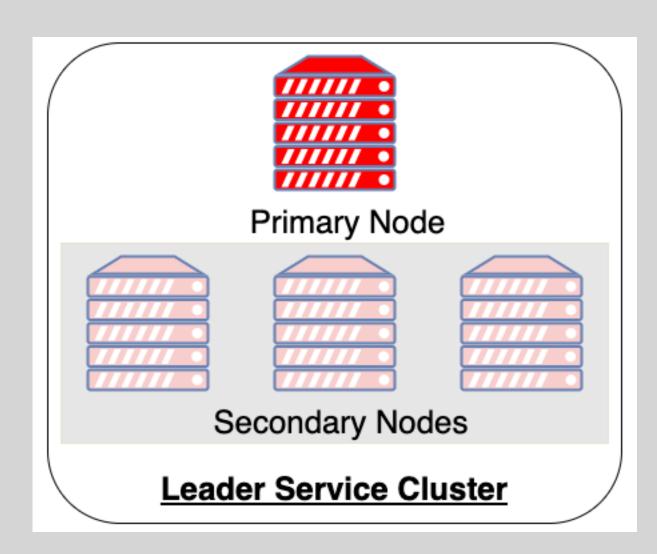
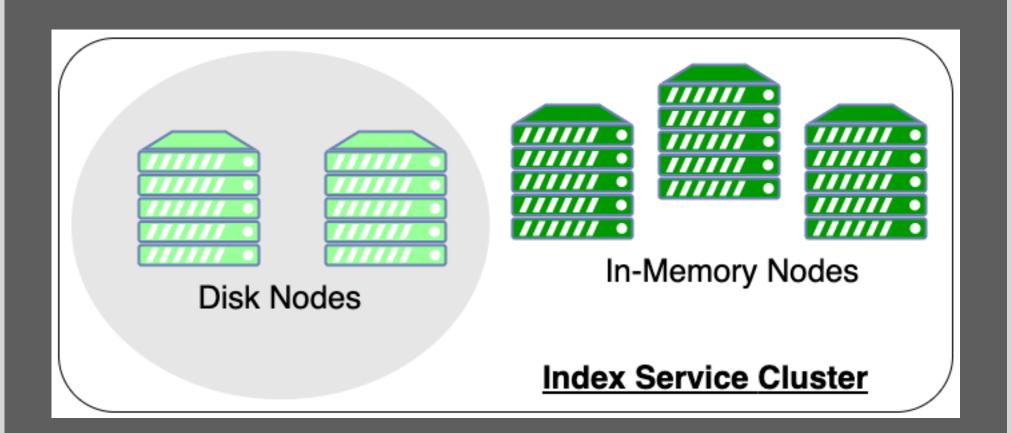
Storage System Design

Preliminary Storage Engine Design - v4, pt2



Leader Services

- Multiple nodes, approx/min 4
- One leader and multiple are back-ups
- Each node has a memtable
- Each node has a WAL



