

**64비트 시대의 대용량 데이터 초고속 처리 기술**

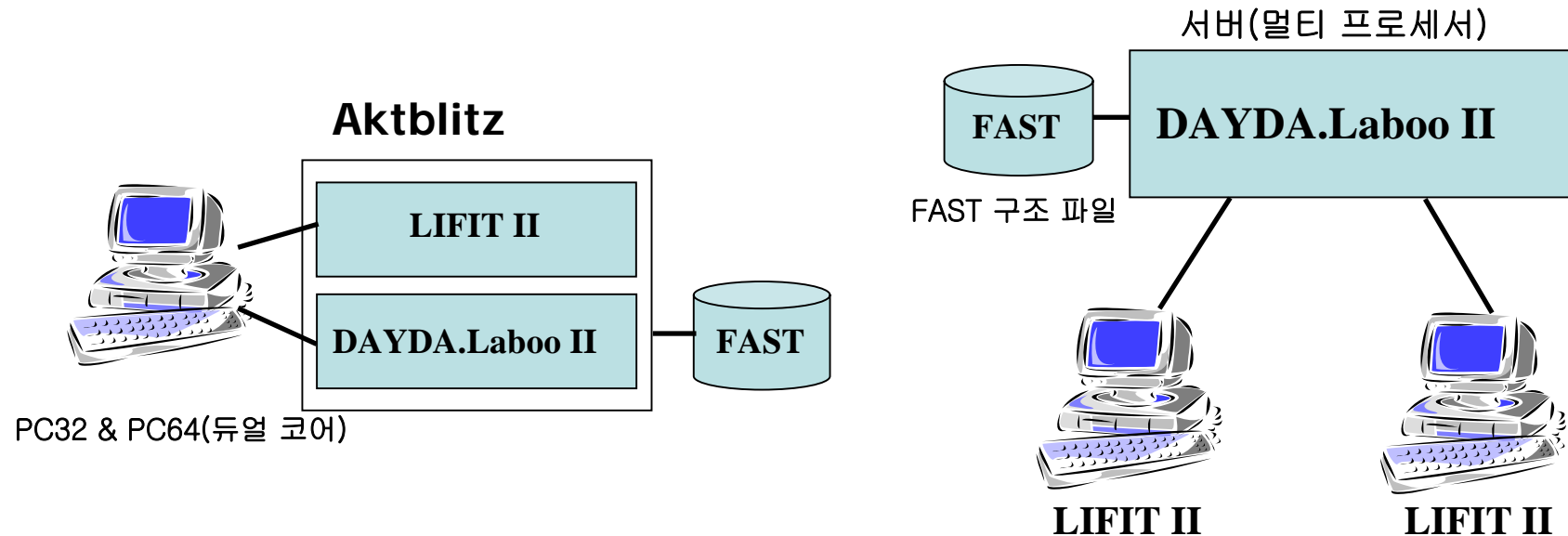
# LFM 처리 기술의 소개

Real Time VIDP : Real time Visual Interactive Database Processing

**주식회사 휴먼에스티**

2007.01.

1. **FAST (Filter Array Structure)**  
DAYDA.Laboo의 특징인 데이터 성분 분해법을 채용한 데이터 구조
2. **LFM (Linear Filtering Method)**  
FAST 구조를 처리하는 알고리즘군의 총칭 (특히 취득)
3. **DAYDA.Laboo II (데이다 라보 2)**  
LFM으로 실현된 멀티 프로세서 대응 초고속 데이터 처리 엔진(1/3 MS모델)
4. **LIFIT II (라이피트 2)**  
DAYDA.Laboo를 사용하기 위한 GUI 툴
5. **Aktblitz**  
스탠드얼론형 데이터 처리 툴 (DAYDA.Laboo + LIFIT)



## LFM 기술이란

기존의 레코드 단위 처리를 전혀 새로운 개념의 **파일 단위 일괄 온메모리 처리**로 구현하고, 메모리 대용량화가 가능한 64비트 시대에 적합한 데이터 처리의 신기술

