DAMA-DMBOK

数据管理知识体系指南CDGA/CDGP认证

第11章 数据仓库和商务智能（完整课程视频请扫描二维码）

第11章 数仓和商务智能

01 引言

定义、业务驱动因素、目标和原则、基本概念

业务驱动因素

数据仓库建设的主要驱动力是运营支持职能、合规需求和商务智能活动。

目标和原则

目标：

1）支持商务智能活动
2）赋能商业分析和高效决策
3）基于数据洞察寻找创新方法

原则：

1）聚焦业务目标
2）以终为始
3）全局性的思考和设计，局部性的行动和建设
4）总结并持续优化，而不是一开始就这样做
5）提升透明度和自助服务
6）与数据仓库一起建立元数据
7）协同（考题）
8）不要千篇一律

1、商务智能

第一层含义：商务智能指的是一种理解组织诉求和寻找机会的数据分析活动。数据分析的结果用来提高组织决策的成功率。

第二层含义：商务智能指的是支持这类数据分析活动的技术集合

2、数据仓库

一个集成的决策支持数据库和与之相关的用于收集、清理、转换和存储来自各种操作和外部源数据的软件程序

广义上来说，数据仓库包括为任何支持商务智能目标的实现提供数据的数据存储或提取操作。

3、数据仓库建设

指的是数据仓库中数据的抽取、清洗、转换、控制、加载等操作过程。

建设流程的重点，是通过强制业务规则、维护适当的业务数据关系，在运营的数据上实现一个集成的、历史的业务环境。

4、数据仓库建设的方法

1）数据仓库存储的数据来自其他系统

2）存储行为包括以提升数据价值的方式整合数据

3）数据仓库便于数据被访问和分析使用

4）组织建设数据仓库，因为他们需要让授权的利益相关方访问到可靠的、集成的数据

5）数据仓库数据建设有很多目的，涵盖工作流支持、运营管理和预测分析

5、企业信息工厂CIF（Corporate Information Factory）

1）面向主题的　2）整合的　3）随时间变化的
4）稳定的　5）聚合数据和明细数据　6）历史的

CIF的组成部分包括：

1）应用程序
2）数据暂存区
3）集成和转换
4）操作型数据存储（ODS）
5）数据集市
6）操作型数据集市（OpDM）
7）数据仓库
8）运营报告
9）参考数据、主数据和外部数据

