Introduction to Database HW4(Final Project) Report

一　組員：0416324胡安鳳　0316323薛世恩　夏頡民

二　貢獻度：各佔三分之一

三

Query2（胡安鳳）:select team\_team\_id,count(player\_player\_id) from team,player where

　　team\_team\_id=player\_team\_id and category='badminton' group by

　　team\_team\_id

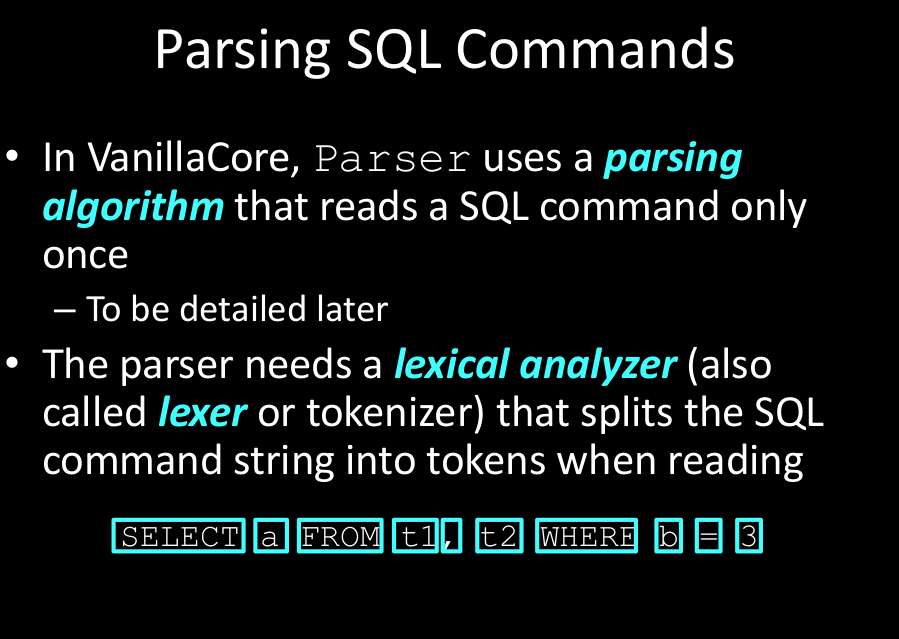
這個query是用來查詢對於所有打羽毛球的球隊中，其隊伍中的球員總數。

當我們輸入的query其實是被當作一個很長的字串載入SQL processor 中

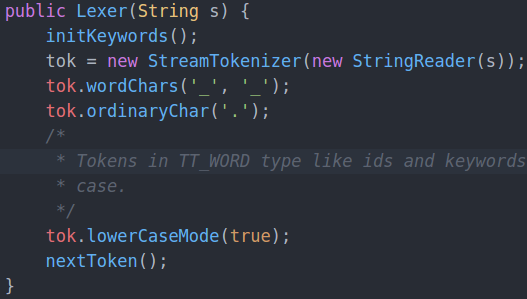
整個程式的執行順序:

●step1首先由Parser/Lexer執行，其中對於字串的切割是交給Lexer，剖析語意則是交給Parser，兩者相輔相成，如圖中可見Lexer引入了package





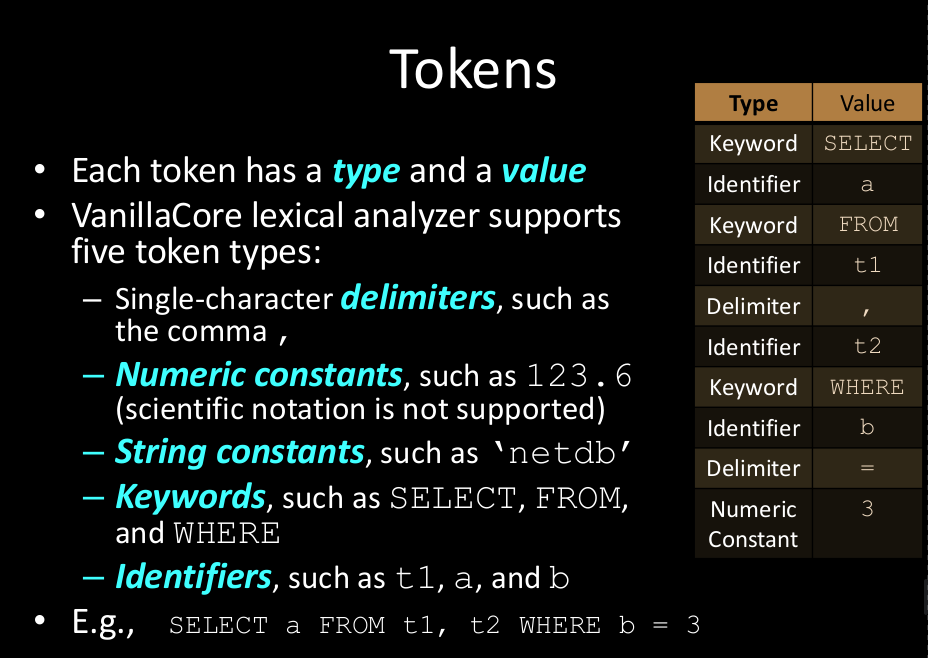
Lexer利用類似c++中StringStream(#include <sstream>)函式庫中切割的方式，同樣都是利用string token(deliminator)這種東西來剖析，如下圖



