LEARNED INDEX

Team 1
組長:105062130 斯楷雯 組員:105062139 林楷宸 105062226 陳評詳

AGENDA

- Paper Summary
- Implementation
- Experiment
- Problem encountered
- Summary

AGENDA

- Paper Summary
- Implementation
- Experiment
- Problem encountered
- Summary

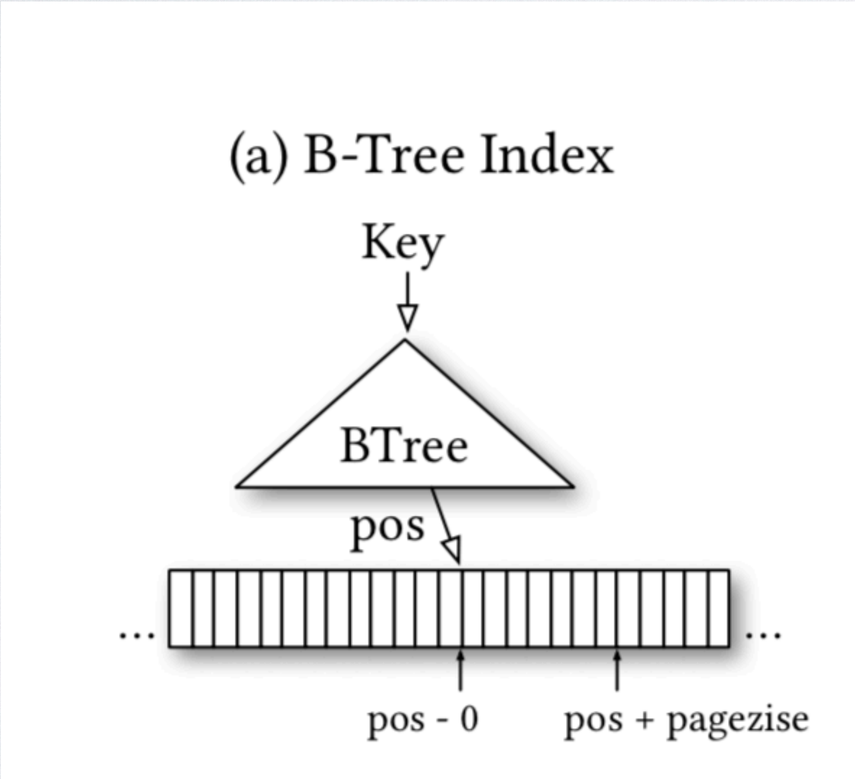
PAPER SUMMARY

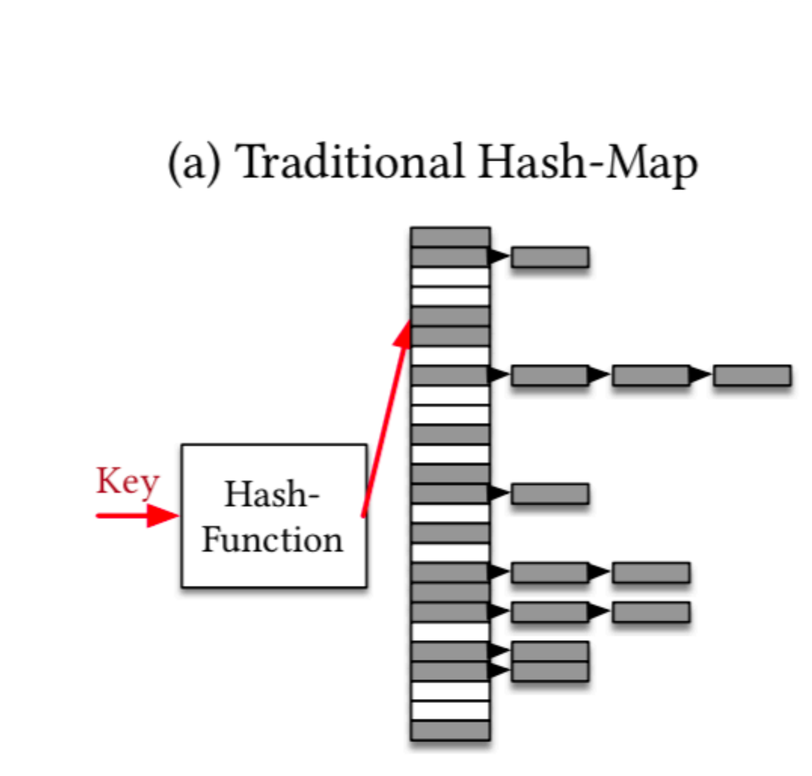
- Title: The Case of Learned Indexed Structures
- Conference: SIGMOD'18
- Published year: June 10-15, 2018
- Authors: T. Kraska, A. Beutel, E. H. Chi, J. Dean, and N. Polyzotis

