

DTCC

2013中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2013

大数据 数据库架构与优化 数据治理与分析

SequeMedia
盛拓传媒

IT168.com

ITPUB

ChinaUnix

Database
BDaas
flowingdata
DB2
NoSQL MySQL
Oracle Big Data

BI/DW 的实施与价值

应用而生的金山快讯 平台

宋运奎@ 金山 西山

居北京运营中心

运营开发部

2013-4-18 北京



2013中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2013
大数据 数据库架构与优化 数据治理与分析

SequeMedia
盛拓传媒



主要议题

BI/DW概述

数据仓库实施

数据仓库价值

BI/DW概述：BI

是帮助企业提高决策能力和运营能力的概念、方法、过程以及软件的集合，其主要目标是将企业所掌握的信息转换成竞争优势，提高企业决策能力、决策效率、决策准确性。

BI/DW概述：DW 特性

1. 面向主题的
2. 集成的
3. 相对稳定的
4. 记录历史变化的

BI/DW: 主要支撑技术

数据库

ETL

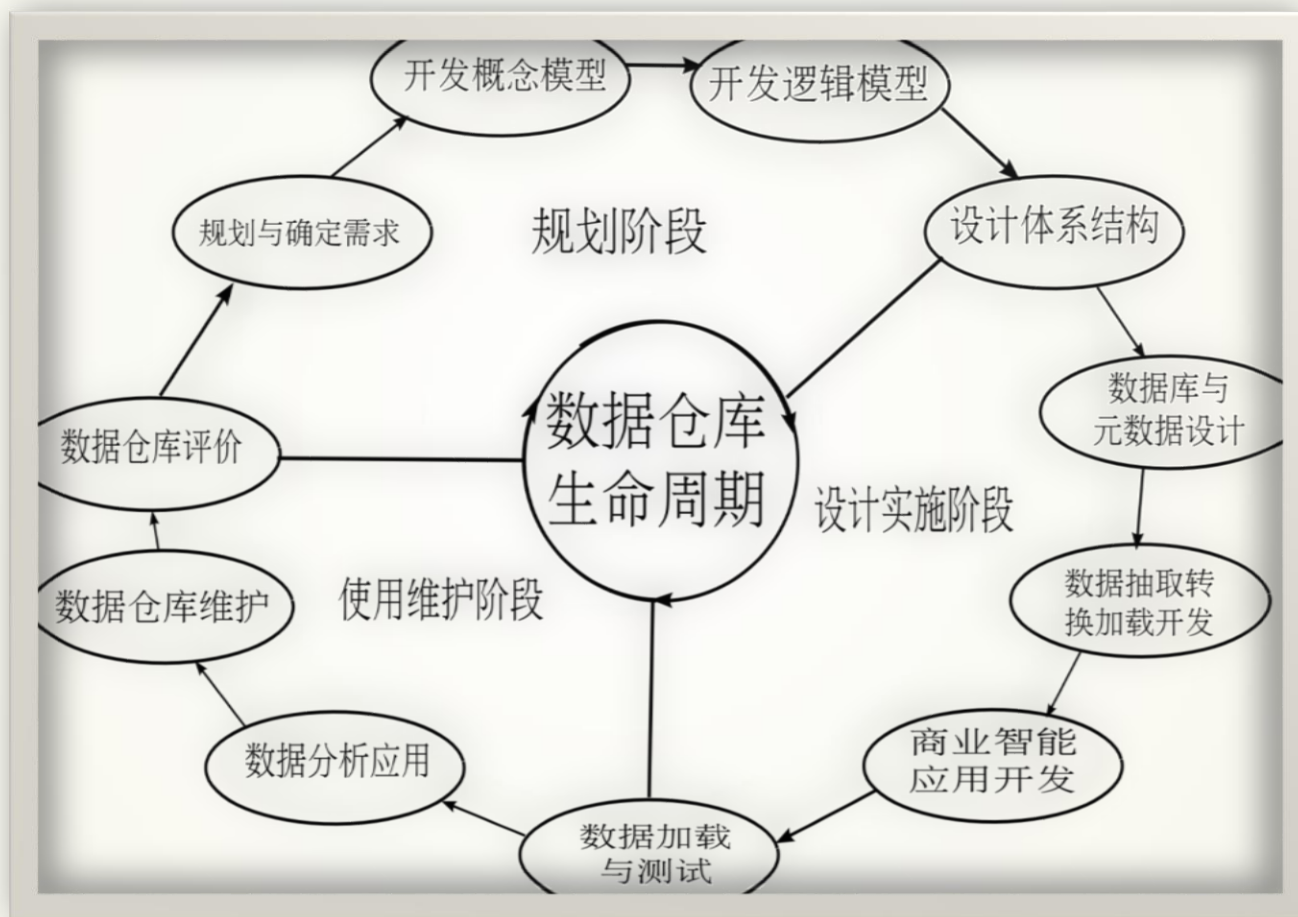
维度建模

在线分析

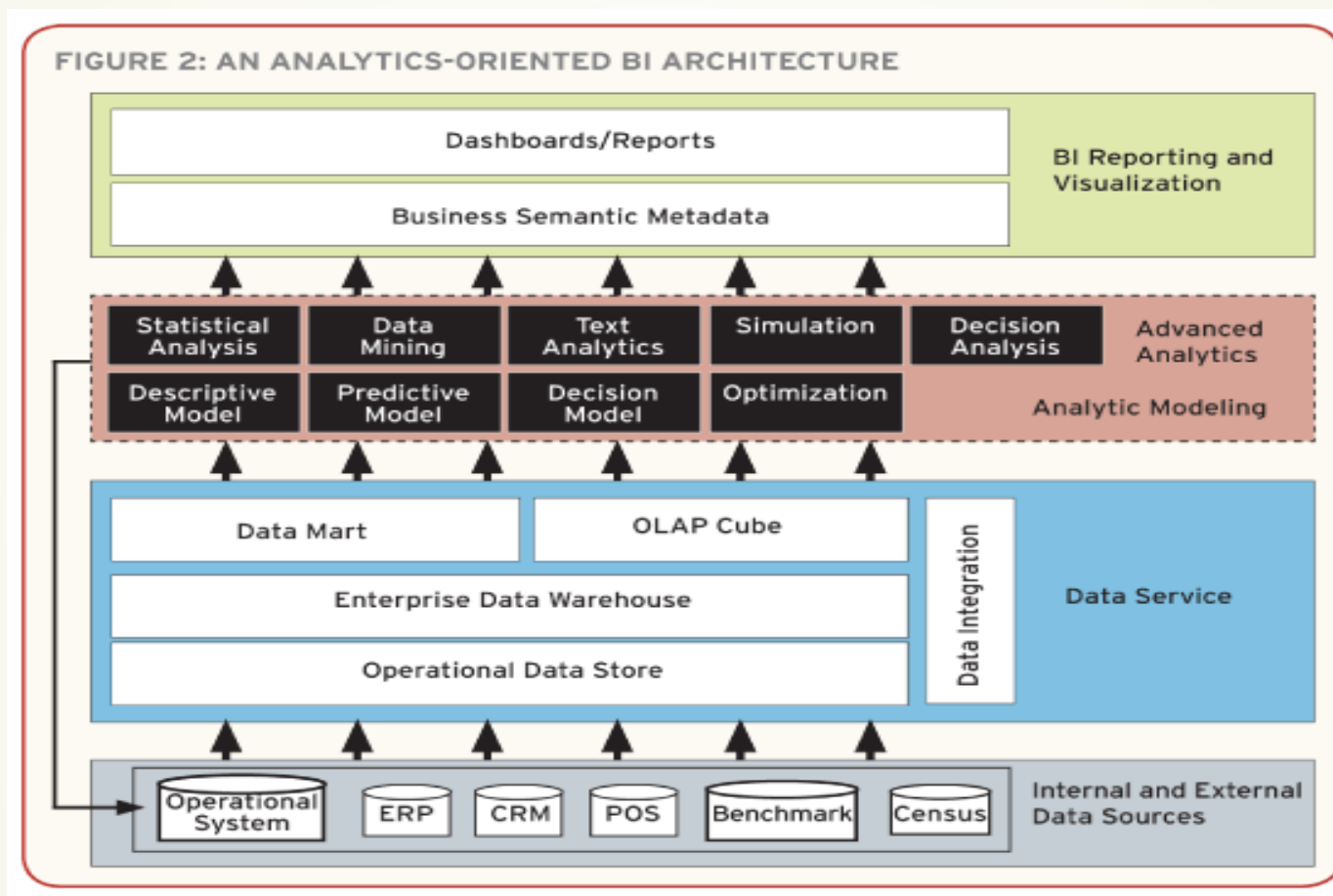
