

# 적극적인 데이터 관리를 요하는 데이터 증가 문제

Merv Adrian, IT Market Strategy 회장

[www.itmarketstrategy.com](http://www.itmarketstrategy.com)

## 전체 개요

데이터 증가는 오늘날 기업이 직면하고 있는 가장 중요한 문제이면서 또한 경쟁력 강화를 위한 좋은 기회가 되기도 합니다. 다양한 형태의 데이터 증가, 분석을 위한 상세 자료의 보관 요건, 증가하는 데이터의 고가용성 보장, 그리고 규제 및 보안 등의 많은 요인들에 의해 데이터는 지속적으로 증가하고 있습니다. 이렇듯 폭증하고 있는 중요하고 새로운 정보를 어떻게 이용할 것인가에 대한 방안을 수립하는 기업은 상대적으로 큰 경쟁 우위를 차지할 수 있습니다. 그러나 스토리지를 전체 정보 수명주기에 걸쳐 효과적으로 관리하기 위해서는 3~10배에 달하는 비용이 필요하고, 데이터 증가에 따른 시스템의 성능 및 기업의 조직 관리에도 추가적인 비용 투자가 필요합니다. 특히 지금과 같은 경제 환경에서는 스토리지 비용 증가에 대한 예산 확보가 어려운 실정입니다.

유감스럽게도, 데이터 증가에 대한 문제를 인식하고 관리하는 일관적면서 계획적인 방안을 가지고 있는 기업은 많지 않습니다. 최근의 IBM CIO 설문조사에 의하면, 자사의 데이터를 올바르게 이해하고, 종합적으로 관리하고 있다고 대답한 CIO는 15%에 불과한 것으로 나타났습니다. 경영진은 데이터 증가 문제의 심각성이나 용량 증가 속도가 앞으로 얼마나 지속적으로 가속화될 것인지를 인식하고 있지 못하고 있을 수 있습니다. 이 같은 문제를 해결하기 위한 많은 하드웨어 및 소프트웨어 신 제품이 나오고 있지만, 정보 사용 요건을 효과적으로 분석하고 해당 요건을 충족하기 위한 비용을 분명히 파악하여 예산 계획을 세우지 않는 기업은 적절한 시점에 의사결정을 제대로 내리지 못해 많은 비용을 부담하게 되는 위험에 처할 것입니다.

더 많은 데이터, 그리고 올바른 데이터는 더 나은 의사결정, 프로세스 개선, 의사소통 향상, 분석 개선에 기여하는 풍부한 상황 정보로 이어져 기업의 성과 개선을 가져올 수 있습니다. 이 모든 것은 경쟁 우위를 확보하기 위한 중요한 요소가 됩니다. 데이터를 좀 더 폭넓게 활용하기 위한 투자는 지금도 많은 기업에서 우선시되고 있지만, 이 같은 목표가 올바른 데이터를 인식, 수집 및 보호하고 가용성을 보장하기 위한 전략과 계획을 수반하지는 못하는 것이 현실입니다. 의료 이미지 데이터, 협업 저장소 및 루틴 자료실(Archive) 같은 응용 애플리케이션에서 발생하는 파일 기반 데이터는 폭증하고 있지만, 관리되지 않는 경우가 많습니다. IDC는 최근에 생성되는 데이터의 50% 이상이 파일 기반 형식으로 되어 있으며 2008년에는 출고된 파일 기반 스토리지 용량이 DB 기반 용량을 처음으로 초과했다고 지적한 바 있습니다. 그러나, 파일 기반 스토리지 관리가 “전통적인” 기존 애플리케이션의 데이터 관리 체계에 포함되는 경우는 드뭅니다.

본 글에서 “데이터 증가”라는 용어는 데이터베이스에 저장되는 정형 데이터 뿐만 아니라 위에서 언급한 막대한 양의 파일 기반 정보까지 포함하는 의미로 사용합니다.

경영진은 데이터 증가가 기술뿐 아니라 인사 및 여타 업무 프로세스에 걸쳐 다양하게 기업의 성과에 영향을 미친다는 사실을 인식하고, IT 부서로 하여금 이 문제에 대응하는 분명한 비전을 제시하도록 요구해야 합니다. 기술 관리자는 전통적인 측정 기준과 하드웨어 예산 프로세스에 의존하고 있을 수 있습니다. 그러나, CFO들은 데이터 관리가 기업 리스크를 완화하는데 있어서 담당하는 중추적인 역할을 인식하고 IT에 보다 정확한 보고를 기대하고 있습니다. 업무 혁신은 우선 업무 부서로부터 “의식적인 데이터 관리”가 이루어지도록 프로세스를 개선하는 작업이 필요하며, 이러한 의식 함양을 통해서 우선과제를 알리고 역할을 정하며 이해관계자들의 참여를 보장하게 될 것입니다.

