

技术词汇

access (访问或存取)—— 在存储单元上查找、读或写数据的操作。

access method (访问方法或存取方法)—— 用于将物理记录从大容量存储设备传入或传出的技术。

access pattern (访问模式或存取模式)—— 访问数据结构的一般序列(例如, 从元组到元组, 从记录到记录, 从段到段等等)。

accuracy (精确度)—— 一种对避免误差的定性估计, 或对误差大小的定量度量, 表示为一个相对误差的函数。

ad hoc processing (特别处理)—— 仅执行一次, 偶尔访问, 并且用从未用过的参数操纵数据, 通常以启发式的迭代的方式进行。

after image (后映像)—— 当完成一个事务后, 放入日志的数据快照。

agent of change (变化动因)—— 大得不能抗拒的驱动力, 通常是系统的老化、技术的变化、需求的根本改变等等。

algorithm (算法)—— 组织好用以在有限步骤内解决问题的一系列语句。

analytical processing (分析型处理)—— 使用计算机为管理决策提供分析, 通常包括趋势分析、向下探查分析、统计分析及概要分析等等。

application (应用)—— 支持一个组织或企业需求的一组相互联系的算法和数据。

application database (应用数据库)—— 组织好用以支持一种特定应用的数据集合。

archival database (存档数据库)—— 包含具有历史特性的数据的数据集合。一般来说, 存档数据是不被更新的。每个存档数据单元都和一个过去的时间点有关。

artifact (人工关系)—— 在DSS环境中用于表示参照完整性的一种设计技术。

atomic (原子)—— (1)存储在数据仓库中的数据; (2)处理分析的最低层次。

atomic database (原子数据库)—— 由原始的原子数据组成的数据库; 一个数据仓库; 一个DSS基础数据库。

atomic-level data (原子层数据)—— 具有最低粒度级的数据。原子层数据存放在数据仓库中并随时间变化(即精确到过去某一时刻)。

attribute (属性)—— 可以对实体或关系取值的特性。实体可以被赋予多个属性(例如, 一个关系中的元组由多个值组成)。一些系统也同样容许关系具有属性。

audit trail (审计跟踪)—— 可用于跟踪活动的数据, 这些活动通常是更新活动。

backup (备份)—— 作为数据库备份活动的基础的一个文件。通常是以前某一时刻的数据库的快照。

batch (批处理)—— 多个程序(通常是长时间运行的、面向顺序的)互斥地访问数据的计算机运作方式, 当活动正在进行时不容许用户干预。

batch environment (批处理环境)—— 一种顺序控制的处理模式; 在批处理中, 收集并存储输入, 而后进行处理。一旦收集好, 批处理输入将顺序地对一个或多个数据库进行处理。

before image (前映像)—— 更新前的记录快照, 通常放在活动日志中。

bitmap (位映射)—— 用于显示一组块或记录存在与否的特殊索引形式。建立和维护位映射是非常昂贵的, 但它能提供快速比较和访问的手段。

blocking (组块)—— 合并两个或多个物理记录, 使它们的物理位置在一起。物理位置相连的结果是可以通过执行单个机器指令来访问和获取它们。

cache (高速缓存)—— 通常在设备级上建立和维护的缓冲区。从高速缓存中检索数据要比从磁盘柱面检索快得多。

cardinality(of a relation) (关系的基数)—— 在一个关系中元组(即行)的数目。

CASE —— 计算机辅助软件工程(computer-aided software engineering)。

checkpoint (检查点)—— 数据库的一个标识的快照, 或使事务对数据库冻结或停顿的一个点。

checkpoint/restart (检查点/重新启动)—— 在程序起始点以外的地方重新运行程序的一种方法, 例如, 当发生故障或中断时。在整个应用程序运行中, 可能会使用 N 个检验点。在每一个检查点要存储足够的信息, 以便程序能够恢复到检查点设定时的状态。

“CLDS” —— 以幽默方式对分析型的 DSS 系统的系统开发生命周期的命名。我们之所以这样命名是因为它实际上是经典的系统开发生命周期 SDLC(System Development Life Cycle)的逆序写法。

