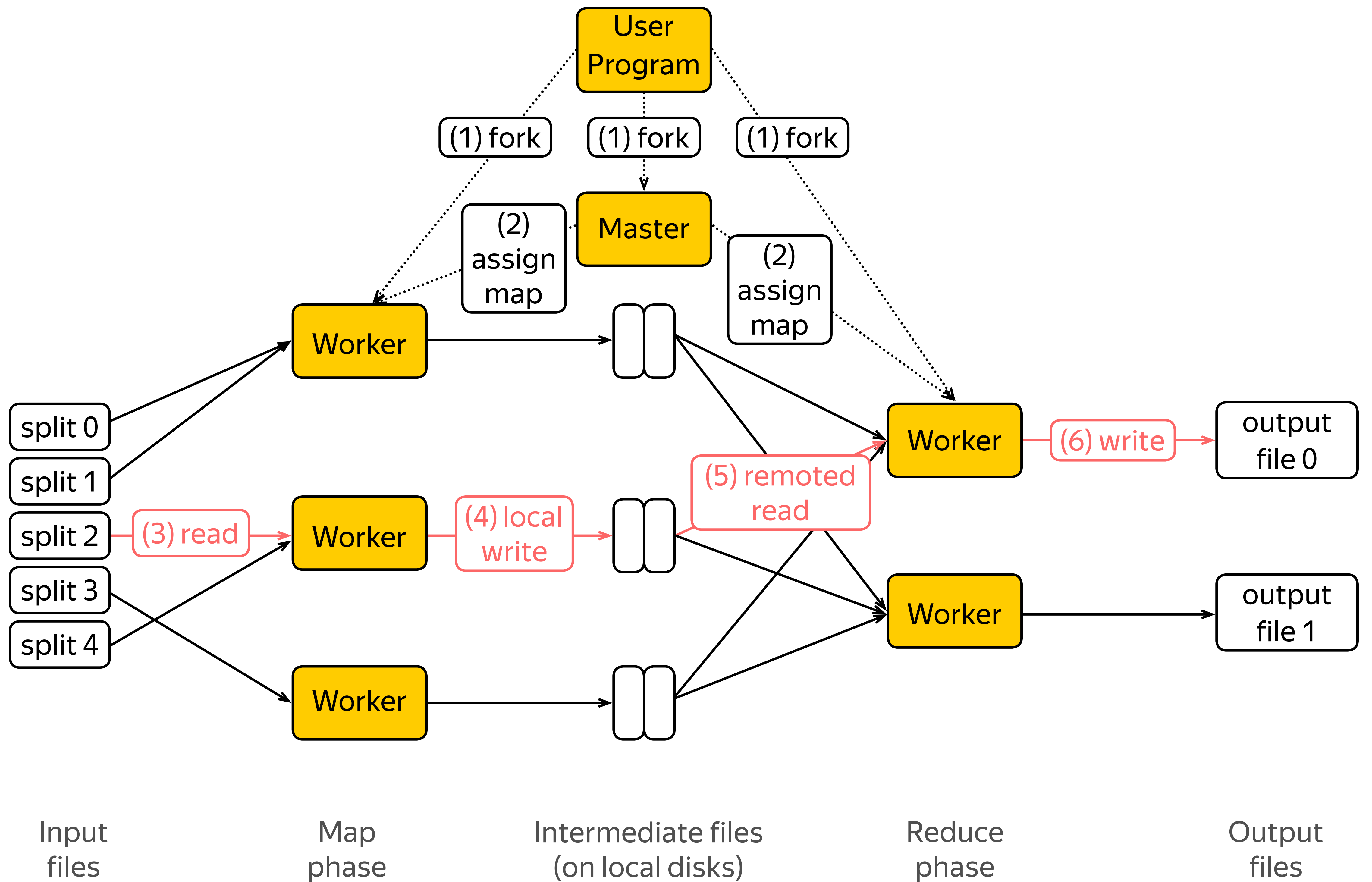
