ILM 구현을 위한 Database 기반의 Archiving System 구축

김동한
고려대학교 컴퓨터정보통신대학원

e-mail : didiland@gmail.com

Archiving System construction based on Database for ILM implementation

Dong-Han Kim
Graduate School of Computer & Information Technology , Korea University

요 약

IT 환경이 급속도로 발전해 가고 있다. 기업 내에서 생성 및 보관되는 데이터의 양은 시간이 흐름에 따라 급격히 증가하고 있으며, 기업의 의사결정에 중요한 데이터의 가치는 변화를 거듭한다. 데이터의 증가는 곧 그것을 관리하는 비용의 증가를 불러오고, 방대한 데이터 속에서 진정한 정보의 가치를 창출하는 것이 어려운 상황에 직면해 있다. 이러한 기업의 비즈니스 프로세스에 맞춰 효과적인 ILM(Information Lifecycle Management)의 구현은 반드시 필요하다.

본 연구는 데이터의 가치를 보다 효과적이며 비용 효율적으로 사용할 수 있는 ILM 의 구현을 위해 Database 기반의 Archiving System 구축 방법론을 제시하고 이를 바탕으로 실제 ILM 이 구현된 기업의 구축 효과 및 성능 평가를 통해 그 타당성에 대하여 연구하는 것을 그 목적으로 한다.

1. 서론

2007 년 IT 전문 조사기관인 IDC 가 발표한 보고서에 따르면 디지털 정보 세상에서 1 년에 인류가 만들어내는 데이터 용량이 약 161 엑사바이트(ExaByte) 정도라고 추산했다. 1 엑사바이트는 1 기가바이트(GigaByte)의 10 억 배의 수치로, 161 엑사바이트의 정보량만으로 계산하면 대용량 아이팟이 약 20 억대가 필요한 어마한 양이다. 더욱이 2010 년에는 전세계적으로 1 년간 생산되는 데이터가 988 엑사바이트에 육박 할 것이라고 한다. [1][2]

그러나 실제로 저장된 데이터의 90%는 일정 기간 이후에는 거의 사용되지 않는다고 분석되고 있다. [3]

또한 IT 환경이 급속히 발전하면서 위와 같은 데이터의 폭발적인 증가는 시스템 운영의 비효율성으로 인한 저장 장치 관련 구매/관리 비용증가를 초래하여 기업은 인프라 관리 비용 통제의 어려움에 직면하게 되었고, Database 기반의 어플리케이션 성능저하, 데이터 백업 및 복구 등의 소요시간 증가로 인하여 IT 서비스 수준의 저하를 가져왔다.

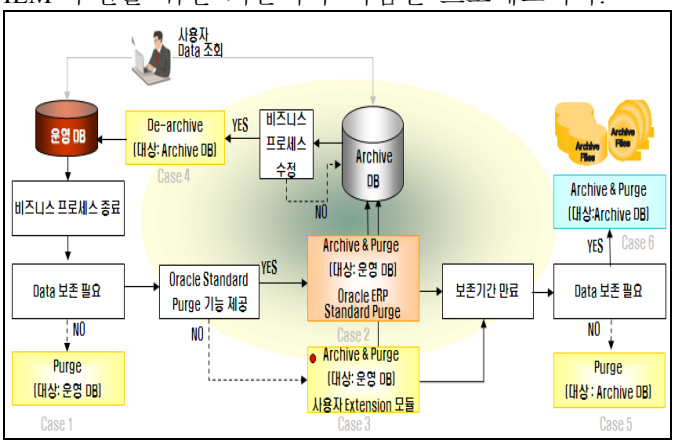
이러한 문제점들에 대해 각 기업은 대량의 데이터를 어떤 방식으로 관리하는 것이 가장 효율적인가에 대해 관심을 기울이게 되었고 ILM(Information Lifecycle Management) 은 충분한 해결 방안이 될 수 있다.

본 논문은 ILM 구현을 위한 다양한 시스템 구축방법 중 스토리지 계층화 기반의 구축방안이 아닌 Database 기반의 Archiving System 구축 방법론을 제시하고, ILM 이 실제 구현된 기업의 분석 및 성능 평가를 통해 구축효과와 그 타당성에 대하여 파악하였다.

2. Database 기반의 Archiving System 구축 방법론

Database 기반의 Archiving System 구축을 위하여 S-MAP(Strategic, Systematic and Simple-Methodology of Archive & Purge) 이란 단계별 구축 방법론을 제시하고자 한다. S-MAP 은 Archiving System 을 구축한 실제 프로젝트 경험을 기반으로 하여 만들어졌다.

첫 번째, 분석 단계에서는 Archive & Purge 대상 시스템의 인프라 아키텍처, 어플리케이션 아키텍처를 이해한 후 Archiving System 아키텍처 및 운영/과거/통합 데이터 조회 방안을 수립한다. Archive & Purge 는 스토리지 계층화 기반의 구현이 아닌 Database 기반의 ILM 구현을 위한 기본이자 핵심인 프로세스이다.



(그림 1) Archive & Purge Process

즉 (그림 1)과 같이 Archive & Purge 는 Database 를 사용하는 환경에서의 사용자가 자주 Access 하지 않는 데이터를 운영 Database 에서 삭제하고 다른 Database 인 Archive Database 에 저장하는 것이다.

조회 방안 마련을 위해서 분리된 데이터는 과거/통합 데이터 접근 환경을 제공하여 어플리케이션 프로그램의 특별한 변경 없이 사용자가 Database 의 요소인 Synonym 과 Merge View 를 통하여 과거 데이터를 조회가 가능하도록 구성한다. [4]

또한 어플리케이션에서의 데이터 생성 흐름을 기준으로 가치에 따라 데이터를 분류한 후 데이터 관리 기준을 수립하고, Archive & Purge 대상 테이블을 정의한다.

