

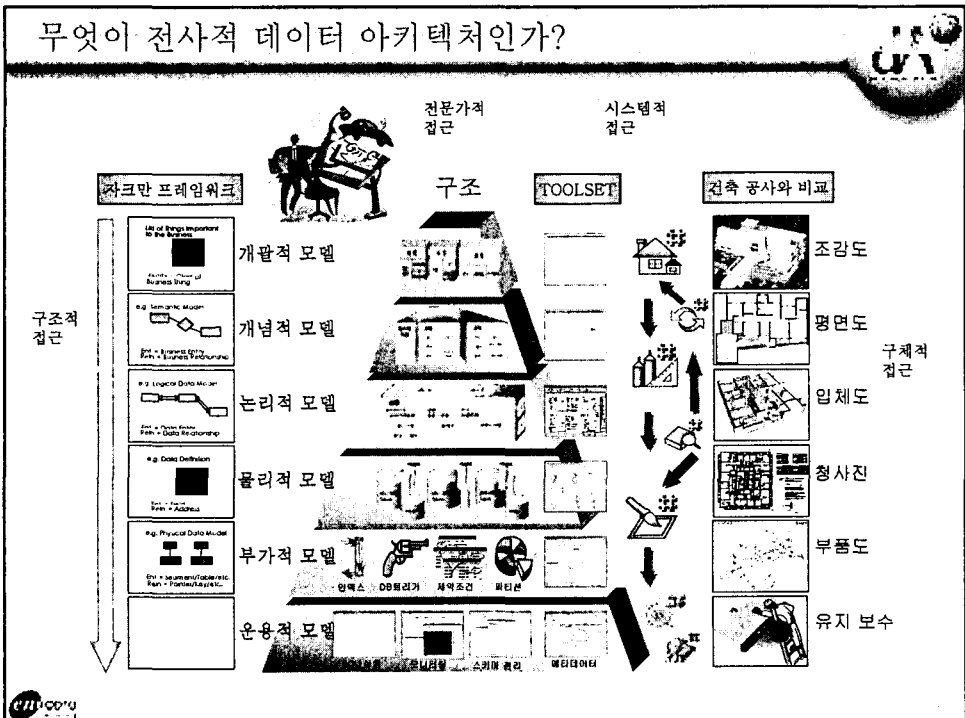
**ENGORE DATA ARCHITECTURE SOLUTION**

**전사적 데이터 아키텍처 구축을 위한  
리버스 데이터 모델링 방법론**







**(주)엔코이앤알농  
수석 컨설턴트 전미영**




## 현실의 데이터 관리 현황은 ?



문서만으로는 시스템 관리의 한계



지금 운영하는 데이터 모두 사용되는 것은 아니다



전체 데이터 간의 관계가 명확하지 않다



각 데이터의 사용 용도에 대한 문서가 따로 관리 된다

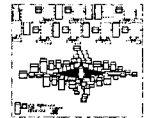
각 데이터의 사용 용도에 대한 문서가 따로 관리 된다



각 데이터의 사용 용도에 대한 문서가 따로 관리 된다



데이터를 호칭하는 용어의 표준화가 정립되지 않았다



ERD가 존재하지 않거나 변경사항이 반영되지 않았다



DA에 대한 시스템적 통제와 Repository 관리 필요




기업 내 데이터에 대한 전체적인 청사진을 보여주기 힘들다



DA에 대한 시스템적 통제와 Repository 관리 필요

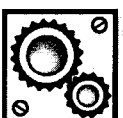


기업 내 데이터에 대한 전체적인 청사진을 보여주기 힘들다




DA에 대한 시스템적 통제와 Repository 관리 필요

## 왜 리버스를 해야 하는가 ?




**AS-IS System**



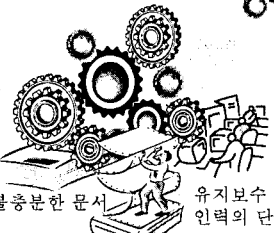
간헐적 요구사항

### 溫故知新

신사각 DA 구축의 디딤돌은 기존 시스템의 리버스로부터 시작!!