PAPER SUMMARY

- Searching a specific index **cost a lot of times**
- Once the database grows, the searching time rises significantly
- Index searching varies since different data distributions

PAPER SUMMARY





PAPER SUMMARY

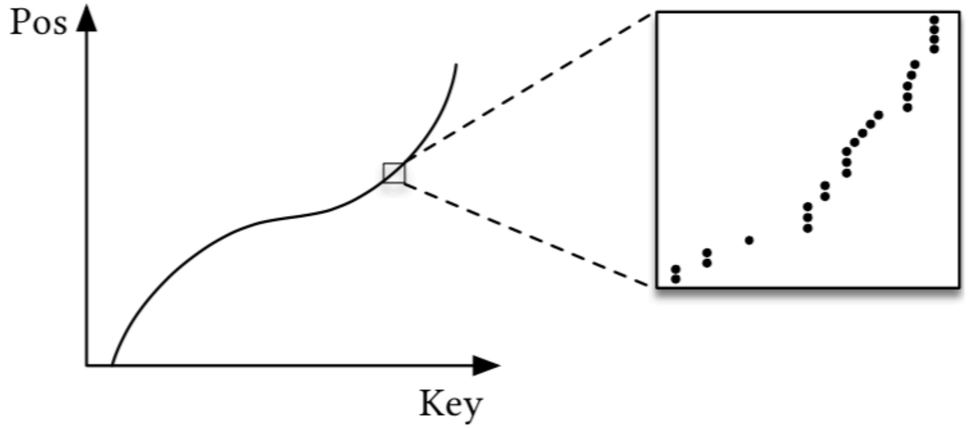


Figure 2: Indexes as CDFs

$$p = F(\text{Key}) * N$$

PAPER SUMMARY

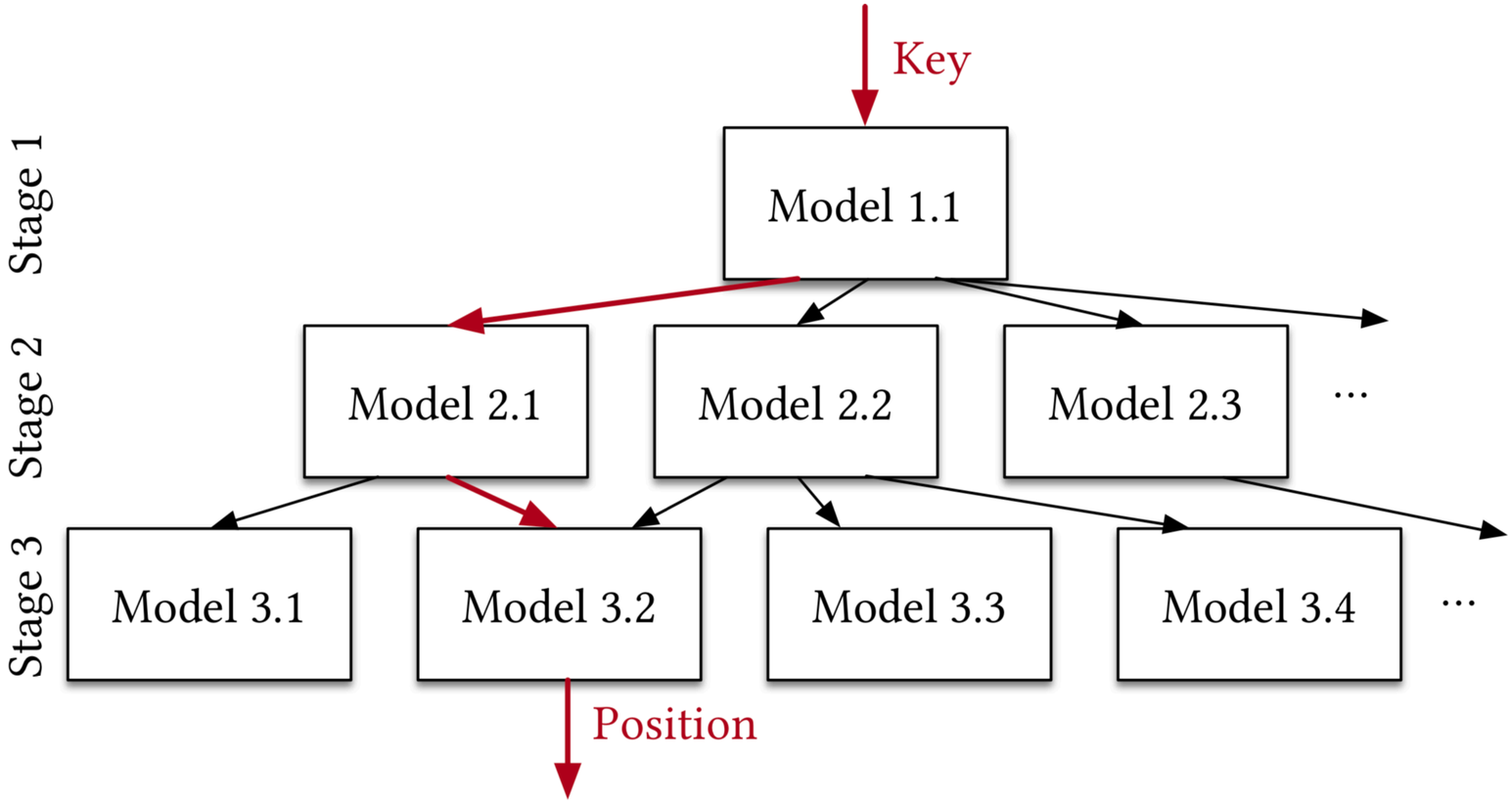


Figure 3: Staged models

PAPER SUMMARY

Algorithm 1: Hybrid End-To-End Training

Input: int threshold, int stages[], NN_complexity
Data: record data[], Model index[][]
Result: trained index

```
1  M = stages.size;
2  tmp_records[][];
3  tmp_records[1][1] = all_data;
4  for i ← 1 to M do
5  |   for j ← 1 to stages[i] do
6  |   |   index[i][j] = new NN trained on tmp_records[i][j];
7  |   |   if i < M then
8  |   |   |   for r ∈ tmp_records[i][j] do
9  |   |   |   |   p = index[i][j](r.key) / stages[i + 1];
10 |   |   |   |   tmp_records[i + 1][p].add(r);
11 for j ← 1 to index[M].size do
12 |   index[M][j].calc_err(tmp_records[M][j]);
13 |   if index[M][j].max_abs_err > threshold then
14 |   |   index[M][j] = new B-Tree trained on tmp_records[M][j];
15 return index;
```

