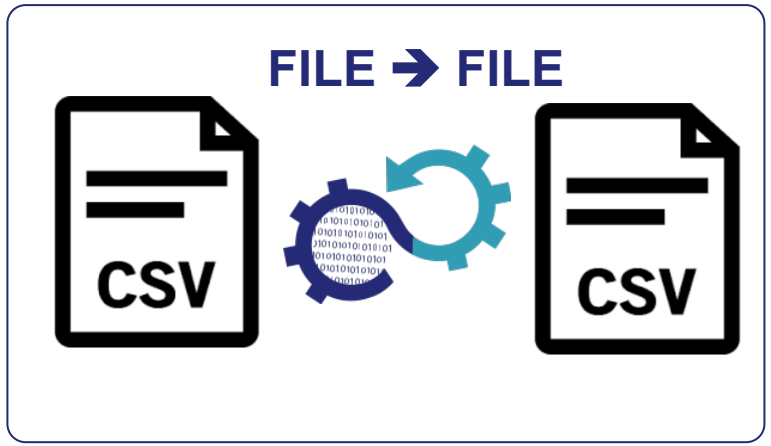
REST API 예제 – 복수 처리

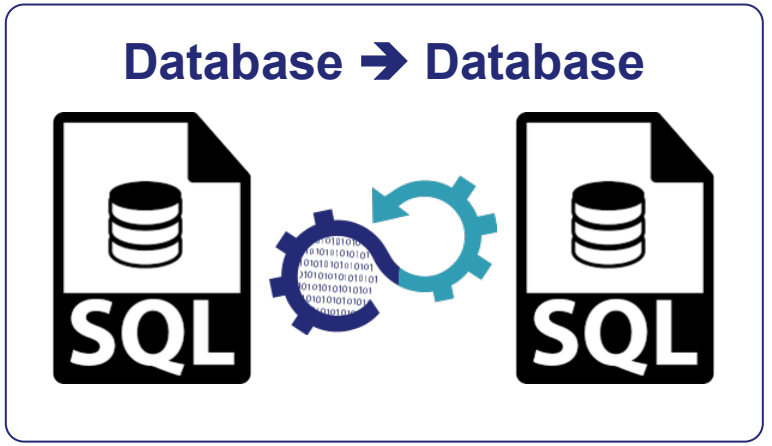
가명화 대상 정보를 파일에 입력한 후 한 번에 처리

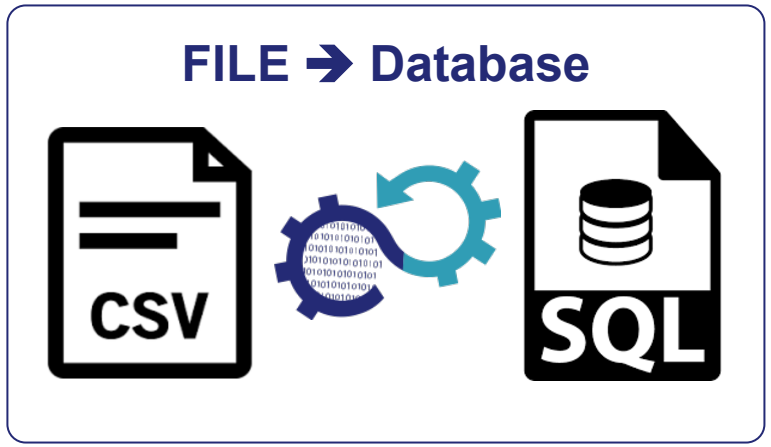
	Example
File data.txt	[{"tokengroup" : “t-group", "data" : "0111-0222-0333-0444", "tokentemplate" : “testTemplate“}, {"tokengroup" : “t-group", "data" : “2323-4545-6767-8989", "tokentemplate" : “testTemplate"}, {"tokengroup" : “t-group", "data" : “9999-8888-7777-6666", "tokentemplate" : “testTemplate"},
Request & Return	# curl -k -X POST -u vtsUser1:Ssl123! --data-binary @data.txt https://192.168.10.41/vts/rest/v2.0/tokenize {“token” : “6029-5413-1453-7206”, "status":"Succeed"} {“token” : “9453-6776-2900-8564”, "status":"Succeed"} {“token” : “4465-9767-9211-6170”, "status":"Succeed"}

