

제1부 BI의 이해

AGENDA

- **BI 개요**
- **BI 히스토리**
- **데이터 웨어하우징과 BI**

BI 개요 - 정의

- BI(Business Intelligence)는 비즈니스 의사결정권자에게 유용하고 적절한 시기에 신뢰할 수 있는 데이터를 제공하는 것
- 최종 사용자가 다차원 정보에 직접 접근하여 대화식으로 정보를 분석하고 의사결정에 활용하는 과정(조재희/박성진 1996)
 - 다차원 정보
 - 가장 큰 BI의 특성, 다차원 정보분석
 - 직접 접근
 - 전산 부서와 같은 정보 매개자를 거치지 않고 자신이 원하는 정보에 직접 접근
 - 대화식 분석
 - BI환경에서는 사용자는 시스템과의 상호작용을 통하여 정보를 분석하고 원하는 결과를 원할 때까지
 - 의사결정에 활용
 - 기업의 방향을 설정하는 역할
- 과거에는 다양한 이름으로 사용했다
 - OLAP(On-Line Analytical Processing)
 - MIS(Management Information System)
 - DSS(Decision Support System)

BI 개요 - 정의

- BI(Business Intelligence)는 의사결정 지원에 사용되는 기술, 프로세스, 스킬, 응용프로그램을 의미하는 포괄적인 용어
- 요구 사항에서 추후의 의사결정을 돕기 위해 과거 데이터, 이미 사용된 데이터를 다루는 것
- 전략적인 계획에 적용되는 경영 문제의 내부 정보에 초점을 맞춤
- 정보는 일반적으로 하나 혹은 여러 방식으로 구조화된 관점과 현재의 프로세스에서 수집됨
- 비즈니스 목표 설정을 위한 요청의 초기 핵심성과지표(KPI)를 기반으로 함
- KPI 요소는 기본적으로 조직의 운영 시스템이라 이러한 운영 시스템으로부터 얻어지는 값들의 집계를 통해 얻어지며 이는 팩트로 알려진 수치로 세분화될 수 있다

BI 개요 - 정의

- 일반적으로 **BI** 환경에서의 분석을 개별분석보다 종합적인 분석이 수행됨
예) 기업은 개별 고객의 구매 이력보다는 종합적으로 얼마나 많은 고객이 있는지, 살고 있는지, 특정 지역에서 구매하는지 알고 싶어함
- 좋은 **BI**는 정확성, 시기적절성, 고부가가치, 실행성을 지님
- **정확성**은 신뢰할 수 있는 데이터를 의미
- **고부가가치**는 비즈니스 사용자에게 의미 있게 사용되는 것
- **실행성**은 정보가 비즈니스 의사결정에 사용됨을 의미

BI 개요 - BI의 가치

- BI는 비즈니스 의사 결정자에게 데이터 자체를 쿼리하는 기능을 제공
- 1970년대 말 컴퓨터 업계에서는 비즈니스 사용자가 데이터 액세스가 요구되어 질문해야 할 때, IT 보고서에 리포트를 요청하는 메모를 사내메일로 보내야 했다.
- 오늘날 BI는 비즈니스 사용자에게 데이터 직접 쿼리가 가능하게 한다. 이 가시적은 작은 성과가 몇 일이나 주 단위로 소요되던 의사결정 과정을 분 단위로 줄였다.
- BI는 사용자에게 IT와 관련된 단계에 직접 참여할 필요 없이 원하는 정보를 바로 쿼리하게 하는 기능을 제공

BI 개요 - BI 성공 요소

- 성공적인 **BI** 구축에 영향을 주는 요소가 **DW** 요소와 일치하지 않는다.
- **DW** 시스템은 일반적으로 사용자에게 **BI**로 통칭되지만 **DW**는 비즈니스적인 것보다 데이터 측면에서 더 중점을 둔다.
- **BI**는 큰 **DW** 시스템의 하위 구성요소로 볼 수 있음
- **DW**는 비즈니스에 직접적이거나 바로 사용되기보다는 특정 기준으로 데이터를 모으고, 모은 데이터를 목적에 맞게 정리하는데 노력함
- 비즈니스 중심의 방법론과 프로젝트 관리
 - 분명한 비전과 계획
 - 열정적인 경영지원과 후원
 - 데이터 관리와 품질 문제
 - 사용자의 요구와 솔루션의 적합 여부
 - **BI** 시스템의 성능 고려사항
 - 강력하고 확장 가능한 프레임워크

