H2database源码分析

陶封邑 2020K8009937014

**功能分析与建模**

1.什么是H2database？

H2database是一个开源的、纯Java编写的嵌入式数据库引擎。它包含了JDBC的API，拥有ODBC的驱动程序，并且提供了一个Web控制平台来对数据库进行操作。得益于其纯Java编写的代码组成，H2database不受平台的限制，可以运行在Windows、Linux等不同的操作系统上。

H2database可以实现内存模式与硬盘模式两种数据库运行模式。内存模式中数据仅在内存中运行，每一个用例执行完随即还原到初始状态，这种方式适合用于测试。硬盘模式中数据被持久化的保存为单个文件，可以储存少量的结构化数据。

H2database拥有嵌入式和服务器两种使用模式。嵌入式指将H2database嵌入进Java项目中，与项目应用同时运行，这种方式的H2database性能最高。服务器模式指H2 server将H2database启动，应用通过与H2 server连接进而访问、使用数据库。

H2database最大的特点在于“轻量”上，它的所用功能与实现被包含在一个2MB左右的jar包中，可以十分轻易的嵌入需要数据库的项目中。“麻雀虽小，五脏俱全”，H2database也包含各种功能，如：数据加密、全文搜索等等。

2.H2database的主要模块与功能。

（1）API包：用于用户定义扩展的接口，例如触发器和用户定义的聚合函数。

（2）BNF包：BNF范式的解析器和相关工具，包含为BNF自动填补上下文的类。

（3）COMMAND包：SQL语句的解析器和它的基础类，包含实现DDL语法、DML语法的两个类以及实现查询的类。

（4）COMPRESS包：用于数据无损压缩。

（5）CONSTAINT包：用于数据库约束，如检查约束、惟一约束和引用约束。

（6）ENGINE包：包含数据库的高级类和不适合放在另一个子包中的类，如用户以及用户相关的设置、远程链接的工具、管理员。

（7）EXPRESSION包：数学式、简单值的表达式，包含实现聚集函数、数据分析操作和窗口函数、状态表达式、其他函数（包含表值函数）的类。

（8）FULLTEXT包：实现原生全文搜索与Lucene全文搜索。

（9）INDEX包：实现各种表索引，以及用于索引中导航的标签。

（10）JDBC包：实现JDBC API，包含实现JDBC数据库的元数据API的类。

（11）JDBCX包：实现拓展的JDBC API，如JDBC的连接池。

（12）JMX包：实现JAVA的管理拓展。

（13）MESSAGE包：实现跟踪(日志记录工具)功能、与错误消息相关的工具。

（14）MODE包：实现与其他数据库(例如MySQL)兼容的实用程序。

（15）MVSTORE包：实现树形持久储存，包含用于cache、h2database中mvstore的相关助手、R-树、事务处理事件中mvstore的相关助手、数据类型和序列化/反序列化的类。

（16）RESULT包：实现行与内部结果集，如远程结果集、抽象提取的结果集。

（17）SCHEMA包：实现模式和存储在模式中的对象，如表示用户定义的函数或别名、序列和常量。

（18）SECURITY包：实现数据安全与加密（例如加密安全哈希算法），包含身份验证相关的类。

（19）SERVER包：实现PostgreSQL、H2 Console、TCP客户端。

（20）STORE包：实现存储抽象、抽象文件系统，例如带有缓存的文件、将值转换为字节数组，包含用于将文件储存到硬盘（并通过JAVA语句访问）的系统、抽象的加密文件系统、将文件完全保存在内存（非JAVA堆）的系统（可选择压缩）、记录所有写操作并能重复调用这些操作的文件系统、当文件被关闭重新打开并重新尝试操作的文件系统、将文件分割成多个小文件的文件系统、基于ZIP的文件系统的类。