시스템 인력의 이적




불충분한 문서 유지보수 인력의 단순 경험

Data와 프로세스가 혼합된 프로그램

정보의 분산과 잦은 변경으로 현재의 정보시스템은 혼수 상태



그래도 AS-IS System은 그간의 정보가 녹아있는 정보의 보고(寶庫)



개산할 수 있는 정보의 발굴이 관건

녹아있다?  
- 숨어 있다.  
- 그러나 꺼낼 수 있다!

## 리버스를 하지 않는다면 ?



전사적 DA 구축을 위하여  
현재의 사용자에게 정보를  
얻기 위한 시도



그러므로 EDA 구축을 하려면  
리버스를 하는 것이 훨씬 유리



그러나 모든 현업 담당자를  
인터뷰 할 수 있을까?



그렇다고 대표되는  
현업 담당자가 모두 알고 있는가

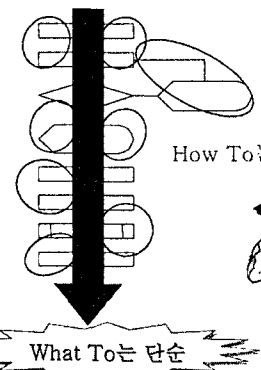


결국 알고자 하는 정보에 대한  
오류와 누락으로 구축하고자 하는  
TO-BE 프로젝트 갈팡질팡



## 전사적 데이터 아키텍처 구축의 시발점 ?

### 프로세스 중심의 리버스 ?

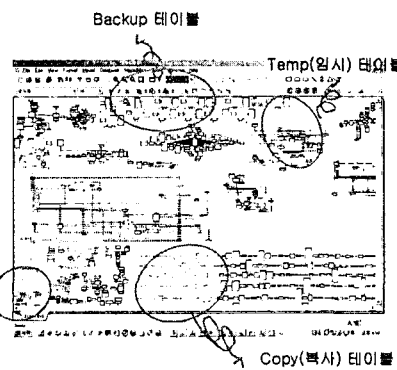


How To는 복잡

What To는 단순

복잡한 How To의 분석으로  
원래의 모습을 찾기 힘들다!

### 데이터 중심의 리버스 ?




Backup 테이블

Temp(임시) 테이블

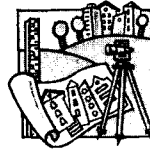
Copy(복사) 테이블

현재의 모습이 적절하지 못하더라도  
없어져야 할 것은 그리 많지 않다

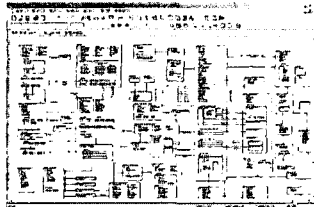
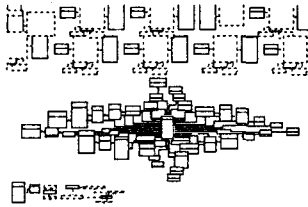


## 리버스 데이터 모델링이란 ?

- ▶ 이미 존재하는 것의 구체적이고 입체적인 설계도를 재현 하는 것
- ▶ 리버스 데이터 모델링을 거친 현행 논리적 모델은 목표 데이터 모델의 60~70% 이상이 이미 완성되어 있는 형태
- ▶ 단순한 1:1 전환이 아닌 현실적인 제한이나 고려할 사항들을 무시하고 오로지 본질적인 내용 중심으로 접근



	리버스 엔지니어링	리버스 데이터 모델링
엔티티	모든 데이터, 류 (백업, 복사, 임시, 통계, 요약, ...)	본질적으로 실제로 정의하여 관리할 대상
속성	모든 컬럼	정규화, 원자화된 속성으로 구조적인 속성 유형 정의
식별자	Primary Key Constraints (일련번호가 식별자인가 ?)	본질식별자(의미상 주어), 식별식별자 구분
관계	Foreign Key Constraints로 매우 제한적 설정	선정된 엔티티간의 비인적 1:1 참조 관계 설정