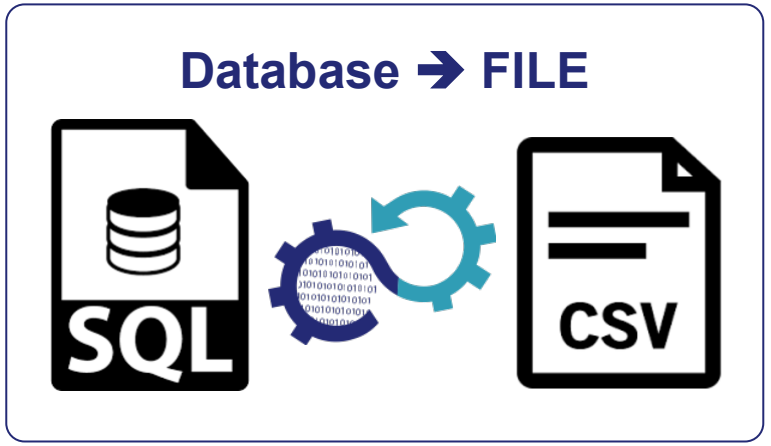
배치 암호화

- 초기 암호화 시 신속한 토큰화 배치 처리
- 빅데이터 활용을 위한 비식별화
- 지원 환경
  - CSV 파일
  - Oracle
  - Microsoft SQL Server
  - MySQL
  - DB2