（21）TABLE包：实现表和相关表元数据，例如JBDC相关的链接表、列解析器。

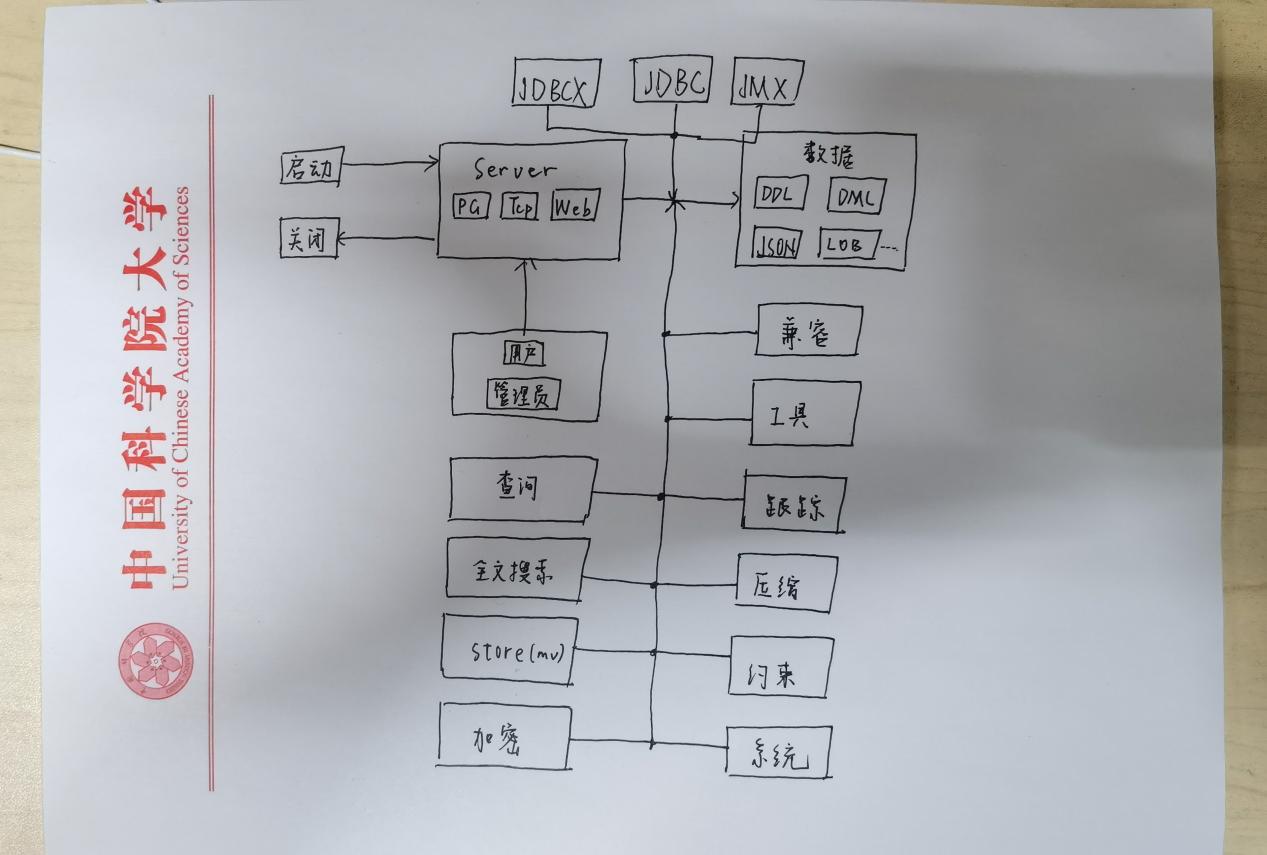
（22）TOOLS包：实现各种工具，例如CSV工具、shell工具。

（23）UTIL包：实现内部实用工具，例如计算MVTable的关键词、值和页占用的内存量的工具，包含用于解析处理GEOMETRY、JSON 数据的类。

（24）VALUE包：实现数据类型和数据值，例如数据转换和比较的方法，包含用于LOB数据值的类。

（25）DRIVER包：实现一个JDBC驱动。

3.H2database的主要流程。



4.着重分析的模块。

接下来将主要对查询功能进行分析。

**核心流程设计分析**

H2database作为一个数据库，它应该包含组织、存储和管理数据的基本功能。这一节我们将从管理数据入手，分析H2database的核心功能“查询”的具体实现流程。

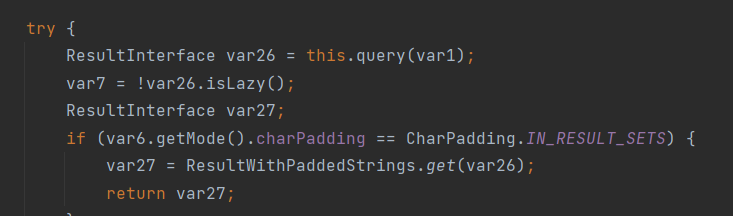
在此之前首先介绍用户与数据库交流的工具：结构化查询语言SQL（STRUCTURED QUERY LANGUAGE）。SQL一种是用于访问和处理数据库的ANSI的标准的计算机语言，它在1986年10月由美国国家标准局（ANSI）通过的数据库语言美国标准，接着，国际标准化组织（ISO）颁布了SQL正式国际标准。可以把SQL分为两个部分：数据操作语言（DML）和数据定义语言（DDL）。

H2database在接收到SQL语句之后，会先将该语句进行解析，再将解析之后的语句用于对数据库的操作。SQL语句由org.h2.command.Parser类解析，其中解析DML语言的相关方法在org.h2.command.dml包中；解析DDL语言的相关方法在org.h2.command.ddl包中。

