



MapReduce и Fault Tolerance

Драль Алексей, study@bigdatateam.org

CEO at BigData Team, <https://bigdatateam.org>

<https://www.facebook.com/bigdatateam>



**BIGDATA
TEAM**

MapReduce

**User
Program**

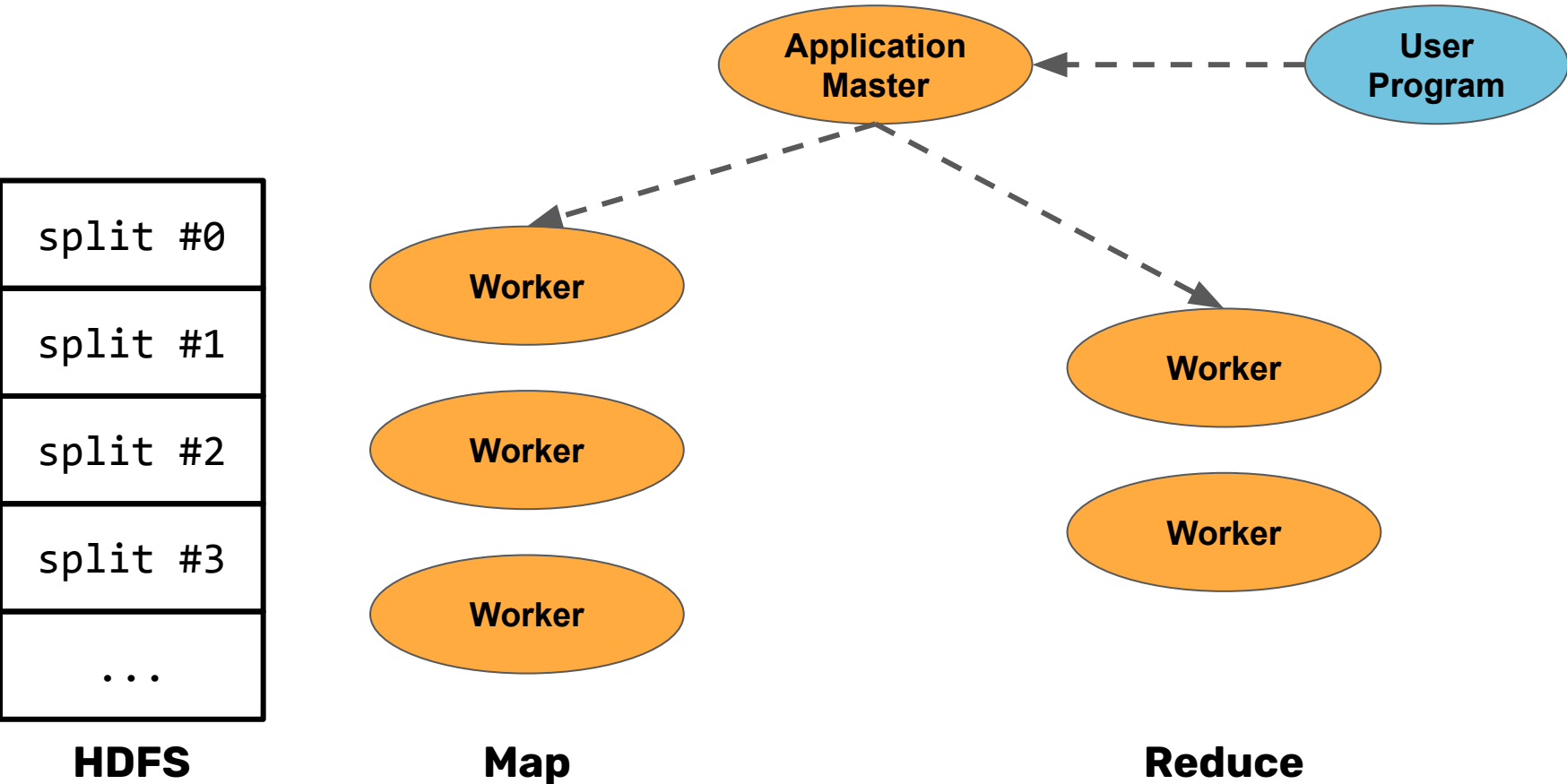


MapReduce