## 데이터 증가의 원인

데이터의 증가는 불가피하고 막을 수 없다는 말은 이제 상투적인 말이 되었습니다. IBM의 연구에 따르면, 기업은 평균적으로 20~40개의 데이터 사본을 보관하고 있는 것으로 나타났습니다. 이 같은 데이터의 폭발적인 증가 규모를 관리함으로써 얻을 수 있는 이익은 상당합니다. 데이터를 증가시키는 여러 요인은 쉽게 생각할 수 있는 것이지만, 쉽게 눈에 띄지 않는 것들도 있습니다. 일상적인 트랜잭션, 비즈니스 문서, 이메일 등의 비즈니스 활동을 수행하는 간단한 행위를 포함하여 주요 데이터의 증가 요인은 다음과 같습니다.

- 애플리케이션 및 테스트 데이터의 급증.** 같은 용도로 사용되는 복수의 제품은 종종 각기 특수한 형식의 자체적인 데이터 인스턴스를 갖고 있습니다. 기존 제품과의 합병 인수, 신제품을 도입하는 사업부의 추진 과제 이니셔티브, 기존 제품의 버전 마이그레이션, 새로운 아키텍처 이니셔티브 그리고 웹 및 모바일 기기와 같은 전달 채널의 증가는 모두 신규 데이터를 발생시키고, 이 중 대다수는 중복의 형태로 존재합니다. 일부 테스트 및 개발 데이터는 필요에 따라 생성된 후 작업이 끝날 때 삭제되어야 함에도 불구하고 “영원히 사라지지 않습니다.”
- 분석 데이터 저장소의 확산.** 많은 사업장에서는 분석 데이터 마트, 데이터 큐브, 예측 모델링을 위한 테스트 데이터 등과 같이 분석 애플리케이션을 새로 개발할 때마다 새롭게 도출한 데이터 테스트 환경을 만드는 일이 일상처럼 되어버렸습니다. 이 중 상당수는 분석 도구가 특정 양식의 데이터를 요구한다는 이유로 생성되지만, 그렇지 않은 경우도 의외로 많습니다.
- 법규 준수를 위해 필요한 데이터.** 새로운 정보 법규는 오래된 데이터를 더 많이 보관해야 하는 원인이 되고 있습니다. 이는 새로운 것이 없지만, 다른 여러 이니셔티브와 마찬가지로 규제 보고 프로젝트는 과거 이력을 식별하는 경우가 드뭅니다. 여러 시스템은 비활성화된 후에도 계속 남아 있으며, 이 같은 시스템을 위해 데이터를 만들고 저장하는 “생성(Feeder)” 프로세스는 중단되지 않습니다.
- 고객 및 파트너 연계는 자체적인 데이터를 받아 고객 및 파트너에 의한 셀프 서비스를 지원합니다.** 이 같은 요구사항을 충족하는 가장 손쉬운 방법은 중복되는 하위(Subset) 데이터 저장소를 만들어 위험을 최소화하는 방법입니다. 하지만 이는 효과적인 허가 및 인증 전략이 있을 경우 불필요할 수 있으며, 분리할 필요가 있을 만큼 성능에 미치는 영향이 크지도 않습니다.
- 파일 시스템의 확장은 또 다른 문제입니다.** Gartner는 데이터의 85%가 “비정형” 데이터라고 추정하고 있습니다. 기존 데이터베이스 외의 파일 콘텐츠 증가는 일반적으로 데이터베이스의 확장을 정확하게 추적하기 위해 예측됩니다. Enterprise Strategy Group이 최근에 실시한 조사에서는 설문에 응한 기업의 36%가 이메일 자료만 매년 41% 넘게 증가할 것으로 예상되고 있습니다.

## 데이터 증가로 인한 비용 이슈

단순히 “더 많은 데이터”를 저장하는 방법은 많은 비용을 초래합니다. 데이터 용량이 증가하면 그에 따른 비용도 매년 증가합니다. 데이터 증가로 인해 가시적으로 발생하는 직접적인 비용은 다음과 같습니다. 스토리지 하드웨어는 영구적이고 피할 수 없는 비용입니다. 외장 스토리지 하드웨어는 매년 수백억 달러의 자본 지출을 초래합니다. Forrester의 Andrew Bartels는 스토리지 하드웨어가 일반적인 하드웨어 예산에서 차지하는 비중을 약 9%로 추정하고 있습니다. 데이터 증가 속도는 계속하여 스토리지 하드웨어의 발전 속도를 증가하고 있습니다. 그 밖의 기술 비용으로는 네트워크 통신 비용, 전력 및 냉각 비용과 공간 비용 등이 있습니다. 일각에서는 이 두 가지 비용이 향후 5년 간 3배나 증가할 것이라 예상하고 있습니다. IDC는 “데이터 센터의 평균 디스크 드라이브” 한 대의 전력 및 냉각 비용이 2008년에 36.29달러였다고 말합니다. 그리고, 건립된 지 10년이 넘는 데이터 센터(즉, 대부분의 데이터 센터)가 더 많은 전력과 공간을 얻으려면 더 많은 비용을 지불해야 합니다.