6、多维数据仓库

1）业务源系统
2）数据暂存区域
3）数据展示区域
4）数据访问工具

7、数据仓库架构组件

（1）源系统　　（2）数据集成ETL

（3）数据存储区域

1）暂存区

2）参考数据和主数据一致性维度：　　存储在单独的存储库中

3）中央数据仓库：完成转换和准备流程后，数据仓库中的数据通常会保留在中央或原子层中。该区域的数据结构是根据性能需求和使用模式来设计和开发的。

数据结构的设计元素包括：

①基于性能考虑而设计的业务主键和代理主键之间的关系

②创建索引和外键以支持维度表

③用于检测、维护和存储历史记录的变更数据捕获

4）操作型数据存储（ODS）：中央持久存储的一个解决方案，支持较低的延迟

5）数据集市：用于支持数据仓库环境的展示层，还用于呈现数据仓库的部门级或功能级子集，以便对历史信息进行集成报表、查询和分析。

6）数据立方体：　　存在三种经典的支持在线分析处理系统（OLAP）实现方法：基于关系数据库的、基于多维数据库的混合型存储结构的，它们的名称与底层数据库类型有关。

8、加载处理的方式

（1）历史数据

数据仓库的一个优势是它可以捕获所存储数据的详细历史记录。有多种不同的方法来捕捉这些详细信息，想要获取历史数据信息，组织应该根据需求进行针对性设计。

另一种方法称作DataVault，作为数据暂存处理的一部分，同样进行数据清洗和标准化，历史数据以规范化的原子结构存储，每个维度定义了代理键、主键、备用键。

（2）批量变更数据捕获

（3）准实时和实时数据加载

准实时的两个关键设计概念是变更隔离和批处理的替代方案

批处理的替代方案三种：

1）**涓流式加载（源端累积）**：是以更频繁的节奏或者以阈值的方式进行批量加载，允许白天就做批处理操作

2）**消息传送（总线累积）**

极小的数据报发布到消息总线时，实时或近实时的消息交互

3）**流式传送（目标端累积）**

用缓冲区或队列方式收集数据并按顺序处理。

02 活动

1、理解需求

构建一个数据仓库与开发一套业务系统不同。业务系统的开发取 决于精确的、具体的业务需求。

1、在收集数据仓库/商务智能项目的需求时，**首先，要考虑业务目标 和业务战略，确定业务领域并框定范围**；

2、**确定并对相关的业务 人员进行访谈**，了解他们想做些什么和这么做的原因，记录他们当下 关心的具体问题和想要询问的数据，以及他们如何区分和分类重要信 息。在可能的情况下，界定并书面记录关键的性能指标和计算口径。 这些信息可以揭示业务规则，为数据质量自动化奠定基础。

3、**将需求进行分类并排出优先级。**与生产上线相关的排在前面，将 与数据仓库相关的和那些可以等的排在后面。寻找并快速启动那些简 单且有价值的项目，以便在项目初始发布阶段就能获得产出。数据仓 库/商务智能项目需求描述应该包括业务领域及其范围内流程的完整业 务背景。

定义和维护数据仓库/商务智能架构

1、确定数据仓库/商务智能技术架构

最佳架构将提供一种能够以原子化的数据处理方式支撑交易级和运营级报表需求的机制，这种机制可以避免数据仓库存贮每一笔交易细节。

2、确定数据仓库/商务智能管理流程

标准的发布计划

有效的发布流程

3、开发数据仓库和数据集市

三条并存的构建轨迹

1）数据
2）技术
3）商务智能工具

1、将源映射到目标

源到目标的映射为从各个源系统到目标系统的实体和数据元素建立转换规则。

最困难的是确定多个系统中数据元素之间的链接有效性或等效性，考虑将多个计费或订单管理系统的数据合并到一个数据仓库中的工作，可能包含等效数据的表和字段用的不是相同的名字或结构。

2、修正和转换数据

强化数据修正或清理活动的执行标准，并纠正和增强各个数据元素的域值。

乐观加载策略：可以包括创建维度记录以容纳事实数据，这样的过程必须考虑如何更新和处理这些记录。

悲观加载策略：应该考虑一个事实数据的回收区域，并在以后重新加载。实际处理的时候应考虑首先加载回收区的记录在处理新内容。

4、加载数据仓库

确定数据加载方法考虑的关键因素是：

数据仓库和数据集市所需的延迟要求
源可用性
批处理窗口或上载间隔
目标数据库及时间帧的一致性
变更数据捕获的过程检测源系统中的数据变更

5、实施商务智能产品组合

1、根据需要给用户分组
2、将工具与用户要求相匹配

6、维护数据产品

1、发布管理

2、管理数据产品开发生命周期

3、监控和调优加载过程

性能瓶颈和性能的依赖路径

数据库调优技术、分区、备份调优和恢复策略调整、数据归档是一个难题

4、监控和调优商务智能活动和性能

最佳实践是定义和显示一组面向客户满意度的指标，如平均查询响应时间，每天、每周或每月的用户数就是有用的指标。

定期审查使用情况的统计数据和使用方法非常重要

透明度和可见性是推动数据仓库/商务智能的关键原则

03 工具和方法

1、元数据存储库

（1）数据字典和术语

（2）数据和数据模型的血缘关系

1）调查数据问题的根本原因

2）对系统变更或数据问题的影响分析

3）根据数据来源确定数据的可靠性

2、数据集成工具

（1）过程审计、控制、重启和调度

（2）在执行时有选择地提取数据元素并将其传递给下游系统进行审计的能力

（3）控制哪些操作可以执行或不能执行，并重新启动那些失败或中止的进程。

3、商务智能工具的类型

1、运营报表：指的是业务用户直接从交易系统、应用程序或数据仓库生成报表。

2、业务绩效管理（BPM）：绩效管理是一套集成的组织流程和应用程序，旨在优化业务战略的执行。应用程序包括预算、规划和财务合并。包括对组织目标一致性的指标的正式评估，此评估通常发生在高管层面。使用战略上午智能工具支持企业的长期目标。