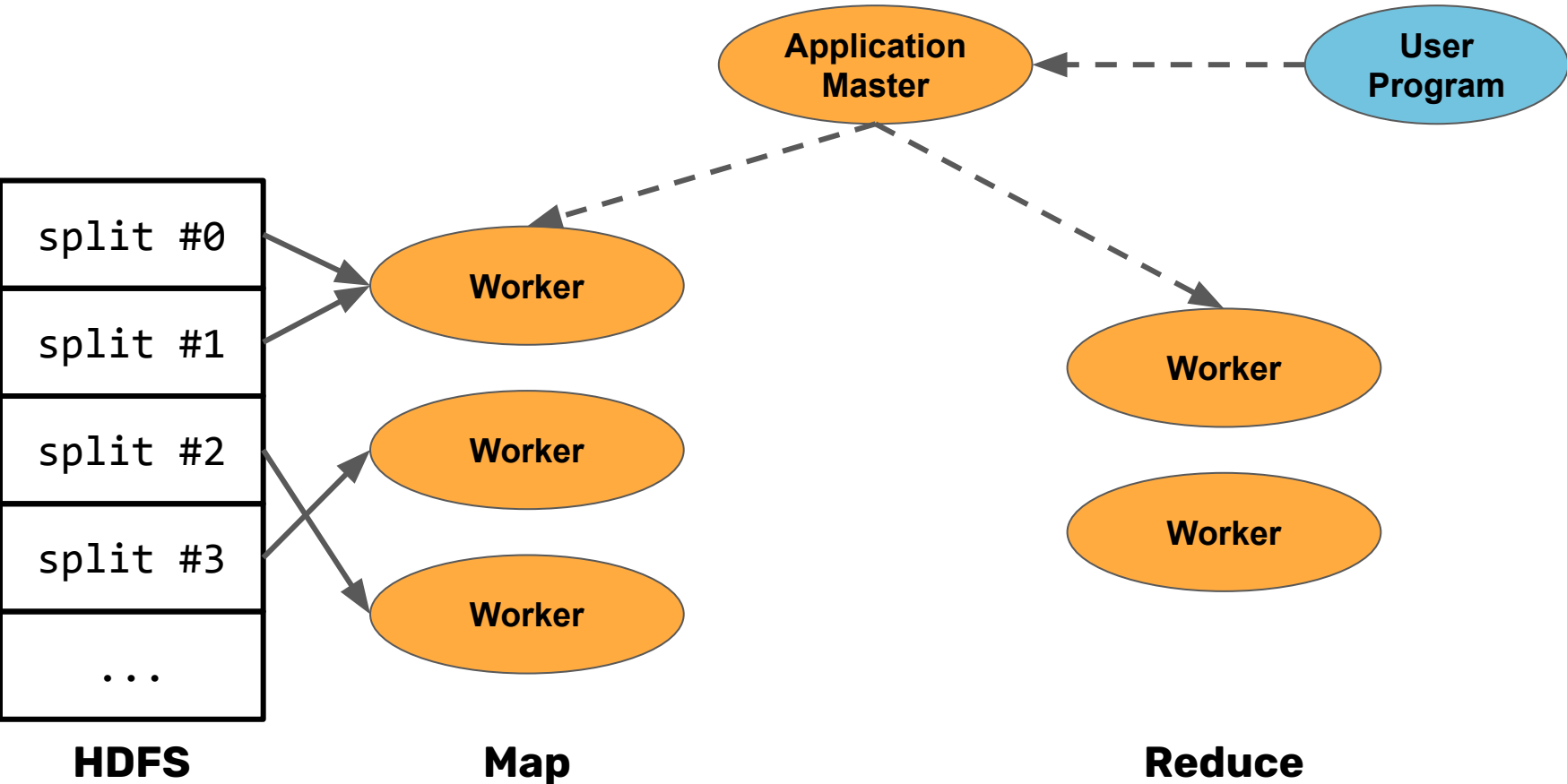


MapReduce



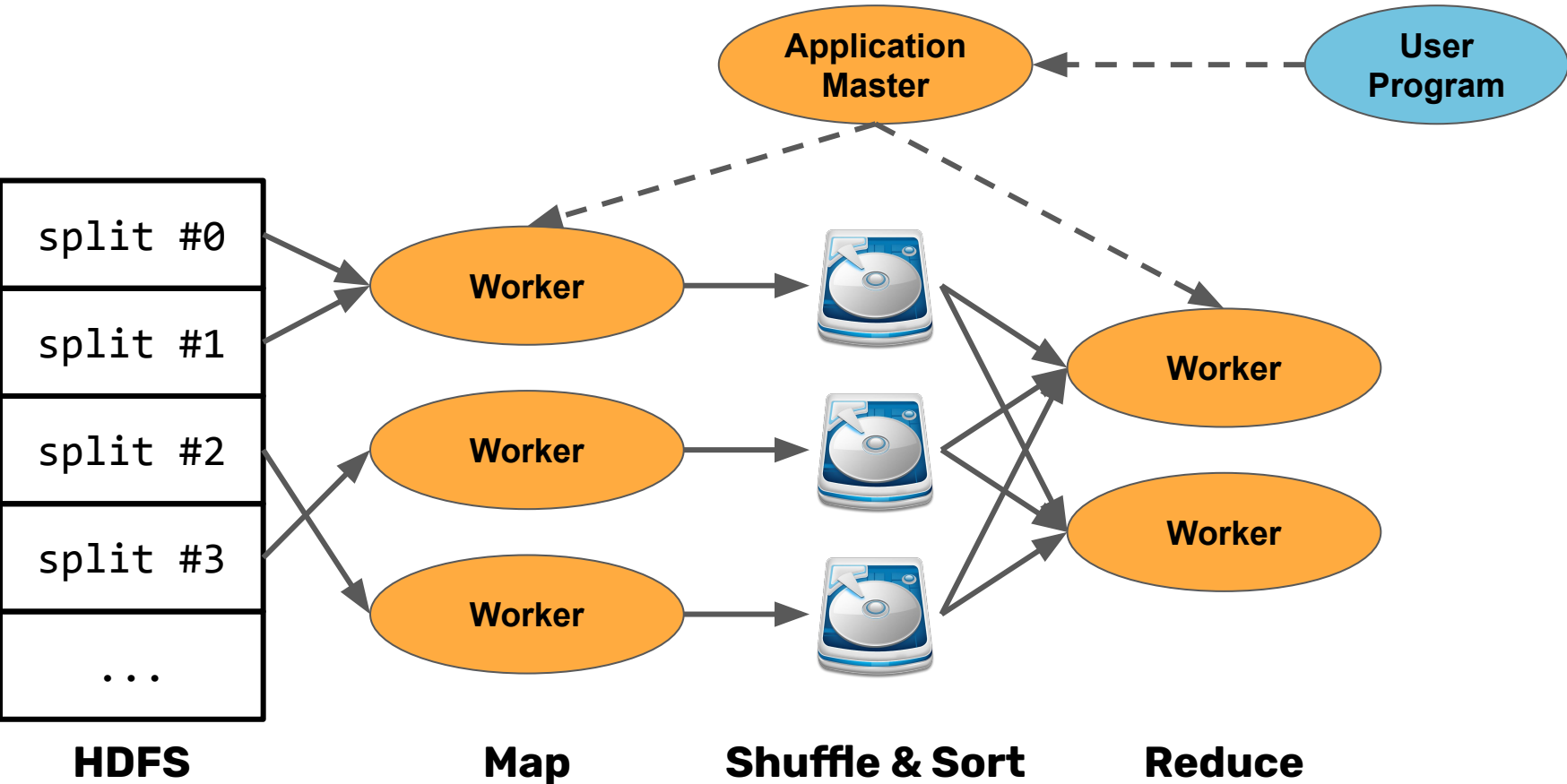


MapReduce