数据挖掘

可视化

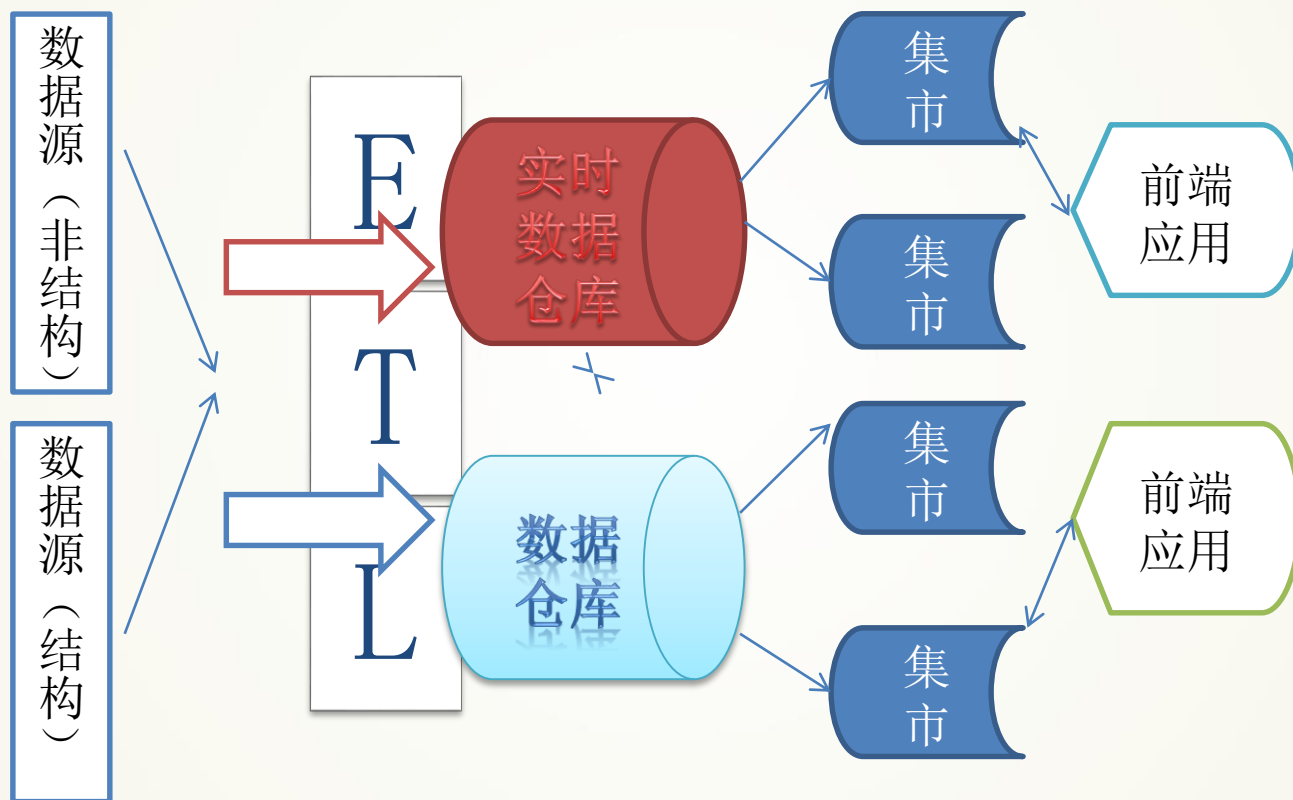
数据仓库实施：它的生命周期



数据仓库实施：架构（标准的、传统的）



数据仓库实施：架构（当前的、实时的）



数据仓库实施：抽象模型

概念模型（CDM）

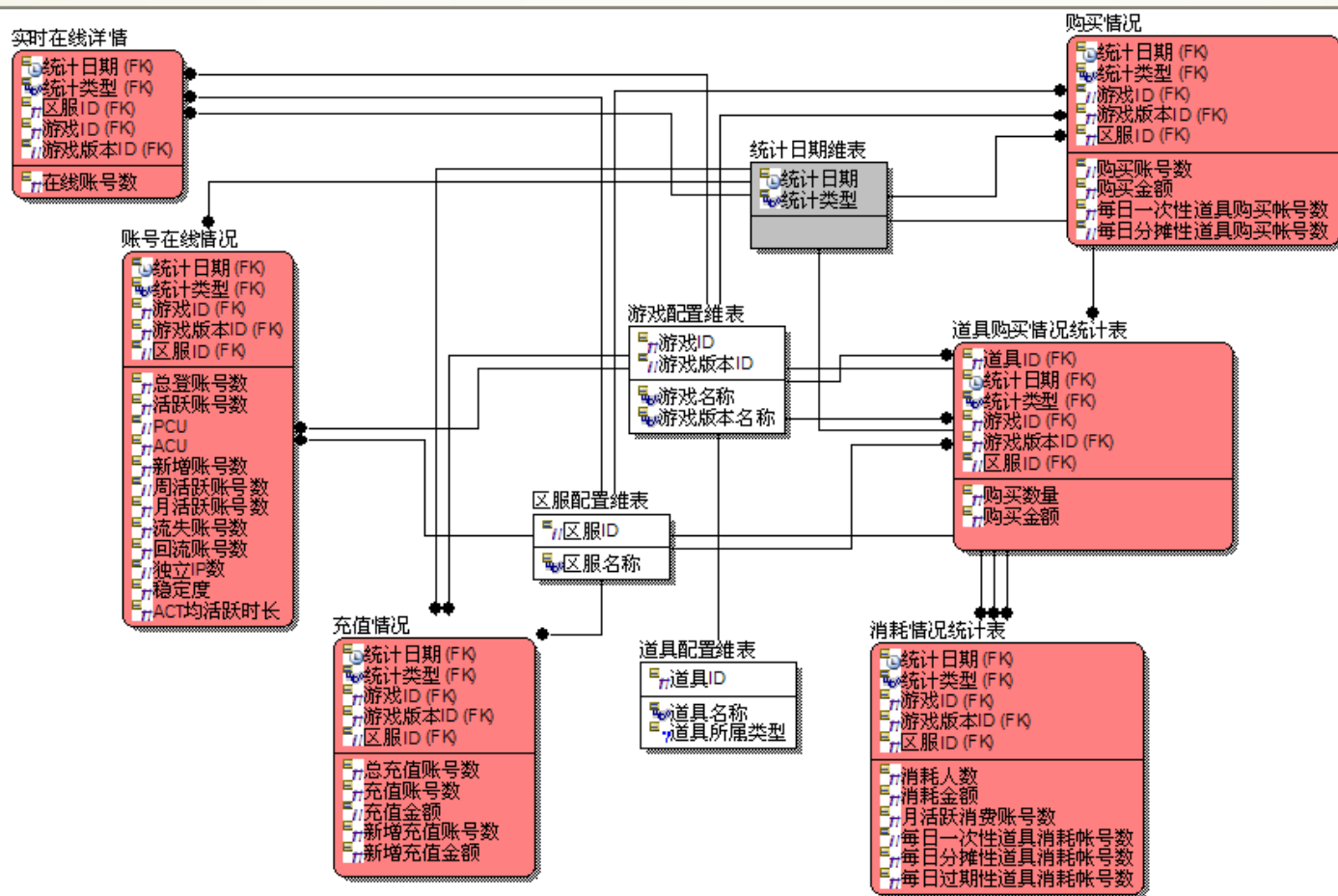
逻辑模型（LDM）

物理模型（PDM）

数据仓库实施：具体模型