쉽게 눈에 띄지 않는 간접적인 비용 중 대표적으로 관리 비용을 들 수 있습니다. 모든 데이터베이스 라이선스는 돈을 지불하고 구입해야 합니다. 데이터를 사용이 아닌, 관리하기 위해 즉, 백업 및 복원, 자료보관, 튜닝, 설계, 데이터 이동, 마스터 데이터 관리 등을 위해 사용되는 소프트웨어는 상당한 추가 비용을 발생시킵니다. Forrester의 Bartels는 소프트웨어 관리 소프트웨어에 평균 소프트웨어 예산의 7%가 투입되는 것으로 추정하고 있습니다. 이 같은 도구는 종종 여러 업체의 제품으로서, 기능은 중복되는 반면 서로 다른 기술력을 요구합니다. 여기에 희소한 기술력을 보유한 인력을 포함하여 인건비를 더 하면 상황은 더욱 심각해집니다. 데이터 용량이 클수록 데이터 관리를 전담하는 인력도 많아지는데 경험이 풍부한 숙련된 인력은 커다란 비용이 소요됩니다.

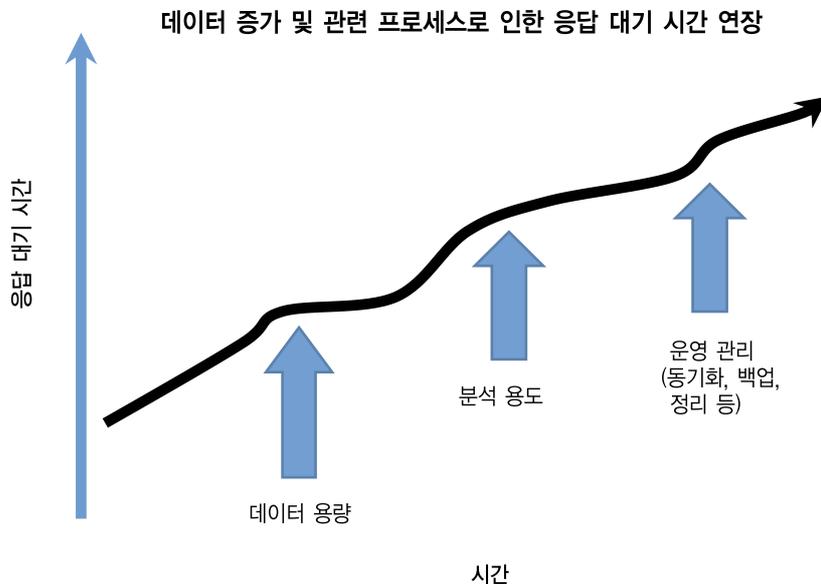
2009년에는 불황으로 인해 IT 환경에 대한 비용 지출이 여의치 않았음에도 불구하고 Gartner가 실시한 조사에서는 선진 시장에서 설문조사에 응한 기업의 57%와 신흥 시장 기업의 75%가 2009년 스토리지 지출이 2008년보다 증가할 것으로 예상된다고 답했습니다. 선진 시장에서 스토리지 지출이 감소할 것으로 예상한 기업의 비율이 (43% 대 25%) 더 높았던 이유가 불황의 타격을 더 많이 받아서인지, 아니면 더 효과적인 데이터 관리 방법 때문인지는 확실하지 않지만, 데이터 관리 개선으로 인한 효과임을 증명하는 증거는 아직 제시된 바가 거의 없습니다.

## 데이터 증가로 인한 성능 이슈

비즈니스 목표의 달성은 비즈니스 프로세스를 지원하는 애플리케이션의 효과적인 성능에 의해 좌우됩니다. 그리고, 이 같은 성능은 데이터의 가용성에 의해 좌우되며 데이터의 용량에 영향을 받습니다. 따라서, 데이터의 가용성을 보장하고 데이터 증가량을 예측하는 기능이 대부분의 중요한 비즈니스 애플리케이션의 설계 및 배포에 항상 포함되는 것은 우연이 아닙니다. 하지만 이 같은 방법은 역설적으로 많은 기업에서 데이터 관리의 책임이 분리되는 결과로 이어졌습니다. IT 부서의 서비스 기반 관리로의 전환은 전체 애플리케이션 포트폴리오에 걸쳐 공통된 데이터 가용성 메커니즘으로 확대 적용되지 못했고, 공유 스토리지와 공유 백업 및 복구 관리, 그리고 비즈니스 요건에 따른 적절한 스토리지 기술에 대한 비전은 여러 사업부나 IT 직능의 우선순위가 서로 상충됨에 따라 종종 부재하게 됩니다.