COBOL—— 面向商业的通用语言(Common Business Oriented Language), 商业界的一种计算机语言。一种非常通用的语言。

column (列)—— 从同一域中选择值的一个纵向表。一行可以由一个或多个列组成。

commonality of data (数据通用性)—— 出现在不同应用或系统中的相同或相似的数据。数据通用性的识别与管理是数据库概念设计和物理设计的基础之一。

compaction (压缩)—— 一种用于减少数据表示的位数而不丢失数据内容的技术。采用压缩技术, 使重复的数据可非常简明地表示。

condensation (紧缩)—— 减少管理数据量而不降低数据逻辑相容性的过程。紧缩与压缩有本质上的区别。

contention (争用)—— 当两个或多个程序试图同时访问同一个数据时出现的情况。

continuous time span data(连续时间跨度数据) —— 一种数据的组织形式, 使得在一个时间跨度上, 连续的数据定义由一个或多个记录表示。

CPU—— 中央处理机(central processing unit)。

CPU-bound (CPU-界限)—— 计算机不能产生更多输出的处理状态, 因为处理器的 CPU 部分被 100% 地使用。当计算机达到 CPU - 界限时, 存储处理部件通常没有被 100% 地使用。使用现代的 DBMS, 计算机更可能出现 I/O-界限, 而不是 CPU-界限。

current value data (当前值数据)—— 与随时间变化的数据相反, 在执行时准确度有效的数据。

DASD —— 参见“直接存取存储设备”。

data (数据)—— 在存储介质上记录的事实、概念或指令, 用于通信、检索, 用自动方式处理, 并表示为人可以理解的信息。

DA, data administrator (数据管理员(DA))—— 负责数据管理软件的说明、获取和维护,

以及负责文件或数据库的设计、验证和安全性的个人或组织。数据模型和数据字典通常是 DA 负责的。

database (数据库)——按照一种模式存储相互关联的数据集合(通常是受控的、限制冗余的)。一个数据库能够服务于一个或多个应用。

DBA, database administrator (数据库管理员(DBA))——负责每天监控和管理数据库的组织职能。DBA比DA职能与物理数据库设计的关系更密切。

database key (数据库键码)——数据库中每个记录存在的一个唯一值。这个值经常是被索引的,虽然它是可以作随机处理或散列的处理。

DBMS, database management system (数据库管理系统)——一个基于计算机软件系统,用于建立和管理数据。

data-driven development (数据 - 驱动开发)——以识别数据通用性为中心的开发方法,它是通过一个数据模型和建立比直接应用更广泛范围的程序来实现的。数据 - 驱动开发不同于传统的面向应用开发。

data element (数据元)——(1)实体的一个属性;(2)一个唯一命名和很好定义的数据类别,由数据项组成并包括在一个活动记录中。

DIS, data item set (数据项集合)——一组数据项,每个数据项直接与数据项所属数据组的键码有关。数据项设置可以在中间层模型中见到。

data model (数据模型)——(1)逻辑数据结构,包括由DBMS为有效进行数据库处理提供的操作和约束;(2)用于表示数据的系统(例如,ERD或关系型模型)。

data structure (数据结构)——数据元之间的逻辑关系,设计用来支持特定数据处理功能(树、列表和表)。

data warehouse (数据仓库)——用来支持DSS功能的、集成的、面向主题的数据库的集合,每个数据单元与某个时刻有关。数据仓库包括原子数据和轻度汇总数据。

DSS, decision support system (决策支持系统)——一个用于支持管理决策的系统。通常,DSS涉及以启发式方式进行的许多数据单元的分析。一般,DSS处理不涉及数据更新。

decompaction (解压缩)——压缩的反过程;一旦数据以压缩的形式存储,必须进行解压缩才能使用。

denormalization (反规范化)——将规范化数据放在物理地址中以优化系统性能的技术。

derived data (导出数据)——依赖于企业的一个主要主题的两个或多个取值的数据。

derived data element (导出数据元)——无须存储但需要时就能产生的数据元(年龄、当前日期、出生日期)。

design review (设计评审)——在编码之前对系统的所有方面进行公开评审的质量保证过程。

direct access (直接存取)——通过直接引用数据的卷地址来检索和存储数据的机制。这种机制与存取数据直接相关,因为它通常需要联机使用数据。也称为随机存取或散列存取。

DASD, direct access storage device (直接存取存储设备)——一种数据存储设备,其上的数据能够被直接存取,而不需要通过一个如磁带文件的顺序文件。磁盘是一种直接存取存储设备。

download (下载)——根据在一个数据库中发现的数据内容,向另一个数据库卸出数据的

过程。

drill-down analysis (向下探查分析)—— 一种分析形式，其中对总数的检查导致对总数的分量的探测。

dual database (双重数据库)—— 从决策支持数据中分离高性能的、面向事务的数据的实践。

dual database management systems (双重数据库管理系统)—— 用多个数据库管理系统来控制数据库环境不同方面的实践。

dumb terminal (哑终端)—— 在所有处理于远程计算机上进行的地方，用于与最终用户直接交互的装置。哑终端只作为搜集数据和显示数据的装置。

EIS, Executive Information Systems(高级管理人员信息系统) —— 为高级管理人员设计的系统，主要用于向下探索分析和趋势分析。

encoding (编码)—— 数据值物理表示的缩写或简写(例如，男 = “ M ”，女 = “ F ”)。

entity (实体)—— 在最高抽象层上数据模型设计人员所感兴趣的人物、地点或事物。

ERD, entity-relationship diagram (实体 - 关系图)—— 一种高级数据模型，以图式表示在集成范围内的所有实体以及实体之间的直接关系。

