통합 데이터 관리(IDM) 솔루션
2009년 3월

Information Management

데이터 프라이버시 접근제어 :
비운영 환경에서의 개인 정보 보호

차례

전체 개요

본 백서에서는 민감한 정보를 보호하고 프라이버시를 보장하는 것이 왜 중요한지를 설명합니다. 정보 도난 및 신분 도용 사건을 다루는 뉴스 기사가 점점 많아지면서 사생활 침해와 그 결과에 대한 의식이 제고되고 있습니다. 정보 보호 문제에 대응하기 위해, 개인정보보호 법규가 전세계에서 시행되었습니다. 이 같은 법규의 구체적인 내용은 각기 다를 수 있지만 개인정보보호법을 준수하지 않을 경우 수백만 달러의 벌금과 징역형을 감수해야 하는 결과를 가져오게 됩니다. 기업은 단골 고객을 잃게 되고 브랜드 가치의 하락 또한 예상할 수 있습니다. 그 타격은 회사가 문을 닫아야 할 만큼 심각할 수도 있습니다.

기업은 ERP, CRM 및 Custom 애플리케이션에 의존하여 일상적인 비즈니스 운영을 지원하기 때문에, 프라이버시를 보장하고 애플리케이션 데이터를 보호하는 것은 데이터의 저장 위치와 관계 없이 필수적입니다. 그러나, 실운영 환경에 있는 데이터를 보호하는 방법은 비운영 환경(개발, 테스트 및 교육 서버 등)의 독특한 요구사항을 충족하지 않을 수 있습니다. IT 조직은 어떻게 직원 및 고객 정보를 포함한 민감한 데이터와 함께 기업의 기밀 데이터 및 지적 재산을 보호할 수 있을까요? 데이터에서 식별 정보를 제거하는 것을 뜻하는 데이터 마스킹은 프라이버시를 보호하는 최선의 방법으로서 권장됩니다. 하지만 데이터 프라이버시 솔루션을 선택할 때 요구되는 사항에는 어떤 것이 있을까요?

이상적인 데이터 프라이버시 솔루션은 가장 간단하고 가장 복잡한 프라이버시 요건을 충족하기 위해 필요한 데이터 마스킹 기법을 제공해야 합니다. 이 같은 기법은 애플리케이션 로직을 반영하고 데이터의 무결성을 보존하는 방법으로 상황에 민감한 데이터를 정확히 마스킹하는 결과를 가져와야 합니다. 또한, 솔루션에는 마스킹 된 데이터 요소를 여러 애플리케이션 및 운영 환경에 걸쳐 일관성 있게 전파하여 유효한 결과를 제시할 방법도 포함되어야 합니다.

IBM® Optim™ Data Privacy Solution은 개인정보보호 법규의 준수에 관한 요건을 지원할 수 있는 일련의 광범위한 데이터 마스킹 기법을 제공합니다.

- 애플리케이션 인식 마스킹 기능은 이름과 상세 주소 같은 마스킹 된 데이터가 원 정보의 형태와 느낌을 유지하도록 보장하는 데 도움이 됩니다.
- 상황을 인식하는 미리 패키지화된 데이터 마스킹 루틴은 지불카드 번호, 사회보장 번호 및 이메일 주소와 같은 요소를 쉽게 마스킹 할 수 있습니다.
- 영구적인 마스킹 기능은 마스킹 된 대체 값을 여러 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼에 걸쳐 일관되게 전파합니다.

Optim을 통해, 기업은 개발, 테스트 및 교육 환경에서 사용하기에 적합하면서 데이터 프라이버시를 보호하는 방법으로 데이터를 마스킹 할 수 있습니다.

데이터 프라이버시가 중요한 이유

정보의 폭증으로 인해 공공 및 개인 정보의 이용은 일상 생활의 일부분이 되었습니다. 필수 비즈니스 애플리케이션은 일반적으로 이 같은 정보를 합법적인 목적을 위해 수집합니다. 그러나, 인터넷과 정보 시스템, 그리고 ERP, CRM 및 Custom 비즈니스 애플리케이션의 상호 연결된 성격을 감안할 때 민감한 데이터는 쉽게 도난 및 남용의 대상이 될 수 있습니다.

신분 도용, 프라이버시 침해 그리고 민감한 정보와 관련된 사기 사건에 대한 보도는 계속해서 머리기사를 장식하고 있습니다. 데이터 프라이버시의 필요성을 인식하는 기업과 더 강력한 보호를 요구하는 소비자가 점점 많아짐에 따라, 정부는 이 같은 안전을 보장하기 위해 점점 엄격한 법률 및 법규를 제정하고 있습니다. 모든 기업은 직원에 대한 기밀 정보, 기업의 비즈니스 인텔리전스 그리고 민감한 고객 데이터를 보호하는 책임을 짐으로써 정보 관리 법규를 준수하고 고객 및 비즈니스 파트너의 신뢰를 얻어야 합니다.

데이터 프라이버시 문제에 대해 전세계가 관심을 가지고 있습니다. 데이터 프라이버시는 여러 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼을 망라하는 전세계에서 발생하고 있는 문제입니다. 아래의 표에는 국제 및 산업별 표준과 데이터 프라이버시 법규의 예가 나타나 있습니다.