BI 히스토리

Year	Event	Comment
1962	Publication of <i>A Programming Language</i> by Ken Iverson	First multidimensional language; used Greek symbols for operators. Became available on IBM mainframes in the late 1960s and still used to a limited extent today. APL would not count as a modern BI tool, but many of its ideas live on in today's altogether less elitist products, and some applications (eg Cognos Planning Analyst and Cognos Consolidation, the former Lex 2000) still use APL internally.
1970	Express available on timesharing (followed by in-house versions later in the 1970s)	First multidimensional tool aimed at marketing applications; now owned by Oracle, and still one of the market leaders (after several rewrites and two changes of ownership). Although the code is much changed, the concepts and the data model are not. The modern version of this engine is now shipping as the MBI engine in Oracle9i Release 2 BI Option.
1982	Comshare System W available on timesharing (and in-house mainframes the following year)	First BI tool aimed at financial applications. No longer marketed, but still in limited use on IBM mainframes; its Windows descendent is marketed as the planning component of Comshare MPC. The later Essbase product used many of the same concepts, and like System W, suffers from database explosion.
1984	Metaphor launched	First RBI. Sales of this Mac cousin were disappointing, partly because of proprietary hardware and high prices (the start-up cost for an eight-workstation system, including 72Mb file server, database server and software was \$64,000). But, like Mac users, Metaphor users remained fiercely loyal.
1985	Pilot Command Center launched	First client/server EIS style BI; used a time-series approach running on VAX servers and standard PCs as clients.
1990	Cognos PowerPlay launched	This became both the first desktop and the first Windows BI and now leads the “desktop” sector. Though we still class this as a desktop BI on functional grounds, most customers now implement the much more scalable client/server and Web versions.
1991	IBM acquired Metaphor	The first of many BI products to change hands. Metaphor became part of the doomed Apple-IBM Taligent joint venture and was renamed IDS, but there are unlikely to be any remaining sites.
1992	Essbase launched	First well-marketed BI product, which went on to become the market leading BI server by 1997.
1993	Codd white paper coined the BI term	This white paper, commissioned by Arbor Software, brought multidimensional analysis to the attention of many more people than ever before. However, the Codd BI rules were soon forgotten (unlike his influential and respected relational rules).