“查询”操作就是对数据库操作中重要的一个，H2database实现查询的Command中的executeQuery方法：



在executeQuery中调用了query：



进一步查看CommandContainer和Query中的query方法，发现其中会出现大量的queryWithoutCache字样：

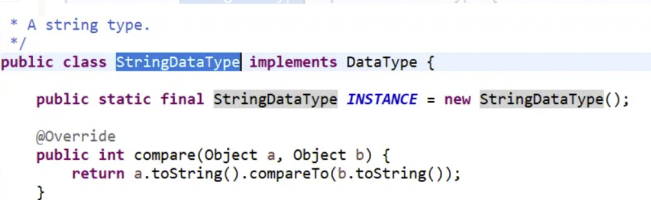


最终可以在Select中找到queryWithoutCache方法：

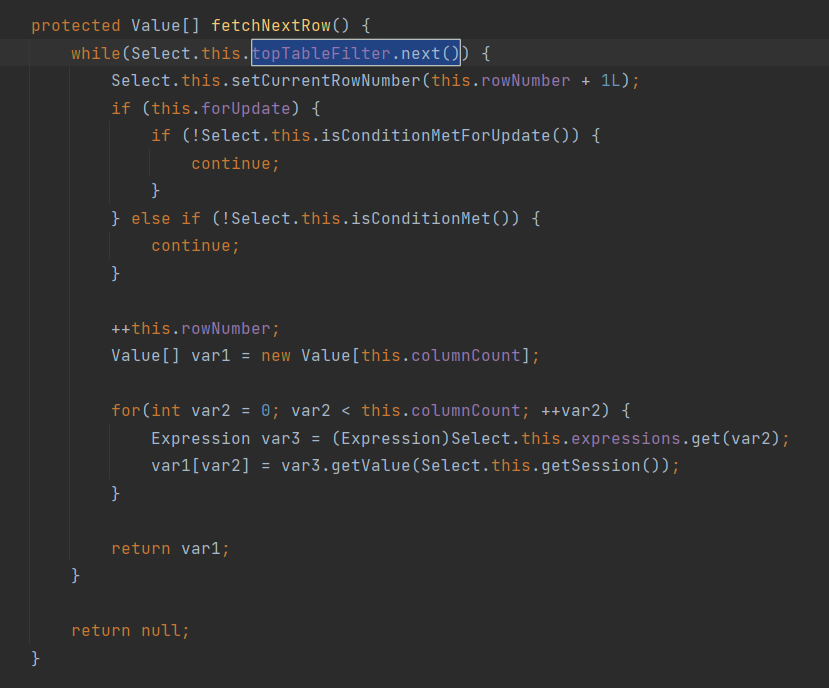


由此我们得到一个具体的查询调用关系：Command.executeQuery→CommandContainer.query→Query.query→Select.queryWithoutCache。

但因为能力所限，并没有在这个过程中找到具体比对关键字的方法，查阅资料得知，真正执行对比的方法为：



Select包含一个TableFilter（猜测也可能是一个过滤条件），负责查询数据库记录，然后isConditionMet负责判断条件是否符合，符合条件后，拼接成一个row，然后返回。



总结一下一个用于查询的SQL语句在H2database中具体执行过程。首先数据库调用parser对SQL语句进行解析，发现它是一个查询语句，即该select语法会创建一个Select对象。此后就会调用executeQuery进行查找，这个查找过程中包含TableFilter用于过滤与记录，StringDataType用于比对。

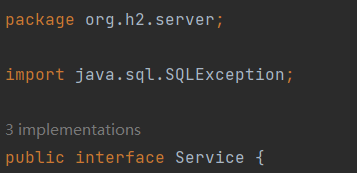
到此，H2database的查询过程告一段落。

**高级设计意图**

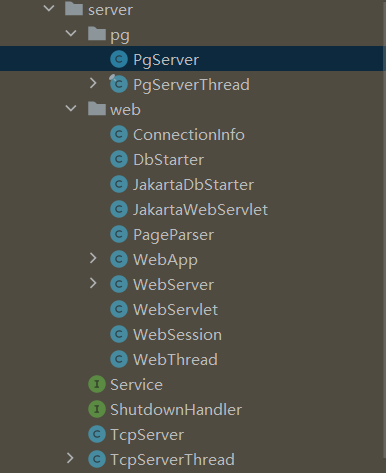
这里介绍H2database中体现的两种高级设计意图：工厂模式、策略模式，当然H2database在整个设计实现过程中还使用了其他设计意图。

工厂模式是一种创建设计模式，它在父类中提供了创建对象的接口，但允许子类更改将要创建的对象的类型。工厂方法模式建议将直接对象构造调用（使用new操作符）替换为对特殊工厂方法的调用。对象仍然是通过new操作符创建的，但它是从工厂方法内部调用的。