산업	**프라이버시 표준**
금융/소매업	PCI DSS (Payment Card Industry Data Security Standard)
국가	**개인정보보호 법규**
호주	2000년 수정 개인정보보호법
캐나다	개인정보보호 및 전자 문서법
유럽 연합	1999년 개인정보보호지침
뉴질랜드	1993년 개인정보보호법
홍콩	1995년 홍콩 개인 정보(프라이버시) 법령
영국	1998년 정보보호법
미국	1999년 Gramm-Leach-Bliley법 1999년 HIPAA (Health Insurance Portability & Accountability Act)

구체적인 내용은 서로 다를 수 있지만, 이 같은 개인정보보호 법규에는 몇 가지 공통점이 있습니다. 먼저, 이들 법규는 개인 정보의 남용 및 오용으로부터 개인을 보호하기 위해 고안되었습니다. 둘째, 개인정보보호법은 복잡하여 준수하기가 어려우며 기업의 정책 및 운영 절차를 바꾸고 새로운 기술을 채택하도록 요구하는 경우가 많습니다.

마지막으로 집행 노력은 홍보 및 교정에 초점이 맞춰질 수 있지만, 법은 위반 시에, 특히 형사상 부정 행위의 경우에 강력한 처벌을 선고하도록 규정하고 있습니다. 하지만 흥미롭게도 많은 법률에는 법규 준수를 위한 기술적 요구사항이 규정되어 있지 않습니다. 그러나 이 같은 법규를 준수하지 않을 경우 커다란 타격을 받을 수 있습니다.

개인정보보호 법규의 준수는 비즈니스에 영향을 미칩니다. 데이터 프라이버시의 보호는 중요한 비즈니스 이니셔티브입니다. 데이터 침해의 상당 부분은 사실 내부적인 약점에서 비롯됩니다. 그 예로는 지불카드 번호 및 기타 민감한 정보를 오용할 수 있는 직원에서부터 기밀 정보를 저장한 노트북을 도난 당한 직원까지 광범위합니다. 마지막으로, 애플리케이션 데이터를 해외 프로세싱 환경에 아웃소싱하면 보안이 보장되지 않는 민감한 데이터에 대한 액세스를 제어하고 면책(Safe Harbor) 지침을 준수하기가 어려워집니다.

여기에는 많은 것들이 포함됩니다. 기업과 그 임원은 사건 하나에 미화 50만 달러에 이를 수 있는 벌금형을 선고 받거나 징역형에 처해질 수도 있습니다. 무거운 처벌은 기업이 타격을 받을 수 있는 한 가지 예에 불과합니다. 다른 부정적인 영향으로는 투자자 동요로 인한 주가 하락과 정보 노출로 인한 부정적인 보도 등이 있을 수 있습니다. 브랜드에 대한 인식이 돌이킬 수 없을 정도로 훼손되면 신뢰받지 못하는 기업이 됩니다.

민감한 정보를 보호하기 위한 조치를 취하지 않는 기업은 고객과 수익을 잃을 위험뿐만 아니라 회사 문을 닫을 위험까지 감수하게 됩니다. Ponemon Institute에서 시중은행 고객을 대상으로 실시한 조사에서는 소비자가 자신의 프라이버시를 보호할 수 있는 은행하고만 거래할 의향이 있는 것으로 나타났습니다. 이 조사에서는 개인정보보호 문제가 한 번이라도 발생하면 해당 은행과 거래를 끊겠다고 답한 고객이 34%에 이르는 것으로 확인되었습니다. 문제가 2번 발생하면 그 비율은 45%로 올라갑니다. 이 조사에 따르면, "신뢰는 브랜드에 대한 높은 충성도로 이어지지만, 은행의 정보 보안 능력에 대한 신뢰가 떨어지면 상당한 고객 이탈로 연결될 수 있다"고 합니다.[1]

고객을 잃고 기업의 재정이 타격을 받지 않는다 해도, 개인정보보호 문제를 조사하는 데 수백만 달러의 비용이 소요될 수 있다는 사실을 생각해 보십시오. Ponemon Institute의 또 다른 벤치마크 조사인 2008년 연례 조사 : 데이터 침해가 미국에 초래하는 비용에서는 몇 가지 놀라운 결과를 발표했습니다. "조사에는 17가지 사업 부문에 걸쳐 각각 4,200에서 113,000 건에 이르는 기록에 이르는 광범위한 침해 사례가 포함되었습니다[2]." 주요 조사 결과는 다음과 같습니다. "데이터 침해 건의 총 평균 비용은 보안이 훼손된 기록 1건 당 202달러로 상승하여 2007년(기록 1건 당 197달러)보다 2.5% 증가하고 2006년(기록 1건 당 182달러)보다 11% 증가"한 것으로 나타났습니다. 데이터 침해는 기업에게 많은 비용을 초래하는 사건입니다. 조사에 응한 한 기업 당 총 평균 비용은 침해 1건 당 660만 달러를 넘었으며(2007년 630만 달러에서 2006년 470만 달러), 그 범위는 613,000달러에서 3,200만 달러에 이르렀습니다.[3]

비운영 환경에서 데이터 프라이버시 접근제어

산업, 운영 및 애플리케이션의 유형에 따라, 많은 실운영 및 실운영 외 데이터베이스는 민감한 정보를 처리하게 됩니다. 문제는 충분한 보호를 보장하면서 비즈니스 요구를 충족하고 데이터가 "알 필요"를 기준으로 관리되도록 보장하는 것입니다.

