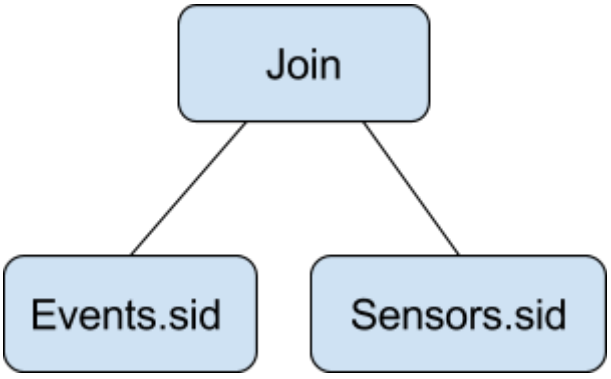
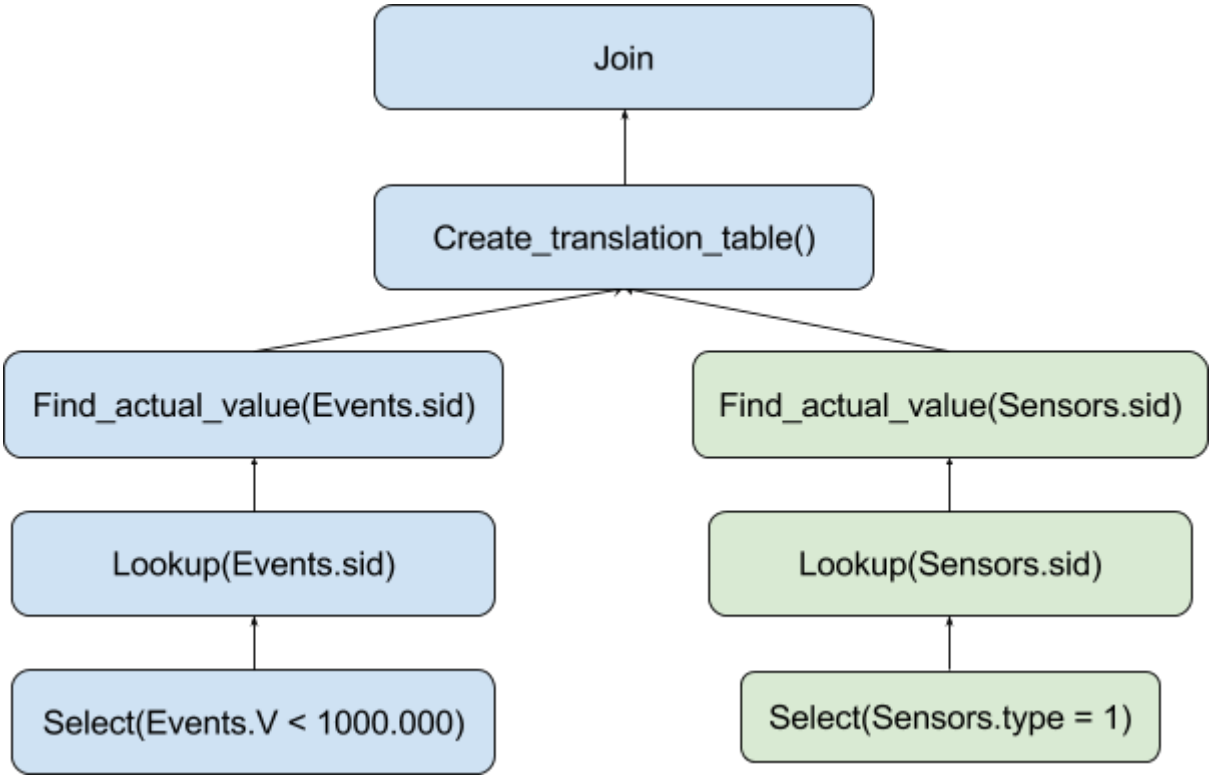
**Homework 3**