3、运营分析应用（描述性的自助分析）

在线分析处理（OLAP）是一种多维分析查询提供快速性能的方法。OLAP和OLTP

传统的应用程序是财务分析，分析师希望反复遍历已知的层次结构来分析数据。

构建数据立方体以提供所需的功能需求，可能需要将较大的维度拆分为单独的数据立方体，以适应存储、加载或计算要求。

在数据立方体中配置基于角色的安全性或多语言文本，可能需要额外的维度、附加功能、计算或创建单独的数据立方体结构。

1）切片：多维数组的子集，对应不在子集中的维度的一个或多个成员的单个值

2）切块：切块操作是数据立方体上两个以上维度的切片

3）向下/向上钻取：在不同数据级别之间导航

4）向上卷积：需要先定义计算关系或公式

5）透视： 更改页面的展示维度

3、商务智能工具的类型

3、运营分析应用（描述性的自助分析）

三种经典的OLAP实现方法如下：

1）关系型联机分析处理（ROLAP）

通过在关系数据库（RDBMS）的二维表中使用多维技术来支持OLAP。常用星型架构

2）多维矩阵型联机分析处理（MOLAP）

MOLAP通过使用专门的多维数据库技术支持OLAP

3）混合型联机分析处理（HOLAP）

它是ROLAP和MOLAP的结合。允许部分数据以MOLAP形式存储，另一部分存储在ROLAP中

1、驱动需求的原型

对源数据的状态评估有助于对集成可行性和工作范围进行更准确的前期估算。

2、自助式商务智能

包括消息传递、警报、查看预定的生产报表、与分析报表交互、开发即席查询报表，当然还有仪表盘和计分卡功能。报表可以按标准计划推送到门户。

3、可查询的审计数据

04 实施指南

就绪评估/风险评估、组织和文化变革

1、就绪评估/风险评估

数据仓库应该能够实现以下几点：

1）明确数据敏感性和安全性约束
2）选择工具
3）保障资源安全
4）创建抽取过程以评估和接收源数据

2、版本路线图

建议将数据仓库总线矩阵作为一个沟通和推广的工具在逐步迭代的过程中使用。

3、配置管理

与发布路线图保持一致，并提供必要的后台调整和脚本，以自动化开发、测试和发布到生产，还通过数据库级别的发布来标记模型，并以自动化的方式将代码库与该标记联系起来，以便在整个环境中协调手动的编码、生成的程序和语义层的内容并进行版本控制。

