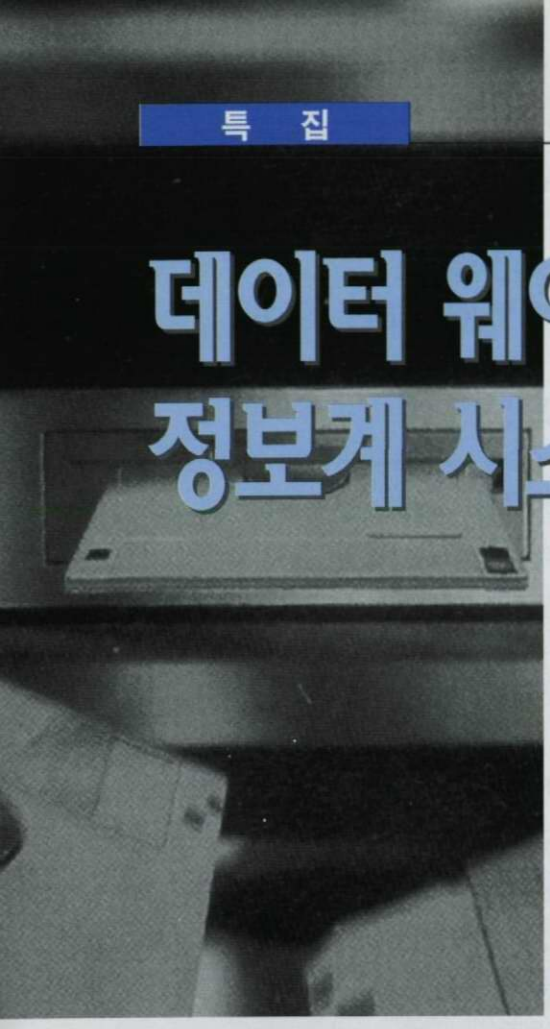

데이터 웨어하우스는 정보계 시스템의 '핵'

한국오라클 · 한국인포믹스 · 한국사이베이스 '솔루션 접전'

최근들어 전산실에서 OLTP시스템이 필수로 자리잡고 있는데, 이것과 동일선상에서 데이터 웨어하우스가 접목될 것이라고 보고 이 시장을 공략하는 업체들이 늘어나고 있다. 한국오라클과 한국인포믹스 등 RDBMS 선두업체는 물론, 한국디지탈과 마이크로소프트 등 하드웨어와 소프트웨어 선두업체들도 서둘러 데이터 웨어하우스 시장에 진입, 올해 시장을 노리고 있다. 뿐만 아니라 연내 SI업체는 모두 데이터 웨어하우스 시장에 들어올 것이라고 관계자들은 예측하고 있다.

OLTP가 어느 회사든지 구축돼 있다 해도, 기구축된 데이터에서 정보를 한눈에 볼 수 있도록 해주는 시스템이 중요하게 부각될 것이라는 전망은 쉽게 해 볼 수 있다. 한국오라클 기술본부의 장동인 이사는 "데이터 웨어하우스는 현실 IS에서 절대 필요한 기술이며, 결국 어떤 기업이든 구축하게 돼 있다"라고 단언하고 있다.

근거는 전산실의 30%가 장표처리 업무이고, 데이터 웨어하우스는 장표를 처리해주면서 다차원 분석까지 해주는 것이므로. 또, 한 회사가 운영계 시스템없이 영업을 할 수 없듯이 정보계 시스템없이는 효율적으로 영업을 할 수 없기 때문이라는 것. 따라서 데이터 웨어하우스는 정보계 시스템의 '핵' 이라고 표현하고 있다.