## 리버스 데이터 모델링 방법론 단계별 목표

### 정보수집 및 환경 설정

- 다양한 물리 테이블 정보의 빠른 수집
- 분산된 시스템 정보와 테이블 정보의 병합
- 대용량 물리 정보의 논리화 작업을 병렬로 수행하기 위한 분석 조직 구성

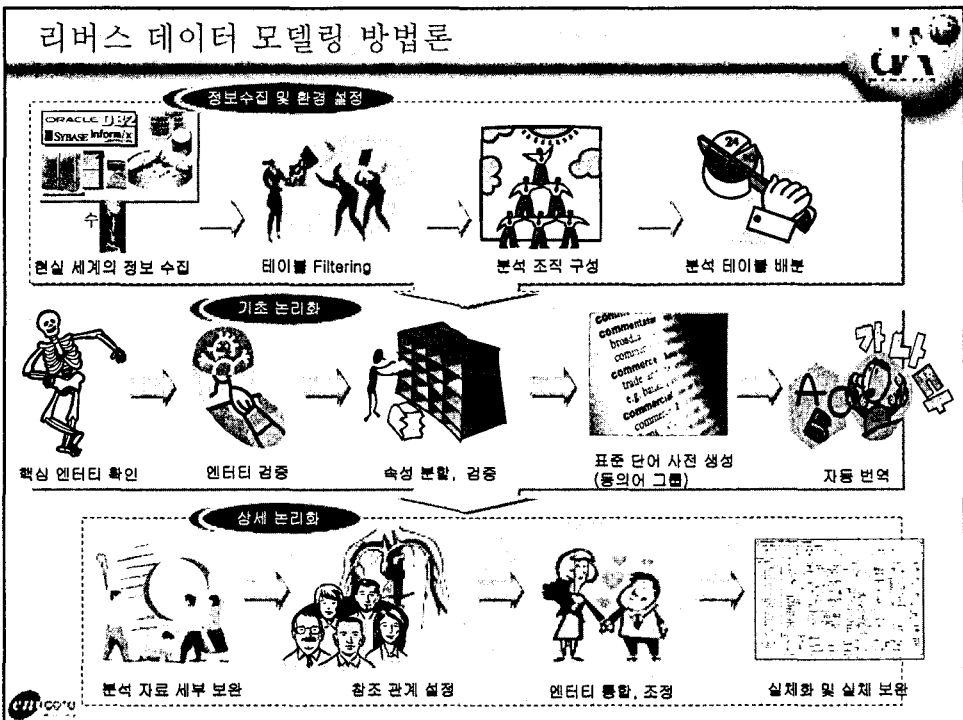
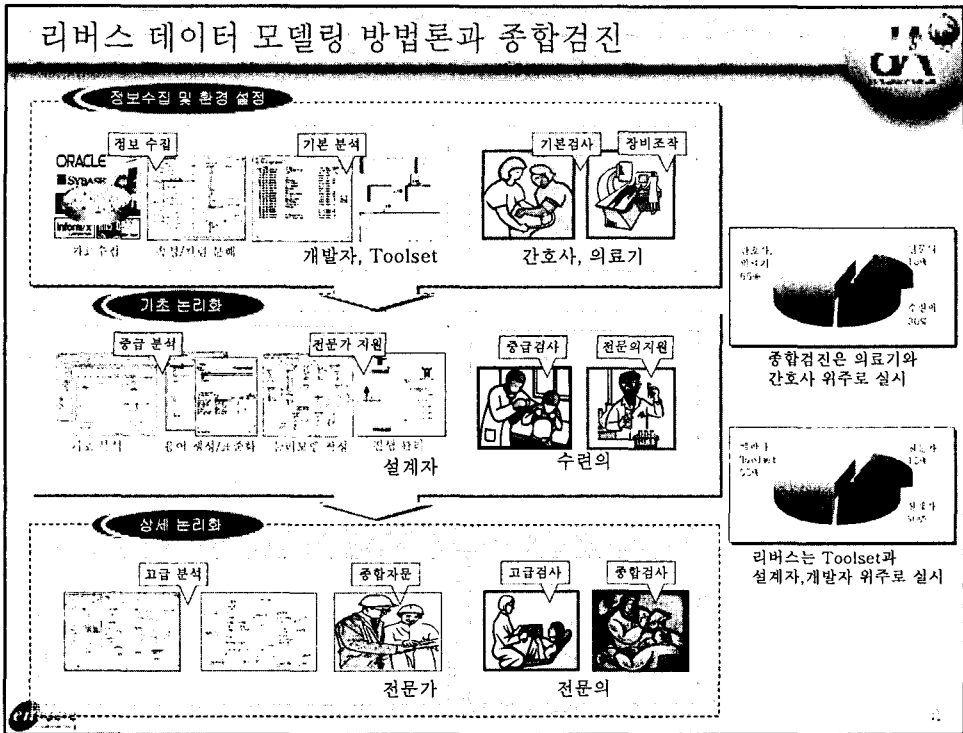
### 기초 논리화