4、组织与文化变革

1）业务倡议　　2）业务目标和范围　　3）业务资源
4）业务准备情况　　5）愿景一致

05 治理

决定事项和度量指标

1、业务接受度

业务对数据的接受程度，包括可以理解的数据、具有可验证的质量，以及具有可证明的数据血缘关系

1）概念数据模型　2）数据质量反馈循环

3）端到端元数据　4）端到端可验证数据血缘

2、客户/用户满意度

3、服务水平协议

4、报表策略

包括标准、流程、指南、最佳实践和程序，它将确保用户获得清晰、准确和及时的信息。策略必须解决如下问题：

1）安全访问

2）描述用户交互、报告、检查或查看其数据的访问机制

3）用户社区类型和使用它的适当工具

4）报表摘要、详细信息、例外情况以及频率、时间、分布和存储格式的本质

5）通过图形化输出发挥可视化功能的潜力

6）及时性和性能之间的权衡

度量指标

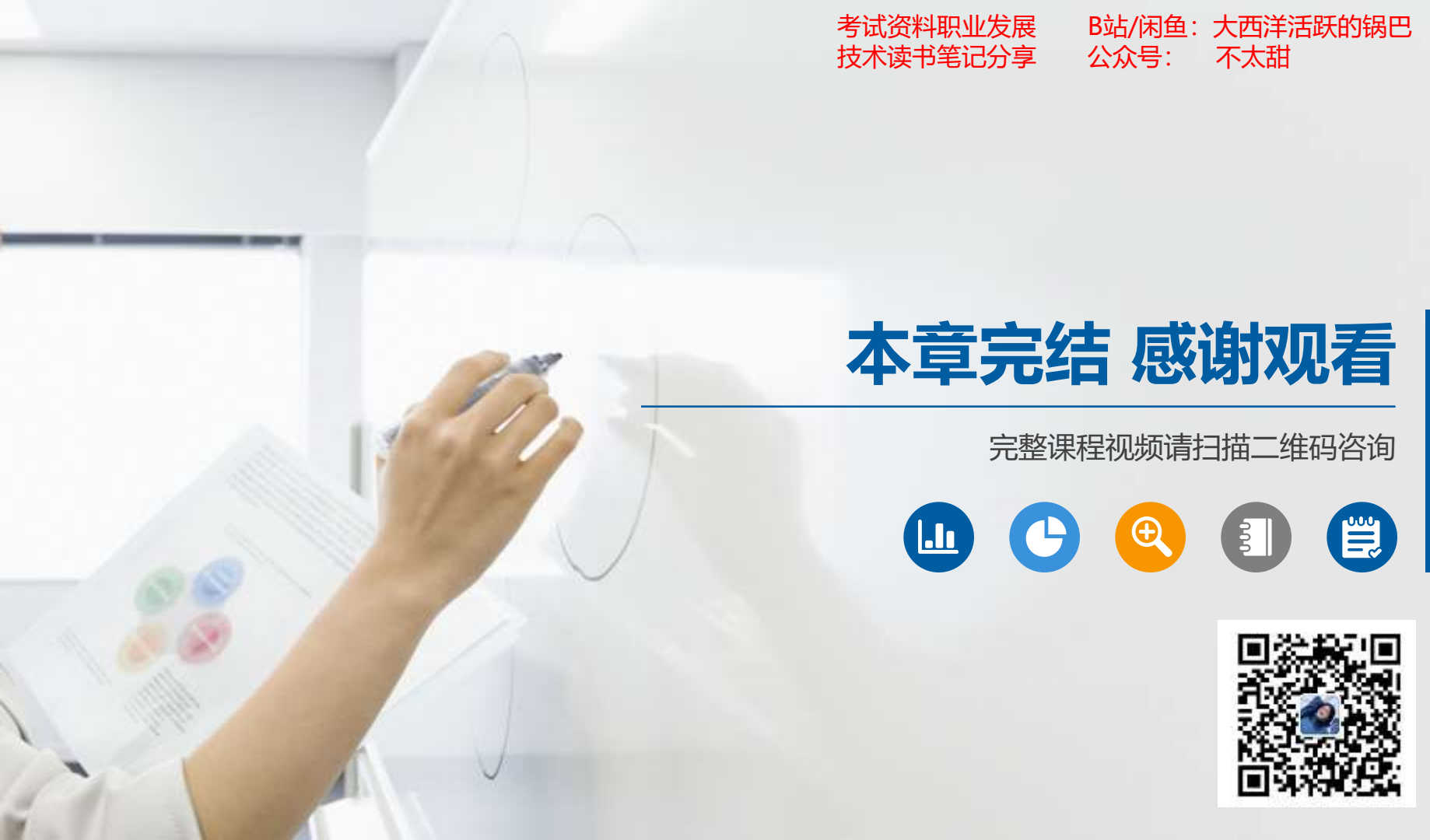
1、使用指标

包括注册用户数、连接用户数、并发用户数；审核用户、已生产的用户查询量和使用用户

2、主题域覆盖率

衡量每个部门访问仓库的程度（从数据拓扑的角度来看），还强调哪些数据四跨部门共享的，哪些还不是但也可能是共享的

3、响应时间和性能指标



本章完结 感谢观看

完整课程视频请扫描二维码咨询