애플리케이션 자체의 성능은 처리되는 데이터의 엄청난 양에 의해 좌우됩니다. 트랜잭션 처리 및 비즈니스 분석 성능은 모두 데이터 용량을 근거로 하여 산업 표준 척도로 사용됩니다. 실시간 분석과 동향 분석의 정확도와 모델링과 예측을 개선하는 장기 시나리오 및 모델에 대한 수요가 증가함으로 인해, 처리할 데이터의 양은 앞으로도 오랫동안 지속적으로 증가할 가능성이 큽니다. 요구사항이 매우 다른 여러 이해관계자가 동일한 데이터를 재사용하는 여러 애플리케이션을 만들에 따라, 이들을 지원하기 위한 중복 사본들이 생성되었고, 이 시스템들은 동기화를 유지해야 하기 때문에 정확도와 가용성에 영향을 미칩니다. 그리고 이와 같은 프로세스는 또 다른 성능의 부담 요소가 됩니다.

또한, 애플리케이션 데이터를 백업하는 데 사용할 수 있는 시간은 데이터 용량처럼 쉽게 확장할 수 없으며 새로운 데이터를 생성하는 새로운 애플리케이션에 대한 관리도 필요합니다. 기업이 성장하고 세계화되면, 사용 가능한 백업 시간은 오히려 줄어들거나 모든 표준 시간대에 있는 사업장이 데이터를 필요할 때 즉시 사용할 수 있어야 하는 데이터 가용성의 요건으로 인해 완전히 사라지는 경우가 많습니다. 이 같은 문제를 해결하기 위한 또다른 전략은 종종 성능에도 영향을 미칩니다.



### 데이터 증가로 인한 관리 이슈

다른 여러 IT 리소스는 주로 물리적 자산으로 취급할 수 있지만, 위험을 수반하는 데이터는 심오하고 복잡합니다. 특히 보안과 법규 준수 영역과 같이 데이터 거버넌스는 일반적으로 구분되어 있습니다. 데이터 거버넌스는 애플리케이션 별로 관리될 수도 있지만, 이 같은 수준에서도 책임감이 결여될 수 있습니다. 기업 차원에서, 법규 준수와 개인정보보호에 관한 고려사항을 관리하기 위한 정책은 이를 관리하기 위한 허가와 권한을 부여 받은 분명한 조직 없이 이론적으로만 존재할 수도 있습니다. IBM이 의뢰한 2008년 Forrester Consulting 조사에서는 심층 면접 조사에 응한 21개 기업 중 대부분의 기업에서 전체 데이터의 수명주기에 걸친 활동이 분리되어 이루어지고 설계 및 개발 부산물, 정책 및 표준에 대한 효과적인 의사소통이 단계별로 거의 이루어지지 않으며 시스템을 개선하기 위한 자동화도 거의 계획되어 있지 않은 것으로 나타났습니다.

사베인옥슬리법이나 Basel II 같은 다양한 국제 표준의 규제 준수 요건을 지원하도록 설계된 시스템의 상당수는 기업 내에서 전사적 시각이나, 더 이상 “실시간 혹은 실시간에 준하는 액세스”가 의무화되지 않고 보고 요건이 적용되지 않는 데이터를 아카이빙 하기 위한 계획이 없이 구현되었습니다. 이로 인해 데이터의 증가는 더욱 심각해지고, 점점 더 해결하기 어렵고, 이슈화되고 있으며, 비용 부담은 날로 커져 기업의 위험요소가 되고 있습니다.

문제를 해결하는 데 있어서 놓치기 쉬운 중요한 단계 중 하나는 데이터를 이해하고 권한을 부여하고 관리해야 하는 이해관계자가 누구인지를 파악하는 단계입니다. 이러한 역할을 담당하는 사람으로는 애플리케이션의 개발, 배포 및 관리에 관여하는 비즈니스 애널리스트, 아키텍트, 개발자, 테스터, DBA 및 시스템 관리자 등이 있습니다. 하지만 이들의 지지자인 애플리케이션 담당자들의 입장도 단지 설계 시뿐만 아니라 지속적으로 충분히 반영되어야 합니다. 가장 중요한 것은 이들의 모든 노력이 기업의 법무 담당자, 보안 책임자 그리고 궁극적으로는 이들 모두의 상사인 경영진들이 규정한 범위 내에서 이루어져야 한다는 것입니다.

데이터 거버넌스 프로세스를 자사의 IT 포트폴리오 내에 설계하는 식견을 가졌던 기업은 많지 않습니다. 대부분의 기업은 긴급한 업무 등 당면한 요건에 대응하는 데 바쁜 나머지 상위의 전사적 지침이 될 정책과 프로세스를 정할 여유가 없었습니다. 대신, 데이터 거버넌스의 부재로 인한 위기가 반복적으로 일어날 때에만 데이터 관리 전략을 뒤늦게 수립합니다. 그리고 이 같은 위기는 반드시 찾아옵니다. 하지만 총체적인 데이터 관리 이니셔티브가 없는 경우에도 데이터 증가의 문제에 대처하기 위한 매우 권고할 만한 사전 조치를 취할 수 있습니다. 이러한 적극적인 사전 조치 과정에서 대부분의 기초를 다지는 계획이 수립될 것이며, 시간과 자금이 적절히 보장되는 시점에 좀 더 폭넓은 데이터 거버넌스 이니셔티브가 추진될 수 있을 것입니다.