데이터 웨어하우스를 어떻게 구축할 것인가는 방법론과 예산에 관련되는 문제다. 전사적인 규모의 데이터 웨어하우스를 한꺼번에 투자해서 장기간으로 이끌어가는 것은 대부분의 기업에서 부담스러울 것이다. 전문가들은 "데이터 웨어하우스는 많은 주제를 여러 개로 나눈 후 여러 단계를 거쳐서 하나씩 하나씩 구축하는 것이 성공의 지름길"이라고 제언하고 있다.

장동인 이사는 "올해가 데이터 웨어하우스 구축의 해라고 해도 과언이 아닐 정도로 많은 기업이 데이터 웨어하우스를 구축할 것"이라고 자신한다. 현재 거의 모든 데이터베이스 공급업체, 하드웨어 공급업체, OLAP 도구 공급업체, 시스템 통합업체 등 많은 업체들이 각각의 솔루션을 가지고 데이터 웨어하우징 시장에 뛰어들고 있다.

신규 프로젝트 '데이터 웨어하우스, 반드시 들어간다'

데이터 웨어하우스는 데이터 추출, 데이터 웨어하우스 모델링, OLAP으로 이루어져 있으며, 이를 한 시스템으로 묶는 컨설팅으로 구분된다. 즉, 데이터 웨어하우스는 제품이 아니라 시스템통합이며 컨설팅의 핵심이다. DSS와 EIS가 운영계 시스템에서 데이터의 요약이나 일부 데이터의 요약인 데서 오는 오류의 가능성을 배제할 수 없지만, 데이터 웨어하우스는 한꺼번에 데이터를 다 집어넣자는 것이며, 하드웨어 가격이 하락하는 추세일수록 데이터 웨어하우스의 가능성이 커지고 있는 것이다.

데이터 웨어하우스 시장은 지난해 초만 해도 그리 낙관적이지 못했다. 지금도 '들이는 공에 비해서 크게 떡고물이 생기지 않는' 힘든 사업부문으로 인식되고 있다.

한편으로 수요가능성이 있는 기업에선 지난 1년간 파일럿 구축을 통해서 데이터 웨어하우스가 무엇인지 알

려고 하는 시기는 지났다고 봐야 한다. 현재 상태는 데이터 웨어하우스에 대해서 구축해야 하는 필요성들은 모두 느끼고 있는 상태이나 '시기가 문제' 라는 입장을 견지하고 있다.

장동인 이사는 올해 금융, 통신, 제조, 유통분야의 최소한 30개의 사이트에서 데이터 웨어하우스 구축이 시작될 것이라고 예견했다. 명확히 매출액과 연결되는 사이트는 10개 정도로 꼽았다. 장 이사의 추론 근거는 '새로운 시스템을 구축하려는 프로젝트에는 반드시 데이터 웨어하우스가 들어간다' 는 것이다.

한국오라클, '원스톱 솔루션' 으로 승부

한국오라클의 데이터 웨어하우스 시장 전략은 DBMS에서부터 기술자문서비스까지 시스템 구축에 필요한 모든 것을 한꺼번에 제공한다는 한마디로, '원 스톱 솔루션' 이다.

한국오라클은 특히 데이터 웨어하우스 시스템 구축의 핵심이라고 할 수 있는 다차원 모델링 분야를 적극 공략할 예정이다. 이를 위해 오라클은 '오라클 7.x' 기반 위에서 '스타 스키마', '스노플레이크 스키마' 등 일반 솔루션과 함께 '서머리 테이블', '테이블 파티션', 'SQL 튜닝' 등 전문솔루션 및 '오라클 게이트웨이', 'K*로더' '디스트리뷰티드 옵션' 등 데이터 추출 솔루션을 함께 제공하고 있다.

또한, 최근 윈도우NT용 워크그룹 서버를 기반으로 데이터 추출 도구 '오픈게이트웨이', 고속 로딩 작업처리 도구 '병렬 로더', 정보 분석 도구 '엑스프레스', '디스커버러' 등 데이터마트 솔루션을 발표, 데이터마트 구축에 필요한 제품과 서비스를 일괄 제공해 저렴한 비용으로도 빠른 시일내에 투자효과를 확인할 수 있도록 해준다.