## 특징

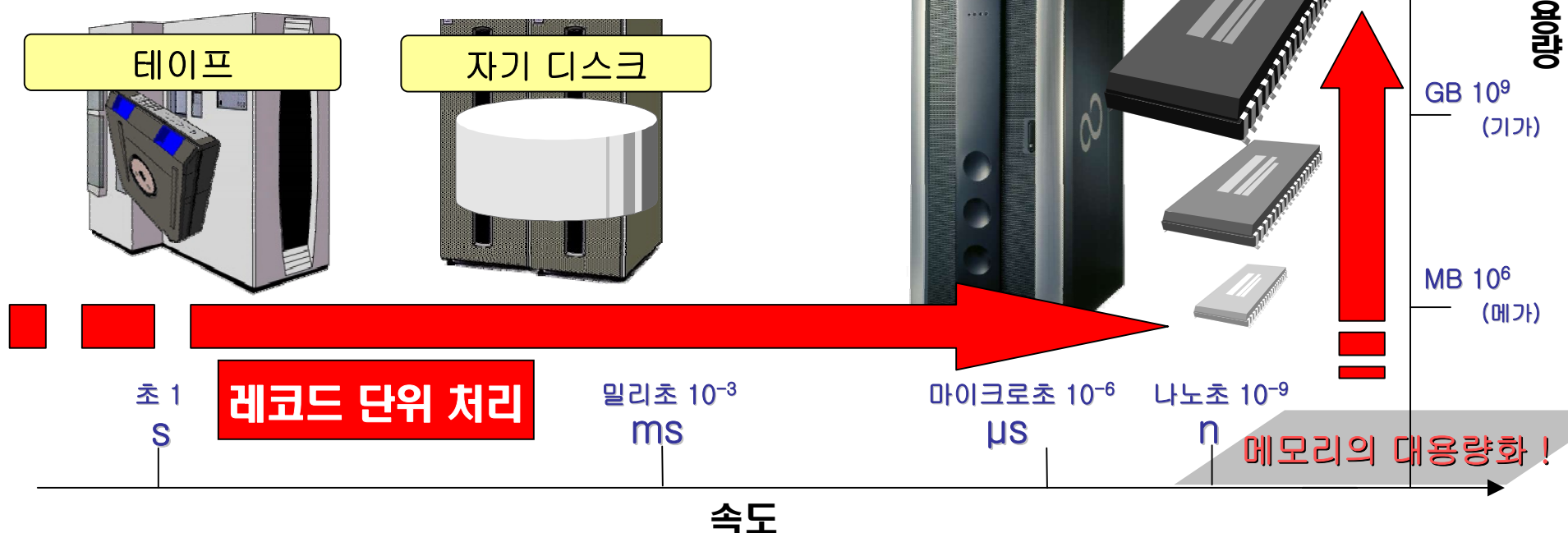
1. 범용성/확장성
  - 노트북, 서버, SMP, CELL 컴퓨터, 패러렐 컴퓨터에 적용 가능
2. 대용량 데이터 처리 가능
  - **20억행의 대용량 데이터**까지 처리 가능
3. 초고속 처리
  - **최대 20억행까지 비주얼하게 대화형으로 초고속 처리**
  - 소트, 조인, 검색, 추출, 집계, 계산 등 각종 데이터 가공 처리를 컬럼 집합으로 일괄처리
4. 데이터 처리 시스템(배치, DW 등)의 개발비, 기간을 대폭 삭감
  - **비용을 1/5 이상 절감**
  - 데이터를 비주얼하게 대화형으로 처리
5. 시스템 비용 대폭 절감
  - 중간 파일용 디스크 불필요
  - 고비용 디스크 드라이브의 대폭 삭감 (1/40 - 1/100)
  - 배치처리 시간의 단축에 의한 시스템 운용 비용의 대폭 삭감 실현

## 64비트 시대는 온메모리 처리의 시대

- 64 bit 서버 환경 본격화
- 대용량 메모리 탑재 가능 (메모리 가격 하락)

**일괄 온메모리 처리의 실현**

메모리를 통한 초고속 처리가 가능한 데이터 구조를 개발하여  
기존 디스크 처리보다 비약적인 초고속성을 실현



## 1. LFM이란

대용량 데이터의 검색, 소트, 머지, 조인 등의 처리 속도를 개선하기 위해 개발된 온메모리 데이터 처리 알고리즘

## 2. 알고리즘의 종류

FAST(Filter Array Structure) 데이터 구조 알고리즘을 고안 하여, 현재 1/3 데이터 모델을 제품화 (초고속 처리 엔진 제품 : DAYDA.Laboo)

## 3. DAYDA.Laboo 탑재 제품

(주)Hitachi의 RH-BOM의 데이터 처리 엔진으로, (주)Fujitsu(전사 조달 관리 시스템) 및 일본의 유력 기업에서 엔진을 탑재하여 제품화  
Fujitsu BSC는 1/3 모델 특허 사용권을 취득하여, 신제품 Oh-Pa 1/3를 발표

## 4. 특허 - 약 30개의 특허에 대해 세계 각국에서 특허 권리 보유

### ■ 데이터 형식과 모델 (실행 환경별)

데이터형식	표형식	트리형식	비고
Single CPU	1/3	2/4	
SMP	1/3MS	2/4MS	CPU간 메모리공유
Parallel	3/5	4/6	메모리 공유 안함
CELL	3/5CELL	4/6CELL	

※ n/N  
n : 오름차순 성분수  
N : 데이터 구조 기술 성분수

연구 개발 중

개발 예정

## 1. LFM의 특징

파일 전체를 메모리상에 D5T라는 형식으로 일괄 전개하고, 기존에 레코드 단위로 처리하던 데이터를 컬럼(항목) 단위의 집합으로 처리함으로써, 디스크로부터 레코드 단위로 읽어들이던 기존 처리에 비해, 비약적으로 빠르고 유연한 데이터 처리가 가능하게 되었습니다.

## 2. 효과

### 2.1 대규모 DW의 실현

대규모 DW 등의 실용화가 가능하게 되었습니다. (Hitachi의 RH-BOM)