기업은 실운영 시스템에 상주하고 비운영환경(개발, 테스트 및 교육 등) 에 상주하는 모든 민감한 정보를 보호해야 합니다. 대부분의 실운영 환경에는 데이터 침해를 보호하기 위한 검증된 보안 및 액세스 제한 장치가 마련되어 있습니다. 기본적인 보안 조치는 네트워크, 애플리케이션 및 데이터베이스 단계에서 각기 적용할 수 있습니다. 물리적인 출입 통제는 키 토큰이나 심지어는 생체인식에 이르는 다인성 인증 체계를 도입함으로써 강화할 수 있습니다. 그러나, 실운영 환경의 데이터를 보호하는 방법이 항상 비운영 환경의 보호를 위한 특수한 요구사항을 충족하는 것은 아니기 때문에 이 같은 보호 조치를 간단히 모든 환경에 걸쳐 똑같이 적용할 수는 없습니다.

그렇다면 비운영 환경이 심각하게 취약한 이유는 무엇일까요? 해답은 비운영환경의 데이터베이스가 작성 및 사용되는 방법에 있습니다. 예를 들면 애플리케이션은 실운영 환경 밖에서 테스트해야 하는데, 이는 테스트에서 애플리케이션 오류가 발견될 경우 실운영 시스템은 영향을 받지 않도록 해야 하기 때문입니다. 애플리케이션 기능을 테스트하고 정확성과 안전성을 보장하기 위해서는 현실적인 데이터가 필수적입니다. 그러나, 실제 데이터를 사용하면 데이터가 무허가자에게 노출되고 보안 침해에 대한 위험이 높아집니다.

예를 들면, 온라인 금융 애플리케이션의 기능을 테스트할 때 테스터는 시스템 사용자로 로그인하고 트랜잭션을 처리하며 고객의 계좌번호와 사회보장번호와 기타 기밀 정보를 볼 수 있습니다. 이 "실제" 데이터는 실운영이 아닌 테스트 환경에 상주하지만, 남용 또는 오용될 위험이 있습니다.

마찬가지로, 애플리케이션 스트레스/부하/성능 테스트는 수백 내지는 수천 명의 사용자가 애플리케이션 데이터베이스를 동시에 액세스하는 상황을 시뮬레이션하도록 고안되었습니다. 이 시나리오에서는 종종 실제 데이터를 실운영 환경에서 추출하여 IBM Rational® Performance Tester 같은 자동 테스트 도구에 로드한 다음 데이터 풀에 저장하여 테스트의 인스턴스화(Instantiation)를 자동화하는 경우가 많습니다. 이 같은 데이터는 보호되지 않는 비운영 환경에서 액세스가 가능한 상태로 남아있을 가능성이 있습니다. 이 같은 데이터가 캡처되어 차후에 무허가자의 손에 들어가게 될 경우, 기업은 심각한 문제를 겪을 수 있습니다.

실운영 데이터 및 데이터베이스는 암호화할 것을 권장합니다. 그러나, 데이터를 실운영 환경에서 암호화한다고 해도 완전히 안전한 것은 아닙니다. 일단 데이터를 실운영 데이터베이스에서 스프레드시트나 다른 파일 형식으로 내보내면 암호화는 더 이상 유효하지 않으며, 데이터를 잃어버리거나 도난 당할 위험이 있습니다.

또한, 적절한 애플리케이션 개발, 테스트 또는 교육 범위를 얻기 위해서는 복수의 환경이 필요할 수 있습니다. 많은 경우, 간단히 실운영 데이터베이스의 사본을 복제하는 방법으로 하나 이상의 비운영 환경을 만들 수 있습니다. 이는 민감한 정보가 안전한 실운영 환경에서 하나 이상의 취약한 비운영 환경으로 전파됨을 의미합니다.

그렇다면 기업은 어떻게 위험을 줄일 수 있을까요? 먼저, "비운영 환경에서 정말로 실운영 데이터를 사용할 필요가 있는가"를 자문해야 합니다. 대답은 "그렇지 않다"는 것입니다. 산업 전문가들은 애플리케이션 개발, 테스트 및 교육과 기타 비운영 환경에서 데이터 프라이버시를 보호하는 것이 필수적임을 인식합니다. 또한, 이들은 데이터를 마스킹하거나 비식별화하는 방법이 최선의 방법이라는 의견에 동의합니다.

비운영 환경에서 데이터의 비식별화란 간단히 개인을 식별하기 위해 사용할 수 있는 데이터 요소를 체계적으로 제거하거나 마스킹하거나 변환하는 프로세스를 의미합니다. 데이터를 비식별화하면 개발자, 테스터 및 교육진행자는 현실적인 데이터를 사용하여 유효한 결과를 얻으면서도 프라이버시 보호 규칙을 준수할 수 있습니다. 이 같은 방법으로 정화된 데이터는 일반적으로 비운영 환경에서 사용할 수 있는 조건을 만족하는 데이터로 간주되며 도난 당하거나 노출되거나 손실될 경우 아무에게도 쓸모가 없습니다.

종합적인 데이터 프라이버시 솔루션의 선택

데이터 프라이버시의 보호는 더 이상 선택사항이 아니라 법적 의무입니다! 기업은 실운영 및 비운영 환경에서 모두 데이터를 보호하여 개인정보 보호 법규를 준수하고 위험을 피하기 위한 절차를 마련해야 합니다. 효과적인 프라이버시 보호 전략은 데이터베이스 환경에서 개인 정보의 기밀을 보장하고 보안을 개선합니다. 하지만, 전사 데이터 프라이버시 솔루션에서는 어떤 기능을 찾아봐야 할까요?