数据仓库实施：（网游）维度模型



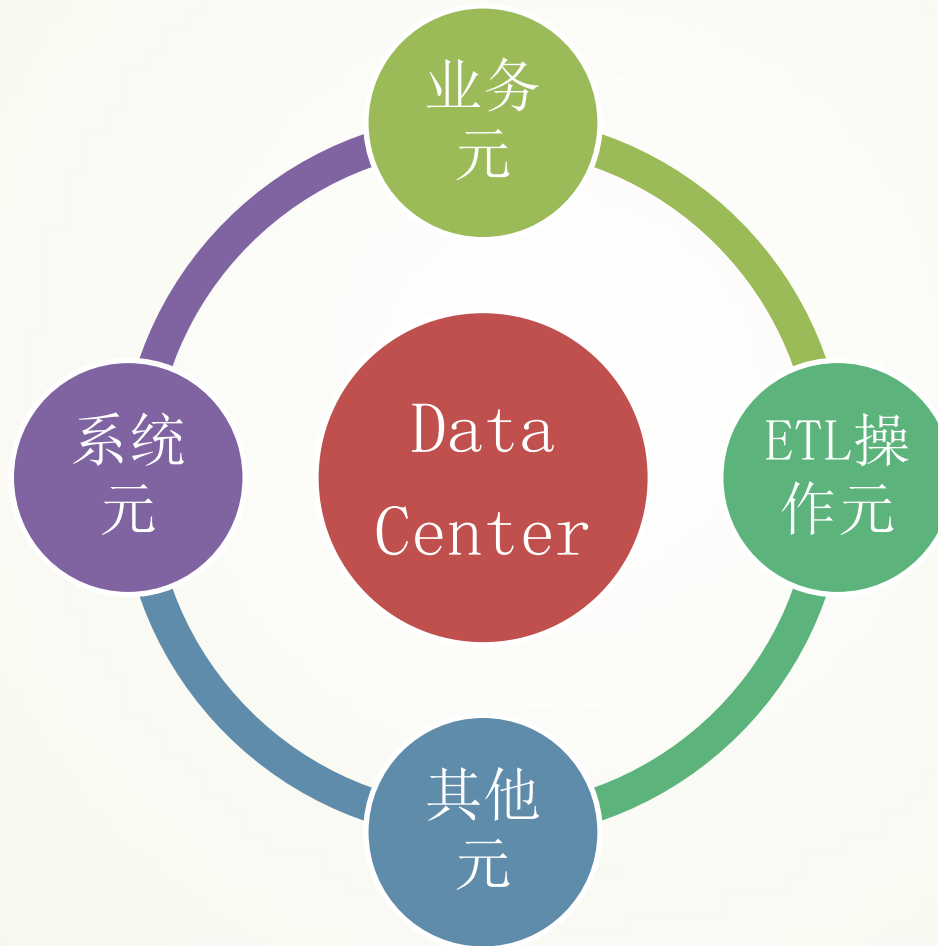
数据仓库实施：策略（传统）

总体策略	自顶向下	自底向上
总体结构	企业平台	数据集市
驱动类型	应用需求	数据过程
整合程度	全局数据	业务部门
复杂难度	相对复杂	较为简单
交付时间	较长等待	较快见效
实施成本	较大	较小
优点 缺点	投资少、见效快、易部署 不一致、不集成、重复多	统一性、规范化程度高 风险大、周期长、见效慢