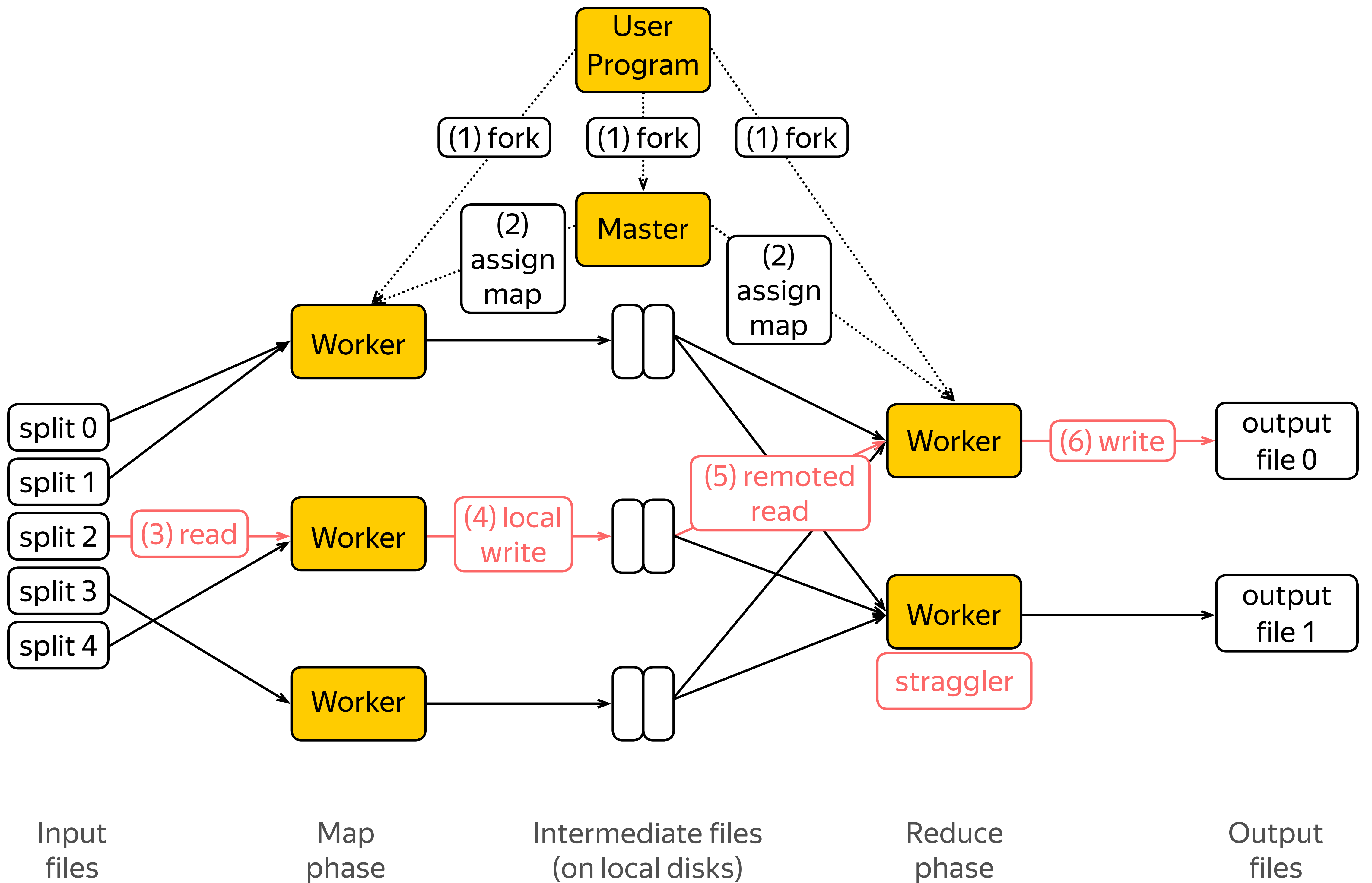


