DB HW4 Report

1. 組員：賴暐倫、賴威仁、張彧豪
2. 貢獻程度
3. 各題說明：
4. 第一題

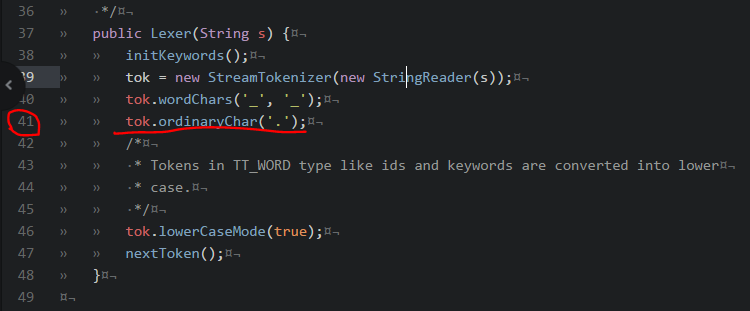
SQL：select certification, count(film\_id) from film group by certification

1. 大致流程：

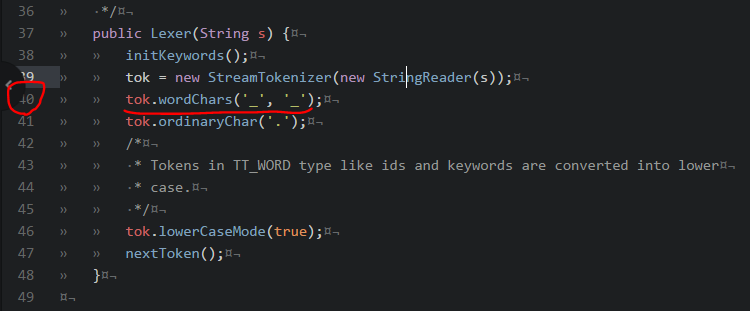
此query會被當成string傳入planner，planner會先將query string做拆解以及解析此query，再來做verify的動作，確認此query沒有語意錯誤，最後才做plan的動作

1. Lexer

lexer是query string第一個會來到的地方，在這邊lexer會以空白為間隔符號，把string給一個一個切開，比較特別的地方是，vanilladb不支援talbe\_name.table\_attribute，是因為這個java寫法的關係：



第41行的意思是會把點當成普通的字元並將它當成一個single-character token，所以會導致talbe\_name.table\_attribute這種語法無法執行，因為table\_name跟table\_attribute會被分成兩個token；取而代之的語法是tablename\_tableattribute，這種語法會被接受是因為這個java寫法的關係：



第40行的意思是把底線當作跟一般英文字母一樣的字元，所以遇到底線時不會被拆成兩個token，而是一個完整的token，所以tablename\_tableattribute這種語法可以正確執行。這邊有個小發現是，每個被割開的token，java有定義不同的描述(description)來描述一個token，像是一般的tablename、keyword之類的都是TT\_WORD，而數字就是TT\_NUMBER。而第46行的lowerCaseMode(true)指的是會把屬於TT\_WORD的token都轉成小寫。Lexer會一直讀直到遇到eof為止。

1. Parser

Parser做的事情就是將剛剛lexer割開的token蒐集起來，要對