한국오라클은 지난 4월부터 2차 프로젝트에 들어간 신세기통신외에 삼성전자, 삼성카드 등 6개 사이트에 데이터 웨어하우스를 구축중에 있으며, 20여명의 데이터 웨어하우스 전담 인력을 두고 국내 데이터 웨어하우스 시장을 적극 공략하고 있다.

신세기통신 1차 프로젝트 구축 완료

한국오라클이 구축중인 데이터 웨어하우스 프로젝트로는 신세기통신과 삼성전자 마케팅부문 등이 있다. 신세기통신은 지난해 11월부터 한국오라클과 포스데이타가 함께 프로젝트를 진행해 지난 3월말 1차 구축을 완료하고 4월부터 2차 프로젝트에 들어갔다.

신세기통신 데이터 웨어하우스 프로젝트는 ▲기존 신세기통신의 고객서비스 및 요금시스템, MIS, 네트워크 관리시스템 등 다양한 운영계 시스템에 축적되는 각종 데이터를 '얼마나 빠르고 유연하게 통합하여', ▲이를 바탕으로 다양한 현업 관련 보고서 및 분석을 위한 전산 환경이 '얼마나 유연하고 신속하게 제공될 수 있는가' 를 검증하는데 목표를 두고 프로젝트가 추진됐다.

데이터의 소스가 되는 운영계 시스템은 오라클 RDB를 사용하는 VAX 클러스터(오픈MVS)로부터 오라클 RDBMS를 사용하는 HP 유닉스 및 디지탈 알파 유닉스까지 다양한 종류의 기종과 OS 및 DBMS를 보유하고 있다. 이러한 다양한 종류의 데이터를 가지고 디지탈 알파 2100과 오라클 RDBMS 7.3을 이용하여 데이터 웨어하우스 서버를 구축하였으며, 대상 업무는 가입자 분석 및 통화 분석으로 한정하여 진행하였고, 마지막으로 네트워크 관리 부문이 부분적으로 추가됐다.

신세기통신은 신규 제 2통신사업자로서 신규가입자에 대한 관심이 매우 높아 가입자 분석 중 신규 가입자 및 해지자에 대한 추이 분석은 일일 분석

전문가들은 "데이터 웨어하우스는 많은 주제를 여러 개로 나눈 후 여러 단계를 거쳐서 하나씩 하나씩 구축하는 것이 성공의 지름길"이라고 제언하고 있다.

〈표 1〉 데이터웨어하우스 시장 규모 추산

연도	총 DW 사이트수 (개社)	총 DW 규모 (원)
1997	10	50억
1998	30	150억
1999	50	250억
2000	100	500억

자료 : 한국오라클

(총 데이터 웨어하우스 규모=하드웨어+소프트웨어+컨설팅+유지보수)

이 가능하도록 매일 데이터 로드를 실시하고 있다. 특히, 데이터 추출 부문에서 운영계 시스템의 부하를 극소화하기 위하여 복잡한 연산 작업이나 데이터 정화 작업은 데이터 웨어하우스 서버에서 실시하고 있다.

이밖에도 가입자의 특성에 대한 분석을 위하여 여러가지 특성을 나타내는 속성들이 추출되었으며, 이들 특성의 시계열적 변경 추이 분석을 위하여 월간 배치작업으로 데이터를 추출하여 데이터 웨어하우스에 로드하고 있다. 또한, 요금 작업이 끝난 후 정산된 데이터를 운영계로부터 추출하여 각종 가입자별 통화 분석을 위한 수단으로 쓰여질 예정이며, 시간대별, 통화 유형별 통화 분석을 위하여 작업이 실시되고 있다.