BI 히스토리

1994	MicroStrategy DSS Agent launched	First RBI to do without a multidimensional engine, with almost all processing being performed by multi-pass SQL — an appropriate approach for very large databases, or those with very large dimensions, but suffers from a severe performance penalty. The modern <u>MicroStrategy 7i has a more conventional three-tier hybrid BI architecture</u>
1995	Holos 4.0 released	First hybrid BI, allowing a single application to access both relational and multidimensional databases simultaneously. Many other BI tools are now using this approach. Holos was acquired by Crystal Decisions in 1996, but has now been discontinued.
1995	Oracle acquired Express	First important BI takeover. Arguably, it was this event that put BI on the map, and it almost certainly triggered the entry of the other database vendors. Express has now become a hybrid BI and competes with both multidimensional and relational BI tools. Oracle soon promised that Express would be fully integrated into the rest of its product line but, almost ten years later, has still failed to <u>deliver on this promise.</u>
1996	BusinessObjects 4.0 launched	First tool to provide seamless multidimensional and relational reporting from desktop cubes dynamically built from relational data. Early releases had problems, now largely resolved, but Business Objects has always struggled to deliver a true Web version of this <u>desktop BI architecture. It is expected finally to achieve this by using the former Crystal Enterprise as the base</u>
1997	Microsoft announced OLE DB for BI	This project was code-named Tensor, and became the ‘industry standard’ BI API before even a single product supporting it shipped. Many third-party products now support this API, which is evolving into the more modern XML for Analysis.
1998	IBM DB2 BI Server released	This version of Essbase stored all data in a form of relational star schema, in DB2 or other relational databases, but it was more like a slow MBI than a scalable RBI. IBM later abandoned its “enhancements”, and now ships the standard version of Essbase as DB2 <u>BI Server. Despite the name, it remains non-integrated with DB2</u>
1998	Hyperion Solutions formed	Arbor and Hyperion Software ‘merged’ in the first large consolidation in the BI market. Despite the name, this was more of a takeover of Hyperion by Arbor than a merger, and was probably initiated by fears of Microsoft’s entry to the BI market. Like most other BI acquisitions, <u>this went badly. Not until 2002 did the merged company begin to perform competitively</u>
1999	Microsoft BI Services shipped	This project was code-named Plato and then named Decision Support Services in early pre-release versions, before being renamed BI Services on release. It used technology acquired from Panorama Software Systems in 1996. This soon became the BI server volume market leader through ease of deployment, sophisticated storage architecture (RBI/MBI/Hybrid), huge third-party support, <u>low prices and the Microsoft marketing machine.</u>
1999	CA starts mopping up failed BI servers	<u>CA acquired the former Prodea Beacon, via Platinum, in 1999 and renamed it DecisionBase. In 2000 it also acquired the former IA Eureka, via Sterling. This collection of failed BIs seems destined to grow, though the products are soon snuffed-out under CA’s hard-nosed ownership, and as The BI Survey 2, the remaining CA Cleverpath BI customers were very unhappy by 2002. By the The BI Survey 4 in 2004, there seemed to be no remaining users of the product.</u>
2000	Microsoft renames BI Services to Analysis Services	Microsoft renamed the second release of its BI server for no good reason, thus confusing much of the market. Of course, many references to the two previous names remain within the product.

BI 히스토리

2000	XML for Analysis announced	This initiative for a multi-vendor, cross-platform, XML-based BI API is led by Microsoft (later joined by Hyperion and then SAS Institute). It is, in effect, an XML implementation of OLE DB for BI. XML for Analysis usage is unlikely to take off until 2006.
2001	Oracle begins shipping the successor to Express	Six years after acquiring Express, which has been in use since 1970, Oracle began shipping Oracle9i BI, expected eventually to succeed Express. However, the first release of the new generation Oracle BI was incomplete and unusable. The full replacement of the technology and applications is not expected until well into 2003, some eight years after Oracle acquired Express.
2001	MicroStrategy abandons Strategy.com	Strategy.com was part of MicroStrategy's grand strategy to become the next Microsoft. Instead, it very nearly bankrupted the company, which finally shut the subsidiary down in late 2001.
2002	Oracle ships integrated BI server	Oracle9i Release 2 BI Option shipped in mid 2002, with a MBI server (a modernized Express), called the Analytical Workspace, integrated within the database. This was the closest integration yet between a MBI server and an RDBMS. But it is still not a complete solution, lacking competitive front-end tools and applications.
2003	The year of consolidation	Business Objects purchases Crystal Decisions, Hyperion Solutions Brio Software, Cognos Adaytum, and Geac Comshare.
2004	Excel add-ins go mainstream	Business Objects, Cognos, Microsoft, MicroStrategy and Oracle all release new Excel add-ins for accessing BI data, while Sage buys one of the leading Analysis Services Excel add-in vendors, IntelligentApps.
2004	Essbase database explosion curbed	Hyperion releases Essbase 7X which included the results of Project Ukraine: the Aggregate Storage Option. This finally cured Essbase's notorious database explosion syndrome, making the product suitable for marketing, as well as financial, applications.
2004	Cognos buys its second Frango	Cognos buys Frango, the Swedish consolidation system. Less well known is the fact that Adaytum, which Cognos bought in the previous year, had its origins in IBM's Frango project from the early 1980s.
2005	Microsoft to ship the much-delayed SQL Server 2005?	Originally planned for release in 2003, Microsoft may just manage to ship the major 'Yukon' version before the end of 2005.