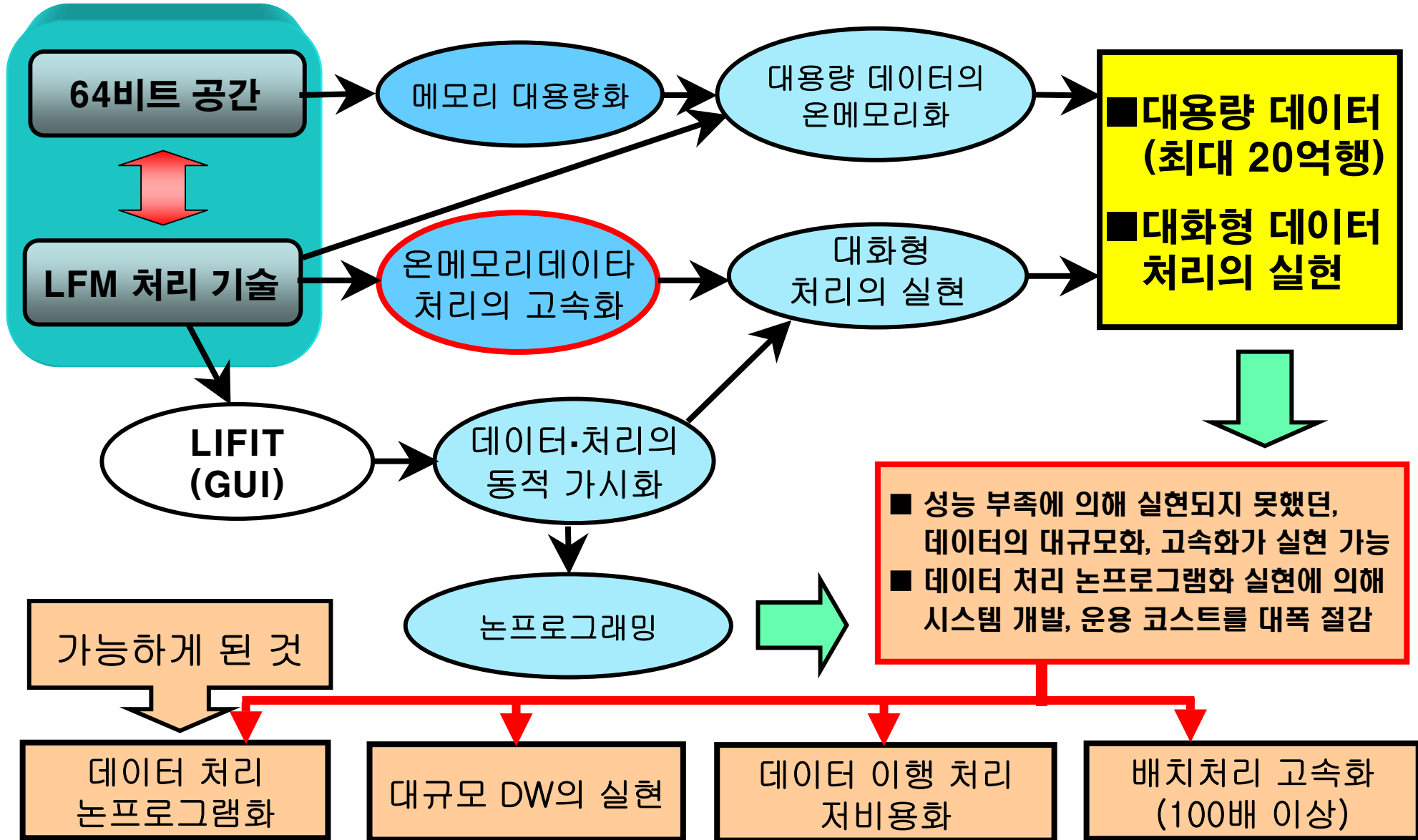
### 2.2 배치처리의 고속화/생산성 향상

데이터 처리를 동적으로 가시화함으로써 개발 언어를 사용하지 않고 데이터 처리가 가능하게 되었습니다.

배치처리 등의 데이터 처리 생산성이 비약적으로 향상합니다.

32비트에서는 탑재할 수 있는 메모리가 제약을 받았지만, 64비트에서는 거의 무제한으로 메모리 탑재가 가능하며, 메모리 가격 하락에 따른 대용량의 메모리 탑재도 가능하게 되어 데이터 처리에 새로운 패러다임의 전환이 가능한 혁신적인 처리 기술입니다.

개발 언어	1/3 모델
개발 기간	1/2
개발비	1/4 ~ 1/5
발생 장애	1/10
시스템 성능	수십~수백배



## 1. 대용량 데이터 처리 시간 단축 (수십배 - 수천배)

- 매일 야간 배치처리 시간 단축
- 메인프레임(32비트, 고성능 I/O 처리 효율, 디스크 의존 시스템)으로부터 오픈시스템(64비트) 이행시의 **I/O 처리 효율 부족 회피**
- 실시간으로 생성되는 대량 데이터의 각종 분석의 비약적 단주기화  
예) **반도체 라인의 모든 제조 장치의 가동 정보를 5분 간격으로 분석 실현**
- 대규모 DW의 실현

## 2. 시스템 구축 생산성 / 비용 절감 (기존의 5배 이상)

- 논프로그래밍화 : **개발 엔지니어 불필요**
- 동적 데이터 처리 정의에 의한 개발, 테스트, 시스템 변경의 단주기화, 배치처리의 비약적 생산성의 향상

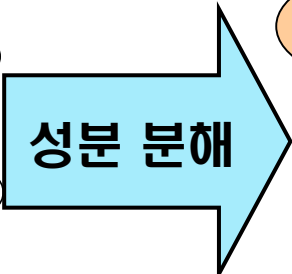
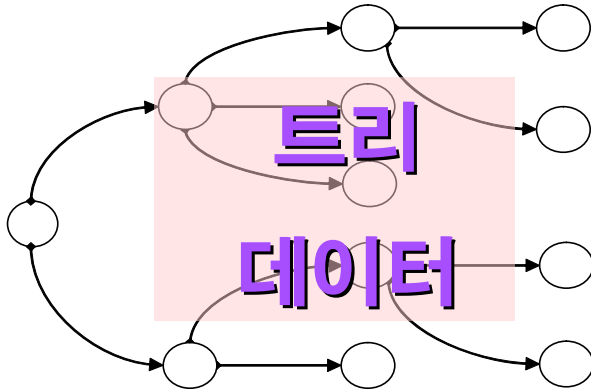
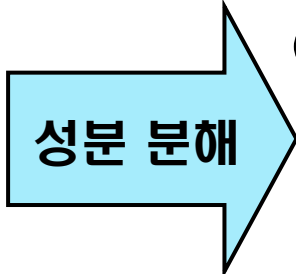
## 3. 기업 IT시스템 부문과 최종 사용자 부문의 책임 분담 명확화

- 전사 시스템의 구축 유지 : 시스템 부문
- 데이터 마트의 작성 유지 : 업무 담당자가 자유롭게 데이터 마트 작성

기존 데이터 처리는 레코드 단위이지만, LFM에서는 필드(컬럼) 단위로 순서, 값테이블, 위치 정보의 3개로 분해해서 관리합니다.

