Corporate/Government Information Factory (CIF, GIF), EII (Enterprise Information Integration)

GPL3 Free Software

B a c k   R o o m / O f f i c e

F r o n t   R o o m / O f f i c e

Production Data ②

- zOS
- UNIX
  - OracleSolaris
  - AIX
  - Linux
  - Windows

XML files
MDM

SCM/ERP/CRM SAP, DW

DB2, Informix Oracle, Mysql SQL-Server FoxPro,Access PostgreSQL ODBC

VSAM, ISAM IDMS, IMS

Archived Data
Dump files
Archives

Internal (Private) Data
Excel, *.CSV Customer Profiles
POS

Unstructured Data
Shadow Systems
Sensors/Video cam Movie/Animation Music/Sound Photo/Pics Maps/GIS
Sniffing communications

Documents libraries
Text Files Transcripts PDF,PPT E-mail Click Streams

Machine Data
Internet web pages
PLC & Machine Sensors
Computer log files

External Data
Government Reports

Web Services (SOA)
Data Sources Operational database layer

---

④ Data Acquisition Management (Staging Area)
ETL (E-LT , CTF (Change Data Captuer, Transform and Flow))

Independent Data Marts

Extraction
Push Process (available data)
Mappings

Connecting to and access data
Reading source-data model
Scheduling the source system
Capturing changed data
Condensation data, Decoding
Data Conversion (Collation/Character Set Conversion)
Data Quality Assurance (Data quality firewall)
Staging the extracted data to disk

Feedback
Mappings

Transformation/Scrubbing/Cleansing/Integration/Summarization/Metadata Updates
Converting from one Hardware Architecture, OS and DBMS to another
Data type conversions, Assigning data types and default values
Applying any form of simple or complex data validation
Reformatting data to Standard/Common data format, Conversion of Units of Measurements
Make consistent units of measure
Encoding free-form values
Creating the metadata that describes the activities of conversion
Different keys same data, Different data same name, Same data different name
Match and Merge Data of different record type
Metadata updates, Referencing source data,
Derive/Calculate/Summarize/Aggregate multiple rows of data
Translating coded values (Semantic integration, Standard data format)
Generating surrogate key values , Insert key of dimension tables to fact tables
Transposing or pivoting
Makes up the "**single version of truth**", **Atomic Data**
Ensure that the data is presented in a **subject-oriented** fashion.
Data Migration (Local/Department View)(Conform and Deduplicate data).
Data Quality Assurance (Data quality firewall)

Profiling, Select, Edit, Combine, Splitting/Jointing , Conversion, Format Revision, Screening, Filtering, Merging, Sorting, Classify/Categorizing Data, Converting Keys, Scrubbing, Enrichment, DBMS Conversion, Cleansing, Soft/Hard Reject, Modeling, Deploying, Populating, Accessing, Analyzing, Presenting, Planing, Administering, Synchronization, Balance Control, Reset, Propagation, Resequencing , Restructuring data, Local/Global View, Data Mutation, Assigning, Realigning, Evaluation, Quality Control, Optimize, Index the Data, Adding Timestamp, Track the change, Summarization/Reduction, Refresh Data warehouse, Sanitization, Maintain/Update Metadata, Deliver/Load data, Visualization.

Enforcing column properties
Enforcing structure, Enforcing data and value rules
Enforcing complex business rules
Building the meta data foundation to describe data quality
Staging the cleaned data to disk
Conforming (Resolve conflict between systems/data)
Conforming business labels (in dimensions)
Conforming business metrics (in fact tables)
Merge/Combine Data of Different sources, Conform and Deduplicate data (**Globally**)
Householding, Internationalizing
Data Integration/Mediation (Global/Enterprise View)
Data Quality Assurance. (Data quality firewall)
Staging the conformed data to disk

Feedback
Mappings

Initial/Incremental(Ongoing) Loading/Refreshing
Data Refreshment
Load flat and snowflakes dimensions/subdimensions
Generate time dimensions
Loading slowly changing dimensions
Handling late-arriving dimensions and facts
Loading multi-valued dimensions
Loading ragged hierarchy dimensions
Loading text facts in dimensions
Running surrogate key pipeline for fact tables
Loading 3 fundamental fact table grains
Loading and updating aggregations
Condense/Compress data (profile or aggregate record)
Post Load procedures (Updating Indexes, ...)
Staging the delivered data to disk

Feedback

Management Monitoring Control

Data Conversion : Engine/System/Manager/Parent Change
Data Migration : World/Child Change
Data Integration **(CIF Level)** : Make a World — Without Boundary, (Enterprise Level)
Data Standardization : (Ministry Level)
Data Globalization (GIF Level) : Make a Universe Without Boundary (Government Level)
Data Modernization : (G2G Level)