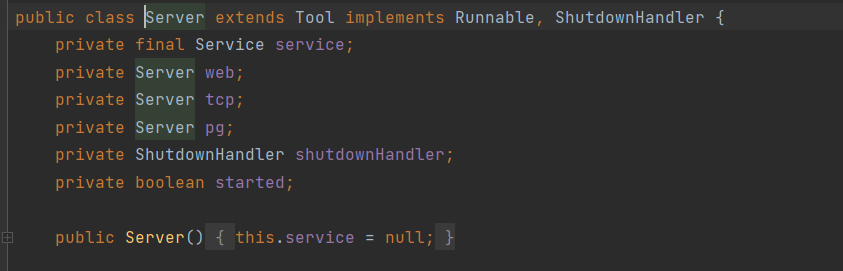
H2database中最能体现工厂模式的是它对Server的实现。在server包里H2database首先实现了一个叫做Service的父类，定义了init()、getURL()、start()、listen()、stop()、isRunning()、getAllowOthers()、getName()、getType()、getPort()、isDaemon()等方法，这些方法的具体实现则交由更具体的Service类型来完成。



H2database中实现了三种不同的Server，分别为PgServer、WebServer、TcpServer，其中Pg、Web、Tcp表示三种网络连接方式，通过实现三种具体子类，就可以启动上面三个Server。

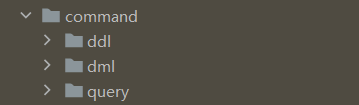


另外，启动Server的调用顺序是init() → start() →listen()，用于启动的类org.h2.tools.Server类。

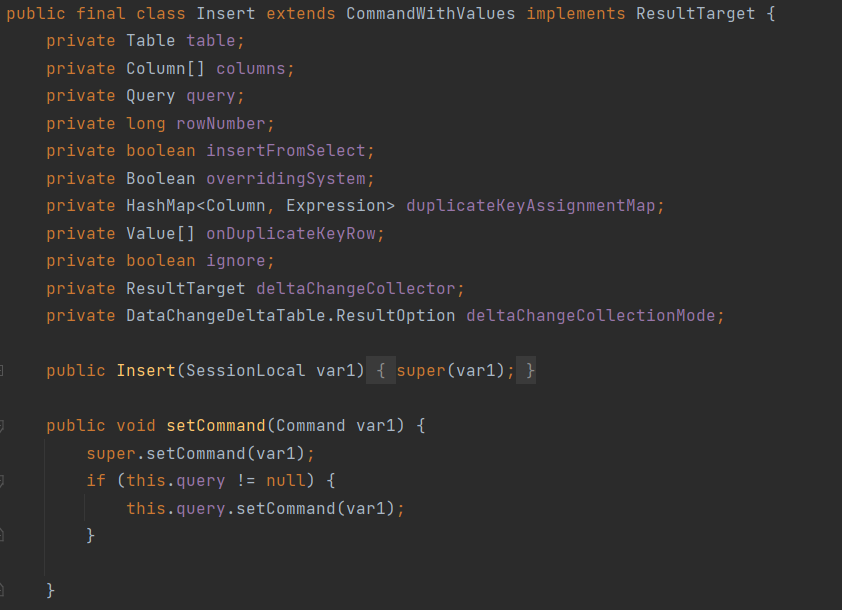


策略模式指针对一组算法，将每一个算法封装到具有共同接口的独立的类中，从而使得它们可以相互替换。策略模式使得算法可以在不影响到客户端的情况下发生变化。一个类定义了多种行为,并且这些行为在这个类的操作中以多个条件语句的形式出现。将相关的条件分支移入它们各自的类中以代替这些条件语句。

H2database中体现策略模式的一点是它对SQL语句的解析过程。SQL语句分为DML语句与DDL语句两种，两种语句的权限、语法等都不相同，因此H2database在实现SQL解析功能的时候将分析DML语句与DDL语句需要的类分别封装到两个包中。



比如CREATE TABLE语句对应org.h2.command.ddl.CreateTable类，INSERT语句对应org.h2.command.dml.Insert类，SHOW语句在Parser类中当成SELECT语句，对应org.h2.command.dml.Select类，相应的CREATE TABLE语句属于DDL语句，INSERT语句与SHOW语句属于DML语句。



**高级设计意图**

总而言之H2database模型层次结构清晰，注释也很完善，是学习java的好工程。但因为java对我而言是一门全新的语言，根据一学期的学习也终究只是掌握了一些皮毛。同时数据库相关知识也是下学期才学习，这里提前查询资料难免有很多疑惑的地方。因此在阅读代码过程中遇见了不少问题，得出的结论也不尽人意，很多问题没有解答。但是总体过程是愉快的，不仅加深了对java和数据库的理解，也了解到了开源工程的有趣之处。