	성별	연령	신장	체중
0	남성	18	168	55
1	여성	21	172	64
2	남성	15	159	48
3	남성	10	172	48
4	여성	28	181	78
5	남성	16	163	66
6	남성	15	158	52
7	여성	33	174	65
8	여성	24	177	64
9	남성	18	170	55

테이블  
데이터



### 표 형식의 데이터 표현

	성별	연령	신장	체중
0	남성	18	168	55
1	여성	21	172	64
2	남성	24	159	48
3	남성	16	172	48
4	여성	28	181	78
5	남성	20	166	66
6	남성	16	168	52
7	여성	33	174	65
8	여성	24	177	64
9	남성	18	170	55

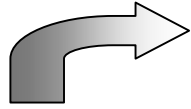
- 컴파일 처리  
표 형식의 데이터를 컬럼마다 성분 분해해, FAST 구조로 변환
- OrdSet(Ordered Set:순서 표현)  
레코드의 오리지널 표시 순서를 표현
- VL(Value List:값테이블)  
컬럼에 포함되는 Unique한 데이터값을 오름차순으로 관리
- VNo(Value Number:위치 정보)  
값테이블에서의 참조값

**FAST 구조 (D5T)**

**컴파일 처리(성분 분해된 실데이터 작성)**

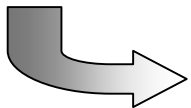


## 검색 처리의 예

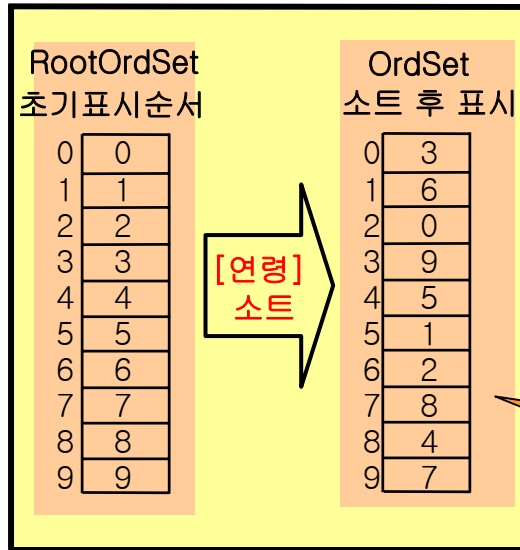
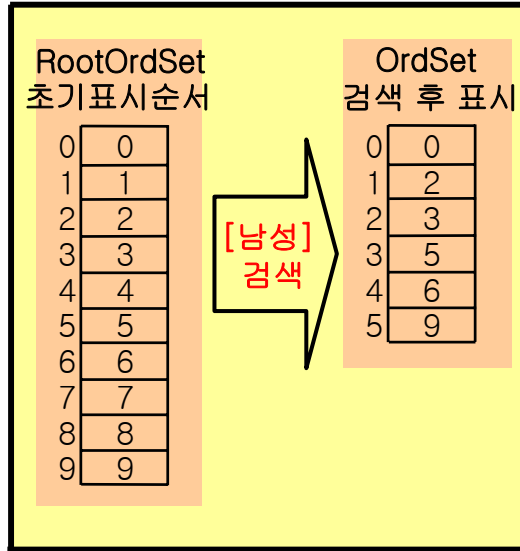


## 표 형식 데이터

	성별	연령	신장	체중
0	남성	18	168	55
1	여성	21	172	64
2	남성	24	159	48
3	남성	16	172	48
4	여성	28	181	78
5	남성	20	166	66
6	남성	16	168	52
7	여성	33	174	65
8	여성	24	177	64
9	남성	18	170	55



## 데이터 소트의 예



- VNo, VL 각 테이블은 1회만 생성
- OrdSet 테이블만 처리에 따라 작성

## 성분 분해된 실 데이터

성별		연령		신장		체중	
VNo	VL	VNo	VL				
0	0	0	1	2	159	2	48
1	1	1	3	4	166	3	52
2	0	2	4	0	168	0	55
3	0	3	0	4	170	0	64
4	1	4	5	7	172	6	65
5	0	5	2	1	174	5	66
6	0	6	0	2	177	1	78
7	1	7	6	5	181	4	
8	1	8	4	6		3	
9	0	9	1	3		2	