이 프로젝트에서의 데이터 웨어하우스 모델링은 기본적으로는 스타 스키마로 구성하였고, 데이터의 양을 고려하여 일부분은 정규화된 ER 모델이 가미되었으며, 각 데이터 추출 및 정제, 로드, 스케줄링, 모니터링, 에러로깅 등의 기능이 기본적으로 구현됐다. 이와 같은 스타 스키마 모델에 저장된 데이터는 OLAP 도구를 이용하여 유연한 보고서 및 분석을 위한 전산 환경을 제공할 수 있도록 하였다.

이 데이터 웨어하우스 시스템 데모를 지켜 본 신세기통신 현업 담당자들은 "기존에는 보고서를 개발, 작성하려면 전산실을 거쳐야 하는 등 최소 3일 정도 걸렸는데, 이번 데이터 웨어하우스 시스템 구축을 통해서는 유연한 보고서를 실시간으로 작성할 수 있게 돼 만족스럽게 생각하고 있다"라고 소감을 피력했다.

특히 데이터 웨어하우스 구축 결과, 새로운 주제 영역의 신속한 수용 및 보고서의 작성으로 기존 운영계에서 지원하던 보고서 체계를 이양시킬 수 있다는 가능성을 발견하였고, 이를 바탕으로 4월부터는 전사적으로 데이터 웨어하우스를 확장하기로 했다. 이에따라 기존 데이터웨어하우스 TFT 요원들은 현재 새로운 주제영역의 설정 및 서버 하드웨어의 선정 작업을 진행하고 있다.

한국인포믹스, '데이터블레이드'와 연계 전략

인포믹스는 미국 MCI에 1억3천만명에 달하는 마케팅 지원 및 고객정보 시스템을 인포믹스 제품군으로 구축, 지난해 8월 현재 4.3테라바이트에 달하는 대용량의 데이터 웨어하우스를 구축완료한 사례로 데이터 웨어하우스 솔루션 및 기술력을 인정받고 있다.

한국인포믹스는 데이터 웨어하우스 시장 공략을 올해의 가장 중요한 사업 분야로 여기고 있으며, 지난해 2월, 삼성SDS와 파일럿 프로젝트를 공동수행한 이후 쌍방울과 현대자동차 프로젝트가 진행중이다. 인포믹스에서는 DBMS '온라인다이내믹서버'와 분석처리 도구 '메타큐브', 데이터로딩 및 필터링 도구인 '하이퍼포머로더', 웹브라우저 '넷스케이프'를 한데 묶은 데이터마트 솔루션을 강조하고 있다.

인포믹스 데이터 웨어하우스 솔루션은 'DB 서버-OLAP 서버-클라이언트/서버 도구' 등 3티어 아키텍처로 구성돼 있다.

인포믹스는 코딩 및 애플리케이션 개발을 최소화하기 위해 DBMS에 직접 접근하여 다차원 분석기능을 제공하는 ROLAP(Relational OLAP) 솔루션과 클라이언트/서버 도구의 인터페이스로 개발시간을 단축시키는 외에 EIS 애플리케이션을 지원하기 위하여 API를 기본적으로 제공한다.

인포믹스에서는 10GB를 초과하는 대용량 데이터 웨어하우스의 경우 ROLAP이 필수적인 선택이라고 권한다. 이는 "확장성과 더불어 데이터 처리를 서버에 두고 다차원분석을 OLAP 도구에서 실시함으로써 뛰어난 성능을 보장한다"는 것. 반면, MOLAP(Multi-dimensional OLAP)의 경우 별도의 MDB를 구축하여야 하는 관계로 10GB를 초과하는 대용량의 데이터 웨어하우스의 경우 성능 및 관리의 제약을 받을 수밖에 없다고 한다.

사용자수가 향후 급격하게 늘어날 경우 웹과의 연계가 불가피하다고 보고, 인포믹스는 Web/OLAP 기능을 제공하여 데이터 웨어하우스와 웹과의 연계를 보장하고 있다.