데이터의 비식별화는 널리 인정되는 최선의 방법으로서, 실운영 환경에서 프라이버시를 보호하고 법규 준수 이니셔티브를 지원하기 위한 가장 효과적인 방법이 됩니다. 기밀 데이터를 비식별화하는 기능은 프라이버시를 보호하면서도 개발, 테스트 및 교육이나 기타 합법적인 비즈니스 목적을 위해 필요한 "현실적인" 데이터를 제공할 수 있어야 합니다. 다음을 제공하는 데이터 프라이버시 솔루션을 찾아보십시오.

- **광범위한 데이터 마스킹 기법.** 이상적인 데이터 프라이버시 솔루션은 사용이 편리하고 다양한 마스킹 기법을 제공해야 합니다. 가장 간단한 기법은 문자 또는 숫자 데이터를 마스킹하거나, 무작위 또는 연속 번호를 생성하는 방법입니다. 보다 발전된 마스킹 루틴은 복잡한 데이터 프라이버시 요건을 지원하기 위해 사용할 수 있습니다.
- **애플리케이션 로직 지원.** 데이터 마스킹 기법은 애플리케이션 로직을 존중하고 결과를 보는 사람에게 이해가 되어야 합니다. 즉, 마스킹된 데이터의 모습은 원 정보와 유사해야 합니다. 숫자 필드는 기능 테스트가 모든 애플리케이션 적합성 검사를 통과하도록 적절한 구조와 패턴을 유지하고 허용되는 값의 범위 안에 포함되어야 합니다.
- **비즈니스 상황 데이터 요소 지원.** 데이터 마스킹 기법에는 특정한 비즈니스 상황을 반영하는 데이터 요소의 기능이 포함되어야 합니다. 예를 들면, 사회보장번호와 지불카드번호와 이메일 주소를 정확하게 마스킹하는 기능이 미리 패키지에 포함되어 있다면 매우 유익할 것입니다.
- **데이터의 무결성을 보존하는 기능.** 데이터 마스킹 기법은 데이터의 참조 무결성을 보존해야 합니다. 데이터 요소를 자동으로 마스킹하고 마스킹된 데이터 요소를 관련 테이블과 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼으로 정확하게 전파하여 유효한 테스트 결과를 보장하는 기능이 있는 지 살펴보십시오. 솔루션이 데이터의 무결성을 보존하지 않으면 테스트 결과는 부정확할 것입니다.
- **유연성, 확장성 및 적응성.** 데이터 마스킹 기능은 실운영 환경에서 데이터를 추출하고 데이터를 비운영 환경의 데이터베이스에 삽입하거나 로드하기 전에 마스킹 할 수 있어야 합니다. 이 기능은 실제 데이터가 절대 보안이 보장되는 실운영 환경 밖으로 전파되지 않도록 보장합니다. 또한, 실운영 데이터베이스가 이미 복제되어 작성된 경우에는 "그 자리에서" 데이터를 마스킹하는 기능이 필요합니다. 이 기능은 민감한 데이터를 그것이 상주해 있는 장소와 관계 없이 보호할 수 있도록 보장합니다. 마지막으로, 데이터 마스킹 기능은 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼에 걸쳐 확장이 용이하여 요구사항이 바뀔 경우 그에 맞게 개조할 수 있어야 합니다.

간단히 말해, 현재와 미래의 전사 데이터 마스킹 요구사항을 충족하도록 확장할 수 있는 데이터 프라이버시 솔루션이 필요합니다.

IBM Optim을 사용한 데이터 프라이버시 문제 해결

IBM Optim Data Privacy Solution은 여러 비운영 환경에 걸쳐 효과적으로 사용할 수 있는 광범위한 애플리케이션 데이터 비식별화 기능을 제공합니다. Optim 데이터 마스킹 기술은 데이터의 무결성을 보존하고 애플리케이션 로직을 반영한 일관되고 정확한 결과를 제시합니다. 마스킹된 데이터는 여러 비운영 환경으로 정확히 전파하여 유효한 데이터 결과를 얻을 수 있습니다. 예를 들면, 실운영 데이터베이스는 IBM DB2® for z/OS®에 구성하고 테스트 환경은 Linux®를 탑재한 Oracle® 데이터베이스에 구성할 수 있습니다. 마지막으로 Optim 데이터 마스킹 기법은 현재와 미래의 요구사항을 충족하기 위해 확장이 용이하고 여러 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼에 걸쳐 배포할 수 있습니다.

효과적인 데이터 마스킹의 기초

Optim은 효과적인 데이터 마스킹을 위한 다음과 같은 기초적인 요소를 제공하므로 기업은 이를 통해 가장 복잡한 데이터 프라이버시 문제까지도 해결할 수 있습니다.

애플리케이션 인식 데이터 마스킹. Optim의 애플리케이션 인식 데이터 마스킹 기능은 마스킹된 데이터가 애플리케이션 로직을 위반하지 않도록 데이터 요소를 정확히 이해하고 캡처하여 처리합니다. 예를 들면, 이름의 성은 무의미한 문자열이 아닌 무작위로 선택된 성으로 대체됩니다. 숫자 필드는 적절한 구조와 패턴을 유지합니다. 예를 들어 진단 코드가 4 자리 수이고 값의 범위가 0001에서 1000 사이일 경우, 마스킹된 값이 2000이라면 애플리케이션을 테스트하는 상황에서 유효하지 않을 것입니다. 체크섬은 유효하게 유지되므로, 기능 테스트는 모든 애플리케이션 적합성 검사를 통과합니다. 가장 중요한 것은 Optim이 모든 마스킹된 데이터 요소를 전체 테스트 데이터베이스에 걸쳐 그리고 기타 관련 애플리케이션 및 데이터베이스로 일관되게 전파한다는 사실입니다.