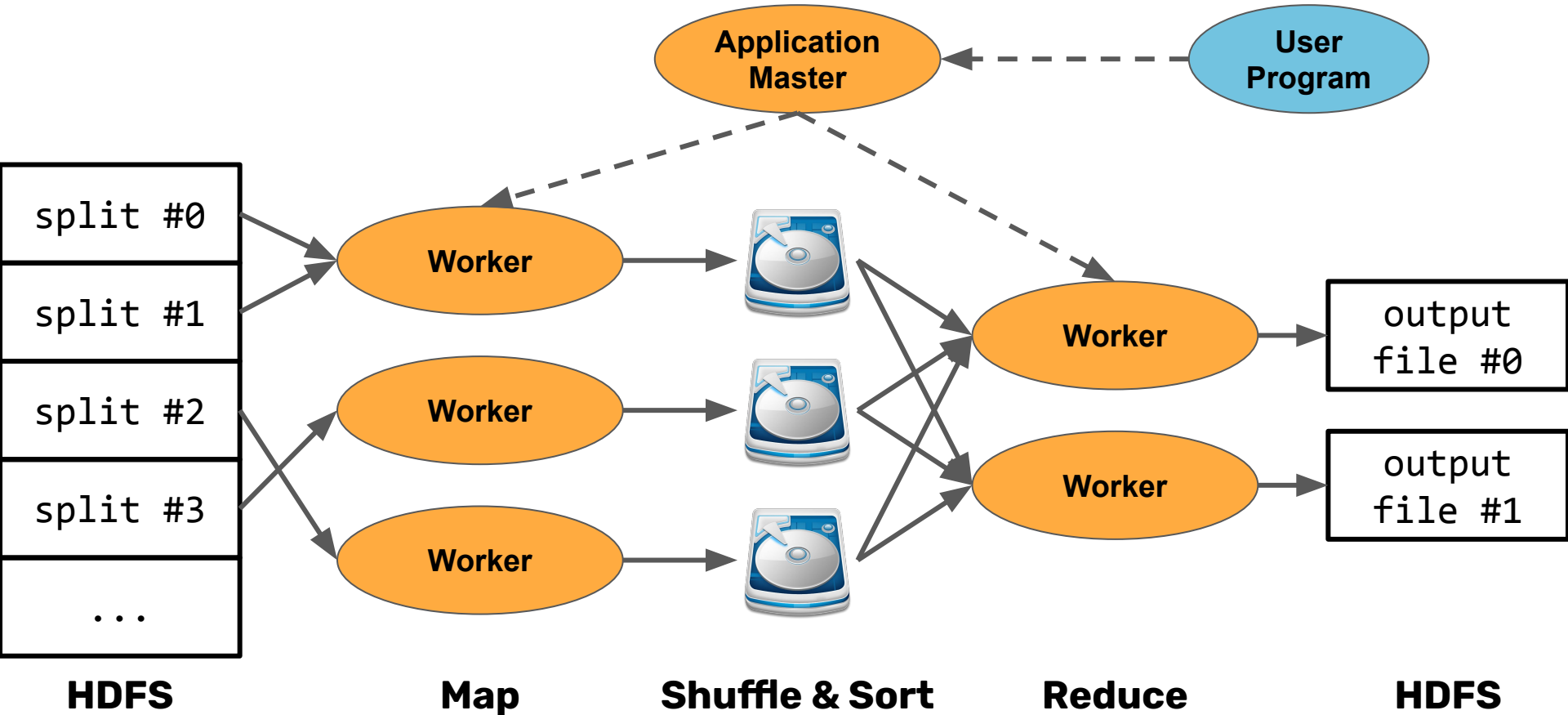


MapReduce



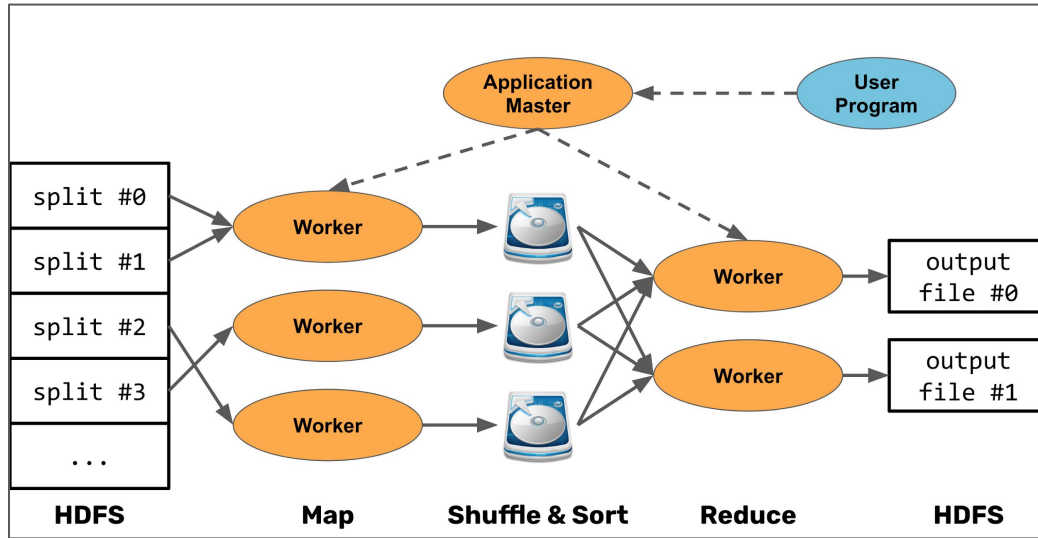


MapReduce





Где храним Shuffle & Sort?



