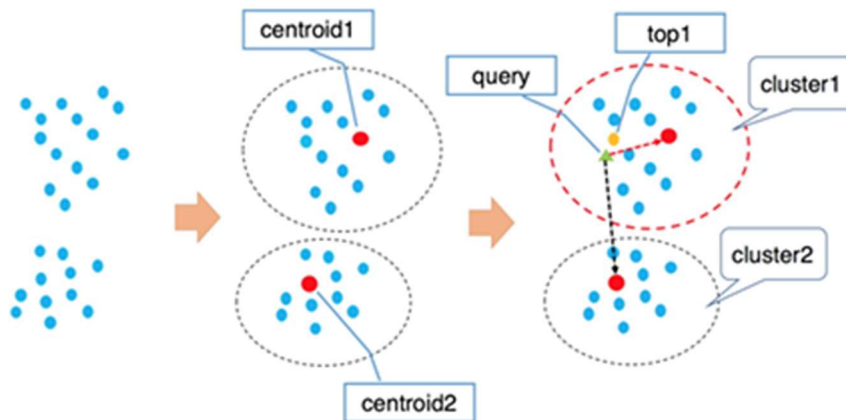


Planning :



參考 PASE 的 paper 上的理論來實作利用 IVFFlex Index 來實作 Vector Database，以 Kmeans 來對 Vector 做 Clustering 以增加 Query 時 Search 的速度。

Implementation :

Vector Support Enhancement

- **新增 VectorConstantRange.java :**
參考原本附的 VarcharConstantRange class 用來實做 SearchRange。
- **VectorConstant.java**
新增新的 constructor 來方便初始化 vector，新增 calcDistance() method 來計算 Vector 間的距離。
- **Parse.java, Lexer.java :**
新增 IVF 的 index 選項來方便實作 Vector indexing。
- **Type.java :**
新增 VECTOR type 方便使用。

Index

- **IVFIndex.java**

大致上以 `hashindex` 作為模板。嘗試實做 IVF 來對 `Vector` 做 Indexing。其中有更改的部份為新增 `getNearestCentroids` 找出離要查詢的 `searchkey` 最近的群。在此 `method` 中，使用前面更改到的 `EuclideanFn.java` 來計算兩個 `searchkey` 的 `vector` 之歐式距離，取距離小的群作為離此查詢的 `searchkey` 的群分類。

- **Index.java**

在 `searchCost` 和 `newInstance` 中新增 `ivf` 辨別選項。

- **SearchKeyType.java**

`getVecIdx()`:將 `type` 類別裡的 `array` 取出。

SIMD

- **EuclideanFn.java :**

引進 `jdk.incubator.vector` 的 `FloatVector` `VectorSpecies` 和 `VectorMask` 進行 SIMD。使用 `VectorSpecies` 定義 `vector` 大小，此用 `VectorMask` 讓某些元素不被更改。

Experiment & Conclusion :