상황 인식 데이터 마스킹. 미리 패키지에 포함되어 있는 Optim의 상황 인식 데이터 마스킹 루틴은 주요 데이터 요소를 비식별화하는 데 도움이 됩니다. Optim은 생일, 은행 계좌번호, (캐나다의 사회보험번호나 이탈리아의 Codice Fiscale 같은) 국가 식별번호, 복리후생 정보, 건강 보험 식별번호 등과 같은 여러 가지 유형의 민감한 정보를 비식별화하기 위해 사용할 수 있고, 다양한 효과가 입증된 데이터 마스킹 기법을 제공합니다.

Optim Transformation Library™ 루틴을 사용하면 사회보장번호와 지불카드번호와 이메일 주소 같은 복잡한 데이터 요소를 정확히 마스킹할 수 있습니다. 내장된 검색(Lookup) 테이블은 이름 및 주소의 마스킹을 지원합니다. 또한, 여러 관련 애플리케이션 및 데이터베이스에서 프로세싱 로직를 통합하고 향상된 유연성과 창의성을 제공하여 가장 복잡한 데이터 마스킹 요건까지도 지원하는 사이트별 데이터 변환 루틴을 결합할 수도 있습니다.

영구적인 데이터 마스킹. Optim의 영구적인 마스킹 기능은 원본 열(Column)을 변환한 대체 값을 생성하고 대체 값을 여러 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼에 걸쳐 일관되고 정확하게 전파합니다. 영구적인 데이터 마스킹 기능은 여러 개발, 테스트 및 교육 환경에 걸쳐 보호하기 위한 프라이버시의 확장성을 보장합니다.

효과가 입증된 데이터 마스킹 기법

Optim은 데이터를 변환하고 비식별화하는 일련의 광범위한 데이터 마스킹 기법을 제공합니다. 사용할 기법은 마스킹하는 데이터의 유형과 얻고자 하는 결과를 기준으로 결정하면 됩니다. 예를 들면, 간단히 무작위 또는 연속 번호 마스킹 기법을 사용하거나 환자 이름을 미리 정해진 문자열로 대체함으로써 고객 식별 번호를 쉽게 마스킹 할 수 있습니다. Optim과 함께 제공되는 몇 가지 마스킹 기법에 대해서는 이어지는 문단에 설명되어 있습니다.

문서 및 숫자 데이터 마스킹. Optim은 문자 및 숫자 데이터를 마스킹하는 몇 가지 기법을 제공합니다. 간단하게는 String Literal을 사용하여 숫자 데이터를 마스킹하기 위해 사용할 값을 지정할 수 있습니다. String Literal은 따옴표로 묶은 문자 또는 숫자의 모든 조합을 사용하여 정의할 수 있습니다. 예를 들면, 자동차 보험 회사에서는 보험금 청구서의 텍스트 설명을 "Code 60"으로 쉽게 대체할 수 있을 것입니다.

마찬가지로, Substring 마스킹 기법은 하위 문자열이나 열의 컨텐츠 중 일부를 반환합니다. 지역번호와 전화번호의 첫 세 자리수를 포함하는 하위 문자열을 사용하면 필요한 상세정보를 제공하면서도 실제 전화번호가 노출되지 않도록 할 수 있습니다.

Sequential 마스킹 기법은 문자 또는 숫자 데이터 유형과 함께 사용할 수 있으며, 연속적으로 증가하는 값을 반환합니다. 예를 들면, 이 기법을 사용하여 간단히 시작 계좌번호를 지정한 다음 각 번호에 7씩 더함으로써 금융 애플리케이션의 당좌예금 계좌번호를 마스킹할 수 있습니다.

Random 마스킹 기법은 사용자가 지정한 값의 범위 내에서 무작위로 선택되어 문자 또는 숫자 데이터를 마스킹하기 위해 사용할 수 있는 값을 반환합니다. 예를 들면, 건강 보험 애플리케이션을 테스트할 때에는 가입자 ID, 그룹 번호, 카드 번호, 카드 날짜 및 지급인 ID를 마스킹할 무작위 번호를 생성할 수 있습니다.

Shuffle 마스킹 기법은 궁극의 무작위 데이터 마스킹 기법입니다. 이 기법은 데이터를 하나 이상의 열(Column)에서 정해진 수의 행(Row)으로 재배분하고 순서 섞기(Shuffle) 차수와 관계 없이 선택적으로 고유성을 갖게 합니다. 이 기법은 거의 모든 데이터 유형에 적용할 수 있으며 성명에서 성 및 이름과 상세주소, 도시, 국가 및 우편번호를 포함한 주소 정보를 마스킹하기 위해 쉽게 사용할 수 있습니다.

그 밖에, 생일 또는 사망일이나 입원 및 퇴원일 같은 날짜 위주 정보, 장치 식별번호 또는 일련번호, 웹 주소 또는 URL, 그리고 인터넷 프로토콜(IP) 주소를 쉽게 마스킹 할 수도 있습니다. 이 같은 데이터 유형과 더불어, Shuffle 기법은 모든 유형의 계좌번호, 의료기록번호, 의료보험 및 수혜자 식별정보, 인증서 또는 면허번호, 차량 식별번호, 직원 ID 번호 및 지불카드 번호를 마스킹하기 위해 사용할 수 있습니다.