## 적극적인 데이터 관리의 5 단계

데이터 증가 관리를 마스터하는 단계는 총체적인 데이터 거버넌스 수립 단계와 유사합니다. IT Market Strategy은 대부분의 대규모 IT 사업장이 데이터 증가규모 관리에 있어서 5단계 성숙도를 확인할 수 있으며, 많은 업무들이 종종 여러 가지 일에 관해 한꺼번에 진행되지만 기업이 상위 단계에서 좀 더 효율적인 관리를 하기 위해서는 초기 단계의 기초적인 요구사항을 완수하도록 하는 것이 중요하다고 믿습니다.

### 1 단계 - 기업 데이터에 대한 자각

1단계의 성숙도를 지니는 기업은 기업 전반에 걸쳐 데이터의 증가규모나 여타 기준에 대한 우선 과제가 거의 정해지지 않은 상태입니다. 데이터는 일반적으로 애플리케이션 별로 혹은 사업부별로 저장됩니다. 분석가, 프로그래머 및 관리자는 단순히 보유하고 있는 시스템을 가지고 작업합니다. 데이터의 중요도에 대한 의식을 제고하기 위해 디스커버리 및 프로파일링 도구를 사용하여 모든 데이터 소스에 대한 액세스와 관계도를 식별 및 문서화하지만, 실행은 정기적으로 계획되기 보다는 일회성인 경향이 있습니다. 이 단계에서는 데이터 증가 규모가 기록될 수는 있지만 이에 대처하는 방법을 결정하기 위해 필요한 정보는 존재하지 않습니다. 예를 들면, 데이터의 중복 처리 여부는 명시되지 않거나, 실행 여부가 계획되지 않는 경우가 대부분입니다.

상위 단계로 발전하기 위해서는, 데이터 관리 조직 및 예산이 구성되고, 인력이 지원되어야 하며 공통 용어를 정의하는 메타 데이터 시스템과 더불어 데이터의 배포 및 가용성을 위한 문서화된 구조가 의무적으로 제공되어야 합니다. 이 같은 노력은 데이터 증가가 “이슈가 되고 있는 부문”에서 제약적이고 제한적으로 수행될 수 있습니다. 그러나 그렇다 하더라도 구성된 기술력과 메타데이터는 더 광범위한 데이터 거버넌스의 기초를 형성하게 될 것입니다.

## 2 단계 - 데이터의 가치와 사용량의 측정

위에서 도출한 구조적 정의를 토대로, 기업은 주요 비즈니스 프로세스의 데이터 의존도를 바탕으로 데이터의 가치를 측정하는 가장 기초적인 방법을 정의하게 되었습니다. 이는 모든 데이터 위험을 정의하기에는 충분하지 않지만 위험을 정의하는 데에도 도움이 됩니다. 그리고 데이터 사용 모니터링을 통해 이해관계자가 누구인지를 위와 유사한 기초적인 수준으로 파악할 수 있습니다. 이러한 기초 자료를 기반으로 수립된 몇 가지 원칙은 이제 확대 적용할 수 있습니다. 구체적인 데이터 전략은 비즈니스 프로세스의 실행 및 최적화와 연계될 수 있습니다. 예를 들면, 특정한 BI 요건을 지원하기 위해 데이터 웨어하우스와 데이터 마트가 구축되었을 수 있습니다. 가용한 모든 데이터는 여전히 보관되고 있고 데이터 증가는 계속 관찰되고 있습니다. 현 단계에서는 데이터의 증가 규모가 관리되지는 않지만 데이터의 출처, 이해관계자 및 상대적인 가치가 파악됩니다.

## 3 단계 - 데이터 품질 및 보안 관리

데이터를 보호하기 위한 데이터 품질 관리, 액세스 제어를 위한 정보 보안 표준, 보호 대상 개인 정보 정의 및 위험 요소 제거는 적극적인 데이터 관리를 위한 마지막 기초 단계입니다.

이제 규제 및 준수 기준을 따르는 거버넌스의 기초 원칙이 수립되고 이해관계자가 파악되고 정책이 수립된 상태에서, 스토리지 전략의 다음 단계로서 데이터 폐기 및 사용량 기반 최적화 정도를 조사할 수 있습니다. 잦은 보안 위반, 일괄 배치 처리 시간을 초과하는 데이터 전송, 표준 용량 기준을 초과하는 디스크 등과 같은 소스 애플리케이션의 작업 반복을 자주 유발하는 데이터 거버넌스 프로세스는 무엇입니까? 이 같은 문제의 원인이 규칙인지, 데이터 자체인지, 아니면 둘 다인지를 파악하면 데이터 증가 문제에 대한 중요한 정보를 얻게 될 것입니다.

