BDA 4-6

1

Hello everyone, I am Haiying Che, from Institute of Data Science and knowledge Engineering

School of Computer Science, in Beijing Institute of Technology , from this session ,

we start to learn Data storing system, and in this session we discuss about **Unified Data Access Interface**.

2

The big data computing system can be summarized into three categories:

Data storing system, Data processing system, Data application system

The data storage architecture is the foundation of big data computing.

In data storing system , there are 4 parts to accomplish different tasks,

which are Data collection and modeling, **Distributed file system, Distributed database/data warehouse and Unified Data Access Interface.**

**we learned data collection and modeling, distributed file system, database and now we look at Unified Data Access Interface**

3、

Unified data access interface is on top of data acquisition and modeling, DFS and DB， It provide the data to upper level data processing system.

UDAI is Based on a unified data interface to support **Cross-platform** **heterogeneous** data in a distributed environment

4

Access to databases and data exchange by applications is an important issue in distributed computing systems. The industry's earliest application programming interface to provide database access is ODBC (open database connectivity), and the structured query language SQL is used as the database access language.

ODBC is essentially a set of database access APIs, composed of a set of function calls, and the core is SQL statements.

An ODBC-based operation on a database does not rely on any DBMS. The user directly transmits SQL statements to ODBC.

At the same time, ODBC does not rely on any DBMS for database operations, and does not directly deal with DBMS.

All database operations are completed by the corresponding ODBC driver. .

However, database connection programming interfaces such as ODBC and JDBC can support SQL access by applications to the database,

but they cannot provide complex functions such as transaction management, concurrent scheduling, buffer management, heterogeneous database conversion and inheritance in a distributed computing environment. This introduces the data access layer. DAL is a layer of software that provides data exchange functions on top of the database. Its function is mainly to realize the persistent storage of application data, that is, write data into the database, and read data from the database and pass it to the application to realize data exchange. Specifically, it provides support for

* the basic operations of CRUD (create retrieve update and delete) of the database.
* Transaction management,
* concurrent processing
* heterogeneous data conversion.

There are many ways to implement the data access layer, the common ones are data access object DAO (data access Object), implementation based on ORM (Object/relation Mapping), service data object (service data object) service middleware

When the system is extended to need to access cross-platform heterogeneous databases, the data type to be accessed may be UNIX, Linux or Windows, which can be forms, mails, XML documents, EJB components, Web services, images, audio/video files or For other unstructured data, it is difficult for a single DAO or ORM to support such cross-platform heterogeneous database access. And the technology of the big data application layer is also diversified and various standards. The design of the data access layer DAL needs to be compatible with various standard technologies and products, which introduces the Unified data access interface

5

**The Definition is:** Unified data access interface is Based on a unified data interface to support **Cross-platform** **heterogeneous** data in a distributed environment

**UDAI Functions include**

1. Unified data display, storage and management
2. Separation of access interface and implementation code, the change of the underlying database connection does not affect the unified data access interface
3. Shielding the **differences of data sources** and the details of database operations, making the application layer focus on data applications
4. Provide a **unified access interface and a unified query language**

6

UDAL -Unified data access layer

Should include the following components

1） Unified data access interface/unified query language;

2） Data model/metadata/service model;

3 ）Data conversion engine/data service engine/data source manager;

4 ）Data source wrapper

All these can help to transform different format data from different data sources to unified data services and enable the data processing system to unified retrieve the needed data.

7

Until now, we learned all the four parts of Data storage system which include Data Acquisition / Extraction / Transforming / Modeling layer, distributed file system, non-relational database (NoSQL), and unified database Access interface. All these 4 parts work together to finish the getting data, cleaning and transforming data, logical and physical storing data, and unifying the data access.

8

In this session, we learned Unified Data Access Interface.

thank you for your attention, if you have any question, feel free to contact me.