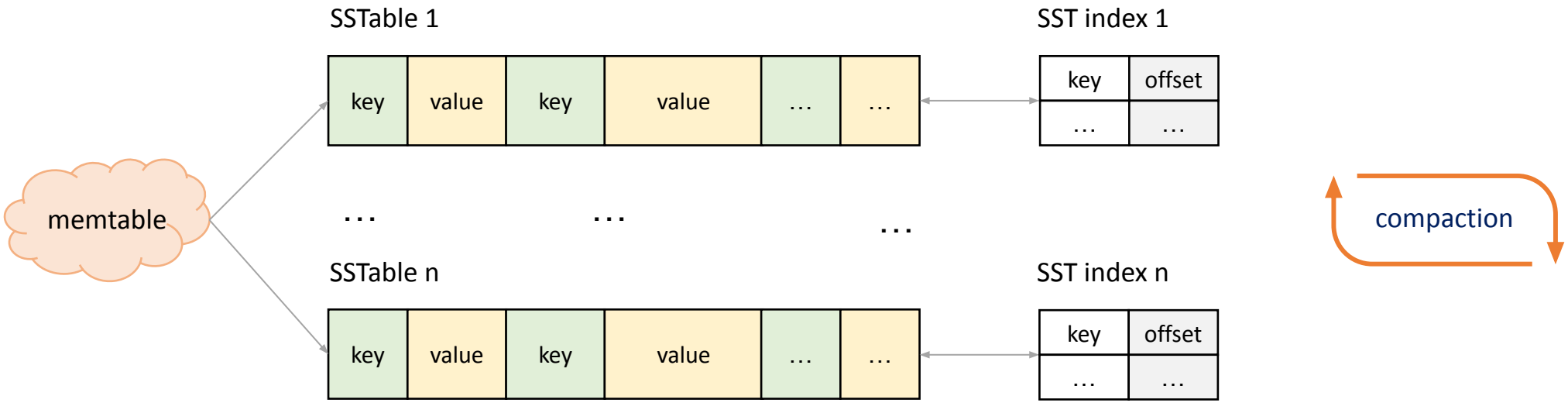
LSM-tree vs B-tree

LSM-tree

log-structured merge-tree



LSM-tree databases

- Google Bigtable
- Apache HBase
- Apache Cassandra
- Google LevelDB
- InfluxDB

B-tree databases

- MySQL
- PostgreSQL
- Apache CouchDB

LSM-tree vs B-tree

LSM-tree		B-tree
✓	faster writes	
✓	less write amplification	
✓	higher write throughput	
	faster reads	✓
	less read amplification	✓
	higher read throughput	✓

write amplification

phenomenon when a single logical write to the database results in multiple writes to disk

read amplification

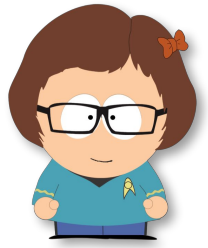
phenomenon when a single logical query results in multiple reads from disk

LSM-tree vs B-tree

For write-heavy systems, I choose an LSM-tree database. And for read-heavy systems, I choose a B-tree database. Right?

software engineer

No.



software engineer

- LSM-tree databases use optimizations to speed up reads (bloom filters, internal cache, read-only memtables)
- sharding helps to increase write throughput for both LSM-tree and B-tree databases
- distributed cache in front of an LSM or B-tree database helps to increase read throughput and reduce latency

LSM-tree vs B-tree

LSM-tree		B-tree
✓	faster writes	
✓	less write amplification	
✓	higher write throughput	
	faster reads	✓
	less read amplification	✓
	higher read throughput	✓
	more predictable performance	✓
✓	less disk fragmentation	