数据仓库架构

2017年4月17日

10:43

**一.基础架构**

immon广义的企业数据仓库

四层:

* 数据库层
* ETL层
* BI层
* 访问层

**二.架构图内容**

**(一)架构**

1.整体架构:两个思路,大平台,面向应用的;

2.技术架构(系统架构):每个模块采用什么技术

3.应用架构:面向分析

4.数据架构:ETL的流转,分层,比较复杂,所以就单独画

5.infrastructure架构: 网络图,主机图

**(二)AS-IS分析和TO-BE分析**

AS-IS: 指现在的做法,如果结合业务流程再造这个观念来看,就是再造之前的作业模式.

TO-BE: 指未来的做法(未来打算怎么做),是规划预计要达成的结果,也就是再造之后的作业模式.

**三.整体架构**

信息消费者

访问层

数据存储层:元数据等

**四.技术架构**

标出产品名称

架构图是用infosphere画图的

**五.应用架构**

**六.数据架构图-etl数据流图**

etl数据流图

数据实体图

Staging:数据登陆图

DW:数据仓库层-三范式3NF

DM:数据集市层-星型图-维度模型-start

小企业,没有DW层

数据理解正确, 架构不是一定的

**(一)注意**

小项目,尽量少分层,

因为数据实时要求比较高,分层,则etl流转浪费时间,而且etl流转,带来数据的血缘关系,影响关系

**(二)Staging层**

1.源系统业务繁忙不让访问-时间窗口短,比如1小时,允许抽取数据

2.上游系统不保存历史,一旦失败,则没有历史数据,BKBDL

ELT--分布式

派生的audit字段:源系统,源表,插入时间

**七.DW层**

各个公司,有自己的数据仓库概念

inmon的思路是,企业中心化,数据仓库

kimball按照星型架构构造数据仓库, 使用公用的维度表, 企业数据总线, 构建数据仓库

inmon的理论产生痛点,关系型数据库不兴起, 各个系统比较多, EIS蜘蛛网, 就有了做集中化,按照主题进行集成,稳定的历史的数据;所以有了,必须做企业集中数据仓库的需求,要整合各个系统的数据,依从数据的本质(3NF)

然后给其他部门的数据集市使用

缺点:周期太长, etl过程太久了

Kimball

Kimball问题: 财务部, 客户部, 集团的维度表设计不一致, 所有企业总线的成本比较高.

inmon的问题,做的模型,三年之后,数据模型变化了,就不稳定

Kimball的一致性维度:要数据稳定, 容易集成, 否者一致性维度, 就很难做.

Kimball主要解决,inmon的实际周期比较长, 面向敏捷, 面向报表的.

inmon提出折中:DW是企业信息工厂, DM用星型模型,但是数据抽取的周期太长,不适合企业发展太快

Kimball希望不用DW层.

DM-需求:star schema,偏向查询

DW-模型稳定性: 如果fdl表改了,则DM的报表都变动了,改动太大了. 改动的流程,必须特别严禁.

DM-模型方便查询数据: 否者报表查询,特别复杂,BI的sql特别多,则sql执行很慢

kimball很难在大企业实施, 因为一个部门太难了,需要很多部门一同参与创建, 所以DW层需要CIO主导

DW-访问:更多是接口,ETL访问,系统访问,访问频率低,但是数据量大.

DM-访问:高频次访问,40s之内出结果.

DW-审计:数据变化审计