## 4 단계 - 정보 수명주기에 대한 인식과 데이터 관리 의식(Stewardship)의 만남

본 단계에서는 단순히 시간 이력 정보를 기준으로 처분해야 하는 데이터 자산, 애플리케이션 및 인프라와, 정책 및 데이터의 중요도를 기준으로 아카이빙 할 수 있는 데이터를 구별하는 작업을 시작합니다. 성공적인 정보 수명주기 관리(ILM)를 위한 핵심 요소인 정보의 수집, 사용, 보관 및 삭제에 관한 정책은 단지 데이터 증가 관리뿐만이 아닌 훨씬 더 많은 이니셔티브에 적용될 수 있습니다. 그 중에서도 데이터 증가 규모에 대한 관리에 이러한 정책이 적용될 경우, 가장 많은 지원이 명확하게 이루어지고, 전사적 관점에서 최고의 제시하게 될 것입니다. 현 단계는 데이터 증가 이니셔티브가 ILM 전략을 지원하는 작업과 밀접히 연결되며, 적극적인 데이터 관리를 위한 마지막 단계를 앞두고 있습니다.

## 5 단계 - 비즈니스 성과에 따른 최적화된 데이터 관리

이제는 분명한 상관 관계에 의해서 정보가 지원하는 주요 업무 의사결정을 기준으로 어떤 정보를 왜 저장하는지에 초점을 맞출 수 있게 되었습니다. 데이터에 대한 신뢰는 정책에 기반한 데이터 관리의 보급, 평가 및 대중화로 인해 극적으로 향상됩니다. 공격적인 중복 제거, 압축 및 삭제 전략은 데이터 사용량을 측정하고 최적화합니다. 적정 사이즈의 데이터 서브셋(Subset)으로 구성된 테스트 데이터는 요구 즉시 생성되고 더 이상 필요하지 않을 때 삭제됩니다. 데이터의 증가는 측정 가능하고 계획적인 공급 전략이 수립됩니다. 이를 통해 의사결정을 지원하기 위해 어떤 정보가 필요한가? 정보는 얼마나 정확해야 하는가? 정보를 수집, 생성 및 공급하는 가장 효율적인 프로세스는 무엇인가? 정보는 언제까지 제공해야 하는가? 등과 같은 다양한 문제들이 해결됩니다.

## 솔루션 : 데이터의 증가규모를 억제하는 적극적인 데이터 관리

통제되지 않는 데이터가 증가함으로 인해 위에서 언급한 관리, 성능 및 비용의 복잡한 상호작용은 더욱 복잡해집니다. 계획과 분명한 책임 목표와 측정 기준이 없는 기업은 데이터의 증가로 인해 자사 기술과 프로세스와 인력에 과중한 부담이 가해짐에 따라 어려움을 겪게 될 것입니다. 이 같은 문제를 해결하기 위해서는 조직적인 변화와 정책 수립 그리고 이를 집행하기 위한 응용 프로그램이 필요합니다. 데이터를 “소유”하는 사업부 관리자 교육을 포함한 교육은 새로운 모델을 중심으로 비즈니스와 IT를 결속시키는 중요한 방법이 될 것입니다.

데이터 관리에 대한 통제력을 장악하려면, 기업은 먼저 데이터가 진단과 계획 없이 계속 증가되지 않도록 하겠다는 의식적인 결정을 내려야 합니다. 위에서 언급한 다섯 단계의 성숙도를 이루기 위해서는 각 단계를 완료할 때마다 다음 3가지의 중요한 태스크를 이행해야 합니다.

**진단** - 문제와 그로 인한 비용을 문서화함으로써 문제를 분명히 정의합니다. 기술 명세서(Descriptive inventory)는 좋은 출발점이 될 것입니다. 프로파일링 및 디스커버리 도구를 사용하여 의존성, 중복, 소유권, 이해관계자, 액세스 패턴 및 노출 위험을 확인하되, 여기서 그치지 않고 필요한 도구 및 방법을 정당화할 근거를 마련하십시오. 문제 개선을 이루기 위해서는 투자가 필요할 것이므로, 현 상황으로 인해 초래되는 비용을 줄이고 시나리오별 데이터 증가 속도를 줄임으로써 얻을 수 있는 이익을 문서화하십시오. 줄일 수 있는 데이터의 양을 수치화하고 불필요하고 중복되는 데이터를 제거하고 값비싼 엔터프라이즈 스토리지 또는 하이엔드 스토리지에 저장해 둘 필요가 없는 데이터를 아카이빙 하여 2차 스토리지에 보관함으로써 얻을 수 있는 이익을 계산하십시오.

