

## 内容大纲

- ▶ 总览
- ▶ 集群复制
- ▶ 快照
- ▶ 导入导出
- ▶ 复制表

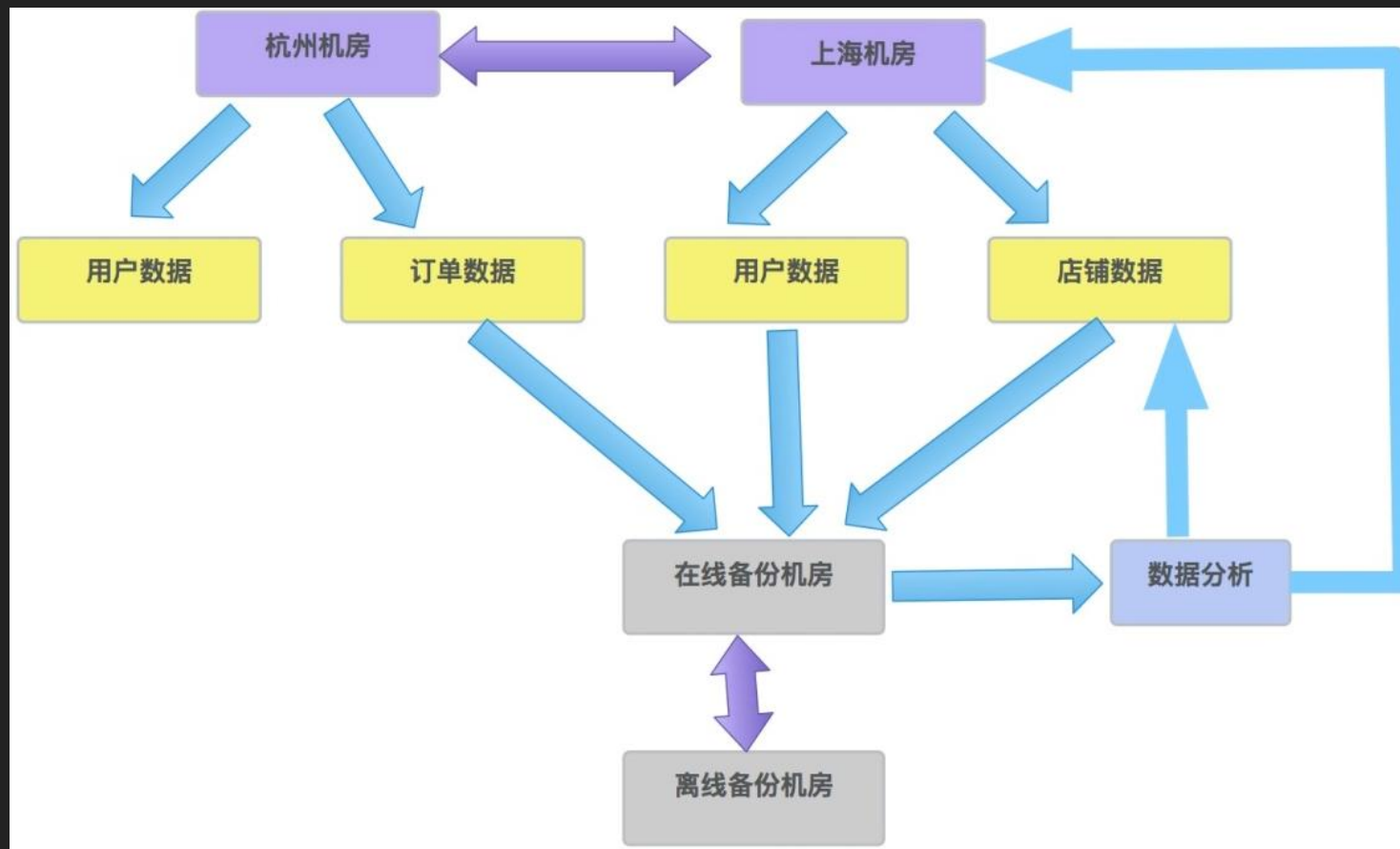
## 总览

方式	性能影响	停机	增量备份	实现难度	恢复时长
复制 (replication)	很小	不需要	实时	较容易	很小
快照 (snapshots)	很小	短暂 (恢复时)	不支持	容易	很小
导出导入 (export/import)	大	不需要	支持	容易	较长
复制表 (copy table)	大	不需要	支持	容易	较长

- 复制：通过Replication实现主从同步，主要用于灾难恢复、单元化等场景。
- 快照：数据的全量、增量备份，比如每天、每周、每月数据的备份，以用于数据损坏恢复，如指标计算错误可以从某个时间点重新计算等。
- 导出导入：可以用来将数据从A集群导出，然后导入到B集群，比如从线上导出到线下以便做模拟测试。
- 复制表：数据的备份与迁移，比如复制特定的数据子集，并将其加载到另一个表中。这对于数据挖掘、数据分析和定期数据快照非常有用。