tok就是用來切割字串的識別符號。其中wordchar這句代表說字詞之間只能用底線連結例如player\_id ，而.當成一般的字符

而不做子查詢(這也是VanillaDB不支援點符號的原因)，其他一律切割。最後將tok一律在小寫的ascii code以免混淆，造成錯誤。

然後以nextToken繼續執行到文末，繼續切割，直到整個輸入的eof(end of file 才會停)



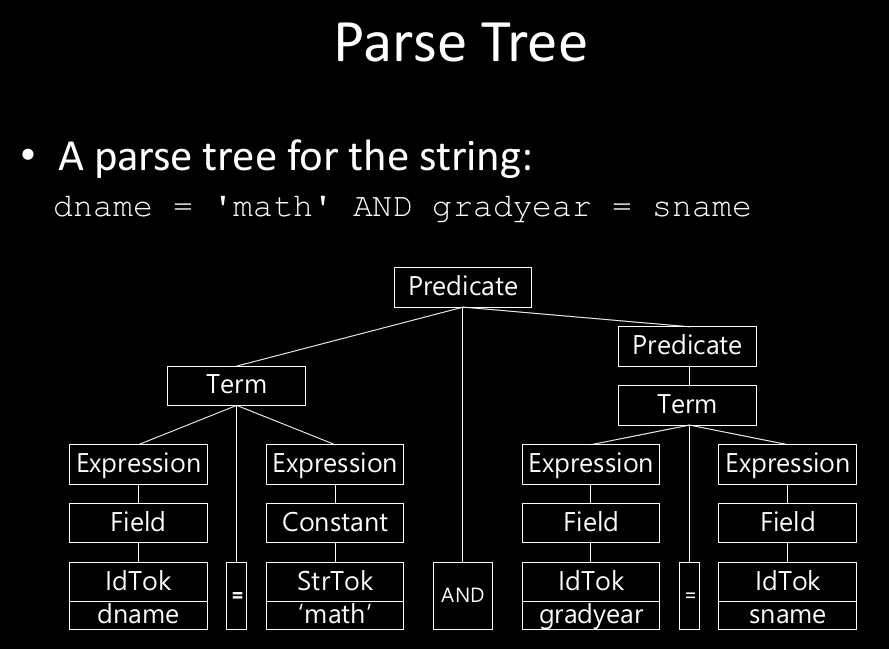
以在同一個Lexer.java中有建立一個字典表，代表會用到的SQL關鍵字

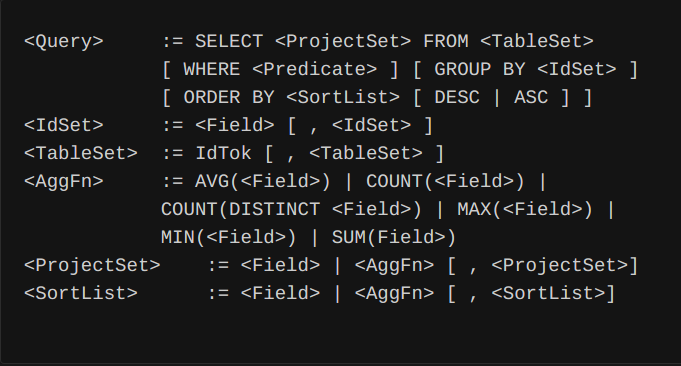
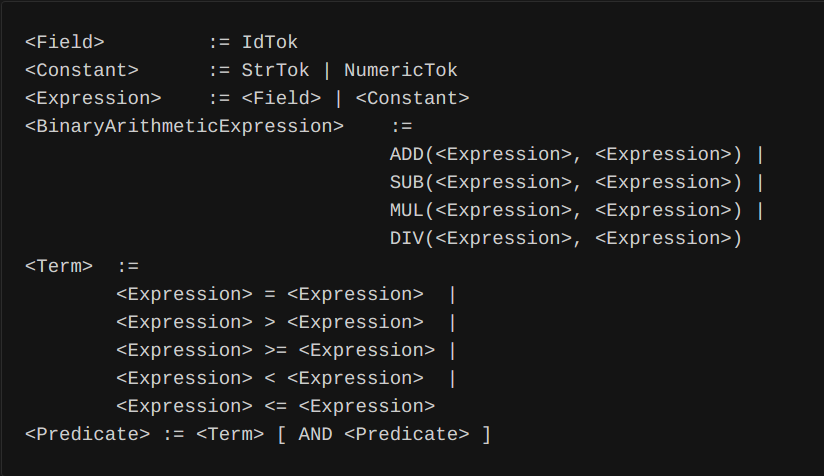