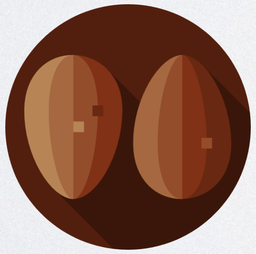
AGENDA

- Paper Summary
- Implementation
- Experiment
- Problem encountered
- Summary

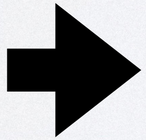
IMPLEMENTATION

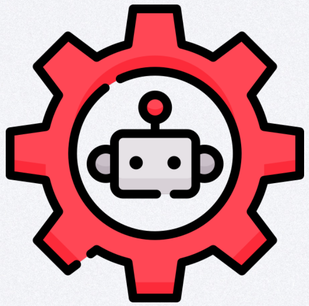
VanillaDB

Super VanillaDB

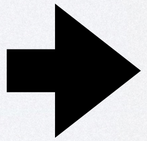


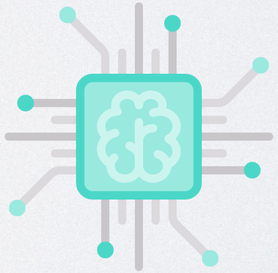
Dataset





Training





Parameters

IMPLEMENTATION

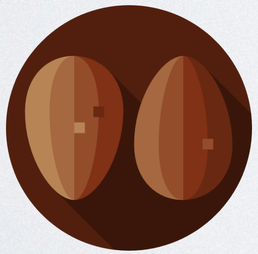
- 2 files constructed
  - ModelIndex
  - NN

ModelIndex

NN

IMPLEMENTATION - ModelIndex

VanillaDB

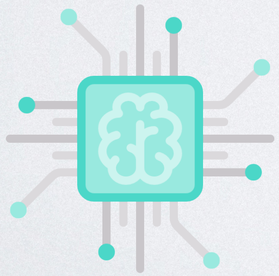


Dataset

- When Loading Testbed
- Use Insert(…) to get Dataset

IMPLEMENTATION - ModelIndex

SuperVanillaDB

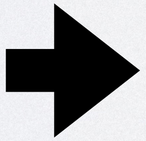


Parameters

- When BenchMarking
- Use Constructor(…) to load parameters
- Store the parameters in NN

IMPLEMENTATION - ModelIndex

ModelIndex



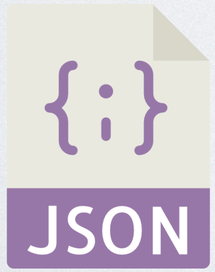
NN.Predict(…) * 2

RecordId

- During getDataRecordId(…)
- Call NN.predict(…) to fetch the position

IMPLEMENTATION - NN

NN



Parameters

- Implement Machine-Learning Model
- Constructor(…) : Load and save the parameters
- Predict(…) : executing the model

IMPLEMENTATION - Others

- Add search time in benchmark result to evaluate the performance
- New DEFAULT_INDEX_TYPE
- New VM Argument