GDPR 요구 사항 – 탈레스 솔루션으로 해결

가명화

데이터 암호화

- 저장되는 모든 파일, 데이터베이스 암호화 및 애플리케이션 암호화

액세스 관리

- 인가된 사용자에게만 액세스 및 복호화를 허용하는 강력한 액세스 관리

강력한 키 보호 및 관리

- 전용 장비를 통한 강력한 암호화 키 보호 및 업계를 선도하는 우수한 키 관리 절차

THALES

탈레스 솔루션 소개 – Vormetric Transparent Encryption

파일 시스템 블록 암호화

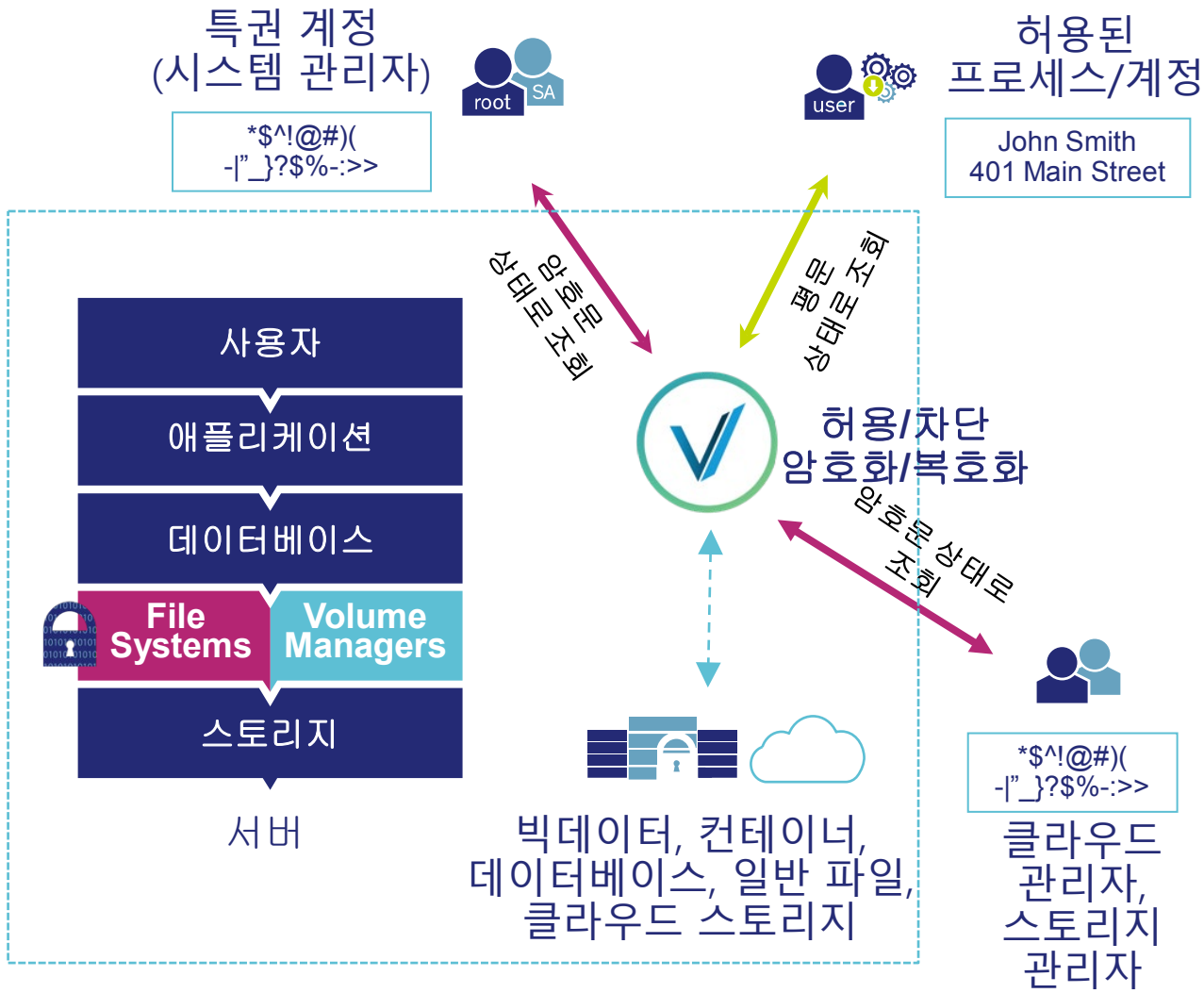
- 모든 유형의 데이터 지원
- 애플리케이션 및 스토리지에 투명하게 동작
- 암호화 전후 동일한 파일 크기