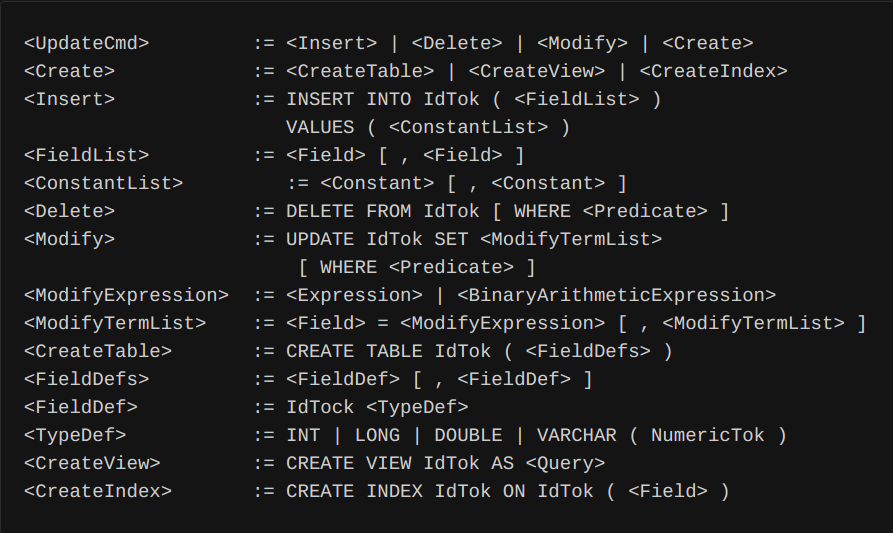
Lexer切完交給Parser後如果有不符合VanillaDBSQL的語法就會報錯，丟出BadSyntaxException這種例外，推斷是運用了『正規語言』的『上下文無關文法』(Context Free Grammar)

用一顆Parsing tree依照SQL的文法剖析輸入的字串

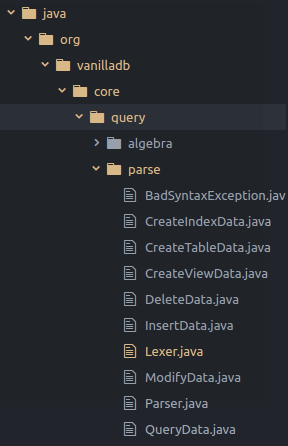
例如:







以上三者分別說明，單一語句(Predicates/左上)，查詢(Queries/右上)，以及更新(Updates/下)的文法規範，在Lexer中便會呼叫Parser中識別輸入的語句是屬於那一種然後找對應以上三者之一的文法歸範(文法剖析樹，如方才提過的Parse Tree)，由以下的檔案樹狀結構照片可以看到，當Lexer呼叫Parser的時候，實際上Parser也會和Lexer互動，由圖片中的檔案結構呼應方才提到兩者相輔相成的概念



每一個都有自己的對應剖析方法

例如創見表格就會對應CreateTableData

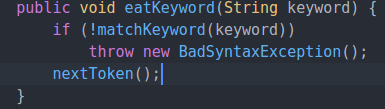
刪除資料對應DeleteData

查詢資料對應QueryData

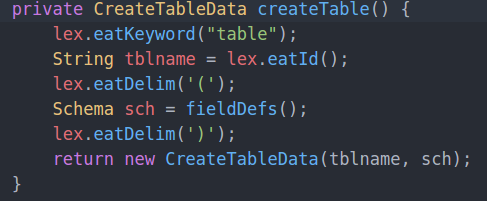
都有自己的剖析方式，由主程式Parser來呼叫他們

就好比C++裡面自己定義函數的模組化原理一樣，相當物件導向

剖析切割後，如果不符合文法便會報錯，BadSyntaxException();　例如eatKeyword吃到不符合規範的關鍵自就會報錯(細部則交由Parser進行)，如下圖所示