분석 단계에서의 주 성공요소는 책임 있는 경영층의 관심 및 지원, 업무지원그룹의 Archive & Purge 에 대한 이해 그리고 적극적/긍정적 참여가 필수적이며, 이를 위해 Archive & Purge 에 의해 영향을 받는 당사자(현업 업무 담당자, 유관 업무 담당자)와 의 원활한 의사 소통이 요구된다.

두 번째, 설계단계에서는 수립된 데이터 관리 기준에 의거하여 데이터 구성 별 Archive & Purge 조건을 SQL Script 형태로 정의하고 Archiving System 구축 후의 Database 논리 모델을 설계한다.

더불어 Archive & Purge 대상 테이블의 데이터 증가량과 향후 일정기간 동안 Archive Database 에 저장되리라고 예상되는 데이터 량을 예측한다.

설계 단계에서의 주 성공요소는 데이터 관리 기준 및 통합/과거 데이터 조회 방안의 변동을 최소화하는 것이며, 업무 지원 그룹과 시스템 지원 그룹의 Archive & Purge 에 대한 적극적 협조가 요구된다.

세 번째, 구현단계에서는 단위 테스트, 통합테스트를 수행하여 수립된 데이터 관리 기준에 의거하여 실제 Archive & Purge 작업을 수행할 수 있는지를 검증하고 Archiving System 구축 후의 변화를 최종 사용자에게 공지한 후 운영 시스템의 데이터 이관 및 초기 Archive & Purge 작업을 수행한다. 또한 Archiving System 사용자 및 운영 담당자를 위한 교육을 수행한다.

구현 단계에서의 주 성공요소는 업무 지원 그룹과 시스템 지원 그룹의 통합테스트 및 운영이관에 대한 적 시성 있는 협조와 최종 사용자의 변화 내역에 대한 이해이다.

3. 구축 효과 및 성능 평가

ILM 구현을 통하여 산출된 효과는 아래와 같다.

데이터 가치평가에 의한 분류로 효율적인 데이터 관리가 가능해졌고, 데이터 중요도에 따른 차별화된 백업 전략이 가능해져 백업 및 복구 소요시간이 단축되면서 유지보수의 효율화가 가능해졌다.

또한 스토리지 투자/관리 비용 감소와 더불어 잠재적인 하드웨어(메모리, CPU 등) 증설의 부담 감소로 인한 비용 절감 효과를 얻게 되었다.

본인이 프로젝트 구성원으로서 Archiving System 을 구축한 K 사의 경우 대상 운영 시스템에서는 약 600G Byte 를 Archive & Purge 하여 Clone 시스템까지 포함하여 총 2,400 기가바이트(660G Byte * 4)의 1 차 스토리지 절감 효과를 보았다.

앞으로 약 5 년 후에는 약 15 테라바이트 (TeraByte)의 1 차 스토리지 절감 효과를 얻을 것으로 예측된다. (Archive & Purge 전: 8TeraByte * 4 = 32TeraByte, Archive & Purge 후: 4.4TeraByte *4 = 17.6TeraByte)

그리고 <표 1> 은 Archive & Purge 작업 후 테이블 관련 프로그램의 성능이 평균 15% 향상된 측정치 값을 나타내고 있다.

즉 데이터 Access 와 가용성 향상으로 인한 시스템 성능 개선효과를 확인 할 수 있다.

<표 1> K 사 Trace 분석을 통한 응답 속도 개선효과

프로그램	Archive & Purge 전		Archive & Purge 후		응답속도 비율분석	
	Total	Main SQL	Total	Main SQL	Total	Main SQL
1	1.98	1.84	0.76	0.07	61.62%	96.20%
2	29.45	29.2	25.7	25.56	12.73%	12.47%
3	255.35	248.92	128.31	127.53	49.76%	48.77%
4	259.86	259.41	136.77	136.53	47.37%	47.37%
5	68.98	68.87	48.12	47.94	30.25%	30.40%
6	0.23	0.14	0.17	0.01	26.09%	92.86%
7	9.47	9.27	0.85	0.59	91.02%	93.64%
8	0.7	0.25	0.71	0.14	-1.43%	44.00%

4. 결론

본 연구를 통하여 ILM 구현을 위한 Database 기반의 Archiving System 의 필요성과 구축 방법론을 제시하고 이를 기반으로 실제 ILM 이 구현된 기업의 구축 효과와 성능 개선을 일정 부분 검증하였다.

데이터의 폭발적 증가로 고민하고 있는 기업들의 고충을 해결할 수 있는 방안 제시와 함께, Archiving System 구축을 추진 중인 기업 들이 참고할 수 있는 기준자료 제공이 되었다.

그러나 본 연구의 한계점 또한 간과할 수 없다.

본 연구에서 제시한 구축방법론인 S—MAP 은 특정 기업들의 프로젝트 경험을 토대로 만들어졌기에 모든 기업의 다양한 조건들을 만족시킬 수는 없기 때문이다.

향후 단계별 구축을 통하여 발생할 수 있는 위험 요소들을 최소화하고, 효율적인 방안을 확대 제시해 가면서 그 완성도를 더욱 높일 수 있는 연구가 요구된다.

참고문헌

[1] John F. Gantz, “The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010", IDC, March 2007

[2] 채수용, “디지털 정보에 묻힌 지구”, 디지털데일리, 2007

[3] 정현용, “market Focus – 정보 수명주기 관리(ILM)-ILM 등장 배경 및 전망”, NETWORK TIMES, 2005

[4] Solix White Paper, Facilitating Compliance and Intelligent Information Management with Effective Database Management, March 2007