인포믹스 김용성 이사는 "데이터 웨어하우징은 기존의 OLTP와는 다른 데이터 모델링 방법론 즉, 스타 스키마나 스노우플레이크 스키마 등이 각종 컨설팅 기능과 더불어 지원될 때만이 프로젝트의 성공을 보장받을 수 있다"면서 "인포믹스는 MCI 및 삼성 프로젝트에서의 구축경험을 기반으로 다양한 모델링 기법 및 노하우를 보유하고 있다"고 말한다.

인포믹스에서는 기본적으로 별도의 애플리케이션 개발이 필요없는 EUC(Enduser Computing) 환경을 제안하고 있으며, OLAP 서버로 충족이 될 수 없는 사용자에 한하여 EIS 애플리케이션 개발을 권장한다. 사용

자별로 환경을 보면 표 3과 같다.

인포믹스는 사용자 요구분석단계에서부터 모델링, 데이터 추출에 걸쳐 최종 EUC/EIS 구축단계에 이르기까지 데이터 웨어하우스 구축에 관한 총체적인 솔루션을 제시하고 있다. 또, '유니버설 서버'의 출시로 기존의 솔루션을 데이터블레이드의 형태로 데이터베이스 엔진내에서 구현하여 데이터 웨어하우스와 연계시키는 전략을 제시하고 있다.

인포믹스의 데이터 웨어하우스 전략에서 특이한 것은 일러스트라에 데이터블레이드 방식으로 데이터 웨어하우스의 기능을 한층 강화한 데이터 마이닝(데이터 웨어하우스 Minnig) 솔루션을 제공해 사용자 선택폭을 넓힌다는 것이다.

쌍방울, 데이터 웨어하우스 구축 완료

한국인포믹스는 쌍방울에 데이터 웨어하우스 시스템을 구축, 올초부터 본격적인 가동에 들어갔다.

쌍방울 데이터 웨어하우스 시스템은 각 업무 영역에서 이뤄지는 실적과 계획을 대조해 분석하기 위한 도구로 주요 소프트웨어로는 '인포믹스 7.13' DBMS와 '메타큐브 2.1' 온라인분석처리도구(OLAP)를 사용했다.

쌍방울 데이터 웨어하우스 시스템은 제품 생산, 판매, 재고, 매출 현황, 상품 판매 정보, 입금 정보, 잔액 정보 등의 6개 업무 영역을 제품명, 제품번호, 제품 유형, 거래처, 브랜드 등 20개 변수에 따라 분석, 실시간으로 제공해준다. 쌍방울은 영업 및 판매 관리에 국한된 이 시스템을 생산관리까지 확대해 전략정보시스템으로 활용할 예정이다.

쌍방울社의 현업요구는 빈번한 배치작업으로 작성되는 판매 및 생산결과를 분석할 수 있도록 해달라는 것이 가장 많았다. 다양한 요구를 해결하는 데 기존의 IBM 호스트 기반 시스템으로는 힘든 점이 있었다. 현업의 불평을 개선시키고 전산실의 백로그를 줄이기 위한 타개책으로 찾은 것이 인포믹스의 데이터 웨어하우스 솔루션이다.

쌍방울은 올초부터 기존의 IBM 시스템을 클라이언트/서버 시스템으로 내리기 시작했는데, 유닉스 기종에서 인포믹스 DB를 사용했으므로 자연히 데이터 웨어하우스도 인포믹스로 결정한 것이다. 정보시스템팀의 김인성 과장은 "데이터 웨어하우스 안내 책자를 봤을 때 이를 도입하면 배치에 의해 발생되는 백로그를 줄일 수 있고, 현업의 요구에도 대처할 수 있을 것으로 판단했다"고 전한다.