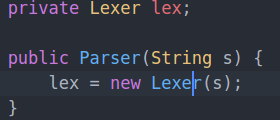


/Lexer



/Parser

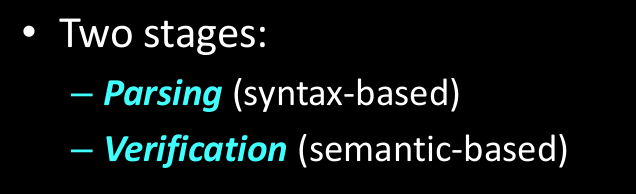
(lex代表呼叫Lexer又再一次呼應了Lexer&Parser兩者相輔相成)



/Parser

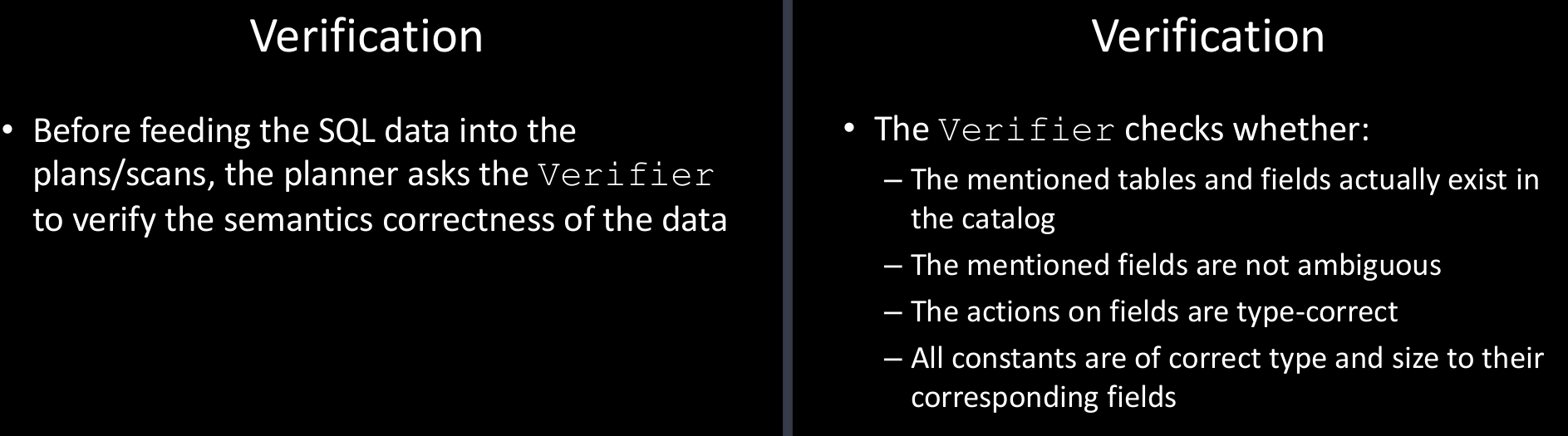
切割完關鍵字後，如果沒有BadSyntaxException，便會是一個沒有語法錯誤的SQL語句，此時

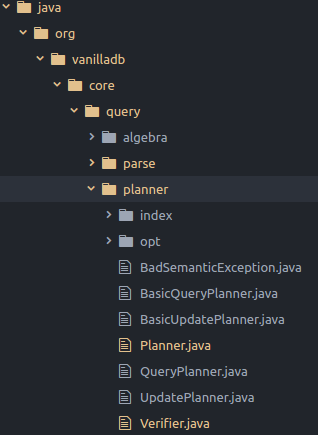
可以送到Verifier.java中執行，進行要真正做的事情。



●step2切割剖析後當然就會把切割剖析好的東西送去分類以及查詢，在此已經準備進入

執行階段了，因此改由Planner資料夾中的Verifier負責語意正確的判斷(注意語法正確並不代表語意正確，一個是SyntaxCorrect，另個是SemanticCorrect兩者不同，方才Parsing procedure只有確認語法正確，也就是文法正確性，好比英文的時態或是主動詞搭配是否正確，現在是要看有沒有合乎邏輯的內容，合理的內容)





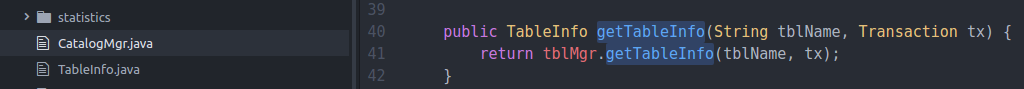
首先會創立兩個陣列分別存放要查詢的表格以及表格內要查什麼，在此可以看到使用List的資料結構