**계획** - 진단 과정에서 설계된 시나리오를 선택하여, 현재 작업에 참여하지 않은 관련자를 참여시킬 기회로 활용합니다. 그들에게 분석 결과를 제시하고 다음 단계에 그들을 참여시킵니다. 진단 단계에서 “문제가 무엇인지”를 조사하여 작성한 문서로는 충분하지 않습니다. 문제의 해결을 위해 “무엇이 필요한가”에 대한 비전을 동반해야 하는데, 그러기 위해서는 그들을 참여시켜야 하며, 새로이 파악된 이해관계자로 하여금 문제가 “어떻게 해결되어야 하는지”를 정의하는 프로세스에 참여시켜야 합니다.

계층형 스토리지 구조에서 서로 다른 응답 속도를 제공하는 것과 같이 비즈니스 실행 및 최적화 관점에서 재정의를 데이터 요건 및 대안도 존재하게 될 것입니다. 즉시 사용 가능해야 하는 데이터는 무엇인지, 다음 날이나 다음 달이나 분기 말까지 기다릴 수 있는 데이터는 무엇인지, 이를 문서화하여 카탈로그/메타데이터의 기초를 만드십시오. 보다 완전한 거버넌스 수립을 위한 노력을 기울이면 훨씬 더 풍부한 메타데이터가 수집될 것이지만, 이처럼 매우 제한적인 작업도 향후 전사적 메타데이터 수집 시에 필요한 기초 자료로 다시 사용할 수 있게 될 것입니다. 현재 데이터의 용량과 증가 추이를 측정할 기준을 정하고, 이를 증가 예측 자료와 비교하여 추적할 프로세스를 마련하십시오. 이 같은 노력을 위해 필요한 부서를 만들어 인력을 총원하고 정해진 역할을 수행할 인력을 교육시키십시오. 새로운 데이터 관리 구조를 위한 “데이터 관리인(Steward) 인증” 단계를 수립하고 이를 실무에 적용하십시오.

즉각적인 데이터 감소 노력의 범위와 비용에 대한 합의를 도출하는 동안 이를 실행하기 위한 방법에 대한 전략 및 계획을 수립합니다. 이러한 전략적인 기술로는 스토리지 가상화의 구현 혹은 확대적용, 구형 스토리지 하드웨어의 적절한 폐기와 재사용, 필요한 새로운 스토리지의 구입, 데이터 삭제, 데이터 중복 제거 및 아카이빙이 있습니다.

**실행** - 마지막은 가능한 다양한 도구 및 기술을 사용하여 임무를 완수하는 매우 간단한 작업입니다. (다음 '적극적인 데이터 관리를 위한 기술, 기법 및 도구' 단락 참조.) 이를 위해서는 직원이 적극적으로 참여하고, 충분히 훈련되고, 목표 기준이 설립되고, 이해 당사자들에게 적절히 보고되어야 합니다. 예를 들면 이해관계자와의 정기적 검토를 통해 아카이빙 된 데이터는 현업이 필요한 적정 조회 성능을 보장하게 됩니다. 또한 예산 수립 기획부서와 달성된 절감 비용 및 계획된 구매 절감 효과 등을 의논함으로써 구매 과정에서의 문제가 야기되지 않도록 계획할 수 있습니다.

보다 효과적인 데이터 관리는 신규 지출을 줄이고, 기존 투자를 활용하고, 과부하를 줄여서 시스템의 성능을 개선하고, 보안 및 규제 준수 관련 위험을 완화하는 등의 여러 가지 방법으로 기업의 성과를 개선할 것입니다. 데이터 증가 규모를 관리하는 의식적이고 적극적인 방법의 기초가 될 프로세스 방침과 인프라를 개발하는 일은 데이터 거버넌스 수립을 위한 중요한 단계이며, 투자 대비 막대한 효과를 가져올 것입니다.

## 적극적인 데이터 관리를 위한 기술, 기법 및 도구

모든 데이터 관리 전문가에게는 도구가 필요합니다. 이에 필요한 몇 가지 중요한 도구와 IBM 제품의 예를 아래와 같이 확인해 보았습니다.