BI 솔루션 현황

Top 10 BI Software Leaders (Out of 374 BI Software Products)

GET THE IN-DEPTH REPORT



BI 솔루션 현황

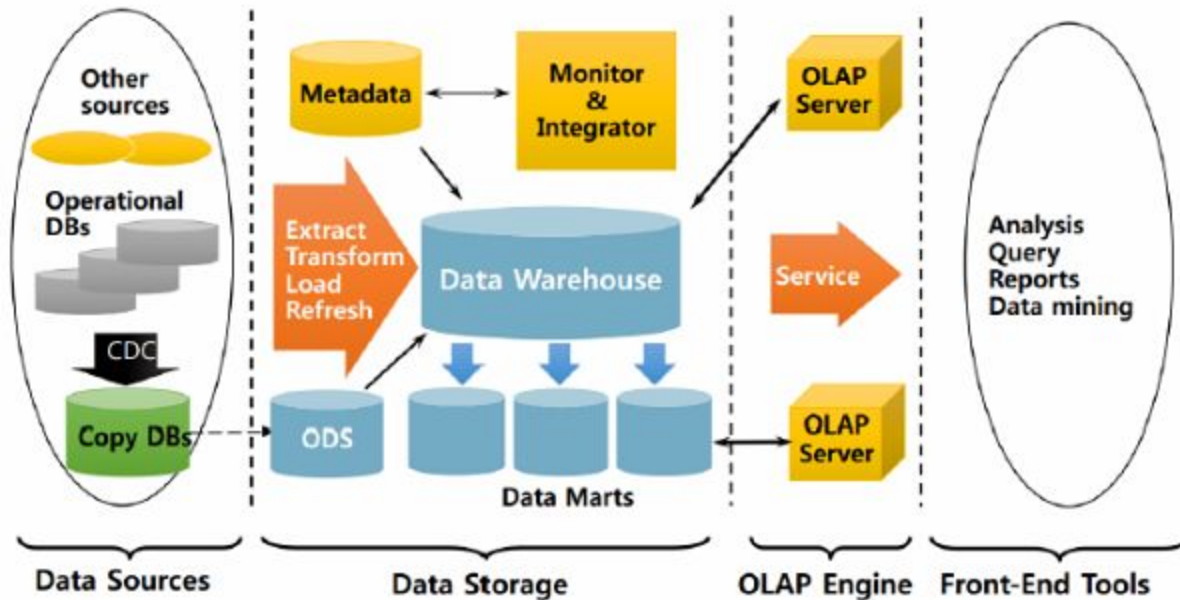
Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2018)

