Final Project

CS471000

NTHU

Outline

- Goal
- Timeline
- Final Presentation
- Submission

Outline

- **Goal**
- Timeline
- Final Presentation
- Submission

Goal

- Can you extend a relational database system to support storing and querying over vectors?

TA60

- Build any index for approximate nearest neighbor search.

- Options:
  - Inverted File Index (IVF)
  - Scalar Quantization (SQ)
  - Product Quantization (PQ)
  - Locality Sensitive Hashing (LSH)
  - Hierarchical Navigable Small-World (HNSW)

TA80

- Can you make distance calculation faster?

- Hint: Single Instruction Multiple Data (SIMD)
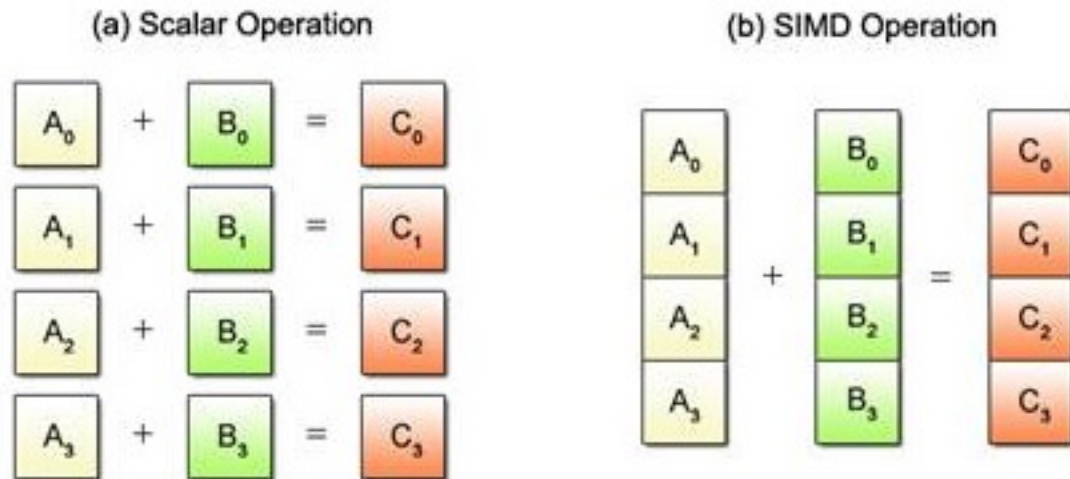
Single Instruction Multiple Data (SIMD) in Java

- Import `jdk.incubator.vector` **module**

- Write your SIMD version of distance calculation

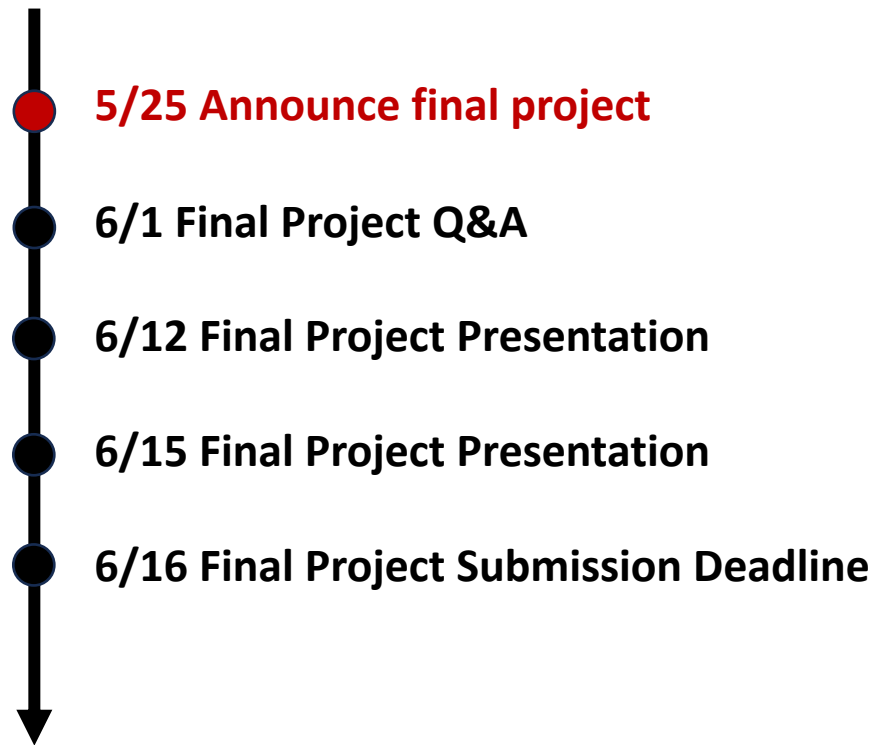- Use `add-modules=jdk.incubator.vector` **flag when running**

```
java --add-modules jdk.incubator.vector -Djava.util.logging.config.file=$PROP_DIR/logging.
properties -Dorg.vanilladb.bench.config.file=$PROP_DIR/vanillabench.properties -Dorg.vanil
ladb.core.config.file=$PROP_DIR/vanilladb.properties -jar jars/server.jar EmbeddingDB
```

Outline

- Goal
- **Timeline**
- Final Presentation
- Submission

Timeline

5/25 Announce final project

6/1 Final Project Q&A

6/12 Final Project Presentation

6/15 Final Project Presentation

6/16 Final Project Submission Deadline

Outline

- Goal
- Timeline
- **Final Presentation**
- Submission

Final Presentation

- 6/12 (Mon) 15:00 - 18:00 Delta 103
- 6/15 (Thu) 14:20 – 15:10 Delta 103

- 4 mins presentation + 2 mins QA for each group
- TAs will announce the order of presentation later

Final Presentation

- Your presentation should cover:
  - Implementations
  - Experiments (Show your throughput and recall)
  - Conclusion

Evaluation

- Throughput
  - How many nearest neighbor search can you finish over a period of time

- Recall
  - A : { Your nearest neighbor result }
  - T : { True nearest neighbors }
  - Recall: |(A ∩ T)| / |T|

Evaluation

- We will evaluate each group based on:
    - Insight
    - Experiment
    - Presentation

Award

- Best Improvement Award
  - Best throughput * recall score

- Best Presentation Award

- Bonus points for teams that win the awards

Outline

- Goal
- Timeline
- Final Presentation
- **Submission**

Submission

- Requirements
  - You have to write a report as usual

- The details of submission will be on GitLab

- Deadline: 2023/6/16(Fri) 23:59

Good Luck !

References

- https://dl.acm.org/doi/pdf/10.1145/3318464.3386131

- https://vksegfault.github.io/posts/java-simd/