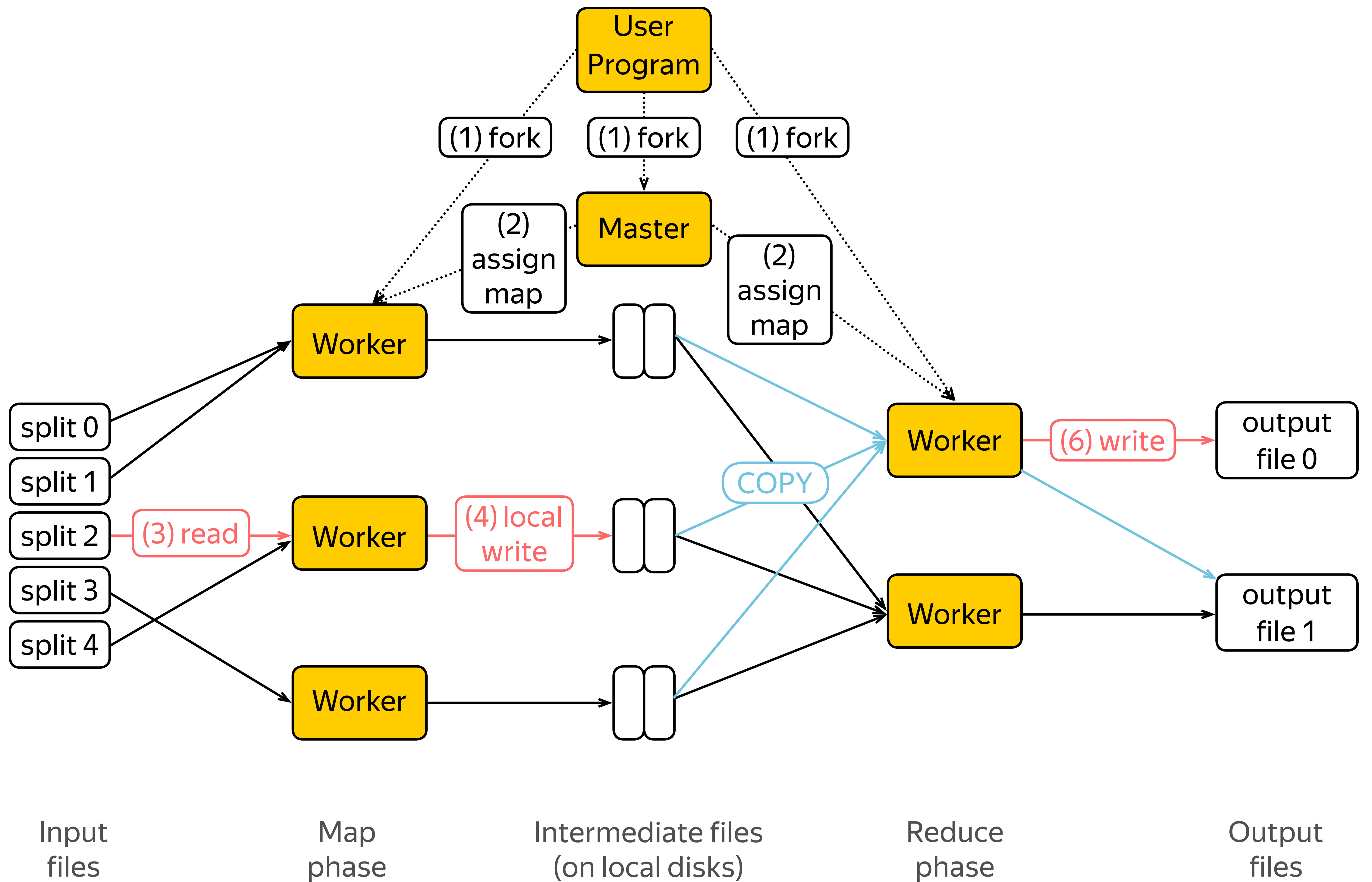
Yandex

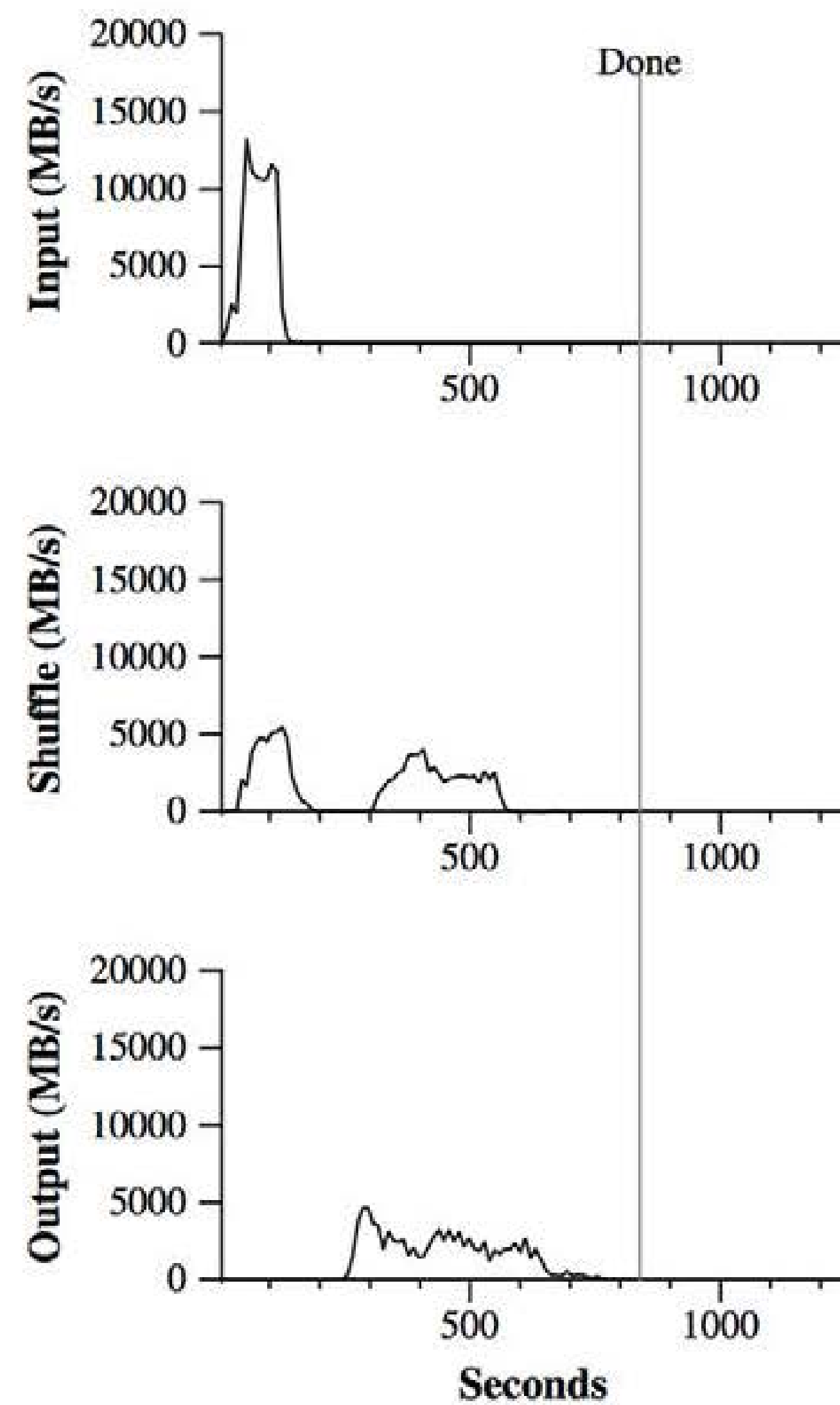
MapReduce

Speculative Execution / Backup Tasks









(a) Normal execution

(b) No backup tasks

(c) 200 tasks killed

Figure 3: Data transfer rates over time for different executions of the sort program

“MapReduce: Simplified Data Processing on Large Clusters”
by Jeffrey Dean and Sanjay Ghemawat