AGENDA

- Paper Summary
- Implementation
- Experiment
- Problem encountered
- Summary

EXPERIMENT

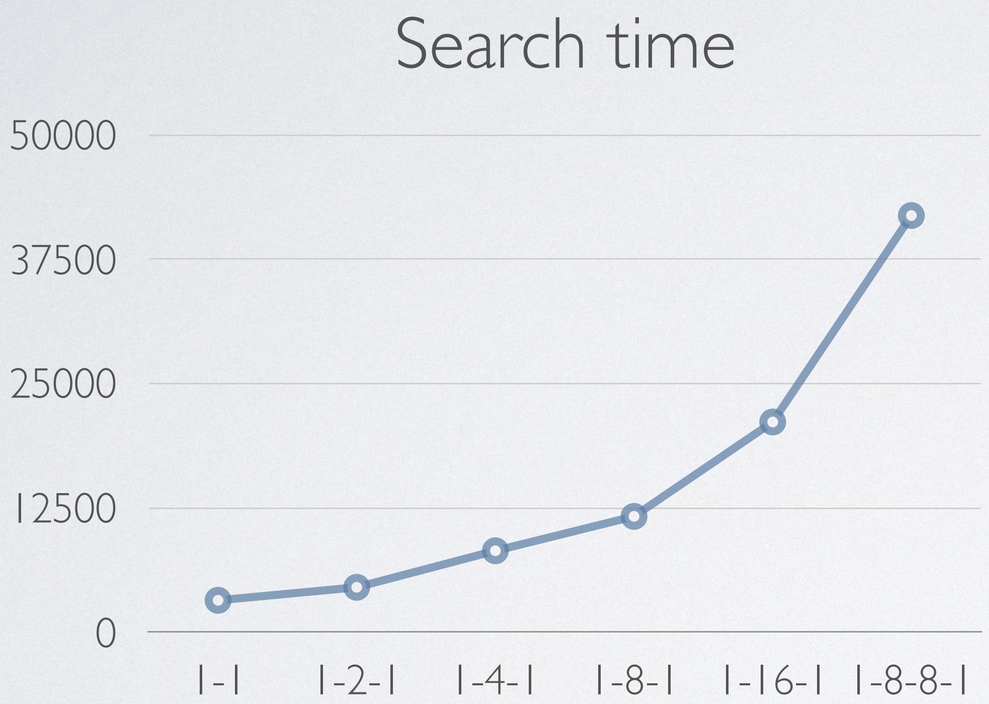
資電 326

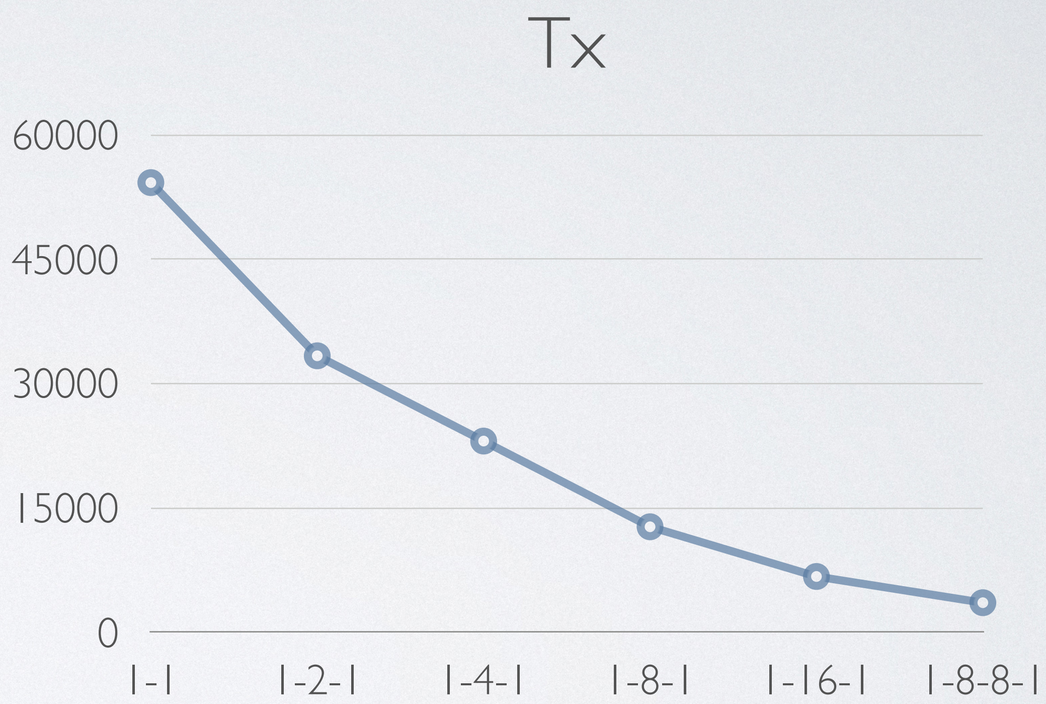
- Intel Core i5-3470 CPU@3.2GHz
- 8GB RAM
- 還原卡+HDD