- Environment
- Intel Core i5-12500H CPU @ 4.6GHz, 24 GB RAM, 190GB SSD, Windows 11

```
WARNING: Using incubator modules: jdk.incubator.vector
6月 16, 2024 10:25:09 下午 org.vanilladb.bench.VanillaBenchParameters <clinit>
INFO: Using SIFT benchmarks
6月 16, 2024 10:25:09 下午 org.vanilladb.core.util.PropertiesLoader getPropertyAsString
WARNING: can't find property: org.vanilladb.bench.VanillaBenchParameters.REPORT_OUTPUT_DIRECTORY, using default value: null
6月 16, 2024 10:25:09 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: checking the database on the server...
6月 16, 2024 10:25:10 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: database check passed.
6月 16, 2024 10:25:10 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: creating 2 emulators...
6月 16, 2024 10:25:10 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: waiting for connections...
6月 16, 2024 10:25:11 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: start benchmarking.
start loading sift dataset...
6月 16, 2024 10:25:31 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: warm up period finished.
6月 16, 2024 10:25:31 下午 org.vanilladb.bench.VanillaBench benchmark
INFO: start recording results...
SiftParamGen: 1000000 vectors loaded.
java.sql.SQLException: java.rmi.ServerException: RemoteException occurred in server thread; nested exception is:
    java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at org.vanilladb.core.remote.storedprocedure.SpConnection.callStoredProc(SpConnection.java:32)
    at org.vanilladb.bench.remote.sp.VanillaDbSpConnection.callStoredProc(VanillaDbSpConnection.java:35)
    at org.vanilladb.bench.rte.TransactionExecutor.executeTxn(TransactionExecutor.java:46)
    at org.vanilladb.bench.benchmarks.sift.rte.SiftTxExecutor.execute(SiftTxExecutor.java:50)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.executeTxnCycle(RemoteTerminalEmulator.java:90)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.run(RemoteTerminalEmulator.java:52)
Caused by: java.rmi.ServerException: RemoteException occurred in server thread; nested exception is:
    java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at java.rmi/sun.rmi.server.UnicastServerRef.dispatch(UnicastServerRef.java:392)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:200)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:197)
    at java.base/java.security.AccessController.doPrivileged(AccessController.java:712)
    at java.rmi/sun.rmi.transport.Transport.serviceCall(Transport.java:196)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint.handleMessage(TCPEndpoint.java:421)
```

```
    at org.vanilladb.bench.rte.TransactionExecutor.executeTxn(TransactionExecutor.java:46)
    at org.vanilladb.bench.benchmarks.sift.rte.SiftTxExecutor.execute(SiftTxExecutor.java:50)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.executeTxnCycle(RemoteTerminalEmulator.java:90)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.run(RemoteTerminalEmulator.java:52)
Caused by: java.rmi.ServerException: RemoteException occurred in server thread; nested exception is:
    java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at java.rmi/sun.rmi.server.UnicastServerRef.dispatch(UnicastServerRef.java:392)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:200)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:197)
    at java.base/java.security.AccessController.doPrivileged(AccessController.java:712)
    at java.rmi/sun.rmi.transport.Transport.serviceCall(Transport.java:196)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint.handleMessage(TCPEndpoint.java:421)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:828)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:795)
    at java.base/java.security.AccessController.doPrivileged(AccessController.java:399)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:794)
    at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1136)
    at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:635)
    at java.base/java.lang.Thread.run(Thread.java:842)
    at java.rmi/sun.rmi.transport.StreamRemoteCall.exceptionReceivedFromServer(StreamRemoteCall.java:304)
    at java.rmi/sun.rmi.transport.StreamRemoteCall.executeCall(StreamRemoteCall.java:280)
    at java.rmi/sun.rmi.server.UnicastRef.invoke(UnicastRef.java:145)
    at java.rmi/java.rmi.server.RemoteObjectInvocationHandler.invokeRemoteMethod(RemoteObjectInvocationHandler.java:215)
    at java.rmi/java.rmi.server.RemoteObjectInvocationHandler.invoke(RemoteObjectInvocationHandler.java:160)
    at jdk.proxy1/jdk.proxy1.$Proxy1.callStoredProc(Unknown Source)
    at org.vanilladb.core.remote.storedprocedure.SpConnection.callStoredProc(SpConnection.java:30)
    ... 5 more
Caused by: java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at org.vanilladb.core.remote.storedprocedure.RemoteConnectionImpl.callStoredProc(RemoteConnectionImpl.java:50)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:77)
    at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.base/java.lang.reflect.Method.invoke(Method.java:568)
    at java.rmi/sun.rmi.server.UnicastServerRef.dispatch(UnicastServerRef.java:360)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:200)
    at java.rmi/sun.rmi.transport.Transport$1.run(Transport.java:197)
    at java.base/java.security.AccessController.doPrivileged(AccessController.java:712)
    at java.rmi/sun.rmi.transport.Transport.serviceCall(Transport.java:196)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint.handleMessage(TCPEndpoint.java:587)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:828)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:795)
    at java.base/java.security.AccessController.doPrivileged(AccessController.java:399)
    at java.rmi/sun.rmi.transport.tcp.TCPEndpoint$ConnectionHandler.run(TCPEndpoint.java:794)
    at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1136)
    at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:635)
    at java.base/java.lang.Thread.run(Thread.java:842)
Exception in thread "RTE-1" exception in thread "RTE-0" java.lang.RuntimeException: java.rmi.ServerException: RemoteException occurred in server thread; nested exception is:
    java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at org.vanilladb.bench.benchmarks.sift.rte.SiftTxExecutor.execute(SiftTxExecutor.java:87)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.executeTxnCycle(RemoteTerminalEmulator.java:90)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.run(RemoteTerminalEmulator.java:52)
java.lang.RuntimeException: java.rmi.ServerException: RemoteException occurred in server thread; nested exception is:
    java.rmi.RemoteException: Cannot invoke "java.lang.Integer.intValue()" because "this.items[i]" is null
    at org.vanilladb.bench.benchmarks.sift.rte.SiftTxExecutor.execute(SiftTxExecutor.java:87)
    at org.vanilladb.bench.rte.RemoteTerminalEmulator.executeTxnCycle(RemoteTerminalEmulator.java:90)
```

實作時，我們遇到了許多不好解決的 bug 加上一個組員失蹤與時間分配的不足，導致我們最終沒有順利地讓我們的 System 運行起來。