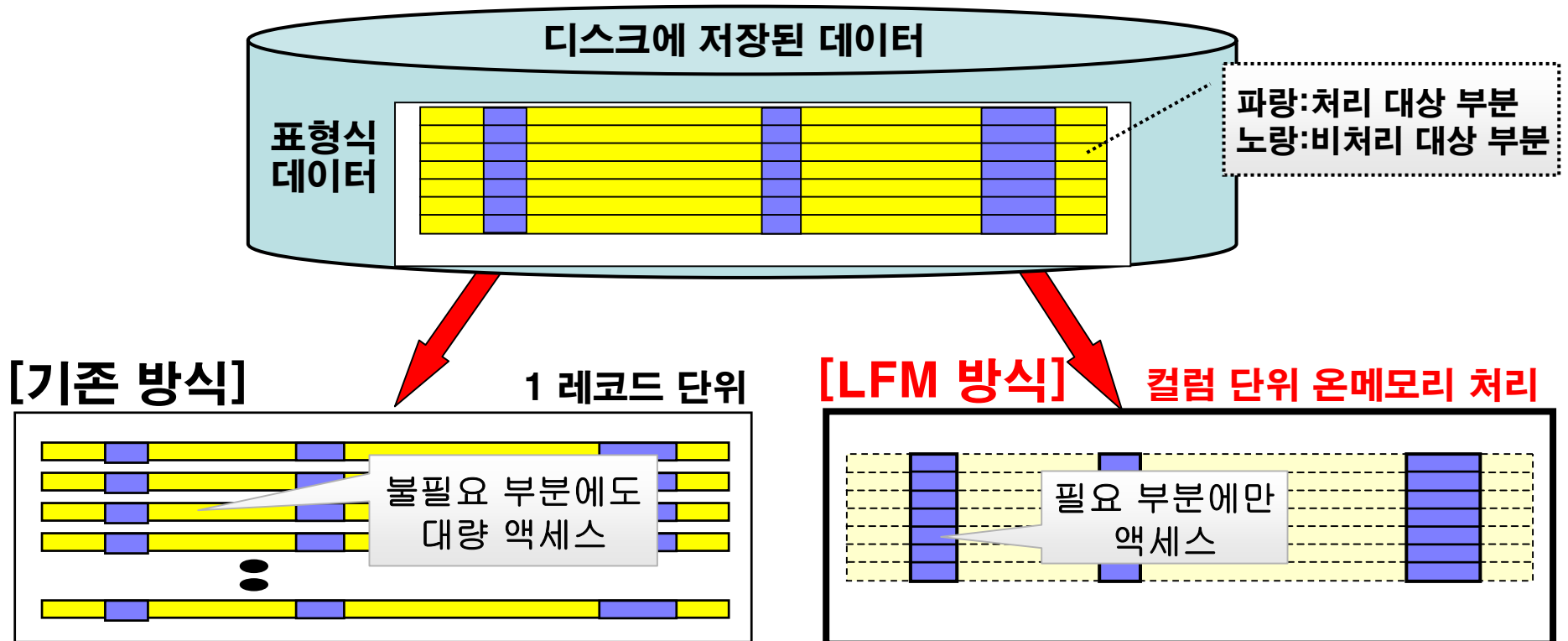
VNo: 위치 (Value List 번호 참조값)

VL: 값 (항목중의 유니크 데이터) 오름차순

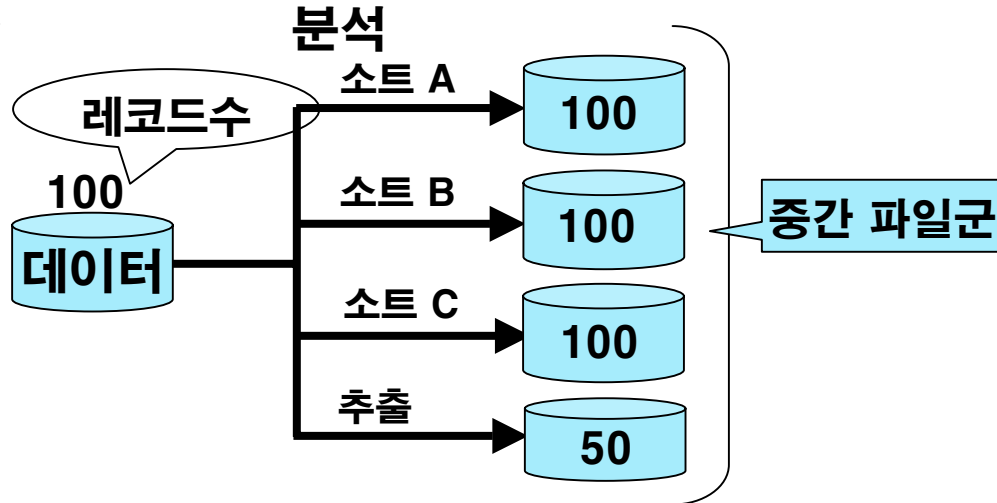
OrdSet: 순서 표현 (데이터 표시 순서)

OrdSet : Ordered Set  
 VNo : Value Number  
 VL : Value List

- **64비트 시대의 대용량 메모리를 체계화**
  - 레코드 단위의 순차 처리를 컬럼 단위의 일괄처리로
- **멀티 코어 시대의 병렬처리를 실현**
  - 검색·소트·JOIN·매칭·집합 연산...모두 병렬화