- 수작업을 최소화하기 위한 1차 논리화 작업 시스템 구축
- 물리 테이블 정보에 실체명과 속성명 지정으로 1차 논리화 완성
- 현행 시스템 분석을 기반으로 한 전사 용어 표준화 구축

### 상세 논리화

- 참조관계 설정 및 실체, 속성 정보 상세 규명으로 현행 논리 모델 구축
- 완벽한 현행 논리화는 신규 목표 시스템의 70% 이상 제공 가능







### 동의어 그룹 생성

속성명, 컬럼명을 분할하면 그룹내의 단어 사용 양상 확인할 수 있다

같은 의미의 업무에 대하여 서로 다른 명칭을 사용하는 동의어 그룹 존재

DECLOAMNY (담보물권)    APPDATE (여신계정)    APLCODE (용자신청인)

여신

현재의 동의어 그룹으로 현재의 정보에 대한 다양한 검색 가능!

승인    결정

APVL    DEC    APP

새로운 업무 정의 시 용어 선택에 혼란

여러 의미로 사용되는 용어로 인한 산출물의 신뢰성 저하 유발

대출    여신

동의어 그룹의 생성은 리버스 단계 → 궁극적으로는 표준화를 목적으로 한다

### 검색의 다양성을 위한 단어 그룹 및 관련성 정의

사원에 대한 모든 정보를 검색하자!

사원이라는 용어로는 별다른 정보가 없네...

사원의 동의어 그룹으로는 꽤 정보가 많군

근로자    노동자    직업인    산업직군    노동자    worker

의사    의관    덕터    개업의    군의관    doctor    Dr

남자    사내    아들    사나이    남    소년    어숙어이    신사    man    male    boy

사람    인간    인류    인자    person    people    human

여성    숙녀    여인    소녀    딸    가사녀    개집어이    female    woman    girl

직업사관    차관    실무자    담당자    근무자    담당    사원

이취구성    할당    배정    임무    근무    assign    assignment

반의어    정규직    경력사원    정규사원    임시직    일용직    촉탁직    인부    노동자    Part timer

의미구성    임의    경영진    이사    집행부    총괄    director    executive    manager

유의어    경원자    대표이사    대표    이사    회장    대표    회장    회고경영자    CEO

남자의 반의어는 여자? 그렇다면 최초승인일이 있다면 최종승인일도 있을까?

사원의 유의어 그룹, 의미구성으로 확장하면서 정보 검색의 범위를 확장하자!