쌍방울에서는 지난해 4월부터 데이터 웨어하우스를 검토하기 시작해 6월에 결정, 7월부터 파일럿 프로젝트에 들어갔으며, 9월부터 실제 구축해서 11월말에 완료시켰다. 총 2억원 규모의 작업이었고 20유저짜리로 현재 10GB(2년치 데이터)를 사용하고 있으나 20GB까지 잡아놓은 상태이며 곧 40GB로 확장 도입할 계획이다.

우선 영업관리와 판매관리 부분에서만 작업하고 있으며, 판매 거래처의 통계만 데이터 웨어하우스로 한정 구축해 총금액은 적은 편이다. 하부구조는 이더네트, 호스트 DB는 CA-데이터콤 DB, 워크그룹 서버는 인포믹스, 클라이언트 개발도구는 델파이를 사용한 환경이다.

데이터 웨어하우스의 구축효과는 영업 및 판매 · 관리부서의 사용자들이 직접 단말기를 통해 옵션넣고 필터링 해서 자료를 뽑아볼 수 있으므로 전산실에 자료를 요청하는 횟수가 현격하게 줄어들었고, 엑셀이나 로터스 등과 연계되므로 자료의 활용폭이 넓

〈표 2〉 인포믹스의 데이터 웨어하우스 솔루션

솔루션/제품명	역할	인포믹스 협력관계	특기사항
데이터 모델링 -스타 스키마s -스노우플레이크 스키마	데이터 웨어하우스 등 분석업무를 위해 지원되는 모델링 기법	인포믹스 자체 솔루션 보유	스노우플레이크 스키마는 인포믹스만이 지원
DBMS -XPS 8.1 -DSA 7.2 -워크그룹 서버 -일러스트라 -유니버설 서버	데이터베이스 엔진 -MPP 플랫폼 지원 -SMP 플랫폼 지원 -소규모 -멀티미디어 전용	인포믹스 자체 제품	하드웨어 유형 및 데이터의 유형별로 모든 DBMS 제공
OLAP 도구	다차원 분석을 위한 OLAP 분석도구	인포믹스 자체 제품	코딩이 필요없는 관계형 OLAP
뉴에라 -파워빌더/엑셀/비주얼베이직 -SAS 등	-일반 최종사용자용 도구	인포믹스 자체 제품, 협력사 제품 (이태릭체 표시)	클라이언트/서버 도구
데이터 추출 도구 -플래티늄 -ETI, 프리즘 등	데이터 추출 및 변환 도구	협력사 제품	

어진다는 것.

또, 기존에는 전산실 직원 60명중 40명 정도가 백로그 처리하느라 새로운 업무개발을 못하던 상황에서 벗어날 수 있는 전기를 마련하게 됐고, 리포트 출력업무가 감소하므로 OLTP 시스템(IBM 메인프레임)의 부하가 적어지게 됐다는 것 등이 가시적이다.

데이터 웨어하우스 엔진은 메타큐브이고, 사용자들은 엑셀을 이용하고 있으며 호스트와 유닉스간 통신은 '브릭스톤' 게이트웨를 사용하고 있다. 상품별, 거래처 · 상품별, 지역별 · 거래처별 매출을 팩트 테이블화해서 통계를 내보고 있는데, 좀더 세세한 자료를 보기 위해 몇가지 더 만들었고 현재의 2년치 데이터에서 향후 5년치 데이터까지 가져갈 계획이다.

전산팀에서는 판매 데이터 웨어하우스와 생산 데이터 웨어하우스를 모아서 의사결정시스템(DSS)을 만들 계획을 갖고 있다고 한다. 김인성 과장은 "일단 손은 댔는데 완벽한 활성화는 아니다"라면서, 기능을 추가시켜 가고 메타큐브 3.0으로 보완을 해서 DSS로 갈 것이라고 밝혔다.

한국사이베이스, '웨어하우스나우' 솔루션

한국사이베이스는 올해 데이터 웨어하우스분야에서 시스템 구축을 위한 방법론을 중심으로 컨설팅 영업에 적극 나선다는 계획으로 제품에서 기술자문까지 일괄 제공하는 접근 방법을 보이고 있다.