Index Services

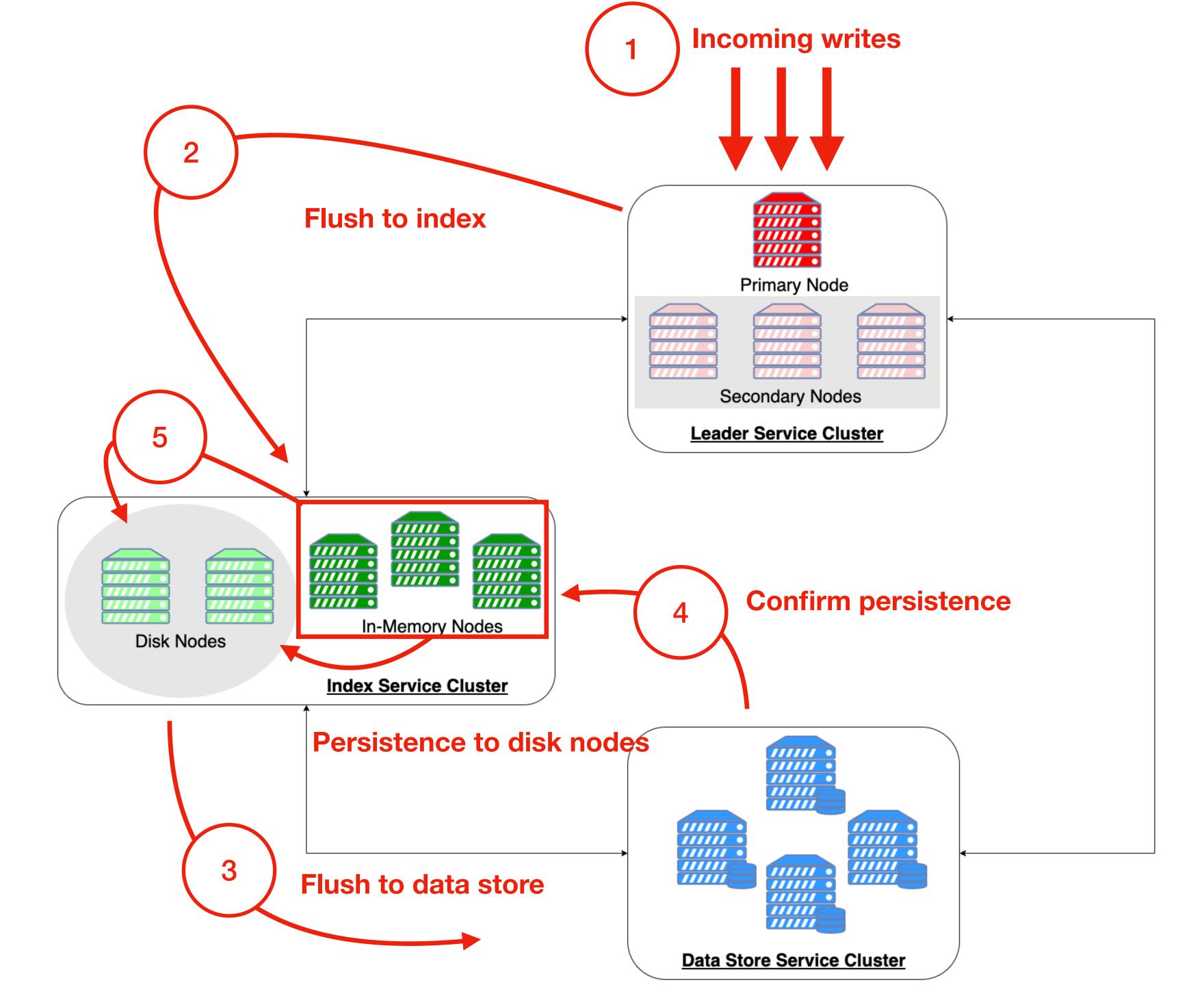
- Multiple nodes, approx/min 5
- Three are exclusively in-memory nodes accepting flushed SSTables from leader service cluster
- Two are exclusively used to persist the indexes into disk from the other three in-memory nodes