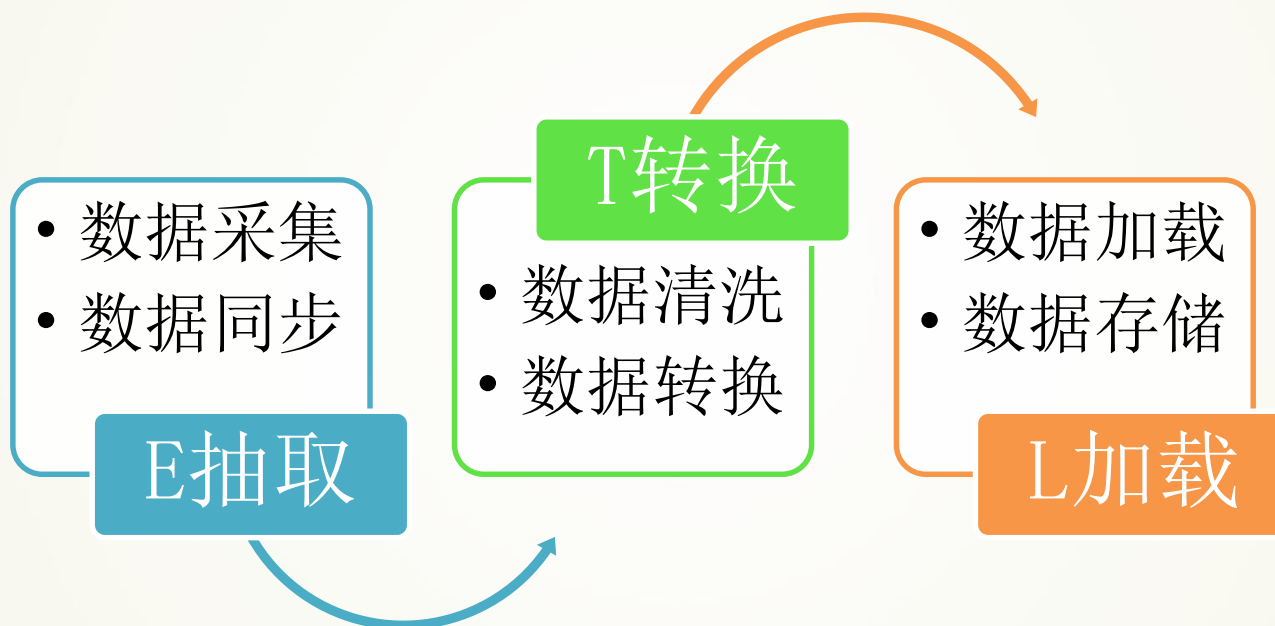
数据仓库实施：策略（目前互联网模式）



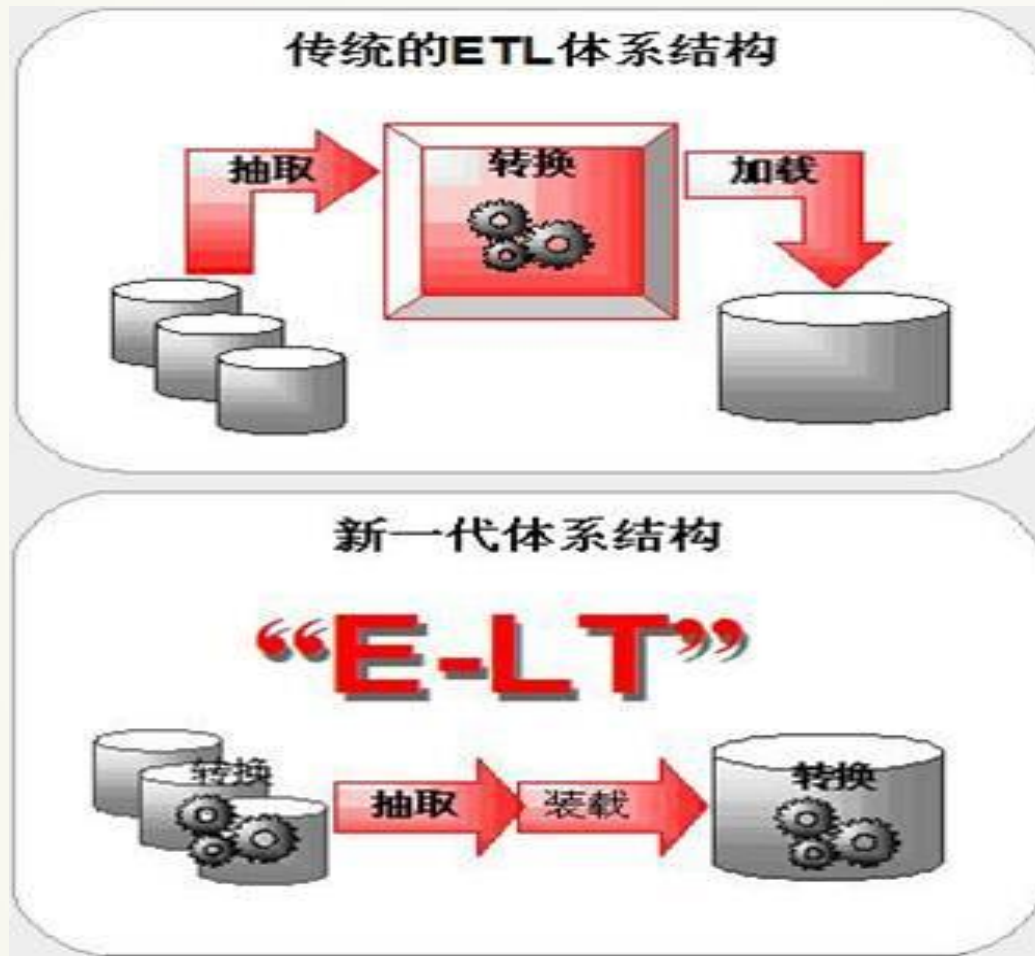
数据仓库实施：元数据管理



数据仓库实施：重中之重 ETL



数据仓库实施：ETL 体系结构



数据仓库实施：重中之重 ETL：重要性

引用业内一位资深架构师的说法：
ETL之于数据仓库有如循环系统于人一样重要，数据平台的健康，新陈代谢都是归功于ETL；

ETL任务的好坏直接决定数据仓库的数据质量，数据产出时间以及很大程度上是否能发挥数据的价值；

数据仓库实施：ETL 几种方法

第一种是借助专业的ETL工具实现；第二种是SQL编程方式实现；第三种是ETL工具和SQL相结合。前两种方法各有优缺点，借助工具可以快速的建立起ETL工程，屏蔽复杂的编码任务，提高速度，降低难度，但缺少灵活性。SQL编程的优点是灵活，提高ETL运行效率，但是编码复杂，对技术要求比较高。第三种综合了前两种的优点，极大的提高ETL的开发速度和效率。

数据仓库实施：ETL工具典型代表产品

Informatica的PowerCenter、
PowerExchange

IBM (Ascential) 的Datastage

Oracle的OWB、ODI

Microsoft SQLServer2000的DTS

Microsoft SQLServer2005的SSIS服务等

开源ETL工具

Pentaho Data Integration (Kettle)