최근에는 의사결정지원(DSS) 데이터베이스 서버 '사이베이스 IQ 11.1'과 데이터 추출 도구를 묶은 '퀵스타트'를 발표했는데, 퀵스타트 데이터 마트는 IBM의 RS/6000을 기반으로 단기간내에 저비용으로 데이터 마트 시스템을 구축할 수 있는 솔루션이다.

〈표 3〉 EUC와 EIS의 사용자별 환경

Knowledge Worker/실무자	EUC	메타큐브 익스플로러
임원층 등	EIS	파워빌더/비주얼 베이직 등과 메타큐브 익스플로러의 연계
마케팅 지원시스템 관리자	EUC/EIS/ 관리기능	메타큐브 에이전트

또한, 사이베이스는 최근 고성능, 분산 시스템을 위한 통합(End-to-End) 데이터 웨어하우스 솔루션인 '웨어하우스나우'를 지원하기 위해 자사의 주요 제품군인 리플리케이션 서버 제품군을 강화시켰다고 발표했다.

이번에 리플리케이션 서버 제품군에 새롭게 추가된 4개의 제품은 MVS용 디스트리뷰션 에이전트, 디스트리뷰션 디렉터, IMS용과 VSAM용 리플리케이션 에이전트이다. 새로이 추가된 이 제품들은 고속으로 데이터를 로드할 수 있으며, 데이터 이동을 위해 구축된 기존의 리플리케이션 서버와 에이전트를 관리할 수 있다.

이러한 새로운 리플리케이션 기능은 데이터를 효과적으로 데이터 마트나 데이터 웨어하우스에 이동시켜야 하는 중요한 요구를 만족시키기 위해 제안된 것이다. 리플리케이션 서버 제품군은 하나의 데이터 소스로부터 트랜잭션 이동 방식과 고속의 대용량 데이터 벌크 이동 방식을 모두 지원한다.

사이베이스의 미들웨어 · 데이터 웨어하우스 부서 최고 관리자이자 부사장인 게리 스틸은 "데이터 마트와 데이터 웨어하우스에 데이터 이동을 관리하는 작업은 엔터프라이즈 인프라스트럭처를 통해 가장 중요한 부분이다"라고 말한다.

Distribution Agent for MVS는 사용자가 단 한번의 데이터 이동으로 동시에 32개의 데이터베이스를 로드할 수 있는 고성능의 데이터 로드 능력을 제공한다. 이것은 MVS 파일에 접근함으로써 수행된다. Distribution Agent for MVS는 파일 전송의 병렬 처리, 체크포인트/리스타트, 문자 변환과 같은 작업을 일괄처리로 로딩할 수 있는 주요한 기능을 제공한다. 더욱이 단일한 인터페이스로 MVS 데이터를 Sybase IQ 11.2, Sybase SQL 11, Flat 파일에 직접 로딩할 수 있다.

Distribution Director는 데이터 웨어하우스의 관리자를 위해 사용하기 간편한 그래픽 사용자 인터페이스 GUI 관리자이다. 더 자세히 말해서 LAN상에서의 타스크 정의와 수행, MVS 작업 지원, 데이터 웨어하우스에 직접 연결을 가능하게 함으로써 Distribution Director는 데이터웨어하우스 관리 능력을 강화했다. Distribution Director에는 Sybase IQ 11.2에 리플리케이션하거나 대용량 데이터 이동을 관리하는 일반적인 데이터 이동 작업과 리플리케이션 작업을 도와주는 마법사를 제공한다. Distribution Director는 Distribution Agent for MVS에 번들되며 Sybase IQ 11.2와 리플리케이션 서버 사용자는 추가비용없이 Distribution Director를 사용할 수 있다.

〈이영애 객원기자〉