## 기존 데이터 처리

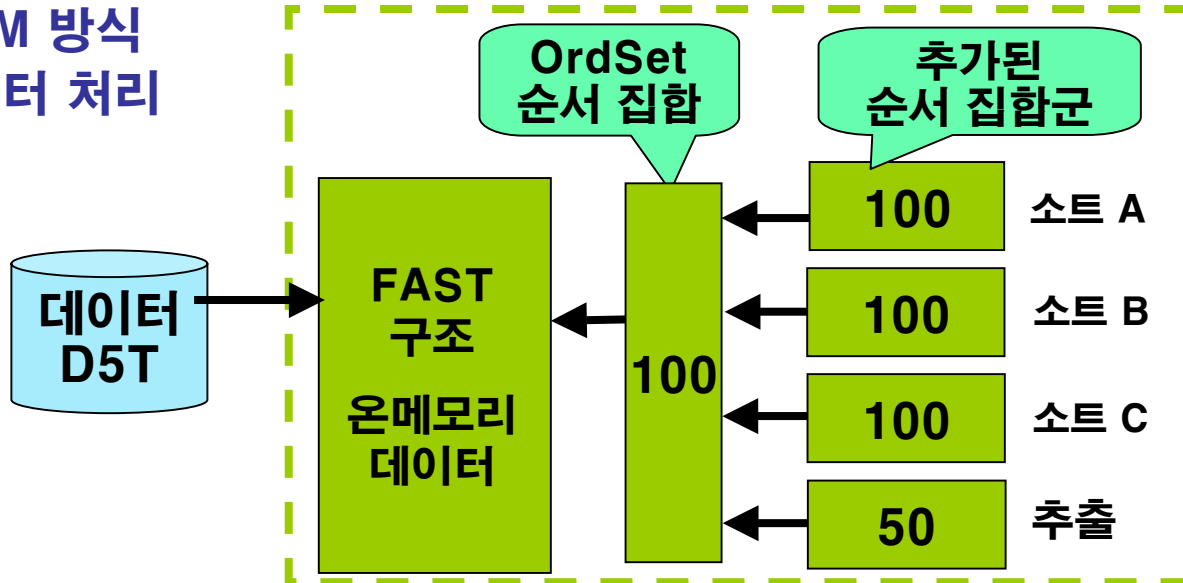


### ■ 디스크 베이스의 배치처리

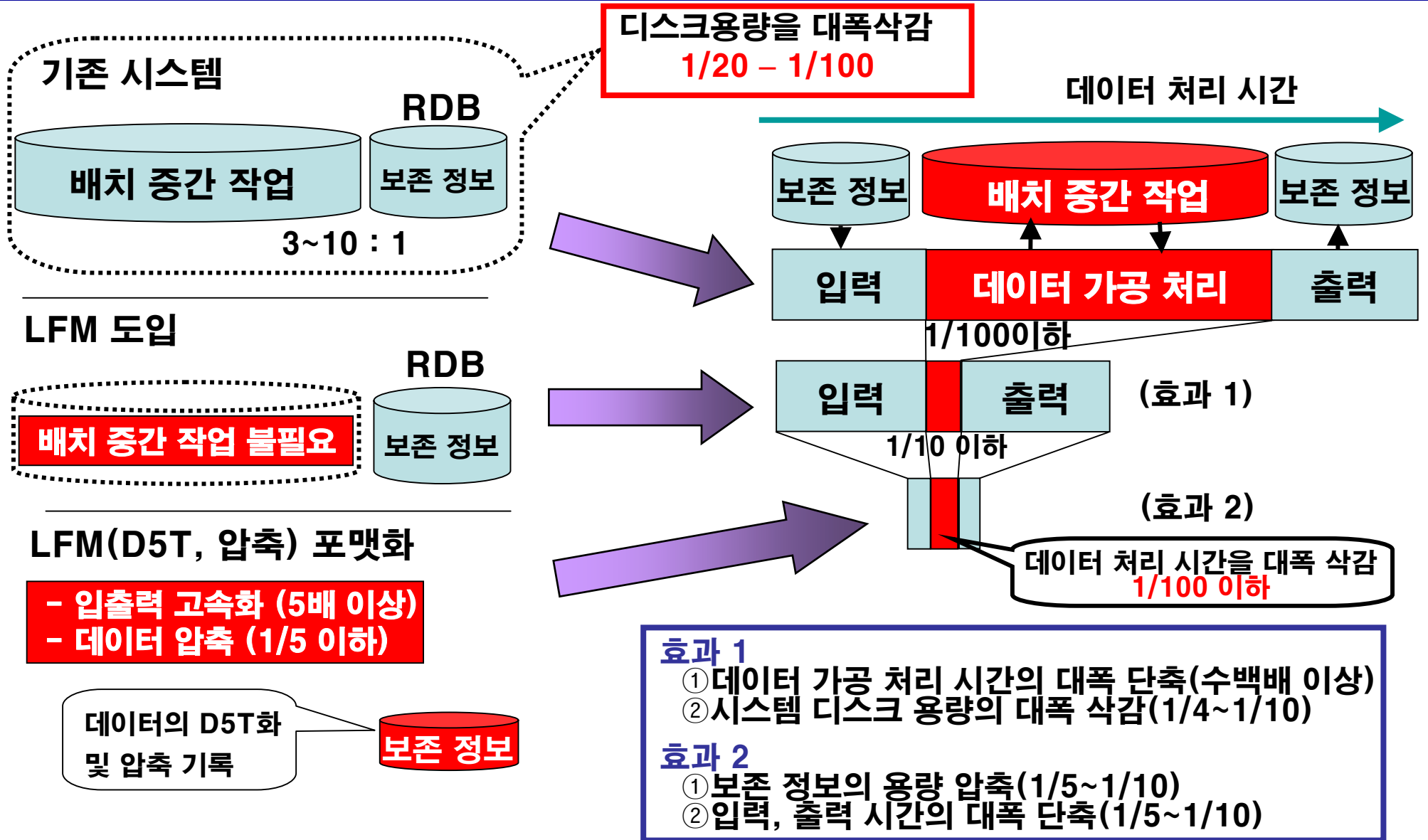
- 중간 파일을 디스크에 레코드 별로 저장한다
- I/O 병목 현상에 의해 처리 시간이 길어진다