### 정보 자동 보완 (컬럼명을 속성명으로 자동 번역)

기초 논리화

● 분석할 대상의 가장 많은 부분은 속성

	테이블 수	컬럼 수
A 은행 - 여신 시스템	798 개	23,549
B 통신회사 - Billing 시스템	153 개	2,709
C 유류회사	437 개	13,094
D Package	581 개	10,487

물리적 컬럼의 논리적 속성명을 찾는 것이 1차적인 정보 보완

많은 테이블과 그보다 10배 이상 많은 컬럼에 대한 논리적 속성명의 부여가 가장 많은 일량을 차지

표준용어를 이용한 자동 번역

테이블	테이블명	컬럼수
<input type="checkbox"/>	TBL_TGBM020	13
<input type="checkbox"/>	TBL_TGBS10	9
<input type="checkbox"/>	TBL_TGBS22	14
<input type="checkbox"/>	TBL_TGBV21	11
<input type="checkbox"/>	TBL_TGBV50	11
<input type="checkbox"/>	TBL_TGBK101	11
<input type="checkbox"/>	TBL_TGBS217	14
<input type="checkbox"/>	TBL_TGBXC00	14
<input type="checkbox"/>	TBL_TGBX000	14
<input type="checkbox"/>	TBL_TGBNC01	14
<input type="checkbox"/>	TBL_TGBNC11	14
<input type="checkbox"/>	TBL_TJUS00	26
<input type="checkbox"/>	TBL_TMCC151	73
<input type="checkbox"/>	TBL_TMCKL40	26

Column	Name	DataType	Len	Not	PK
COL_ESTLADRC		varchar	6	N	T
COL_BRBTBRTGB		varchar	1	N	
COL_BRBTBRTYY		varchar	4	N	
COL_BRBTBRTMD		varchar	4	N	
COL_MLADRNY		varchar	40		
COL_MLADRCD		varchar	1	N	
COL_CURADRCD		varchar	5	N	
COL_CURETCAR		varchar	100	N	
COL_HOMADRCC		varchar	5	N	
COL_HOMETCAF		varchar	100	N	
COL_MBLTELNO		varchar	14		
COL_HOMTELCN		varchar	14		
COL_CURTELCN		varchar	14		
COL_WEDWEDVY		varchar	4	N	
COL_WEDWEDM		varchar	4	N	

Column	Name	DataType	Len	Not	PK
COL_ESTLUNGD	고객 번호	VARCHAR2	6	N	Y
COL_BRBTBRTGB	완납각도분	VARCHAR2	1	N	
COL_BRBTBRTYY	상년	VARCHAR2	4	N	
COL_BRBTBRTMD	상월일	VARCHAR2	4	N	
COL_MLADRNY	이메일주소	VARCHAR2	40	Y	
COL_MLADRCD	무연불수명지	VARCHAR2	1	N	
COL_CURADRCD	한겨주지적	VARCHAR2	5	N	
COL_CURETCAR	한겨주지 기	VARCHAR2	100	N	
COL_HOMADRCC	지역지적코드	VARCHAR2	5	N	
COL_REFBNDND	전 범인번호	VARCHAR2	13	Y	
COL_REWARDAY	상여	NUMBER	13	N	
COL_PAYPAYM	연봉	NUMBER	13	N	
COL_CUROPCPCD	원격연코드	VARCHAR2	4	N	
COL_HOMKNDCI	주거종류	VARCHAR2	2	N	
COL_DBLINCGB	양봉면적부	VARCHAR2	1	N	

APVL APY

DEC APP

APRV

**승인**

속성명을 자동번역으로 수행하면 전체 일량의 70~80%를 사람의 손을 가장 적게 빌리면서 자동으로 빠른 시간에 처리

### 정확한 참조관계의 설정 (1촌 참조관계 식별)

상세 논리화

ENTITY\_A  
#A

ENTITY\_B  
#B

ENTITY\_D  
#A  
#B  
#C

ENTITY\_C  
#A  
#B

ENTITY\_E  
#UID  
A  
B  
C  
D  
E

중조부    1촌    조부    장조부

정확하게 1촌 종속 관계를 찾아야만 진정한 리버스

**형제형**

아버지

할아버지

...