А.

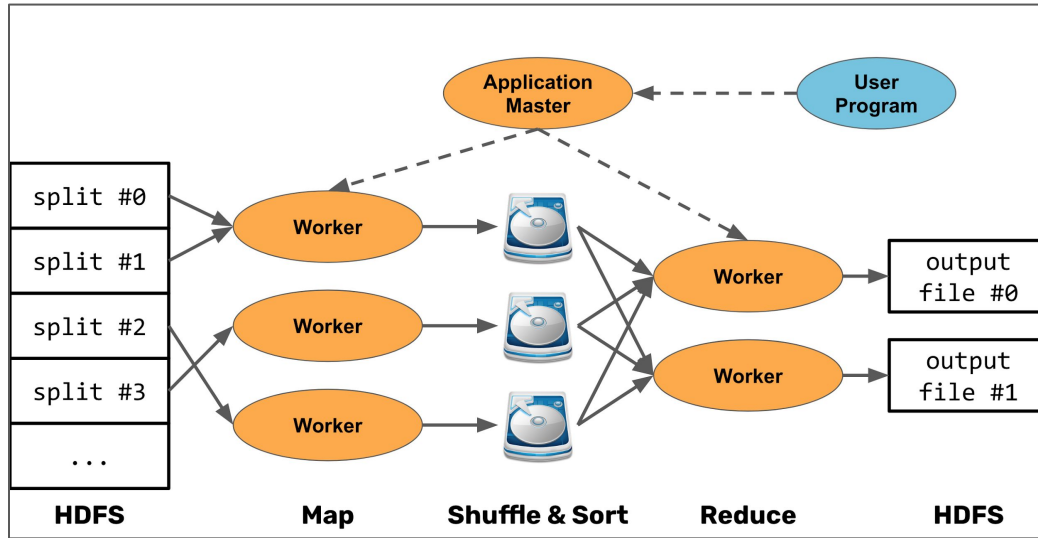
Б.

В.

Г.



Где храним Shuffle & Sort?



A. RAM

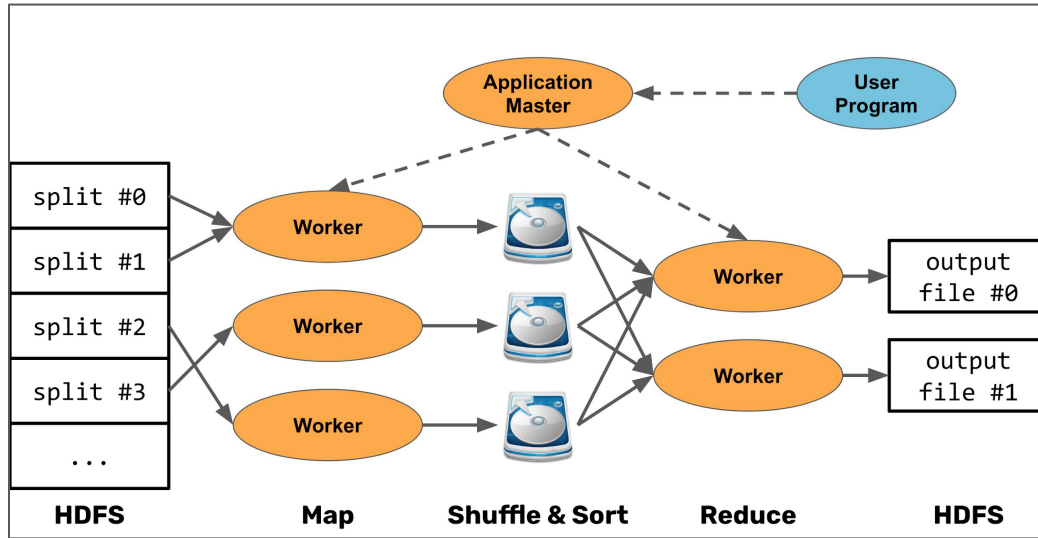
Б.