event (事件)—— 一个重要活动发生的信号。事件是由信息系统解释的。

external data (外部数据)—— (1)来源于企业操作型系统之外的数据；(2)驻留在中央处理系统之外的数据。

extract (抽取)—— 从一个环境中选择数据并将它传送到另一个环境的过程。

flat file (平面文件)—— 不包含数据聚集、嵌套的重复数据项或数据项组的记录集合。

foreign key (外键)—— 一种属性，它不是一个关系系统中的主键，但它的值是另一个关系的主键的值。

fourth-generation language (第四代语言)—— 设计用来允许最终用户随意访问数据的语言或技术。

functional decomposition (功能分解)—— 将操作划分成构成过程基础的层次功能(活动)。

granularity (粒度)—— 包含在数据单元中的细节级别。越细节的数据，粒度级越低。越综合的数据，粒度级越高。

heuristic (启发式)—— 下一步骤由当前步骤的分析结果所决定的分析模式。用于决策支持处理。

image copy (映像复制)—— 为达到备份的目的，将数据库从物理上拷贝到另一介质的过程。

index (索引)—— 当一个记录的索引键项已知时，维护的记录存储结构部分，用来对记录提供有效访问的。

information (信息)—— 人们为了求解问题或作出决策而吸收和评价的数据。

integrity (完整性)—— 确保数据库中包含的数据尽可能地准确和一致的数据库性质。

interactive (交互的)—— 将联机事务处理和批处理的一些特征结合起来的一种处理模式。在交互式处理中，最终用户与他或她单独控制的数据进行交互。另外，最终用户可以启动处理数据的后台进程。

“ is a type of ” (“ 定义类型 ”)—— 在概念数据库设计(例如，“ cocker spaniel ” 是一种

“dog”类型)过程中对数据进行抽象的一种分析工具。

iterative analysis (迭代分析)—— 下一步处理依赖于当前执行步骤所获得结果的处理模式；启发式处理。

JAD, joint application design (联合应用设计)—— 创建和细化应用系统需求的人(通常是最终用户)组成的一个组织。

judgment sample (判断样本)—— 一个数据样本，其中数据由于基于一个或多个参数的样本而被接受或拒绝。

key (键码)—— 用于识别或定位记录实例(或其他类似的数据组)的数据项或数据项的组合。

key, primary (主键)—— 在数据库中用于识别单个记录的唯一属性。

key, secondary (辅键)—— 在数据库中用于识别一类记录的非唯一属性。

living sample (活样本)—— 一个代表性的数据库，它通常代替大型数据库以用于启发式的、统计的、分析的处理。从超大型数据库定期地、有选择地提取数据，这样产生的活样本数据库代表超大型数据库在某一时刻的一个断面。

load (装载)—— 将数据值插入原先为空的数据库。

log (日志)—— 有关活动的日记。

logging (记日志)—— 自动记录与数据访问、数据更新等有关的数据。

loss of identity (特征丢失)—— 当数据是从外部数据源装入，以及外部数据源的特征丢失时，便发生特征丢失。常见的情况是用微处理器数据。

magnetic tape (磁带)—— (1)与顺序处理紧密相关的存储介质；(2)存储和检索图象的大容量磁性带子。

master file (主文件)—— 为给定的数据集(通常是由应用限定范围)保持记录系统的文件。

metadata (元数据)—— (1)关于数据的数据；(2)对数据的结构、内容、键码、索引等等的描述。

microprocessor (微处理机)—— 满足单个用户需要的小型处理机。

migration (迁移)—— 将经常使用的数据项移到比较容易访问的存储区域，以及将不常用的数据项移到比较难访问的区域的过程。

mips(million instructions per second, 每秒百万指令)—— 为小型计算机和大型计算机设定的处理器速度的标准度量单位。

online storage (联机存储器)—— 数据能以直接的方式访问的存储装置和存储介质。

operational data (操作型数据)—— 用于支持公司日常业务处理的数据。

operations (操作部门)—— 负责运行计算机的部门。

optical disk (光盘)—— 与磁性装置相对的一种使用激光的存储介质。光盘通常只写，单位字节的费用比磁性存储设备要便宜，并且可靠性高。

overflow (溢出，溢出区)—— (1)记录或数据段因为其地址已被占用而不能存储于其驻留地址的情况。在这种情况下数据被放在另外地方，这种地方称为溢出区；(2)DASD的区域，当溢出情况发生时，数据被送往该区域。