Bui Tien Cuong
2017-20722

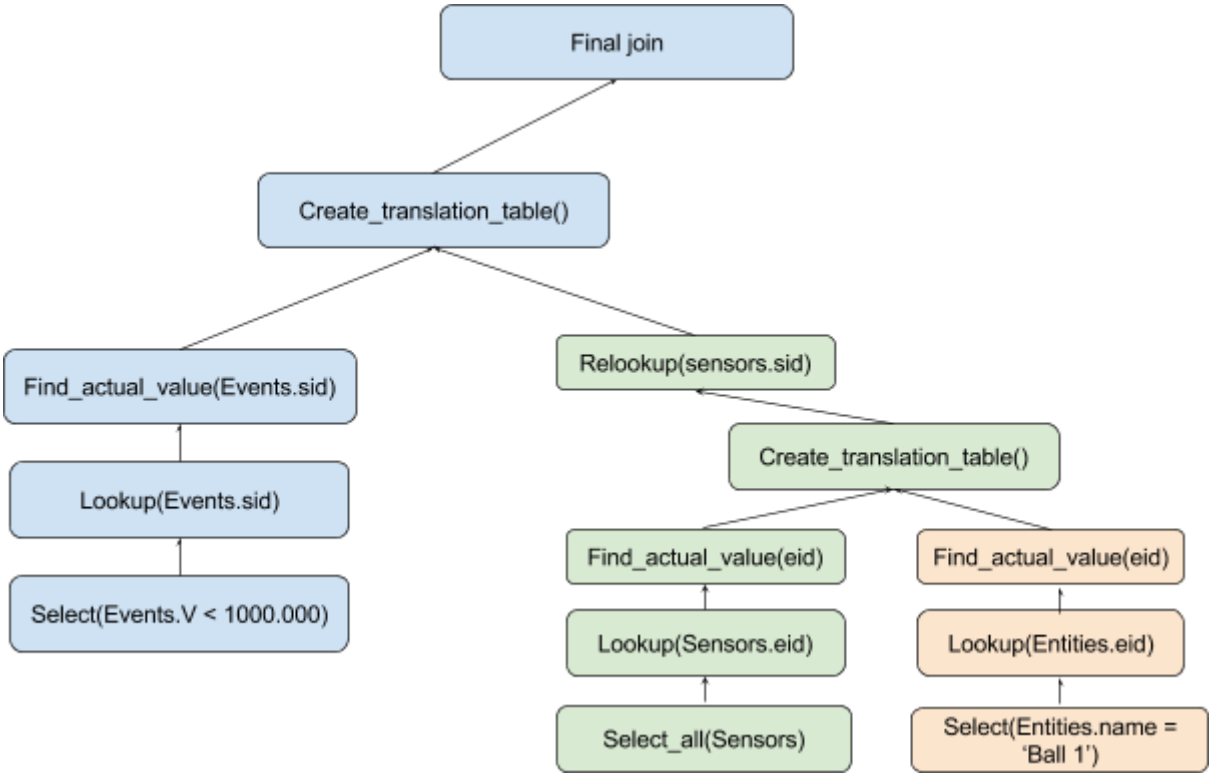
1. Hardware specification
    a. OS: Ubuntu 16.04 x64
    b. Ram memory: 8Gb
    c. CPU: Intel(R) Core(TM) i5-6600 CPU @ 3.30GHz
    d. HDD: 256Gb
    e. C/C++ compiler: g++ 5.4.0
    f. Spark version: v2.2.0
2. Query strategies
    a. Query 1




    b. Query 2



c. Query 3




3. Results and comparison to Spark

| | Properties | Spark | Single Thread | Multiple thread |
|---|---|---|---|---|
| Loading | Running time | 13 s | 20 s | 20.4s |
| | Memory consumption | 386 Mb | 320.066 Mb | 320.109 Mb |
| Q1 | Running time | 11s | 8.17 s | 7.9s |
| | Total records | 1,.048,576 | 1,048,576 | 1.048.576 |
| | Memory consumption | 272 Mb | 59.5117 Mb | 59.4336 Mb |
| Q2 | Running time | 10 s | 7.17 s | 6.6s |
| | Total records | 446,885 | 446,885 | 446,885 |
| | Memory consumption | 34 Mb | 69.5156 Mb | 85.207 |
| Q3 | Running time | 12s | 7.06 s | 6.6s |
| | Total records | 138,206 | 138,206 | 138,206 |
| | Memory consumption | 109 Mb | 0.21 Mb | 0.26 Mb |

Evaluation:
Basically, the running time of multiple thread strategy is lower than the single thread. However, as you can see, multiple thread will consume memory more. I will improve the performance of partitioned join over the next few days and add the updated results to the final report.
For more detail results, I attached results_homework2.txt (C++) and results_spark_hw2.txt (Spark) in docs folder.