30-40% faster

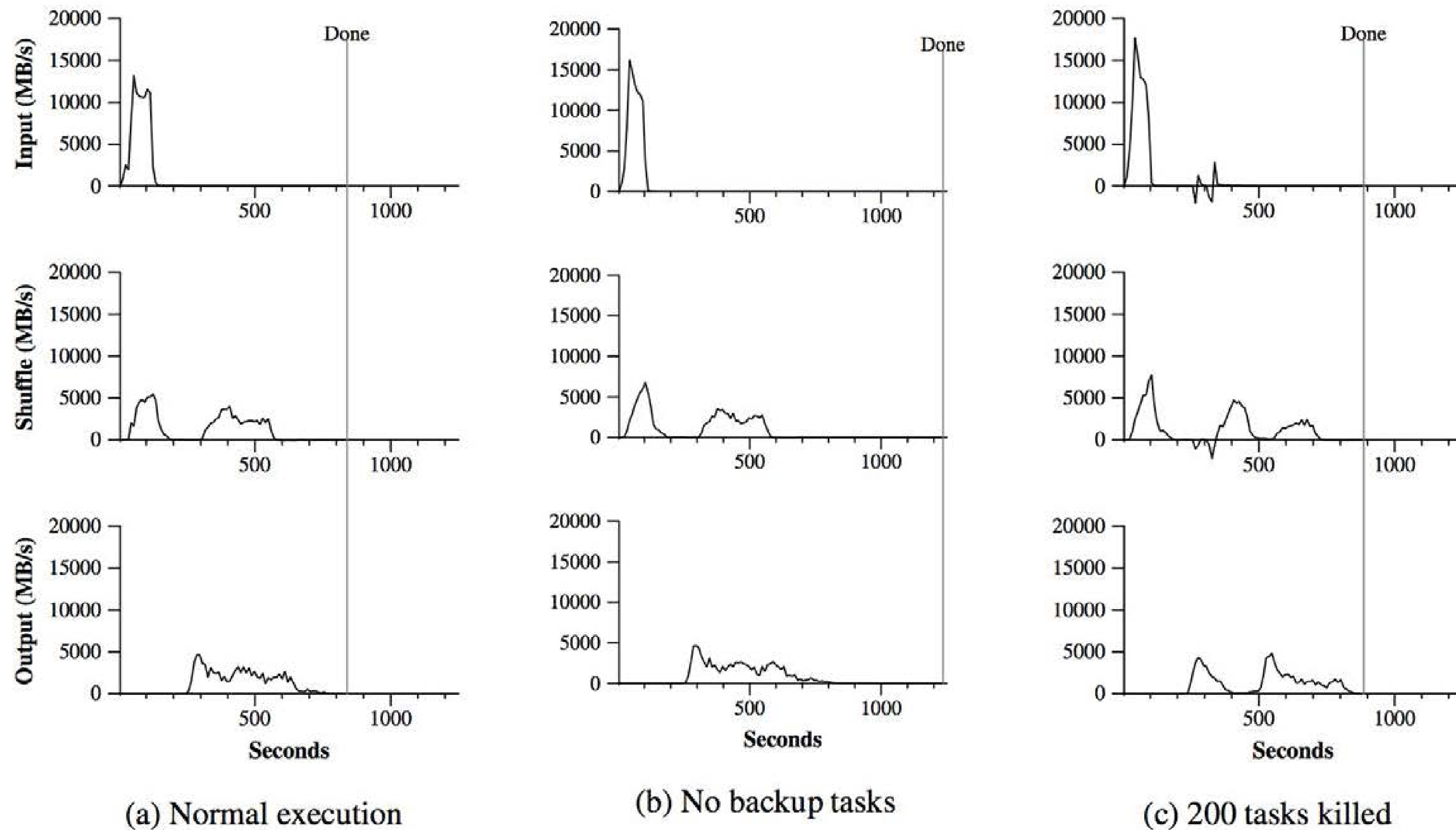


Figure 3: Data transfer rates over time for different executions of the sort program

“MapReduce: Simplified Data Processing on Large Clusters”
by Jeffrey Dean and Sanjay Ghemawat

```
yarn jar $MR_STREAMING_JAR \  
-D mapreduce.map.speculative="true" \  
-D mapreduce.job.speculative.speculative-cap-running-tasks=0.99 \  
-D mapreduce.job.speculative.speculative-cap-total-tasks=0.99 \  
-D mapreduce.job.speculative.retry-after-no-speculate=10 \  
-files flaky_mapper.py \  
-mapper 'python flaky_mapper.py' \  
-output flaky_mr \  
-numReduceTasks 0 \  
-input /data/wiki/en_articles_part
```



```
yarn jar $MR_STREAMING_JAR \  
→ -D mapreduce.map.speculative="true" \  
  -D mapreduce.job.speculative.speculative-cap-running-tasks=0.99 \  
  -D mapreduce.job.speculative.speculative-cap-total-tasks=0.99 \  
  -D mapreduce.job.speculative.retry-after-no-speculate=10 \  
  -files flaky_mapper.py \  
  -mapper 'python flaky_mapper.py' \  
  -output flaky_mr \  
  -numReduceTasks 0 \  
  -input /data/wiki/en_articles_part
```