실메모리 공간

## LFM 방식 데이터 처리

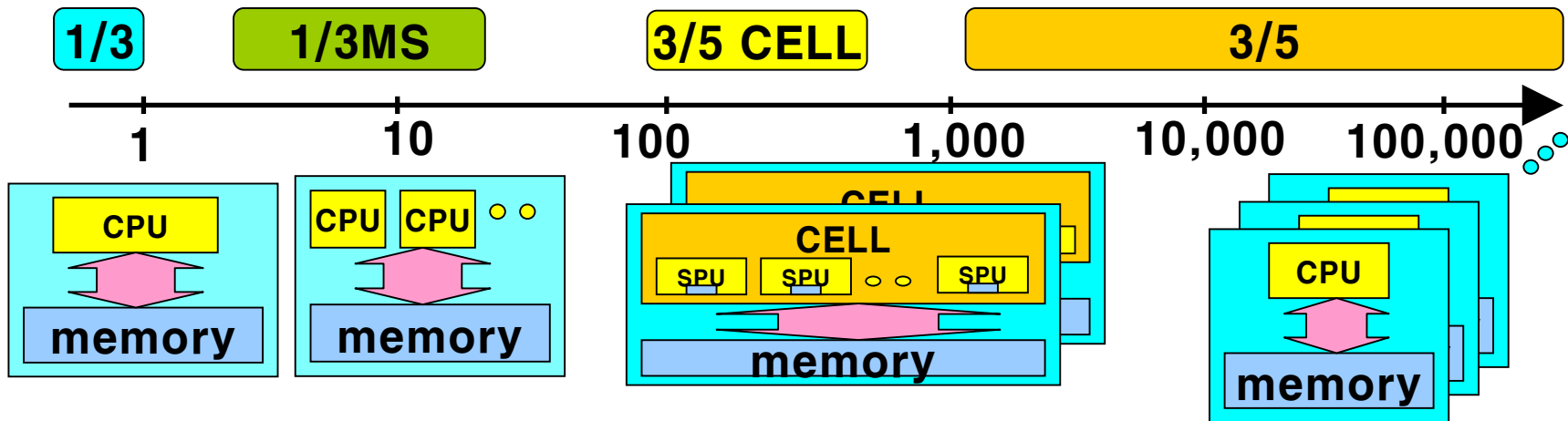


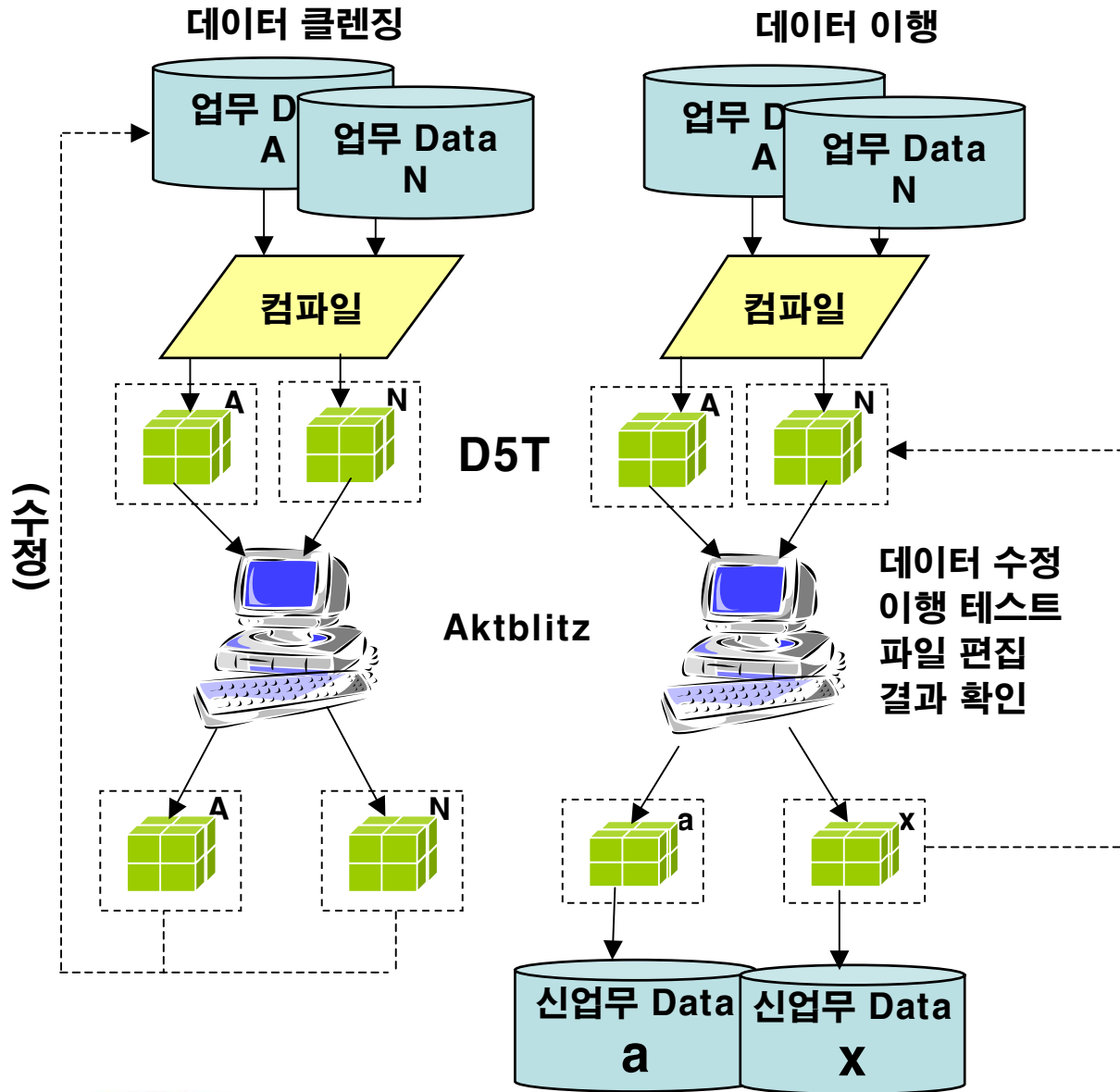
- ① 데이터는 FAST 구조로 온메모리화하고,
- ② 중간 파일은 순서 집합으로 저장하므로 데이터의 복사는 발생하지 않고,
- ③ 기존 처리에 비해 비약적으로 빠르다.