아들

- 할아버지, 아버지, 아들은 형제인가?
- 직접/간접 종속의 혼돈
- 직접, 절대 종속이 내 부모인 의미상의 주어

잘못된 참조 관계

**속보형**

경주이씨

전주이씨

.....

시중공파

주부공파

.....

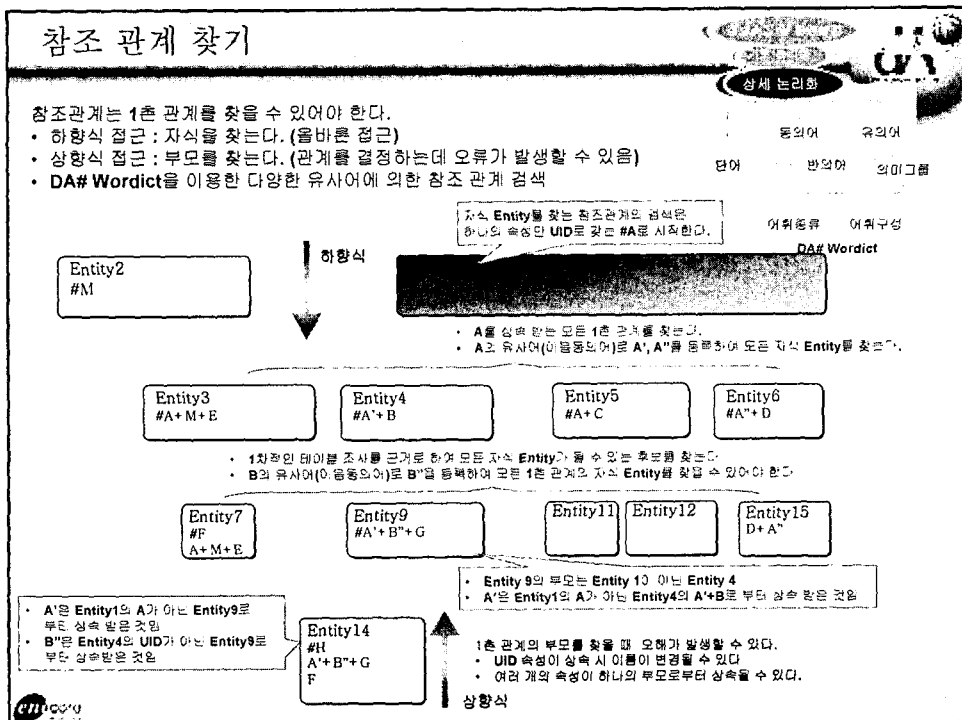
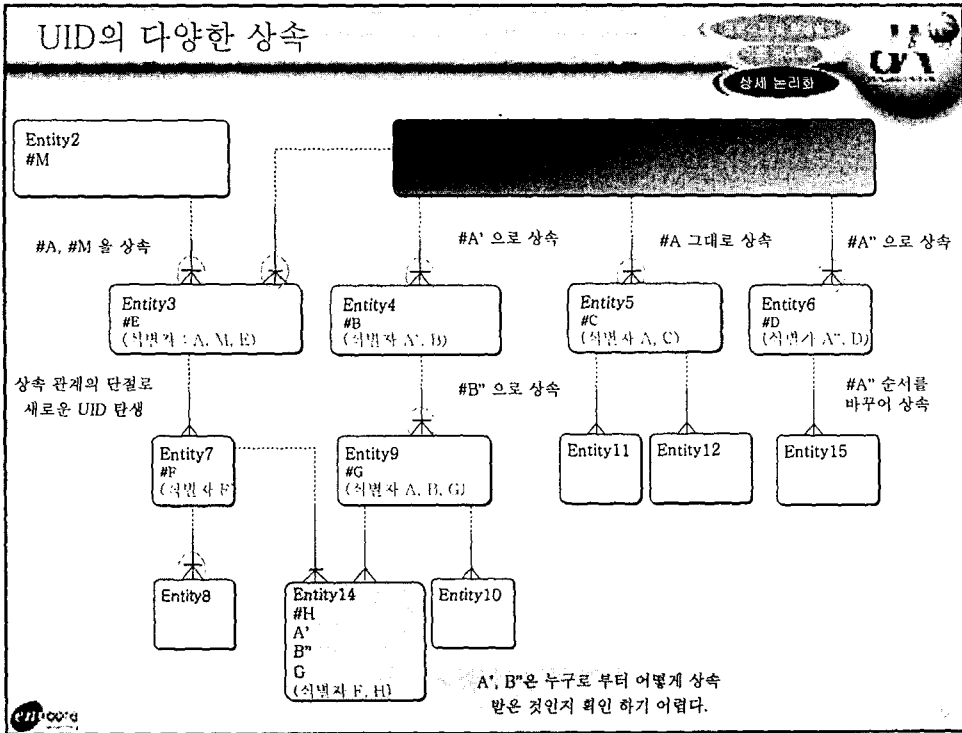
조부