에이전트 방식

- 운영체제 커널에서 동작하는 업계 최고 암호화 성능
- 간편하고 신속한 설치

국내 주요 인증 획득

- 국가사이버안전센터 암호모듈 검증필 (KCMVP)
- GS 인증



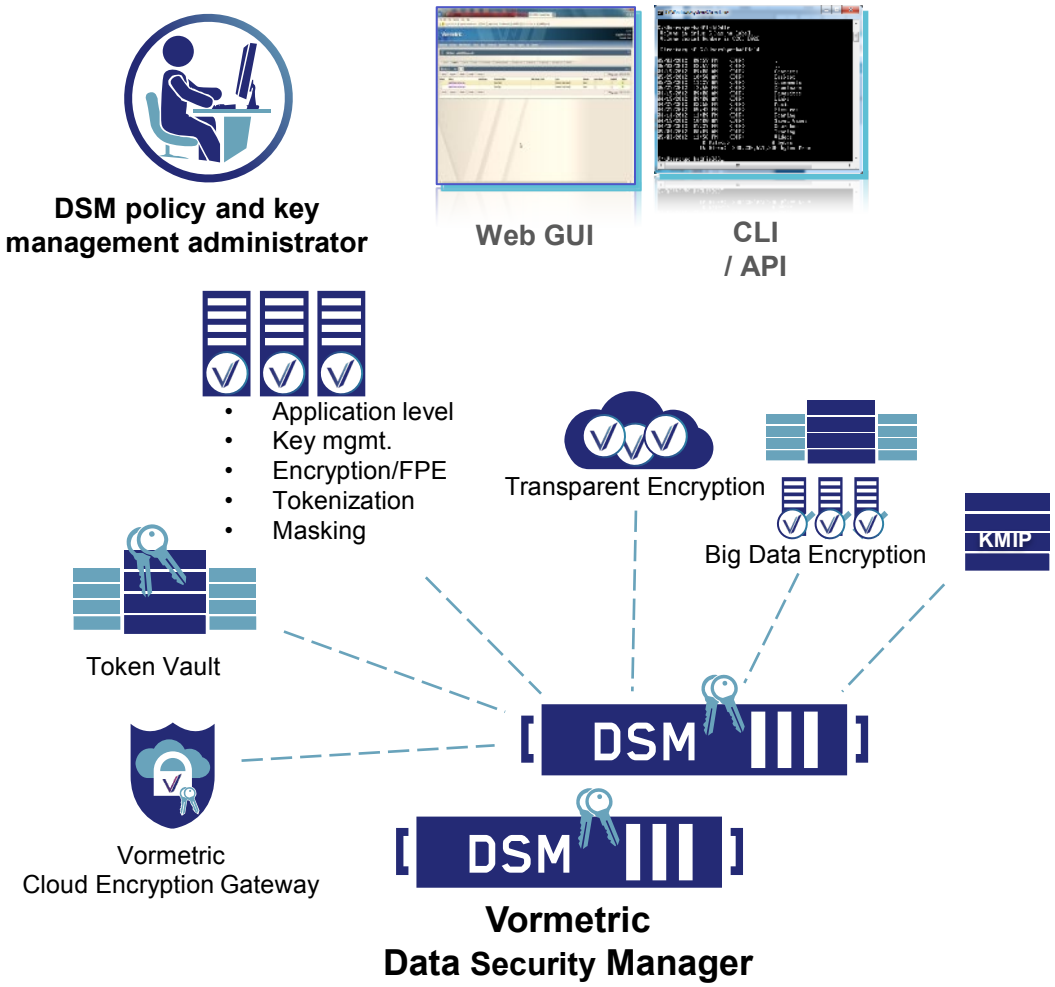
THALES

Live Data Transform – 무중단 암호화 기능

- 기존 저장 데이터 암호화 솔루션의 한계: 초기 암호화를 위한 서비스 중단
- 암호화 키 변경의 어려움: 키가 유출되어 변경하려면 다시 초기 암호화 과정 필요
- LDT의 기능
  - 서비스 중단 없이 암호화 적용 및 암호화 키 변경
  - 암호화를 위한 Quality of Service 제어 기능: 초기 암호화 실행 시간 및 사용 CPU 조정

기능	Live Data Transformation	기존 초기 암호화
무중단 암호화	O	×
Quality of Service 관리: 실행 중지 및 재개 기능, 리소스 (CPU) 사용량 조절	O	×
자동/예약 암호화	O	×
복구 기능	Automatic	None
통계 및 예상 완료 시간	O	×