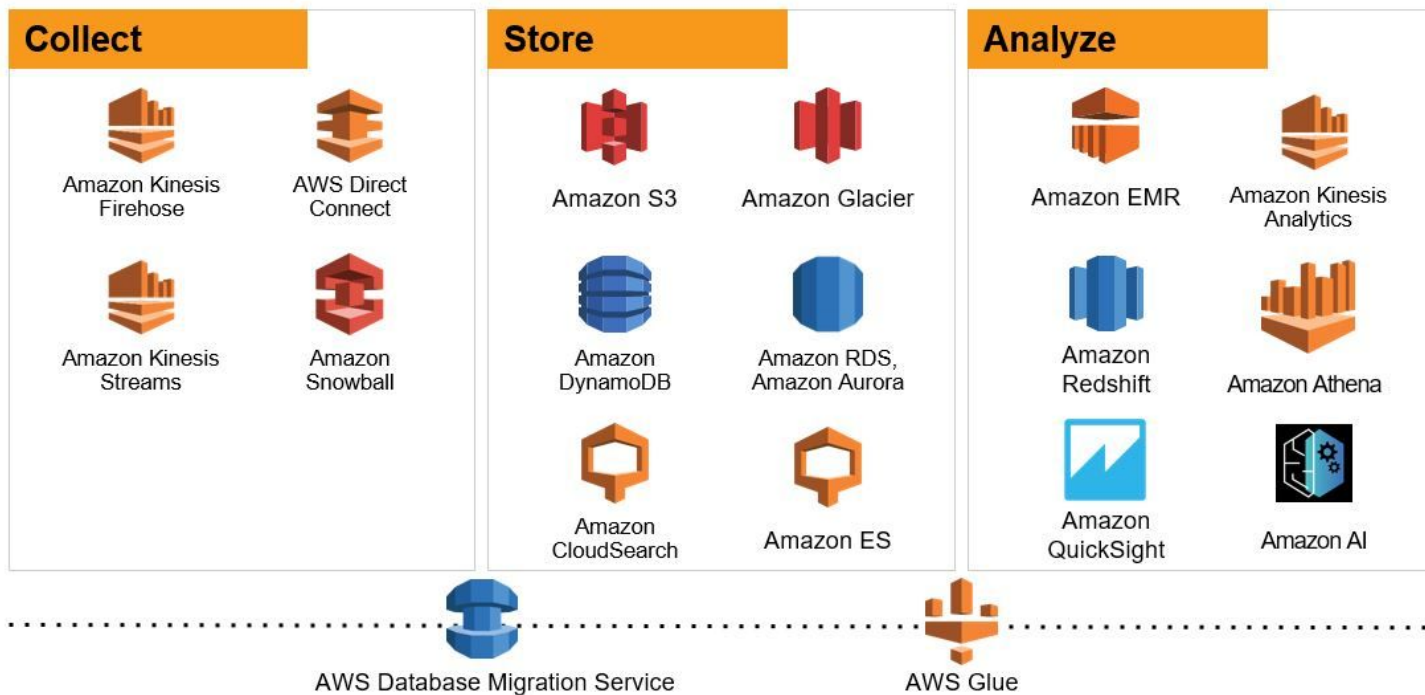
© Gartner, Inc

데이터 웨어하우징과 BI - 운영시스템과 의사결정시스템



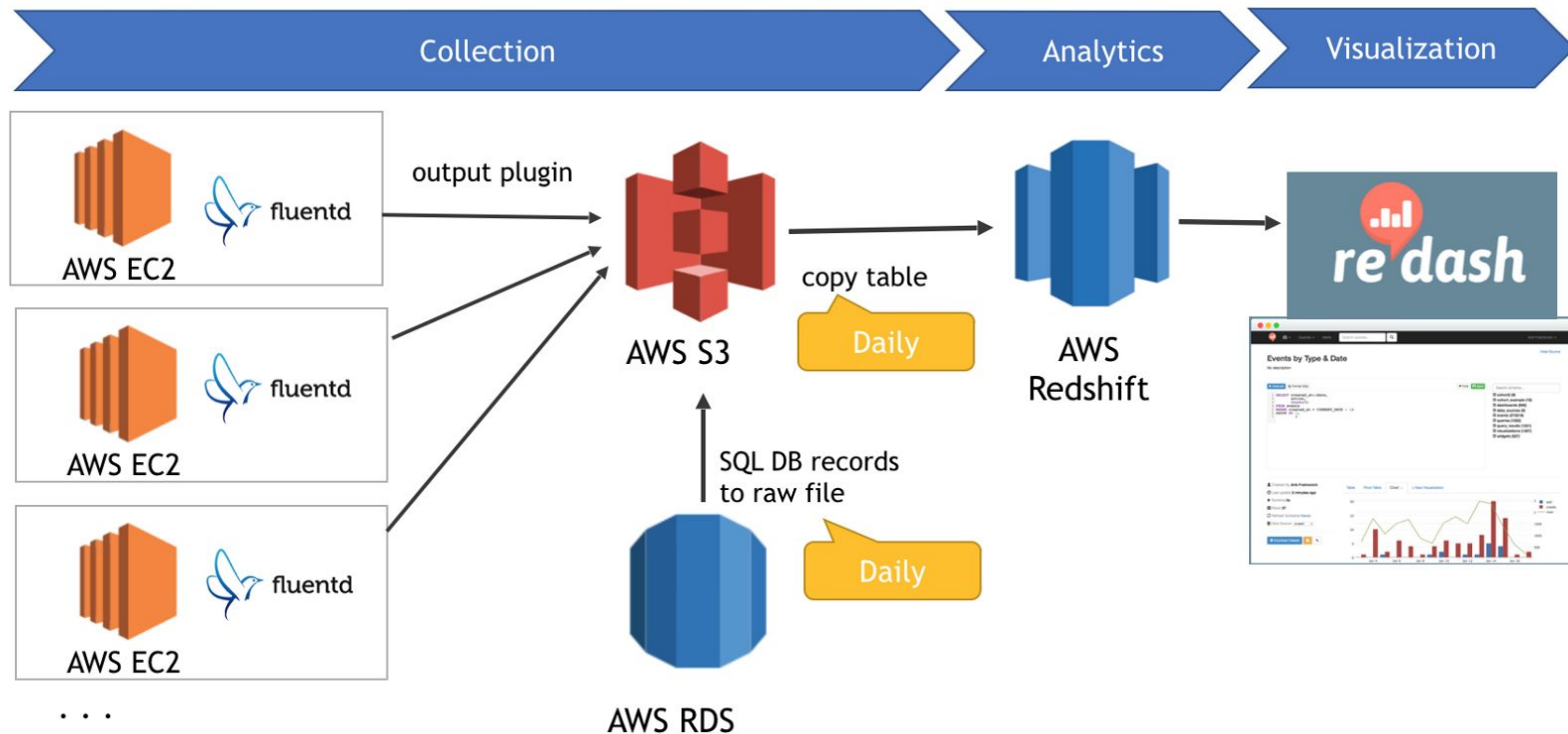
전통적인 데이터 웨어하우스 아키텍처

데이터 웨어하우징과 BI - 운영시스템과 의사결정시스템



AWS의 데이터 웨어하우스 아키텍처

데이터 웨어하우징과 BI - 운영시스템과 의사결정시스템



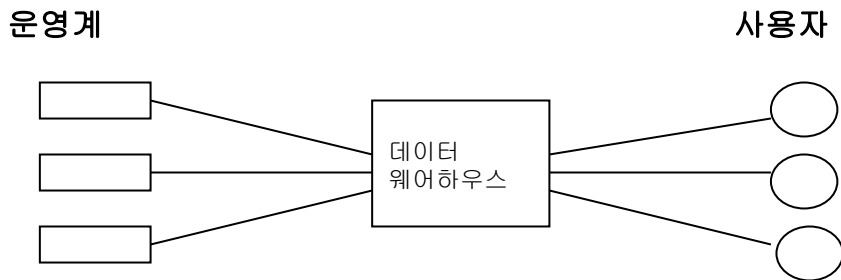
AWS의 데이터 웨어하우스 아키텍처

데이터 웨어하우징과 BI - 운영시스템과 의사결정시스템

- OLTP시스템은 짧은 시간에 끝나는 간단한 트랜잭션들이 대량으로 발생한다. 일반적으로 이 시스템의 업무는 많은 사용자들이 동시에 데이터베이스에 접근하여 몇 개의 레코드를 읽거나 갱신하므로, 단위 시간에 처리되는 트랜잭션의 수와 응답시간이 매우 중요하다.
- 의사결정지원시스템은 OLTP에 비해 동시에 데이터베이스에 접근하는 사용자 수가 적은 반면, 사용자는 다양한 관점에서 데이터를 분석하기를 원한다. 따라서 질의내용이 복잡하고 대화식으로 수행되며, 시스템은 사용자 질의에 응답하기 위해 한번에 매우 많은 양의 데이터를 처리해야 한다.