数据仓库实施：ETL考虑点

增量处理

时间戳、自增主键、trigger, mv、数据比对、数据复制、数据库日志分析

性能

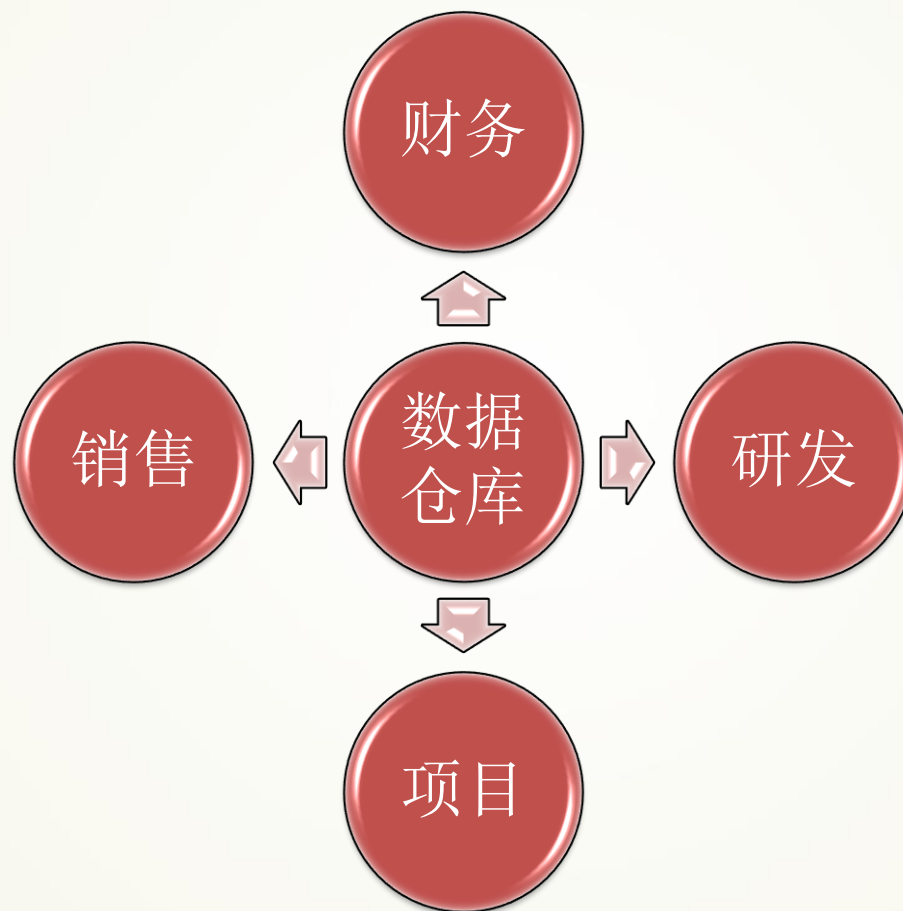
索引、分区、并行、集合操作、逆规范化
当传统方法无效时（拆分业务、优化流程）

数据质量保证

数据稽核模块：维度稽核（OLAP展现）、总量稽核、指标稽核

监控机制 多样的

数据仓库的价值



数据仓库的价值

数据仓库的真正价值体现在于基于其的数据应用上，如果没有有效的数据应用也就失去了构建数据仓库的意义。

正如：金山游戏快讯平台 itd.xoyo.com

The figure displays two mobile application interfaces side-by-side for comparison. The left interface is for 'Jinshan Game Express' (金山游戏快), showing a login screen with fields for email and password, a 'Remember Password' toggle, and a 'Login' button. Below the login section is a recommendation to use Safari/Chrome or Xiaomi browsers. A data chart is overlaid on the right side of the login screen, showing '充值' (Recharge) and '消耗' (Consumption) trends. The right interface is for 'Anti-Terrorism Action' (反恐行动), showing a 'Real-time Online' status and a 'Real-time Login' button. Below this is a detailed data chart showing '充值人数' (Recharge Users) and '消耗金额' (Consumption Amount) over time. The chart includes a legend for '今天' (Today), '昨天' (Yesterday), '上周同' (Same time last week), '消耗金额' (Consumption Amount), and '购买金额' (Purchase Amount). The right interface also features a '以表格查看' (View as Table) option.

谢谢大家！



2013中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2013
大数据 数据库架构与优化 数据治理与分析

SequeMedia
盛拓传媒



欢迎莅临

2013中国数据库技术大会

Database
BDaaS
flowingdata
DB2
NoSQL MySQL
Oracle Big Data