EXPERIMENT

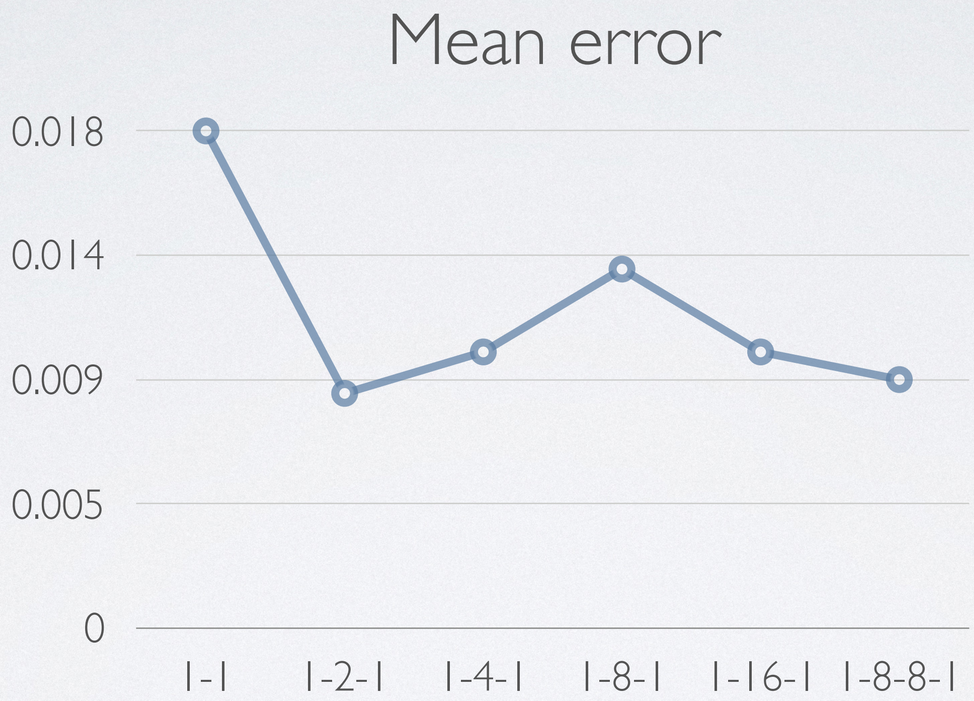
- Model Structure influence
- B-Tree v.s. Model
- # of Model in 2nd stage effects