부

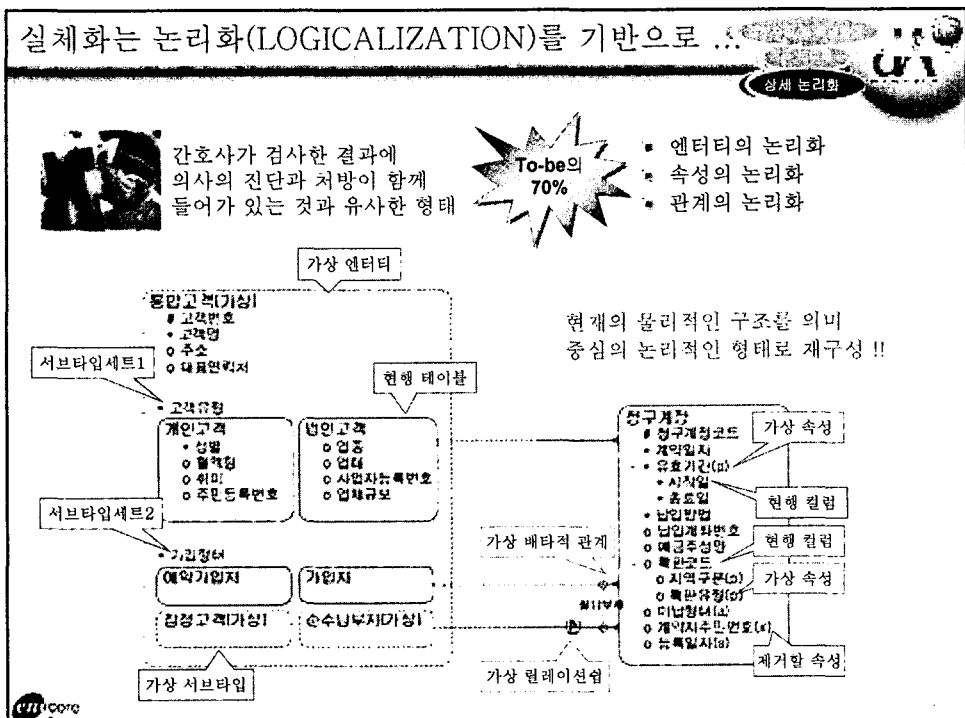
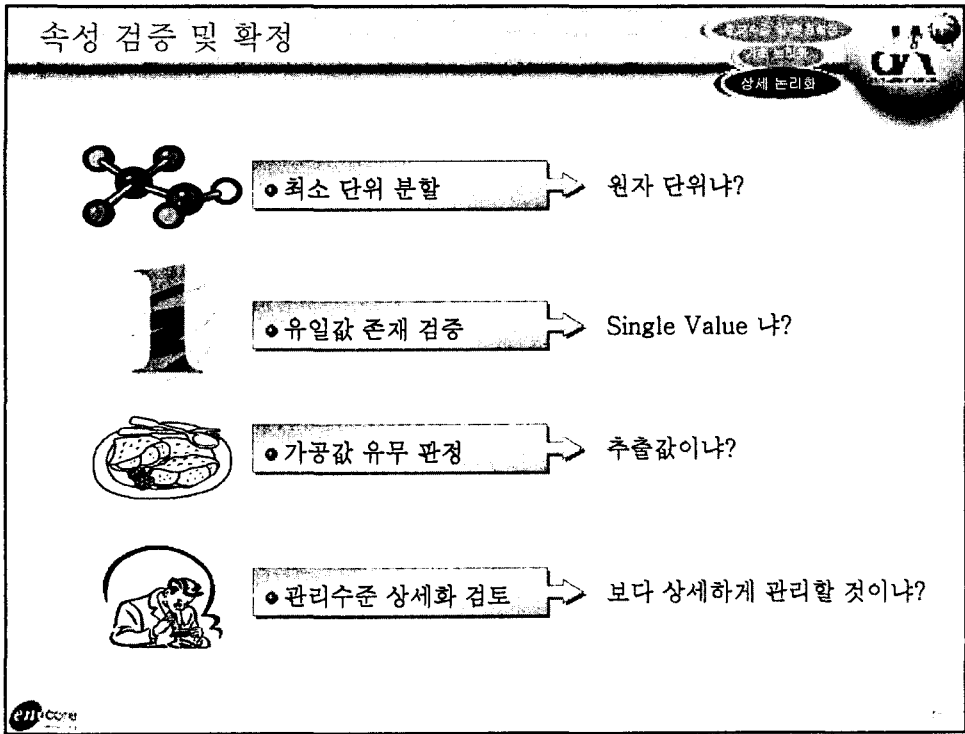
나

