

# RainStor for Teradata EDW

Teradata의 활용도를 높여주는 RainStor 솔루션



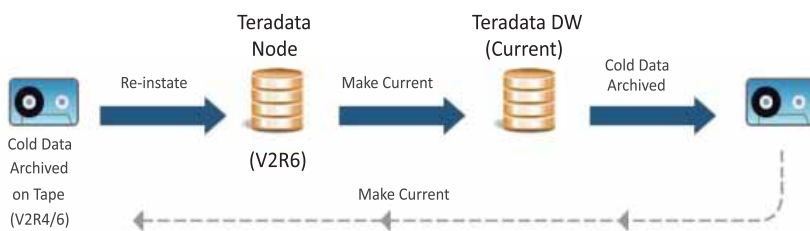
**Teradata**는 현재 가장 많이 활용되고 있는 Enterprise Data Warehouse (EDW)용 DBMS 중 하나입니다. 그러나 최근 들어 급속한 데이터의 증가, 어플리케이션 및 비즈니스의 다양성으로 지금까지의 사용하고 있는 DW 운영환경에 여러가지 도전에 직면하고 있습니다. 그 중에서도 가장 큰 문제점은 데이터 증가에 따른 투자비용의 증가입니다.

과거 데이터를 지속적으로 요구하는 비즈니스 유저 또는 향후 잠재적으로 기업의 매출 및 이익에 영향을 미칠 수 있는 시계열성 트랜잭션 데이터의 특성 때문에 대부분 기업들의 데이터 관리자들은 과거의 데이터를 삭제하기가 현실적으로 매우 힘든 상황입니다. 더욱이 업무 특성상 대규모 데이터가 끊임없이 생성되고 관리되어야 하는 은행을 비롯한 금융관련 그리고 커뮤니케이션 서비스 제공 기업들은 대규모의 비용을 지불하고 있는 실정입니다.

따라서 기업들은 데이터 관련 비용을 줄이기 위하여 과거 데이터를 일반 플랫폼 파일로 관리하거나 또는 오프로드 기반의 테이프 장치 기반의 아카이브를 관리하는 등, 여러가지 방법을 도입하여 EDW에서 관리하는 데이터의 총량을 줄이는 노력을 하고 있습니다.

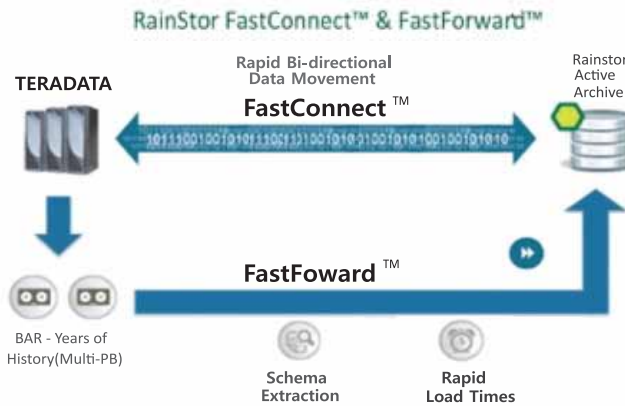
이러한 상황은 Teradata를 사용하는 고객들도 예외는 아니고 주로 Teradata Backup, Archive, Restore(BAR) 를 선호하여 사용하고 있습니다. 그러나 이러한 테이프 기반의 아카이브 시스템은 필요한 시점에 필요로 하는 데이터를 온라인으로 즉시 사용할 수 없다는 것입니다. 이러한 문제를 해결하기 위하여 RainStor와 Teradata 기술진은 RainStor **FastConnect**와 RainStor **FastForward** 솔루션을 공동으로 개발하게 된 것입니다.

## Teradata의 테이프 기반의 아카이브 작업 흐름도(현재)



이 방법의 문제점은 해당 데이터를 온라인 형태로 만들기 위하여 시간과 노력이 많이 요구된다는 것입니다. 물론 테이프 기반의 아카이브는 저비용으로 초기에는 매우 매력적인 솔루션이지만 최근의 컴플라이언스 규정 준수 또는 과거 수년 전 데이터를 요구하는 케이스가 빈번히 발생하고 있어 이제는 리스토어가 거의 없다고 판단된 데이터에 만 제한적으로 사용해야 할 솔루션이 되고 있습니다.

## RainStor 솔루션을 이용한 액티브 아카이브 작업 흐름도(제안)



### FastConnect

- Rapid Data Movement
- Ongoing Query Access
- Improved Analytics Response Times
- Zero Manual Effort to Move Data

### FastForward

- Limitless Historical Archive
- Efficient Tape Data Restoration & Schema Extraction
- Reduced Resource Costs associated with Tape Restoration
- Extended Version Support - V2R4, 5, 6 & TD12, 13
- Integrated with RainStor FastConnect

## Teradata와 RainStor – EDW의 온라인 데이터와 콜드 데이터를 위한 상호 보완적 솔루션

RainStor **FastConnect**는 지속적으로 관리가 요구되는 기업의 가치 있는 대용량 데이터를 온라인 또는 접속 가능한 형태로 관리가 가능하도록 설계되어 제작되었고, 특히 Teradata Parallel Transport(TPT)를 사용하여 제작되어 Teradata와 RainStor 간에 고성능 양방향 데이터 이동을 가능하게 지원합니다.