- 이 두 시스템의 차이로 인해 의사결정시스템은 OLTP와 분리되어 구축되어야 했다. 따라서 기업들은 OLTP로부터 필요한 데이터를 추출하고 수정, 요약해서 의사결정을 지원할 수 있는 새로운 데이터베이스를 구축하였다. DSS는 각 부서별로 산발적으로 구축되어 통합적이지 못하고 다양한 시스템 상에 산재되어 존재하게 되어 활용을 거의 하지 않았다.

데이터 웨어하우징과 BI - 데이터 웨어하우징

- 사용자의 의사결정을 지원하기 위해 기업이 축적한 많은 데이터(Time Variant)를 사용자 관점에서 주제별로(Subject-Oriented)통합하여 별도의 장소에 저장해 놓은 데이터베이스로 이해할 수 있다.
- 의사결정을 위한 별도의 통합된 공간을 가짐으로써 다음의 장점을 얻을 수 있음
 - 첫째, 운영시스템을 보호하고 사용자 질의에 신속한 응답성능을 제공할 수 있다.
 - 둘째, 여러 시스템에 산재된 데이터들이 웨어하우스로 취합되고 통합되므로 사용자는 자신들이 필요로 하는 데이터가 어디에 있는지 신경쓰지 않고 필요한 데이터를 쉽게 가져다 쓸 수 있다.
 - 셋째, 데이터는 웨어하우스로 옮겨오기 전에 정제 및 검증과정을 거치게 되며, 따라서 사용자는 양질의 데이터를 사용할 수 있다.