- 우리 몸 속에는 아버지 피만 흐르는가?
- 나는 정말 '이'씨의 순수 혈통인가?
- 모계는 아무런 영향을 미치지 않았는가?








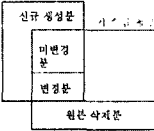


## 리버스 데이터 모델링을 지원하는 TOOL의 기능


▶ 다양한 기초정보 수집




▶ 정보변경에 대한 정합성 검증



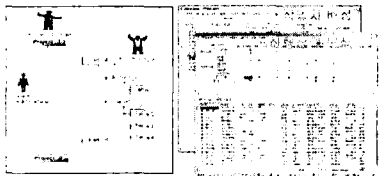
▶ 정보 자동 보완




▶ 고급 분석 및 고급 검색



▶ 종합적 진행관리



▶ 언어학적인 접근 방법 지원



## 엔코아 리버스 방법론 및 TOOL 적용 이점 (Benefit)

요구 기능	대응 기능	부가 설명
대량의 테이블 리버스	• Inverter Repository로 대상 테이블 정보 생성	• Data Dictionary나 Excel File로부터 테이블 정보 로딩
빠른 정보 분석	• N명의 분석자 지정 • 테이블 분배 • 병렬로 기초 정보 분석 입력	• 개발자, 유지보수인력, 현업직원과 공동으로 기초 정보 수집
건사 표준화 지원	• 용어사전과 연계 • 테이블 정보로부터 표준 용어사전 생성	• 이음동의 속성명이나 속성명이 없는 컬럼은 표준 속성명으로
논리화 (Logicalization)	• 참조관계 설정 • 서브타입을 이용한 통합	• 모델러는 논리화를 진행하면서 엔티티를 검증
리버스 진행상황 통제	• 분석상황 통제 • 이슈관리	• 관리자는 전체 진행상황을 확인 • 미결 사항 등을 통해 이슈관리
Data Migration 대응	• Old Table ↔ TO-BE Table	• Inverter → Modeler → Designer → Meta
편리한 사용	• 직관적인 사용법	• 다양한 수준의 검색 기능 지원