ownership (所有权)—— 更新操作型数据的责任。

page (页面)—— (1)在DASD之上的基本数据单元；(2)主存储器中的基本存储单元。

parameter (参数)—— 作为数据限定条件使用的基本数据值，通常用在数据搜索或模块控

制中。

partition (分割)—— 将数据分成不同物理单元的一种分割技术。分割可以在应用层或系统层上进行。

populate (载入)—— 在原先空的数据库中放置数据的具体值。

primary key (主键)—— 一个属性，其中包含的值唯一地确定具有该键的记录。

primitive data (原始数据)—— 数据的存在依赖于企业主要主题领域的唯一一次出现。

processor (处理机)—— 处于计算机程序执行的中心位置的硬件。一般来说，处理机分成三种：大型机、小型机和微机。

processor cycles (处理机周期)—— 驱动计算机(例如：启动I/O，执行逻辑运算、移动数据、算术运算)的硬件的内部周期。

production environment (生产环境)—— 运行操作型的、高性能的处理过程的环境。

punched cards (穿孔卡)—— 存储数据和输入的早期存储介质。现在穿孔卡已很少见了。

query language (查询语言)—— 能够让最终用户与DBMS直接交互的语言，用以检索并可能修改DBMS下存储的数据。

record (记录)—— 用数据与公共键码的关系组织的数据值的集合。

record-at-a-time processing(每次一个记录的处理方法) —— 一次一个记录、一次一个元组的数据存取方法。

recovery (恢复)—— 将数据库复原到初始位置或状态，经常是在物理介质发生较大的破坏之后进行的。

redundancy (冗余)—— 存储数据不止一次出现的设计。当数据可以更新时，冗余会带来严重的问题。当数据不被更新时，冗余经常是一种有价值的必要的设计技术。

referential integrity (参照完整性)—— 确保预定义关系有效性的一种DBMS功能。

reorganization (重组)—— 在差的组织状态下卸载数据，然后在好的组织状态下再装入这些数据的过程。在一些DBMS中，重组用于重新构造数据的结构。重组经常称为“reorg”或“卸载/再装入”过程。

repeating groups (重复组)—— 在一个给定记录取值中，能出现好几次的数据集合。

rolling summary (轮转综合)—— 一种档案数据存储形式，越新的数据具有细节越多的存储，越旧的数据具有细节越少的存储。

scope of integration (集成范围)—— 对建模系统边界的形式定义。

SDLC, system development life cycle (系统开发生命周期)—— 传统的操作型系统开发生命周期，通常包括收集需求、分析、设计、编程、调试、集成和实施。

sequential file (顺序文件)—— 其记录根据一个或多个键码字段的值进行排序的文件。这些记录能够以这个顺序从文件的第一个记录开始处理，直到文件的最后一个记录。

serial file (串行文件)—— 一种顺序文件，其中按顺序的记录在物理上是相邻的。

“set-at-a-time processing”(“成组处理”)—— 成组访问数据，组的每个成员满足一个选择的标准。

snapshot (快照)—— 数据库转储或在一些时候将数据存档在数据库以外。

solutions database (解决方案数据库)—— DSS环境的一个存储先前决策结果的部件。解决方案数据库用于在进行当前决策时帮助确定行动的适当过程。

storage hierarchy (存储器层次结构)—— 连接起来形成存储子系统的存储单元，其中一些单元较快但小而昂贵，其他一些单元较慢但大而便宜。

subject database (主题数据库)—— 围绕企业主要主题进行组织的数据库。传统的主题数据库是针对顾客、事务、生产、零件和供应商的。

system log (系统日志)—— 与系统事件有关的审计追踪的记载(例如事务登录、数据库变化等等)。

system of record (记录系统)—— 操作型数据的定义性的和单一的数据源。假如数据元 abc 在某个数据库记录中的值为 25，但在记录系统中的值为 45，按数据源定义第一个值是不正确的，并且必须改成一致的。记录系统对管理数据冗余是很有用的。

table (表)—— 由一组具有标题的列和一组行(即元组)组成的关系。

time stamping (标时戳)—— 将每条记录标上某个时刻的操作，通常是记录创建时间或者是记录从一个环境传递到另一个环境中的时间。

time variant data (随时间变化数据)—— 其准确度与一些时刻有关的数据。随时间变化的数据的三种形式是连续时间跨度数据、事件离散数据和周期性离散数据。请参看“当前值数据”。

transition data (转换数据)—— 既具有原始特征又具有导出特征的数据；通常对商业的运作非常敏感。典型的转换数据是银行的利率，保险公司的保险率，厂商 / 销售商的零售率等等。

trend analysis (趋势分析)—— 在时间序列上寻找同类数据的过程。请参看“EIS”。

true archival data (真实档案数据)—— 原子数据库中最低级的数据，通常存储在大容量存储介质上。

update (更新)—— 对存储在数据库中的全部或所选择出的项目、组或属性的值进行修改、增加、删除或替换。

user (用户)—— 发送命令和消息给信息系统的人或过程。