**소프트웨어 자산 관리** : 애플리케이션 인스턴스 감소는 강력한 도구입니다. 애플리케이션 목록을 합리적으로 개선하면 데이터를 줄일 수 있습니다. (예 : IBM Optim Application Retirement Solution)

**애플리케이션 데이터 아카이빙** : 애플리케이션을 처분할 때 몇몇 아카이빙 솔루션을 사용하면, 애플리케이션과 BI 도구는 비즈니스 개체를 추가적인 영구 사본으로 "복원"하지 않고 장기보관 스토리지에서 계속해서 직접 불러올 수 있습니다. (예: IBM Optim Data Growth Solution, IBM Smart Archive)

**데이터 프로파일링 및 디스커버리 도구** - 자동 디스커버리를 도와 분명히 드러나지 않는 데이터 간의 의존성을 밝혀낼 수 있습니다. 이들 도구는 시스템의 기초를 형성하는 메타데이터의 소스가 될 수 있습니다. (예 : IBM InfoSphere Discovery, IBM Information Analyzer)

**메타데이터 카탈로그** : 이는 애플리케이션 이외의 액세스를 지원하고, 레코드 정의의 출처가 됨으로써 완전히 폐기할 수 있는 데이터의 "영구 삭제"와 규제 준수, 개인정보보호 및 액세스에 관한 정책을 문서화할 수 있습니다. 전사적 메타데이터 전략은 수집, 분류, 보존 및 삭제의 유지관리를 포함하며 광의에서 MDM과 중첩될 수 있습니다. 선택한 도구가 선택한 메타데이터 보관소와 통합되도록 보장하거나, 향후 통합을 위한 분명한 로드맵을 정하십시오. (예 : IBM InfoSphere Metadata Workbench, IBM InfoSphere Business Glossary)

마스터 데이터 관리는 종종 더 폭넓은 메타데이터 카탈로그 요건의 일부가 될 수 있습니다. 이 부분은 주요 애플리케이션 영역에서 데이터 용량을 관리하고 데이터 증가 속도를 둔화시키기에 충분할 수 있지만, 많은 데이터가 그 범위 밖에 있기 때문에 완전히 종합적인 솔루션을 제공하지는 않을 것입니다. 예 : IBM InfoSphere Master Data Management Server

**ECM 시스템 통합** - Gartner에 따르면, 기업은 5~20가지에 이르는 별도의 시스템을 운영한다고 합니다. 그 중 특정 시스템은 다른 시스템에 포함될 수 있습니다. (예 : IBM의 ECM 제품 라인에는 SAP 전용 옵션인 IBM CommonStore for SAP이 포함되어 있습니다.)

정도의 차이는 있지만, 데이터 압축은 여러 데이터베이스 제품에서 사용됩니다. DBMS를 업그레이드하여 최신 압축 기능을 이용하십시오. (예 : IBM DB2 Storage Optimization for DB2 9.7. 파일 기반 시스템에서는 소프트웨어와 정교한 데이터 압축 기능을 탑재한 스토리지 어플라이언스를 사용하여 데이터를 압축할 수 있습니다.)

**스토리지 하드웨어 업그레이드** : 데이터 센터에서 더 밀도가 높고 효율적인 하드웨어가 이용되도록 하십시오. 건립된 지 10년이 넘는 다수의 데이터 센터들은 새 하드웨어에 투자함으로써 공간, 전력 및 기타 운영 비용을 절감할 수 있습니다.

이 같은 절약 효과는 특히 유리한 보상 업그레이드 제도를 시행하는 업체와 거래할 경우 새로운 취득 비용을 크게 상쇄할 수 있습니다.

중복 제거 소프트웨어는 종종 하드웨어 어플라이언스 안에서 데이터 집합의 중복을 제거하기 위해 사용됩니다. 중복 제거율은 매우 높을 수 있습니다. 네트워크 연결 스토리지 환경에서, 몇몇 중복 제거 제품의 고객은 최고 7:1에 달하는 감소율을 실현할 수 있었다고 주장합니다. 백업 세트에서 이 비율은 20:1 이상으로 그리고 VMware 및 기타 유사한 가상 환경에서는 최고 100:1까지 상승합니다. 파일 기반 시스템과 함께 사용되는 “단일 인스턴스” 전략은 추가 사본의 필요성을 완전히 배제합니다. 이 소프트웨어는 파일 아카이빙에 사용되며, 일부 전사 애플리케이션은 이를 직접 지원합니다. (예: IBM ProtecTIER)

**Hierarchical Storage Strategy** : 비즈니스 사용 프로필을 기준으로 우선순위를 정하여 계층형 스토리지를 설계. 이 방식을 구현하는 소프트웨어는 기존의 구형 스토리지 장비와 같이 상대적으로 저렴한 하드웨어에 중요도가 낮은 데이터를 저장합니다. (예: IBM Tivoli Storage Manager HSM)

스토리지 가상화는 중복 제거 작업을 촉진시킬 수 있습니다. 사용자가 필요한 모든 데이터와 연결하는 액세스 레이어 존재시, 별도의 추가 사본은 없애고 중복을 제거할 수 있습니다. (예: IBM SAN Volume Controller, IBM Tivoli Storage Productivity Center)

테스트 데이터 관리 솔루션은 “적절한 크기”의 데이터를 실운영 외 테스트, 개발 등의 용도로 제공하여 실운영 데이터 베이스에 있는 데이터 용량이 복제된 사본의 수만큼 배수로 늘어나는 경우를 없애 줍니다. (예: IBM Optim Test Data Management Solution)