데이터 웨어하우징과 BI - 데이터 추출 및 가공

- 단순한 복제가 아닌 정제 및 변형가능을 거침.
- 1세대 추출 툴 - 코드 생성기: 프로그램 소스 생성 방식
- 2세대 추출 툴 - 변환 엔진 방식: 추출도구가 추출 전 과정을 제어하는 방식
- 메타데이터: 추출 및 변형과정에서 소스와 타겟사이의 대응관계설정(mapping), 변환규칙, 추출주기 등과 같은 부가적인 데이터가 발생

데이터 웨어하우징과 BI - 데이터 웨어하우스 와 데이터 마트

- 웨어하우스와 사용자사이에 데이터마트가 존재해야 하는 이유
 - 첫째, 웨어하우스는 최종사용자와의 인터페이스보다는 방대한 분량의 데이터를 효율적으로 통합하고 관리하는 측면에 보다 초점을 맞춘다. 따라서 사용자 측면에서 편리한 형태로 설계되지 않을 수 있다.
 - 둘째, 웨어하우스는 전사적인 용도로 구축되기 때문에 각 개별부서나 사용자 집단에 적합한 형태로 데이터가 저장되지 않는다. 따라서 사용자 질의에 최적의 성능을 제공하지 못할 수 있다.
 - 셋째, 대부분의 사용자들은 웨어하우스의 전체 데이터중 일부분만을 주로 사용할 것이다. 기업의 모든 사용자들이 웨어하우스에 대해 직접 질의를 수행하는 것은 많은 시스템 자원을 필요로 하며 전체 시스템 성능에 심각한 부하를 줄 수 있다.
- 웨어하우스가 먼저 구축된 다음 각 부서나 사용자집단의 필요에 의해 구축되는 것을 하향식 (Top-Down) 접근법이라 한다. 각 데이터마트는 중앙의 웨어하우스로부터 데이터를 공급받는데, 이것을 종속형 데이터마트 (Dependent Datamart)라 한다. 웨어하우스 없이 데이터마트만 존재하는 경우 이를 독립형 데이터마트 (Independent Datamart)라 하며, 이와 같이 데이터마트를 먼저 구축하고 나중에 웨어하우스로 통합하는 구축방식을 상향식 접근법 (Bottom_Up Approach)이라 한다.

데이터 웨어하우징과 BI - 데이터 웨어하우스 액세스 툴

- 데이터 웨어하우스 액세스 툴
 - 쿼리 및 리포팅 툴: 기업데이터에 가장 전통적인 방식
 - BI 툴
 - 데이터 마이닝 툴
- 데이터 웨어하우징과 BI 시스템
 - 초기 데이터 웨어하우징의 초점은 의사결정을 지원하는 정보기반의 구축이었고 BI은 정보의 효과적인 활용 측면에 초점을 맞춤.
 - 데이터 웨어하우징의 관점이 구축보다 활용으로 바뀜으로 BI은 데이터 웨어하우징의 환경으로 급격히 통합됨.
 - 일부 BI은 데이터 웨어하우스가 지원하지 못하는 특성을 요구하기도 함. 예산 편성을 지원하는 BI 어플리케이션은 사용자에게 의한 쓰기 기능이 가능해야 함으로 별도의 데이터 마트를 구축해야 함.
- 메타데이터 관리
 - 데이터에 대한 데이터로서 데이터를 올바르게 사용하고 관리할 수 있도록 하위레벨 데이터를 기술하는 상위레벨 데이터로 이해
 - 제품 마다 고유한 양식이 있으나 메타 데이터를 교환하기 위한 통일노력이 이루어지고 있음.