합계(Combined) 값 마스킹. 데이터 마스킹은 연결식(Concatenated Expression)을 사용하여 이루어질 수도 있습니다. 연결식을 사용하면 둘 이상의 원본 열(Column)의 값을 합(Combine)하거나 열 값을 어떤 다른 값과 합(Combine)함으로써 대상 열의 값을 마스킹할 수 있습니다.

예를 들어 "999-9999" 같은 형식으로 된 은행계좌번호가 있고 여기서 첫 세 자리는 계좌의 유형(당좌, 저축, 머니마켓 등)을 나타내고 뒤의 네 자리는 고객 식별번호를 의미한다고 가정하면, 여기서 계좌 번호는 실제 계좌번호의 첫 세 자리를 사용하는 하위 문자열과 순차적 마스킹 기법을 사용하여 도출한 네 자리수를 연결함으로써 마스킹할 수 있습니다. 이 예에서는 연결(Concatenation)을 통해 계좌번호 열의 올바른 형식을 유지하고 계좌 유형에 대한 중요한 정보를 보존함과 동시에 기밀 고객 정보를 비식별화할 수 있습니다. 그 결과, 애플리케이션 테스트의 맥락에서 여전히 유효한 가상 계좌번호가 얻어집니다.

검색(Lookup) 값을 사용한 데이터 마스킹. 또 다른 비식별화 방법은 대체 값을 사용하여 데이터를 변환하는 것입니다. 검색(Lookup) 기법은 원본 열의 값에 대응하는 마스킹 된 값을 대상 열에 반환함으로써 원본 열의 값을 마스킹하기 위해 사용할 수 있습니다. 예를 들면, 검색(Lookup) 테이블은 의료 진단 코드를 테스트용으로 사용할 가상 코드로 변환할 수 있습니다.

Random Lookup 마스킹 기법은 원본 열의 값에 대응하는 무작위로 선택된 값을 대상 열에 반환함으로써 원본 열의 값을 마스킹할 수 있습니다. Optim은 데이터 마스킹의 편리성을 높이는 몇 가지 사전 정의된 검색(Lookup) 테이블을 제공합니다.

- **이름 검색(First names lookup)** : 개인 정보를 비식별화하기 위해 사용 가능한 5,000개가 넘는 이름이 수록되어 있습니다.
- **성 검색(Last names lookup)** : 개인 정보를 비식별화하기 위해 사용 가능한 80,000개가 넘는 성이 수록되어 있습니다.
- **상세주소/도시/주/우편번호 검색** : 전체 주소 정보를 마스킹하기 위한 10만개가 넘는 미국 주소가 수록되어 있습니다.

향상된 Random Lookup 기법을 사용하면 대상 테이블의 행을 검색 테이블에서 무작위로 선택한 전체 데이터 행으로 대체함으로써 테이블의 모든 또는 아무 열에 있는 데이터를 쉽게 변환할 수 있습니다. 예를 들면, 이 기능을 통해 ZIP 코드를 다른 ZIP 코드로 대체하는 대신 전체 상세주소/도시/주/우편번호를 마스킹 할 수 있습니다.

Optim의 Transformation Library(변환 라이브러리)를 사용한 민감한 데이터 마스킹. Optim의 데이터 변환 라이브러리(Transformation Library)를 사용하면 유효한 마스킹된 값을 생성하여 사회보장번호, 신용카드번호 및 이메일 주소를 비식별화할 수 있습니다.

- **사회보장번호** : 미국 사회보장국에서 사용하는 규칙을 따르는 유효한 변환 번호를 생성합니다. 예를 들면, 이 기능은 실업 보험을 처리하는 애플리케이션을 테스트할 때 사회보장번호를 마스킹하기 위해 사용할 수 있습니다.
- **신용카드 번호** : 신용카드 발급업체가 사용하는 규칙에 따라 유효한 변환 번호를 생성합니다. 예를 들면, 이 기능은 PCI DSS의 준수를 지원하기 위해 고객 명세서를 처리하는 애플리케이션을 테스트할 때 지불 카드 번호를 마스킹하는 데 사용할 수 있습니다.
- **이메일 주소** : 문자열 리터럴(string literal)이나 성/명 열과 도메인을 사용하여 유효한 변환 이메일 주소를 생성합니다. 예를 들면, 이 기능은 신입 사원 교육에 사용되는 DM(디렉트 마케팅) 애플리케이션에서 이메일 주소를 마스킹하기 위해 사용할 수 있습니다.

마스킹 된 데이터의 무결성 보존. 지금까지 설명한 각 방법은 데이터를 마스킹하여 기밀을 안전하게 보호하는 데 효과적입니다. 그러나, 관계 데이터베이스 애플리케이션의 경우는 더 복잡합니다. 구체적으로 말하면, 참조 무결성을 유지하기 위해 마스킹한 데이터 요소를 데이터베이스 안의 모든 관련 테이블에 전파하는 기능이 필요합니다.

