Checkpoints + Recovery

Final Project Presentation

Kaige Liu, Tianlei Pan, Xuanxuan Ge
Overviews

Implement full recovery from WAL + Checkpoints.
→ 1. Separate Catalog Logs
→ 2. Store DataTables
Pipeline

1. Catalog Log Filtering (Per Checkpoint)
Pipeline

2. Saving Data Tables (Checkpoint)
Pipeline

3. Recover Checkpoint -> Continuous Logging
Goals & Progress

→ 75%: implement user table recovery from the most recent checkpoint.

→ 100%: enable catalog and user table recovery from checkpoint + continuous logs recovery.

→ 125%: periodic checkpoints + performance improvement (e.g. parallel checkpoint taking).
Modification
Catalog:
```
terrier/src/catalog/*
```
Logging:
```
terrier/src/storage/write_ahead_log/*
```
Recovery:
```
terrier/src/storage/recovery/*
```
Checkpoint (new):
```
terrier/src/storage/checkpoint/*
```
Validation of the Work

Tests:

test/storage/recovery_test.cpp

Performance Measurement:

benchmark/storage/checkpoint_recovery_benchmark.cpp
Testing Procedure

1. Disk → Catalog Log → Catalog
2. Disk → Arrow Format (Checkpoint) → Data Table
3. Disk → Logs after Checkpoint
Benchmark Results

Recovery with checkpoint:

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Time</th>
<th>CPU Iterations</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>CheckpointRecoveryBenchmark/ReadWriteWorkload/min_time:10.000/manual_time</td>
<td>804 ms</td>
<td>2144 ms</td>
<td>121.425k items/s</td>
</tr>
<tr>
<td>CheckpointRecoveryBenchmark/HighStress/min_time:10.000/manual_time</td>
<td>720 ms</td>
<td>1328 ms</td>
<td>135.604k items/s</td>
</tr>
</tbody>
</table>

Checkpoint time: 6.9 s

Recovery with log:

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Time</th>
<th>CPU Iterations</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecoveryBenchmark/ReadWriteWorkload/min_time:10.000/manual_time</td>
<td>3164 ms</td>
<td>1441 ms</td>
<td>30.8664k items/s</td>
</tr>
<tr>
<td>RecoveryBenchmark/HighStress/min_time:10.000/manual_time</td>
<td>2671 ms</td>
<td>396 ms</td>
<td>36.5672k items/s</td>
</tr>
</tbody>
</table>
Benchmark Result

Recovery Benchmark Time Comparison

Recovery time (ms)

Number of txn

- RW-R
- HighStress-R
- RW-C
- HighStress-C
Future work

→ Further collaborate w/ data compression group to ensure the block compaction correctness.
→ Better parallelization schema when taking checkpoint.
Thanks