# HBase入门与实践

## 复制



### ➤ 开启复制

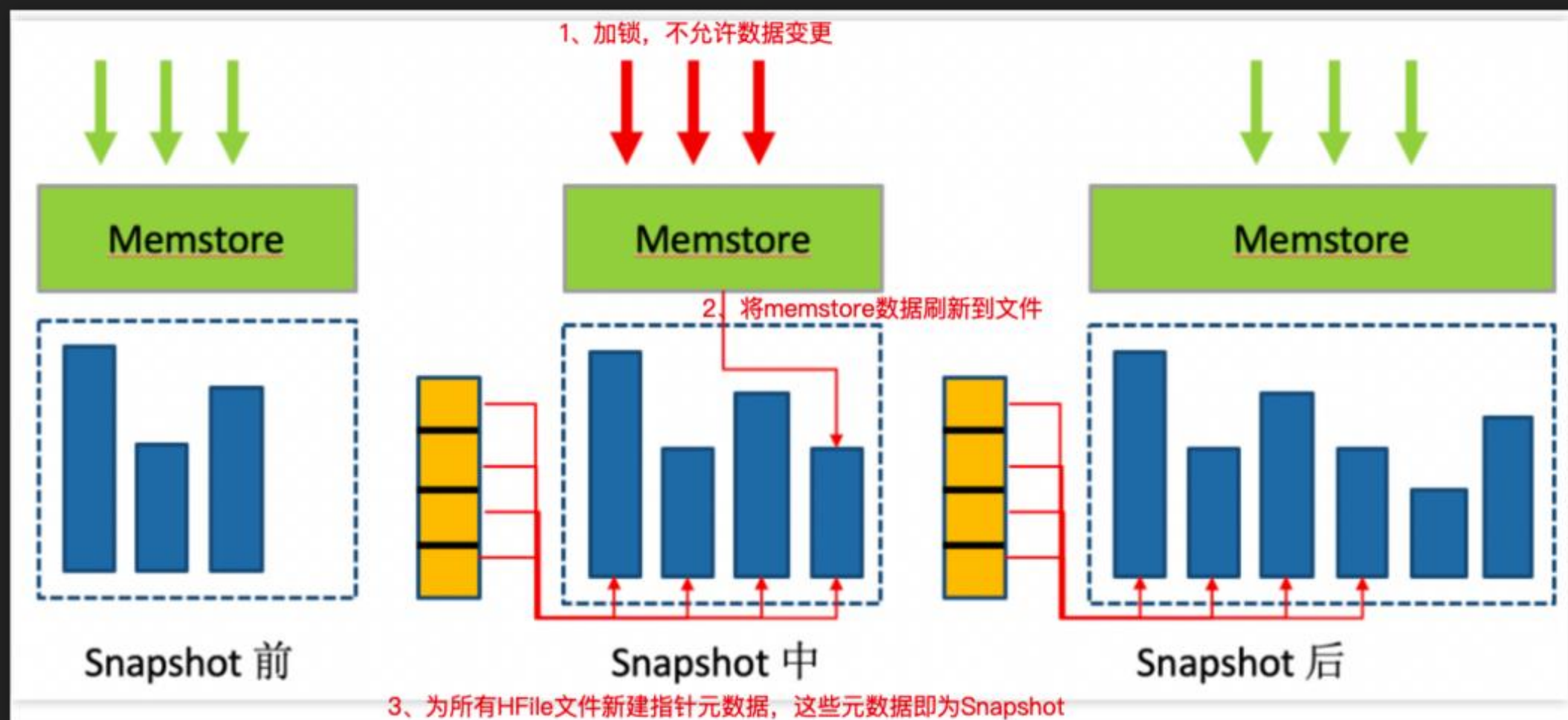
➤ hbase-site.xml配置项hbase.replication=true

➤ 添加集群复制关系: `add_peer '2', CLUSTER_KEY => "master1, master2, slave1:2181:/hbase"`

➤ 开启列族复制:

```
create 'rep_test', {NAME => 'cf', REPLICATION_SCOPE => '1', KEEP_DELETED_CELLS => 'TRUE'}
```

## 快照



### ➤ 开启Snapshot

- hbase-site.xml配置项hbase.snapshot.enabled=true
- snapshot 'tableName', 'snapshotName'

## 导入导出

- 导出：导出到HDFS或者本地文件系统。默认导出到HDFS
  - `./hbase org.apache.hadoop.hbase.mapreduce.Export 'tablename' hdfs:///filepath`
  - `./hbase org.apache.hadoop.hbase.mapreduce.Export 'tablename' file:///filepath`
- 导入：
  - `./hbase org.apache.hadoop.hbase.mapreduce.Import 'tablename' hdfs:///filepath`
  - `./hbase org.apache.hadoop.hbase.mapreduce.Import 'tablename' file:///filepath`

## 复制表

- HBase CopyTable支持同一个集群中不同表之间数据的复制，也可以将一个集群中表的数据复制到另一个集群的表，复制之前需要先创建表
- 同集群复制
  - `./hbase org.apache.hadoop.hbase.mapreduce.CopyTable --new.name=test_copy test`
- 跨集群复制：
  - `./hbase org.apache.hadoop.hbase.mapreduce.CopyTable --peer.adr=omaster1,omaster2,oslave1:2181:/hbase --new.name=test_copy test`