마스킹된 데이터 요소(예: 전화번호)가 데이터베이스 테이블 관계의 1차 또는 2차(외래) 키(Key)일 경우, 이 새로 마스킹 된 데이터 값은 데이터베이스의 모든 관련 테이블로 전파해야 합니다. 키 전파(Key Propagation)는 변환된 데이터의 참조 무결성을 보장합니다. 키 전파 기능이 없으면, 상위 및 하위(Parent & Child) 테이블 사이의 관계는 단절되어 테스트 데이터가 부정확해지는 결과를 초래할 것입니다. 따라서, 애플리케이션 테스트의 결과는 신뢰할 수 없게 될 것입니다.

Optim은 키 전파를 완벽하게 지원하므로, 1차 키나 외래 키에 값을 할당하고 해당 값을 모든 관련 테이블로 전파할 수 있습니다. 지정하는 값은 유효한 열의 이름, 문자열 리터럴, 등식, 또는 다른 마스킹 기법일 수 있습니다. 고객과 주문이라는 두 개의 서로 관련된 테이블로 구성된 간단한 예를 가정해 보십시오 (그림 1 참조.) 고객 테이블은 주문 테이블의 상위 테이블이며, 그 1차 키 열인 Cust_ID는 5자리로 된 숫자 값입니다.

그림 1 : Optim의 키 전파 기능은 데이터를 마스킹할 때에도 참조 무결성을 보장합니다.

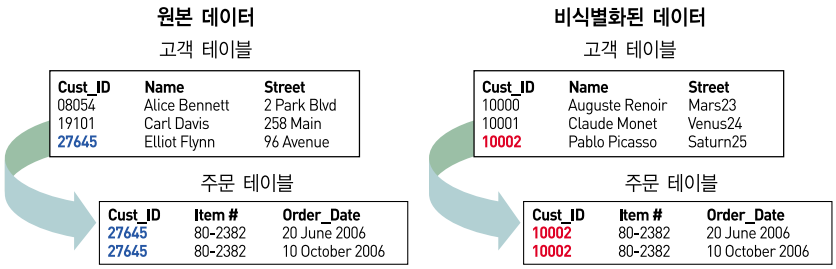
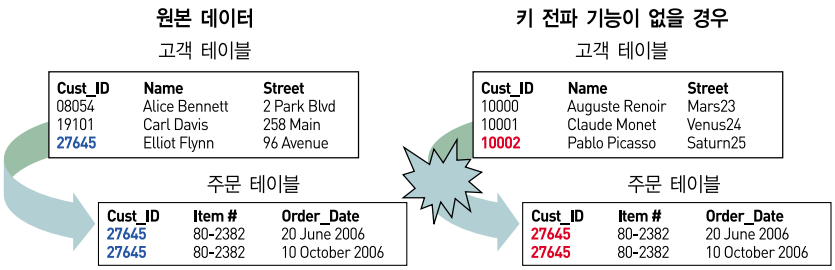


그림 1의 예에서는 Cust_ID, Name 및 Street 열이 마스킹 되어 있습니다. "Elliot Flynn"이라는 이름이 "Pablo Picasso"로 마스킹 되었음을 볼 수 있습니다. 상세주소도 마스킹 되었습니다. 특히, 순차적 마스킹 기법이 사용되어 Elliot Flynn의 원래 Cust_ID가 27645에서 10002로 변경되었습니다. 이 마스킹 된 Cust_ID 값을 고객(상위) 테이블에서 모든 관련 테이블로 전파하면, 테스트 데이터베이스에서 고객과 주문 테이블 사이의 키 관계는 그대로 유지됩니다. 마스킹 된 값을 전파하는 기능이 없다면, 데이터의 참조 무결성이 단절되어 주문 테이블에는 상하위 관계가 없는 줄(Orphan Row)이 만들어질 것입니다.

그림 2 : 키 전파 기능이 없다면 중요한 데이터 관계가 단절될 것입니다.



키 값을 전파할 수 있는 기능은 테스트 데이터베이스의 참조 무결성을 보전하는 데 기여하여 유효한 테스트 결과를 지원합니다. 관련 테이블이 수백 개에 이르고 키를 모든 관련 테이블로 전파해야 한다면 얼마나 복잡할 지 상상해 보십시오. 전파 기능이 없다면 많은 상하위 관계가 없는(Orphan) 테이블이 만들어지고 테스트 데이터베이스는 쉽게 훼손될 것입니다.

사용자 정의 마스킹 루틴. 더 복잡한 데이터 변환 작업을 수행해야 할 때에는 사용자 정의 종료(Exit) 루틴을 작성할 수 있습니다. 이는 간단히 말해 원하는 데이터 변환 작업을 수행하는 프로그램이나 일련의 명령어입니다.

종료 루틴은 특히 다른 방법을 사용하여 정의할 수 없는 대상 열의 값을 생성하는 데 유용합니다. 예를 들면, 테스터는 테스트 데이터를 애플리케이션 보고 프로세스에 제공하기 전에 연누적(YTD) 매출 수익을 모아야 합니다. 또 다른 시나리오에서는 고객의 지리적 위치, 평균 계좌 잔액 및 거래 활동량을 기준으로 고객 ID 코드로 사용할 값을 생성해야 할 수도 있습니다. 이 종료 루틴을 사용하여 생성된 고객 ID 코드는 대상 열을 채워 넣기 위해 사용됩니다.

마스킹 후 이동 또는 제자리 마스킹. Optim를 사용하면 실운영 데이터가 안전한 실운영 환경 밖에서 노출되는 것을 방지할 수 있습니다. Optim의 "Mask and Move(마스킹 후 이동)" 기능을 사용하면 데이터를 추출 및 마스킹한 후 해당 데이터를 하나 이상의 실운영 외 대상 데이터베이스에 삽입 또는 로드할 수 있습니다.