모델	1/3 (1/3)	1/3MS (1/3)	3/5 CELL (3/5 II)	3/5 (3/5)
프로세서 코어수(SPU수)	1	1 ... 8	64 or 72(SPU)	64 ... 2 <sup>20</sup>
테이블 최대행수	21 억	21 억	21 억	1 조
성능 지표(*1)	1	~ 10	100 ~ 300	1200 ~ 400000

\*1 1억행의 테이블을 소트하는데 40초 걸릴 때를 '1'로 한다





## 1. 데이터 클렌징

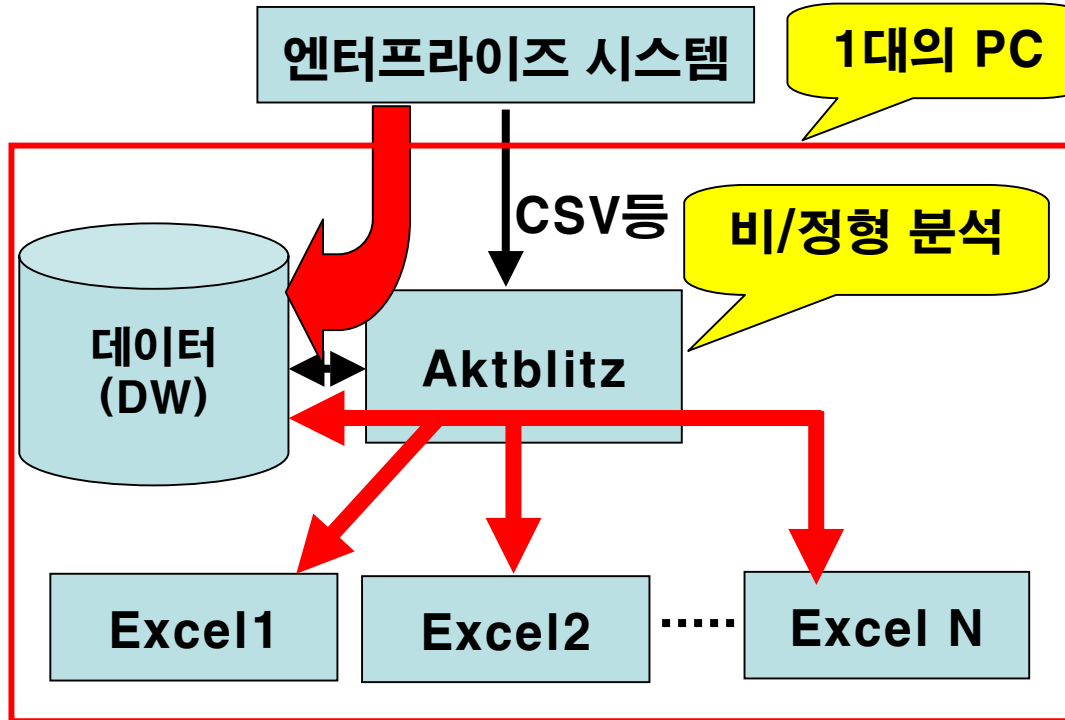
- 대화형으로 컬럼별 체크
- 부정 데이터 검출·수정이 용이
- 조작이 간단하고, 초고속으로 처리하며 정확한 결과값 피드백

## 2. 데이터 이행

- 이행 순서를 대화형 조작으로 확인 가능
- 프로그램 작성 불필요
- 간단하고 빠르게 테스트 및 가동이 가능
- 대화형으로 이행 결과 확인

Aktblitz는 초고속 데이터 처리 엔진 인 DAYDA.Laboo(데이다 라보)와 GUI 틀인 LIFIT(라이피트)를 PC에 설치하여, LFM 처리 기술을 PC 레벨에서 사용할 수 있도록 개발된 툴입니다.

기업의 업무용 이외에도 Excel 등에서 고도의 데이터 처리를 실시하는 파워 유저를 위한 데이터 전처리, 관리툴로서 사용 가능합니다.



- ① 분석 툴 - 정형, 비정형 분석 가능
- ② 데이터 관리 및 처리 툴  
시시각각 변화하는 정보를 취득하여  
일관성 있는 데이터로서 일괄 관리
- ③ Excel 등의 BI툴에 보내는 데이터를  
자동적으로 고속으로 편집 가능

# 제품 문의

주식회사 휴먼에스티

<http://www.humanst.com/Aktblitz/>

대표전화 : 02-2109-8910