Load
Data Firewall
Mappings

Denormal Tables

Independent Data Marts
MOLAP (Cube) Summarized/Aggregated
Department DW (Single Line Of Business) Denormal
Dimensional Data Store **DDS**

Data Warehousing area (a single, universal database)

I Approach
- **Subject-Oriented**
- **Integrated**
- **Non-Volatile**
- **Time-Variant**
- **Support of Management's Decision**

Single Version of the Truth, Atomic Data Model (Inmon Approach).
Business Process Modeling (Kimball Approach).
Long-Term Investment, Return on Investment (ROI).
Large Volume (Detailed/Granular, Summary/Aggregate, Historical) data
Data + Program + Metadata
Non Normalized. Atomic, Separate, Diverse Data (Many Sources)
Enterprise/Corporate Data. Projected and derived data.

1) Business analysis
2) Data source analysis
3) DW logical/physical design
4) ETL design
5) BI reports

Virtual Data Federation

Multi Level DW — Multi Level DW

Distributed DW — Distributed DW — Distributed DW

Time Coupler

Tape Offset Off Site
Offset & delete old unused data
Tape Archive
Aging

Near Online Storage
Staging Request
HSM/CMSM

META DATA
CWM MOF XMI

Flexible Denormalized Relational Data Model

$DW = \Sigma (Integrated\ InDependent\ Data\ Marts)$

Master Data
Loading Slowly Changing Dimensions
Dimensions
Conforming
Daily Fact Summry
Append

ROLAP (Star/SnowFlakes Schema)
Enterprise DW (Whole Line Of Business) **3NF** Normalize Data Store **NDS**

Mappings

MOLAP (Cube) Summarized/Aggregated
Dependent Data Marts
Department DW (Single Line Of Business) Denormal
Dimensional Data Store **DDS**

ETL

OLAP cube

VODS
ODS
Operational Data Store ODS
ODS T4
ODS

Subject-Oriented Integrated Volatile (Read-Write) Current-valued Detailed-only

Purge

ETL

Data Mining DW
Data Mining — Identify Potentially New Opportunities & Hidden Patterns

Data Exploration DW (Statistical Reports) Project-base DW

Pull Process (needed data)

Notification File

Detailed Profile — Click Streams

---

⑤ Business intelligence Service Management Data Delivery
Data (Departemental Level (Data Mart)(OLAP))

Dimension Table
High Performance Business Data Model Star Schema Snowflakes Schema
Dimension Table

Fact Table

Dimension Table
Dimension Table

Dimension Table — Sub Dimension Table
Bridge Table

Dim Date, Dim Resellers Sales Person, Dim Channels, Fact Sales, Fact Shipping, Fact Purchasing, Dim Order Type "Junk", Fact Sales, Dim PlaceLocation Region, Dim Warehousing, Dim Date/Time, Fact Payments, Dim PayType & Currency, Fact Marketing, Fact Delivery, Dim Product, Fact Orders, Dim Promotion

---

① Front Room/Office
Business Intelligence = Structured visualization
Business management
Change Management
Data (Individual Level)
Data Analysis
EIS/DSS
- Power-user Systems
- Informational systems

Power Users Casual Users
Information Navigators
BI /OLAP/Dashboards User Interfaces

Styles of BI
- Business Performance (Scorecards & Dashboards)
- Enterprise Reporting
- OLAP Analysis, Multidimensional Cube Analysis
- Advanced & Predictive Analysis
- Alert & Proactive Notifications
- Business analysis
- Clickstream analysis
- Information visualization
- Forecasting
- Data mining (Web, text, video, and voice)
- Trending analysis
- Query and reporting
- Geo-spatial analysis
- Enterprise portal implementation

Phases
1- Known of Business Requirements
2- Data Resource Analysis
3- Designing Conceptual & Logical Model
4- DW Schema Preparation (Physical model of Dimensions/Facts)
5- Making ETL mapping & process flows
6- Producing the BI Report

5-Customizable Analytical Applications
4-Exploration
3-Data Mining (Data-Driven Analysis)
2-OLAP (Model-Driven Analysis)
1-Queries & Reports

Data Acquisition Phase	Data Warehouse Phase	Data Delivery Phase	Data Consumption
Data Integration Engine (Data Acquisition Metadata Manager)	(Data Storage Metadata Manager)	OLAP/Data Mining Engine (Data Delivery Metadata Manager)	(Informational Access Metadata Layer)

Services Delivery

D a t a   A r c h i t e c t — **I n f o r m a t i o n   F e e d b a c k** — **D S S   A n a l y s t**

D a t a   v e l o c i t y   t h r o u g h   t h e   D a t a   W a r e h o u s e