### RainStor FastConnect™

RainStor FastConnect는 상대적으로 오래된 과거 데이터 또는 거의 사용이 되지 않는 데이터를 Teradata EDW에서 RainStor Database로 이동되어 압축된 상태로 관리되다가 필요할 경우 곧 바로 Teradata로 재로드되어 즉시 사용이 가능하도록 만들어진 솔루션입니다. 물론 필요할 경우 RainStor에서 쿼리를 수행하여 사용도 가능합니다.

FastConnect는 Teradata EDW와 RainStor Database간 양 방향 고성능 데이터 이동 솔루션입니다. FastConnect 솔루션은 Teradata 기술진과 공동으로 개발된 솔루션으로 복잡한 스크립트나 3rd-Party ETL 솔루션이 필요하지 않을 뿐만 아니라 Teradata 고객들이 만약 Hadoop HDFS의 데이터에 대한 접근을 원한다면 과거 시계열 데이터에 대한 쿼리를 어려운 Hive, Pig 그리고 MapReduce가 아닌 표준 SQL로 "SQL 온하둠" 조회를 가능하게 지원합니다.

### RainStor FastForward™

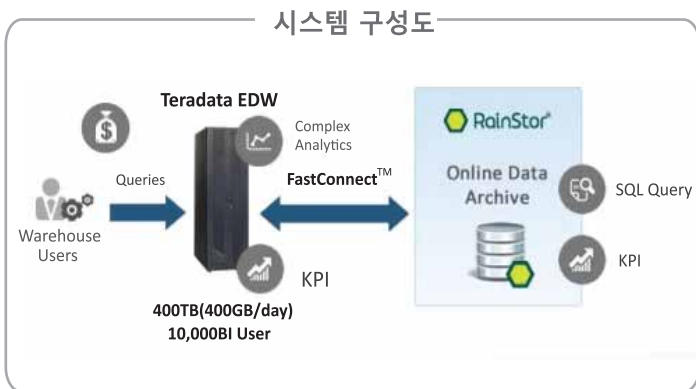
앞서 기술했듯이 많은 Teradata 고객들은 현재 BAR 기반으로 데이터 아카이브를 사용하고 있습니다. 그러나 분석 또는 조회를 위해 BAR 아카이브 데이터를 다시 Teradata로 리스토어가 필요할 때 많은 어려움에 봉착하게 됩니다. 특히 Teradata는 이단계 이전버전 까지만 지원이 가능하기 때문에 그 이전 버전으로 아카이브 된 BAR 데이터는 사실상 리스토어가 불가능합니다.

RainStor FastForward는 Teradata V2R4, V2R5, V2R6, TD12 그리고 V13 까지 거의 모든 버전에 대한 ARCMAN 파일 포맷을 사용하여 RainStor로 이관하여 즉시 SQL 쿼리 분석을 가능하게 지원합니다.

### 제품특징

- 신속한 테이프 데이터 리스토어 지원
- Linux와 IBM ARCMAN 데이터-셋 모두 지원
- ARCMAN 리스트 실시간 스키마 추출
- 비용감소: Teradata 에 대한 투자비용 줄임
- RainStor FastConnect와 연동: 즉시 Teradata로 전송
- 데이터 유지: 컴플라이언스 및 SQL 쿼리 목적
- 스키마 예롤루션 지원

## 적용사례(글로벌 IT 제조사)



### 문제점

- 예상을 뛰어넘는 급속한 데이터 증가 (EDW 시스템 효율적 사용 불가능)
- H/W 및 S/W 업그레이드 불가: 사용년한 남음
- 현재의 EDW의 부하를 감소할 수 있는 방안 마련 필요 (스토리지, IO, CPU 사용률을 줄일 수 있는 방법)

## RainStor for Teradata 적용 장점

Teradata 고객들에 대한 데이프 의존도 제거

### 아카이브 데이터에 대한 쿼리 및 분석 가능

RainStor 장점은 과거 오래된 데이터 또는 변경을 허용하지 않는 데이터를 효율적으로 압축관리하면서 SQL 쿼리 및 분석이 가능하게 지원한다는 것입니다. 기본적인 쿼리는 RainStor에서 수행하고 만약 정교하고 복잡한 분석이 필요한 경우에는 Teradata로 간단하게 이동하여 분석을 진행하고 분석이 완료되면 다시 RainStor로 전송하여 관리하면 되는 것입니다.