- › mapreduce.map.speculative (default:"true")
- › mapreduce.reduce.speculative (default:"true")

```
yarn jar $MR_STREAMING_JAR \  
  -D mapreduce.map.speculative="true" \  
→ -D mapreduce.job.speculative.speculative-cap-running-tasks=0.99 \  
→ -D mapreduce.job.speculative.speculative-cap-total-tasks=0.99 \  
  -D mapreduce.job.speculative.retry-after-no-speculate=10 \  
  -files flaky_mapper.py \  
  -mapper 'python flaky_mapper.py' \  
  -output flaky_mr \  
  -numReduceTasks 0 \  
  -input /data/wiki/en_articles_part
```

- › mapreduce.job.speculative.speculative-cap-running-tasks
 - › (default: 0.1)
- › mapreduce.job.speculative.speculative-cap-total-tasks
 - › (default: 0.01)

```
yarn jar $MR_STREAMING_JAR \  
  -D mapreduce.map.speculative="true" \  
  -D mapreduce.job.speculative.speculative-cap-running-tasks=0.99 \  
  -D mapreduce.job.speculative.speculative-cap-total-tasks=0.99 \  
→ -D mapreduce.job.speculative.retry-after-no-speculate=10 \  
  -files flaky_mapper.py \  
  -mapper 'python flaky_mapper.py' \  
  -output flaky_mr \  
  -numReduceTasks 0 \  
  -input /data/wiki/en_articles_part
```

- › mapreduce.job.speculative.retry-after-no-speculate
 - › (default: 1000, ms)
- › mapreduce.job.speculative.retry-after-speculate
 - › (default: 15000, ms)



Logged in as: dr.who

Attempts for task_1490709912648_0465_m_000001

- Application
- Job
- Task
 - Task Overview
 - Counters
- Tools

Show 20 entries

Attempt	State	Status	Node	Logs	Start Time	Finish Time	Elapsed Time	Note
attempt_1490709912648_0465_m_000001_0	FAILED		/default/virtual-node3.atp-fivt.org:8042	logs	Sun Apr 2 17:04:10 +0300 2017	Sun Apr 2 17:04:15 +0300 2017	5sec	Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 1 at ...
attempt_1490709912648_0465_m_000001_1	FAILED		/default/virtual-node1.atp-fivt.org:8042	logs	Sun Apr 2 17:04:17 +0300 2017	Sun Apr 2 17:04:21 +0300 2017	3sec	Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 1 at ...
attempt_1490709912648_0465_m_000001_2	KILLED		/default/virtual-node1.atp-fivt.org:8042	logs	Sun Apr 2 17:04:22 +0300 2017	Sun Apr 2 17:04:26 +0300 2017	4sec	Speculation: attempt_1490709912648_0465_m_000001_3 succeeded first!
attempt_1490709912648_0465_m_000001_3	SUCCEEDED	hdfs://virtual-master.atp-fivt.org:8020/data/wiki/en_articles_part/articles-part:0+38430992	/default/virtual-node1.atp-fivt.org:8042	logs	Sun Apr 2 17:04:22 +0300 2017	Sun Apr 2 17:04:26 +0300 2017	4sec	

Showing 1 to 4 of 4 entries

First Previous 1 Next Last

attempt_1490709912648_0465_m_000001_2

KILLED

Speculation:
attempt_1490709912648_0465_m_000001_3
succeeded first!

Summary

- › you know **how to use** speculative execution to speed-up MapReduce application

Summary

- › you know how to use speculative execution to speed-up MapReduce application
- › see more configuration options in `mapred-default.xml`

BigDATAteam