또한, Optim의 "Mask in Place(제자리 마스킹)" 기능을 사용하면 타사 도구를 사용하여 추출한 데이터를 비식별화하고 이미 복제된 비운영 환경에 상주하는 데이터를 비식별화할 수 있습니다. 이 같은 대안은 테스트용 데이터를 미리 만들었거나 백업 설비를 사용하여 이 같은 테스트 데이터베이스를 구축하는 기업에게 융통성을 부여합니다. Optim을 사용하여 데이터가 상주하는 곳에서 데이터를 직접 마스킹하면 추가 처리를 위해 데이터를 옮겨야 하는 필요성을 없애면서도 데이터의 참조 무결성을 계속 보존합니다.

최선의 데이터 프라이버시 보호 방법 개요

민감한 데이터의 프라이버시와 기밀을 보호해야 하는 필요성은 여러 산업 및 지리적 경계를 초월한 실운영 및 비운영 애플리케이션 환경을 망라합니다. 그리고 많은 기업은 실운영 환경에서 데이터를 보호하기 위한 효과적인 조치를 이행한 반면, 비운영 데이터베이스 환경의 취약성에는 이제 막 관심을 보이기 시작했습니다.

그러나, 실운영 환경에 적용되는 보호 조치가 항상 비운영 환경의 요구사항을 지원하는 것은 아니기 때문에 많은 문제가 있습니다. 개발, 테스트/품질 보증 및 교육 팀이 각각의 활동을 정확하게 지원하려면 현실적인 데이터가 필요합니다. 데이터 마스킹은 개인을 식별하기 위해 사용할 수 있는 데이터 요소를 체계적으로 제거하거나 가리거나 변환할 수 있는 수단이 됩니다. 마스킹 된 데이터는 유효하고 비운영 데이터베이스 환경에서 사용하기에 적합합니다.

IBM Optim Data Privacy Solution은 상황에 민감한 데이터 요소를 마스킹하기 위한 다양한 데이터 변환 기법과 내장된 검색(Lookup) 테이블을 제공하고 Custom 데이터 마스킹 루틴까지 지원합니다. 변환 라이브러리(Transformation Library)는 사회보장번호, 지불카드 번호 및 이메일 주소를 마스킹한 유효한 값을 생성 및 전파할 수 있는 기능을 제공하여 프라이버시를 보호하면서도 정확한 테스트를 보장합니다. 가장 중요한 것은 마스킹된 데이터 요소를 관련 테이블 간에 정확히 전파하여 데이터베이스의 참조 무결성을 보존할 수 있다는 사실입니다. 더 높은 단계에서는 마스킹한 데이터를 애플리케이션, 데이터베이스, 운영체제 및 하드웨어 플랫폼 간에 전파하여 전사 시스템 전체를 보호할 수 있습니다.

Optim은 주요 데이터베이스 관리 시스템을 지원하고 다양한 실운영 데이터로부터 적절한 데이터를 한 번에 추출하고 마스킹하기 위해 사용할 수 있는 연합 액세스 기능을 제공합니다. 또한, Optim은 현재와 미래의 요구사항에 맞춰 쉽게 수정할 수 있는 유연한 기능을 갖춘 확장이 용이한 단일 데이터 프라이버시 솔루션을 제공하기도 합니다. Optim을 구현하면 개인정보보호 법규를 준수하고 민감한 정보의 기밀을 전사에 걸쳐 보호하는 데 도움이 됩니다.

IBM Optim Integrated Data Management Solutions 정보

IBM Optim Integrated Data Management Solutions는 전사 애플리케이션 데이터를 요구사항 파악 단계에서 처분 단계까지 관리하는 효과가 입증된 통합 기능을 제안합니다. Optim을 통해, 여러 팀은 (모델, 정책 및 메타데이터와 같은) 데이터 결과물을 공유하여 데이터 관리를 비즈니스 목표에 맞추고 협업을 개선할 수 있습니다. 현재 모든 유형의 조직은 Optim을 사용하여 성능을 개선하고, 데이터베이스 관리 능률을 높이고 애플리케이션 개발 속도를 향상시키며, 효과적인 관리를 지원하고 있습니다. Optim은 더 저렴한 비용과 더 적은 위험 부담으로 비즈니스 성과를 개선하면서 전사 애플리케이션, 데이터베이스 및 플랫폼에 걸쳐 확장되는 기능을 제공합니다.

추가 정보

IBM Optim Integrated Data Management Solutions에 대한 자세한 사항은 IBM 영업 담당자에게 문의하거나 ibm.com/software/data/data-management/optim/data-privacy-solution을 참조하십시오.

(135-270) 서울시 강남구 도곡동 467-12
군인공제회관빌딩

한국아이비엠주식회사
고객만족센터

TEL: (02)3781-7114
www.ibm.com/kr

2010년 6월

Printed in Korea

[1] "2006 Privacy Trust Study for Retail Banking," The Ponemon Institute, LLC and Vontu, Inc., January 2006, as referenced in "Ponemon Institute Names Most Trusted Retail Banks," Vontu Press Release, January 26, 2006.

[2] "2008 Annual Study: Cost of a Data Breach," United States, PGP Corporation, 2009. Benchmark research conducted by the Ponemon Institute, LLC. Approved for redistribution by The Ponemon Institute, p. 4

[3] Ibid., p. 4

IMW14008-USEN-01