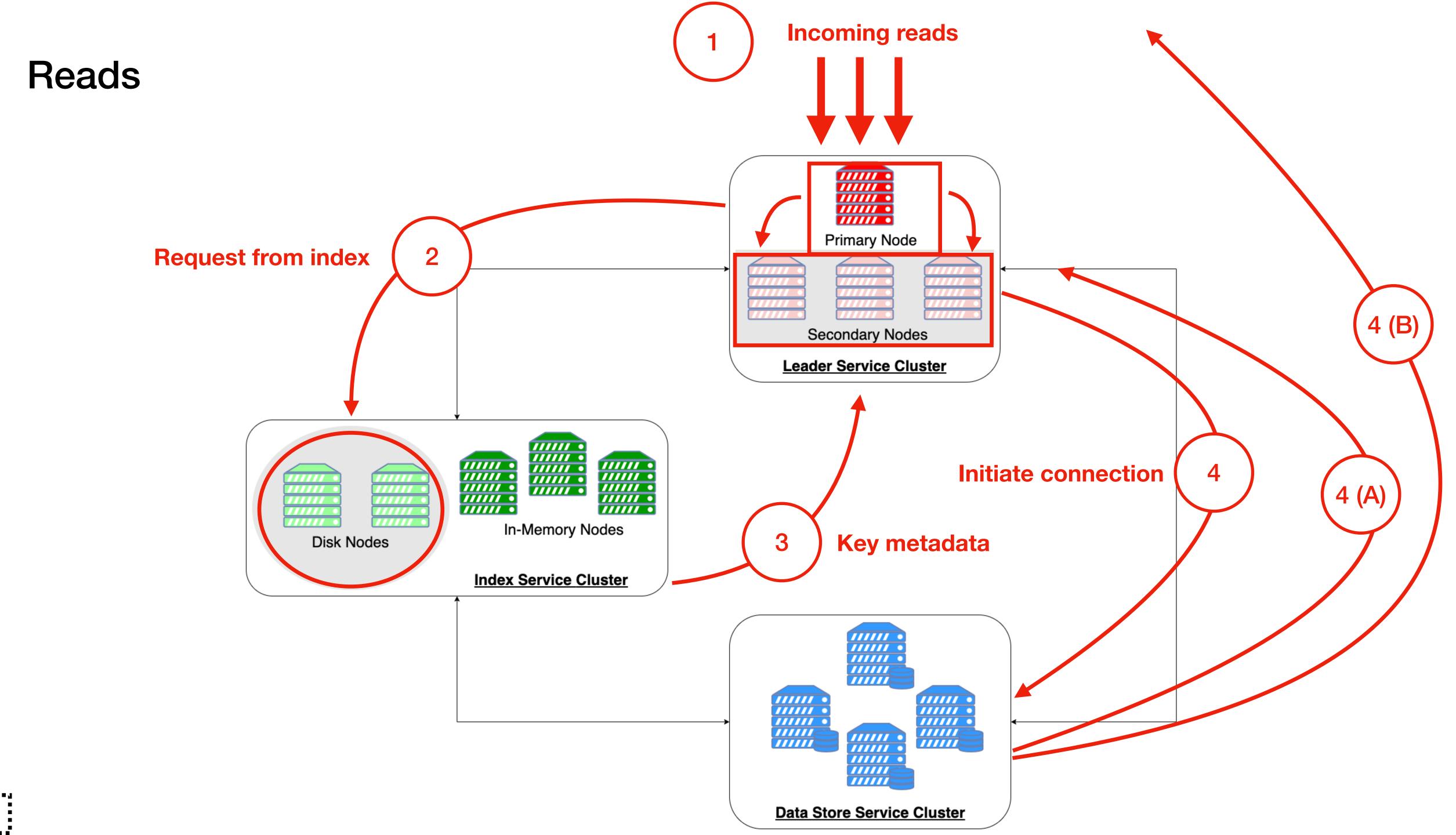


Data Store Services

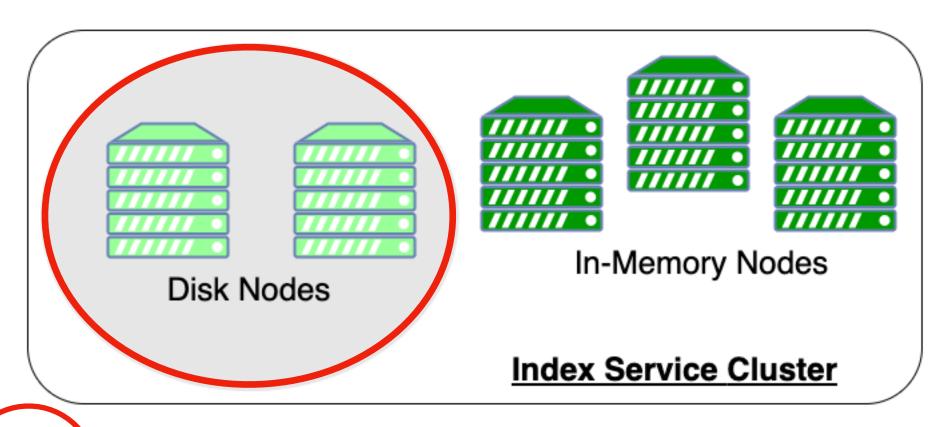
- Multiple nodes, min 4
- All nodes will persist the data object's values received from inmemory index service nodes

Writes/ Updates





Versioning -Merging/Reconciliation



Merging/Reconciliation (Versioning)

Segment 1

1 : [["AA"], 10:30, 1, NULL]

2 : [["BB"], 10:31, 1, NULL]

3 : [["CC"], 10:32, 1, NULL]

4 : [["DD"], 10:33, 1, NULL]

5 : [["FF"], 10:34, 1, NULL]

5 : [["EE"], 10:35, 2, 1]

Segment 2

1 : [["AB"], 10:36, 2, 1]

6 : [["FF"], 10:37, 1, NULL]

1 : [["ABA"], 10:38, 3, 2]

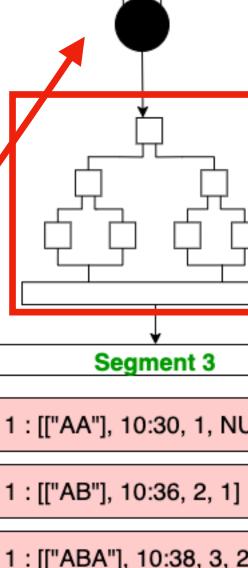
2: [[DEL (X)], 10:39, 2, 1]

7 : [["GG"], 10:40, 1, NULL]

1 : [["AAA"], 10:41, 4, 1]

2

Run keys through Tree DS - two-way sort sequence



1: [["AA"], 10:30, 1, NULL]

1: [["ABA"], 10:38, 3, 2]

1 : [["AAA"], 10:41, 4, 1]

2 : [["BB"], 10:31, 1, NULL]

2: [[DEL (X)], 10:39, 2, 1]

3 : [["CC"], 10:32, 1, NULL]

4 : [["DD"], 10:33, 1, NULL]

5 : [["FF"], 10:34, 1, NULL]

5 : [["EE"], 10:35, 2, 1]

6 : [["FF"], 10:37, 1, NULL]

7 : [["GG"], 10:40, 1, NULL]

Flush keys to new segment

3

Acyclic Graph of Data Object