데이터 웨어하우스와 데이터베이스 비교

데이터 웨어하우스는 대량의 데이터를 읽어 데이터 전반에 걸친 관계와 동향을 이해하는 작업이 포함되는 데이터 분석을 위해 특별히 설계되었습니다. 데이터베이스는 트랜잭션의 세부 사항을 기록하는 것과 같이 데이터를 캡처하고 저장하는 데 사용됩니다.

특징	데이터 웨어하우스	트랜잭션 데이터베이스
적합한 워크로드	분석, 보고, 빅 데이터	트랜잭션 처리
데이터 원본	여러 소스로부터 수집되고 정규화된 데이터	트랜잭션 시스템과 같이 단일 소스에서 있는 그대로 캡처한 데이터
데이터 캡처	대개 미리 결정된 대량 배치 일정에서의 대량 쓰기 작업	트랜잭션 처리량을 최대화할 수 있도록 새로운 데이터가 사용 가능해지면서 연속적인 쓰기 작업에 최적화됨
데이터 정규화	Star 스키마 또는 Snowflake 스키마와 같이 비정규화된 스키마	고도로 정규화된 정적 스키마
데이터 스토리지	컬럼 방식 스토리지를 사용하여 간단한 액세스 및 고속 쿼리 성능에 대해 최적화됨	단일 행 지향 물리적 블록에 대한 고도의 처리량 쓰기 작업에 최적화됨
데이터 액세스	I/O를 최소화하고 데이터 처리량을 최대화하도록 최적화됨	대량의 소규모 읽기 작업

데이터 웨어하우스와 데이터 레이크 비교

데이터 웨어하우스와 달리, 데이터 레이크는 정형 및 비정형 데이터를 비롯한 모든 데이터에 대한 중앙 리포지토리입니다. 데이터 웨어하우스는 분석에 최적화된 사전 정의된 스키마를 사용합니다. 데이터 레이크에서는 스키마가 정의되지 않아 대형 데이터 분석, 전체 텍스트 검색, 실시간 분석 및 기계 학습과 같은 추가 유형의 분석이 가능합니다.

특징	데이터 웨어하우스	데이터 레이크
데이터	트랜잭션 시스템, 운영 데이터베이스 및 LOB(Line of Business) 애플리케이션의 관계형 데이터	IoT 장치, 웹 사이트, 모바일 앱, 소셜 미디어 및 기업 애플리케이션의 비관계형 및 관계형 데이터
스키마	데이터 웨어하우스 구현 전에 설계됨(스키마-온-라이트)	분석 시에 쓰여짐(스키마-온-리드)
가격/성능	고비용의 스토리지를 사용하여 가장 빠른 쿼리 결과 얻음	저비용의 스토리지를 사용하여 쿼리 결과의 속도가 빨라짐
데이터 품질	진실의 중앙 버전 역할을 하는 고도로 큐레이팅된 데이터	큐레이팅되거나 될 수 없는 모든 데이터(예 : 원시 데이터)
사용자	비즈니스 애널리스트, 데이터 과학자 및 데이터 개발자	데이터 과학자, 데이터 개발자 및 비즈니스 애널리스트 (큐레이팅된 데이터 사용)
분석	배치 보고, BI 및 시각화	기계 학습, 예측 분석, 데이터 디스커버리 및 프로파일링

데이터 웨어하우스와 데이터 마트 비교

데이터 마트는 금융, 마케팅 또는 영업과 같은 특정 팀 또는 사업 단위의 요구를 충족시키는 데이터 웨어하우스입니다. 규모가 더 작고, 집중적이며 사용자 커뮤니티에 가장 잘 맞는 데이터 요약물을 포함할 수 있습니다.

특징	데이터 웨어하우스	데이터 마트
범위	함께 통합된 중앙 집중식의 여러 주제 영역	분산된 특정 주제 영역
사용자	전사적	단일 커뮤니티 또는 부서
데이터 원본	여러 소스	단일 또는 몇 개의 소스, 또는 데이터 웨어하우스에 이미 수집된 데이터의 일부
크기	대형이며 수백 기가바이트에서 페타바이트까지 이를 수 있음	소형이며 대개 최고 수십 기가바이트에 이름
설계	하향식	상향식
데이터 세부 정보	완전한, 상세 데이터	요약된 데이터를 포함할 수 있음