EXPERIMENT

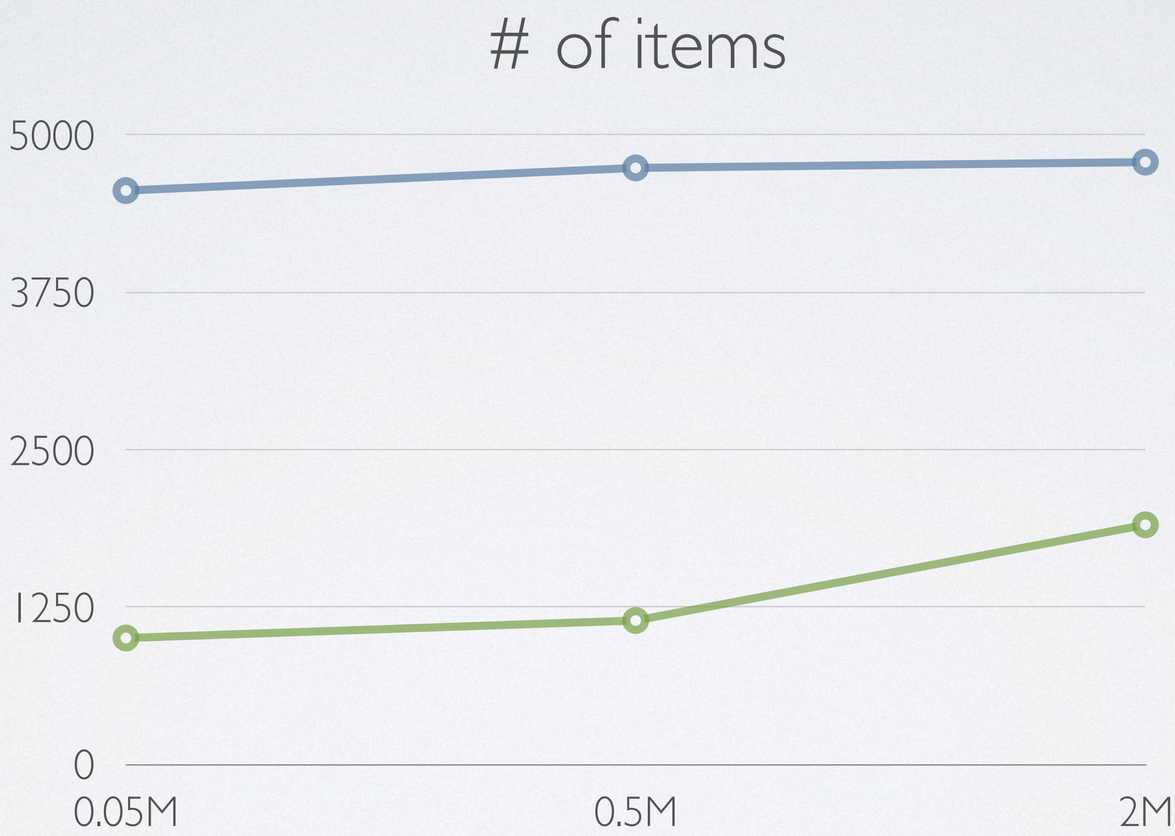

Search time
50000
37500
25000
12500
0
1-1
1-2-1
1-4-1
1-8-1
1-16-1
1-8-8-1


Tx
60000
45000
30000
15000
0
1-1
1-2-1
1-4-1
1-8-1
1-16-1
1-8-8-1

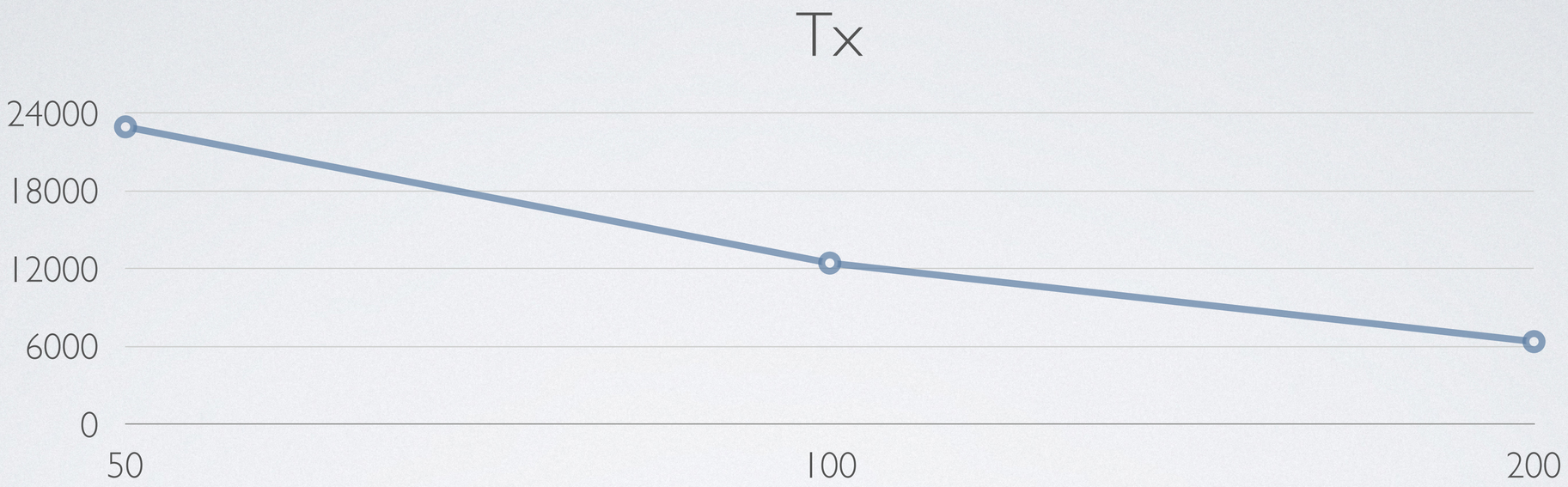
EXPERIMENT



EXPERIMENT



EXPERIMENT



Items per model	50	100	200
Search time(ns)	7623	10746	16094
Mean error	0.01	0.0052	0.0015

AGENDA

- Paper Summary
- Implementation
- Experiment
- Problem encountered
- Summary

SUMMARY

- ML is trending
- Hardware power restriction
- Error tolerance