### 데이터 보안

모든 데이터는 데이터 라이프사이클(DLM) 프레임 내에서 안전하고 보안적으로 관리되어야 하고, 보존기한이 경과한 데이터에 대한 폐기 정책은 자동으로 진행될 수 있어야 합니다. 각 기업들이 엄격한 컴플라이언스 규정 준수가 필요한 카드번호 그리고 주민번호와 같은 민감한 개인정보는 데이터 인크립션 및 마스킹 기능에 의하여 안전하게 관리가 가능합니다.

### 이전 스키마 버전과 호환성 제공

RainStor는 스키마 버전 변경과 같은 어려운 문제를 해결할 수 있는 방법을 제공합니다. 소스 데이터의 스키마는 계속 변경이 이루어지고 테이블과 같은 형태에서 이를 지원하는 것은 매우 어려운 일입니다. FastConnect는 자동으로 소스 Database에서 추출하여 목표 Database에 자동적으로 복제해 주고 따라서 사용자들은 Teradata 또는 RainStor 두 시스템의 스키마가 동일하기 때문에 양쪽 어느 곳에서 데이터를 추출하여 사용할 수가 있습니다.

### 고성능 로드 기능

FastConnect는 고성능 아키텍처임으로 데이터 로드 시간을 줄일 수 있습니다. Teradata와 RainStor의 다이렉트 컨넥션으로 수십억건의 데이터를 하루 안에 추출 및 로딩이 가능하게 됩니다.

## 솔루션

- Capacity Buyback 노력: ILM(Information Lifecycle Management) 구축
- RainStor for Teradata 솔루션 선택
  - "Big Data" 스토리지 및 분석에서 경쟁력을 갖춘 플랫폼
  - Teradata 파트너
  - DBA 팀 PoC 실시후 선택
- RainStor on Hadoop 5.2(CDH4.1.3) 5 node 사용중

## 구축목표

- ILM: Active EDW(Teradata)에서 과거 데이터 또는 사용빈도가 적은 데이터를 저가형 스토리지를 사용하여 관리
  - 생성되고 관리/보관된 이후 보존기한이 지난 데이터에 대한 폐기정책 필요
  - SQL 쿼리 및 컴플라이언스 규정 준수를 위하여 데이터 보존 필요
  - EDW의 Table 사이즈가 감소되면서 EDW 성능 향상 기대
- 투자비용 감소로 타 시스템 자원 업그레이드 기대

## 구축효과

- 현재 약 60TB 데이터 이관완료(Teradata -> RainStor)
- 약 40x 데이터 압축저장: RainStor
- TB당 80% 정도 스토리지 비용감소
- 스키마 변경에도 불구하고 지속적인 SQL 쿼리 지원
- 자동적인 데이터 보존 및 폐기 정책 지원

## RainStor for Sybase IQ, Oracle

Teradata 이외에 SAP Sybase IQ, Oracle 그리고 MS SQL Server와 같은 전통적인 DW용 DBMS와 연동이 가능하여 Pre-Processing 또는 액티브 아카이브 등 DW 사전, 사후 시스템 구축이 가능합니다. 이를 위하여 RainStor는 각 DBMS에 대한 스키마 DDL 변환 기능을 제공합니다. 또 R2ware의 SQL Canvas ILM/Trans와의 통합으로 ILM 구축 또는 데이터 이관/변환 작업을 편리하게 진행하실 수 있습니다.

### 데이터 저장 스토리지 감소

특허기반의 밸류 및 패턴 디플리케이션 알고리즘을 이용한 강력한 압축(30~40배)으로 약 90% 이상의 스토리지 필요량 감소

### 고성능 데이터 쿼리

SQL, 하둡 Map Reduce, Pig, Hive 그리고 표준 BI 툴 지원으로 SQL을 재작성하지 않고 기존의 SQL을 이용하여 유연하고 고성능 데이터 검색 및 분석이 가능함

### 다단계 데이터 보안

강력한 내장 보안 기능 및 오딧 정책으로 해킹, 만료 그리고 권한이 없는 어카운트 사용 등 잠재적인 보안 위협으로부터 데이터를 안전하게 보호  
Kerberos Authentication, Authorization and LDAP, PAM 지원

### 강화된 컴플라이언스 기능

컴플라이언스 규정(예, 개인정보보호법, 바젤, SarbanesOxley, Dodd-Frank, HIPPA 등) 준수를 위한 다양한 기능 제공

### 다양한 스토리지 플랫폼 지원

저가형 NAS부터 SAN, CAS 하드웨어, 클라우드 서비스 그리고 다양한 하둡 HDFS 파일 시스템까지 사용 가능

### 아파치 하둡과 네이티브 연동 지원

오픈소스 아파치 하둡의 분산파일시스템인 HDFS와 MapReduce 기반의 Pig, Hive를 네이티브 연동 지원  
RainStor의 강력한 압축기능과 표준 SQL 사용으로 고성능 "SQL온하둡" 실현이 가능하여 기존대비 약 10배에서 100배까지의 성능향상 가능