4-6-cn

1
本节主要讨论统一数据访问接口。
2
大数据计算系统可归纳为三部分:
数据存储系统，数据处理系统，数据应用系统
数据存储架构是大数据计算的基础。
在数据存储系统中，有 4 个部分来完成不同的任务，
即数据收集与建模、分布式文件系统、分布式数据库/数据仓库和统一数据访问接口。
我们学习了数据收集和建模、分布式文件系统、数据库，现在我们介绍一下统一数据访问接口
3
统一数据访问接口是在数据采集与建模、DFS 和 DB 之上，为上层数据处理系统提供数据。
UDAI 基于统一的数据接口，支持分布式环境下的跨平台异构数据
4
应用程序对数据库的访问及数据交换是分布式计算系统的一个重要问题，业界较早使用的提供数据库访问应用程序编程接口是 ODBC（open database connectivity），使用结构化查询语言 SQL 作为数据库访问语言。
ODBC 本质上是一组数据库访问 API，由一组函数调用组成，核心是 SQL 语句。一个基于 ODBC 对数据库的操作也不依赖任何 DBMS，用户直接将 SQL 语句传送给 ODBC 同时，ODBC 对数据库的操作也不依赖任何 DBMS，不直接与 DBMS 打交道，所有数据库操作由对应的 ODBC 驱动程序完成。
但是 ODBC，JDBC 这样的数据库连接编程接口，虽然能够支持应用程序对数据库的 SQL 访问，
但不能提供分布式计算环境中诸如事务管理，并发调度，缓冲管理，异构数据库转换与继承等复杂功能
这就引入了数据访问层 data access layer。
DAL 是数据库之上提供数据交换功能的一层软件，其功能主要是实现应用程序数据的持久化存储，即将数据写入数据库，另外从数据库中读取数据并传递给应用程序，实现数据交换。
具体提供，对数据库的 CRUD(create retrieve update and delete)基本操作的支持。事务管理，并发处理，异构数据转换。数据访问层的实现方式有很多种，
常见的有数据存取对象 DAO（data Access Object），基于 ORM（Object/relation Mapping)的实现、服务数据对象（service data object）服务中间件
当系统扩展到需要访问跨平台的异构数据库时，运行平台可能是 UNIX，Linux 或 windows 要访问的数据类型可以是表单、邮件、XML 文档，EJB 组件，Web 服务，图像，音频/视频文件或者其他非结构化数据，单一的 DAO 或 ORM 就很难支持这种跨平台异构数据库的访问。
而且大数据应用层的技术也是多样化和各种标准的，数据访问层 DAL 的设计需要兼容各种标准的技术和产品，这就引入了统一数据访问接口 Unified data access interface
5
统一数据访问接口的定义:
基于统一数据接口，支持分布式环境下的跨平台异构数据的统一访问
功能:

1. 统一数据显示、存储和管理
2. 分离访问接口和实现代码，底层数据库连接的更改不影响统一的数据访问接口
3. 屏蔽数据源的差异和数据库操作的细节，使应用层专注于数据应用程序
4. 提供统一的访问接口和统一的查询语言

6
统一数据访问接口应包括如下构件
1）统一数据访问接口/统一查询语言；

2） 数据模型/元数据/服务模型；
3） 数据转换引擎/数据服务引擎/数据源管理器；
4） 数据源包装器
这些都有助于将来自不同数据源的不同格式的数据转换为统一的数据服务，使数据处理系统能够统一检索所需的数据。
7
到目前为止，我们已经学习了数据存储系统的全部四个部分:数据采集/提取/转换/建模
分布式文件系统、非关系数据库(NoSQL)、统一数据库 Access 接口。这四部分共同完成了数据的获取、数据的清理和转换、数据的逻辑和物理存储以及数据访问的统一。
数据存储主要提供数据采集、清洗建模。大规模数据存储管理、数据操作（添加，删除，查询，更新。数据同步等）功能
由于大数据处理的多重数据源/数据异构性/非结构化数据，分布式计算环境等特点，大数据存储系统的设计比关系型数据库系统复杂。
目前的大数据存储架构主要由数据层，分布式文件系统/非关系型数据库（NoSQL）以及统一数据库读取界面组成， 有些设计还会再 NoSQL 数据库之上加一个提供数据挖掘和分析功能的数据仓库层
8
本节的学习就到这里，谢谢大家