在此次作業我們分析Query data 所以用到verifyQueryData，首先檢查欲查詢表格是否存在



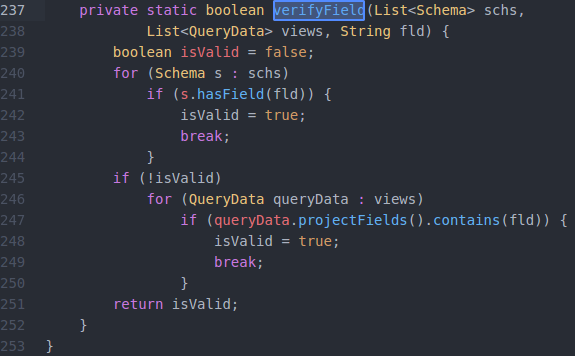
額外呼叫了位於core/strorage/metadata的CatalogMgr(類別管理元，建下面第二章圖片)中的getTableInfo來把table裡面的東西傳出來至Verifier比對，確認是否有那些表格，如果沒有就報錯說無表格(Table does not exist)(在此是沒有採用View虛擬表格的情況下)



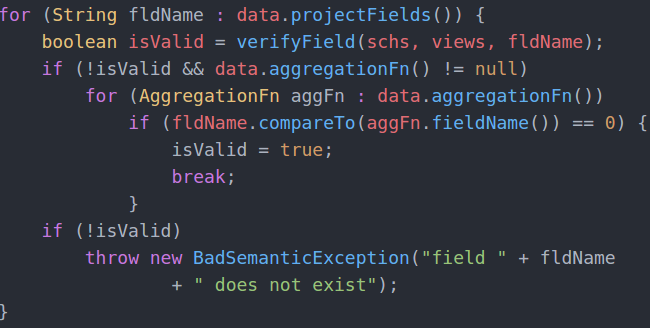
檢查表格後還要看表格的欄位，無論是一般查詢、聚合查詢(aggregation)、分組或是排序

都要檢查有無表格的該欄位才知道能否做事，由Verifier上半部提出各種認證，在交由下半部統一認證

（左為Verifier的上，右為下）



仔細來看可以發現每一個檢查欄位的邏輯都相當類似，皆會先有一個isValid來判斷是否存在，若不存在但是確有聚合函數(aggregationFn!=null)也依然可用(因為對於例：select XXX　YYY ZZZ aggFn(KKK)中剛剛只檢視了XXX但我們還沒檢視到聚合函數中的東西，即aggFn中的KKK因為只要他在表格中視一個有效的欄位亦能查詢，就把isValid設回真)



除非真的什麼都沒有才會丟出BadSemanticException表示邏輯內容無效的句子

其他的也是利用類似的方法檢定，只是Group by 和Sort by不需要聚合函數(正確而言是

SQL不支援這種寫法)

Semantic and Syntax均無誤後變可以開始進行Query了(以下Step3)

●step3

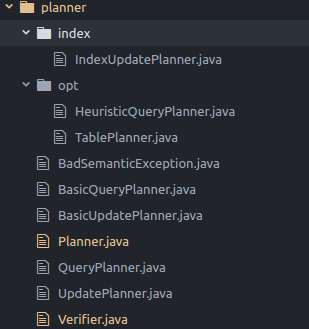
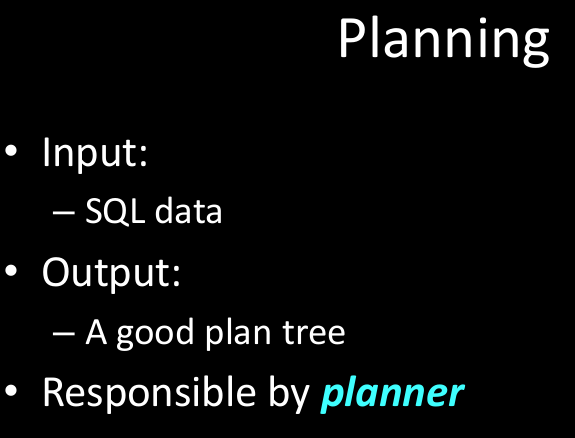
在Query開始前，回顧一下這次作業第二題的Query是

select team\_team\_id,count(player\_player\_id) from team,player where

team\_team\_id=player\_team\_id and category='badminton' group by

team\_team\_id

在正式查詢前會先優化Query ，由Planner負責



檔案架構:HeuristicQueryPlanner便是整個Planner的核心