키관리서버 – Data Security Manager



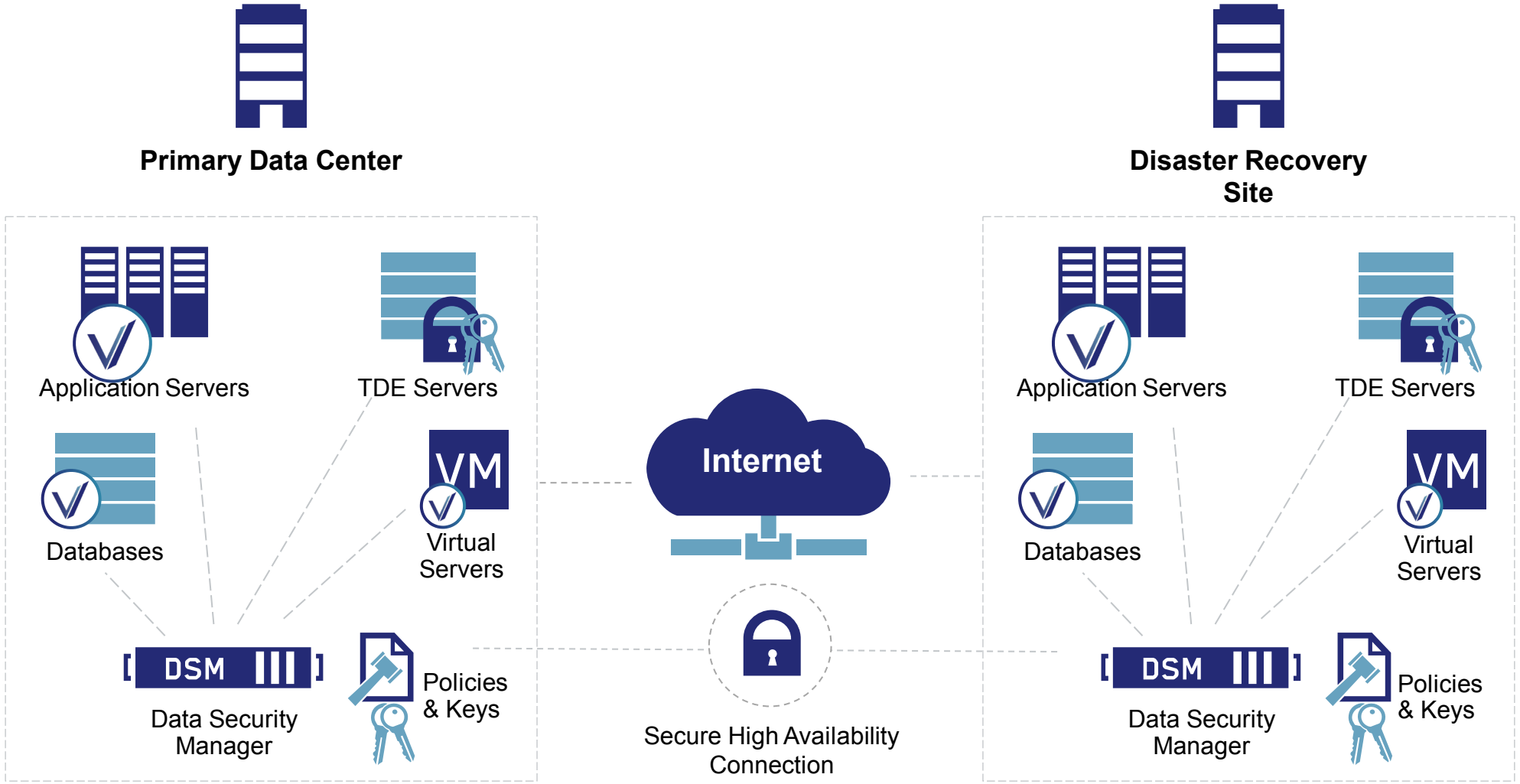
- 중앙 집중형 암호 키 및 정책 관리
- 자체 클러스터링 기능을 통한 장애 대응
- 멀티 테넌트 및 관리자 권한 분리
- 1만대 이상 환경에서 입증된 관리 기능
- 다양한 사용자 인터페이스 지원

FIPS VALIDATED 140-2 ™

Common Criteria

THALES

DSM : 이중화 기능



저장 데이터 암호화 고민 사항

정형데이터 + 비정형데이터 암호화

암호화 이후 성능 저하 없이

암호화 구축

- 기존 시스템 및 운영 환경 변경 없이
- 짧은 기간 내에
- 무엇보다 기존 데이터 암호화를 위한 서비스 중단 없이

강화되는 규제에 대응

- 규제 대상 정보가 지속적으로 확장되고 있음
- 암호화 대상 추가 시 유연하게 대응할 수 있는 솔루션 필요

다양한 환경, 그리고 새로운 기술: 클라우드, 빅데이터, …

보메트릭을 통한 완전한 개인정보 보호