В.

Г.



Где храним Shuffle & Sort?



A. RAM

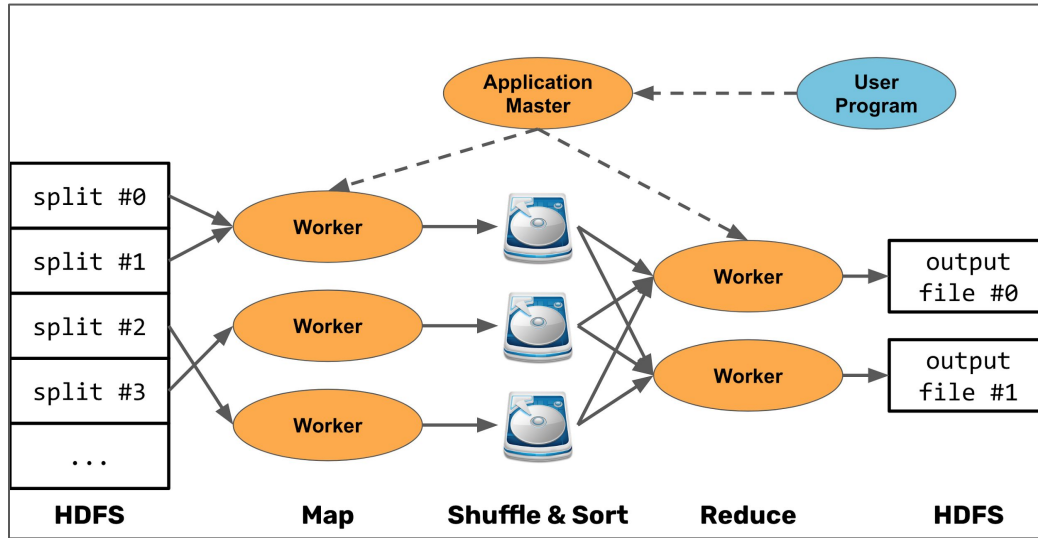
Б. HDFS

В.

Г.



Где храним Shuffle & Sort?



A. RAM

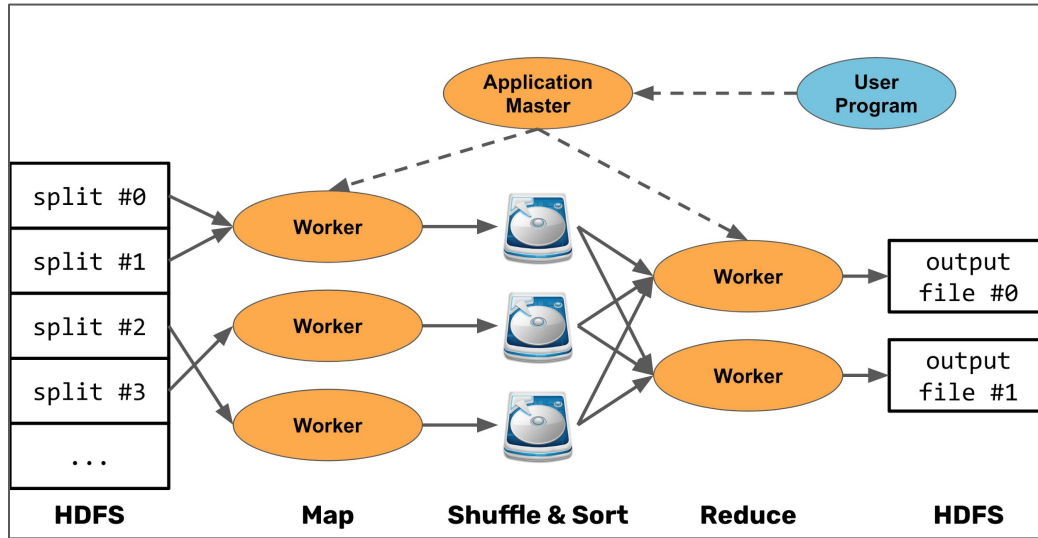
Б. HDFS

В. Local FS

Г.



Где храним Shuffle & Sort?



A. RAM

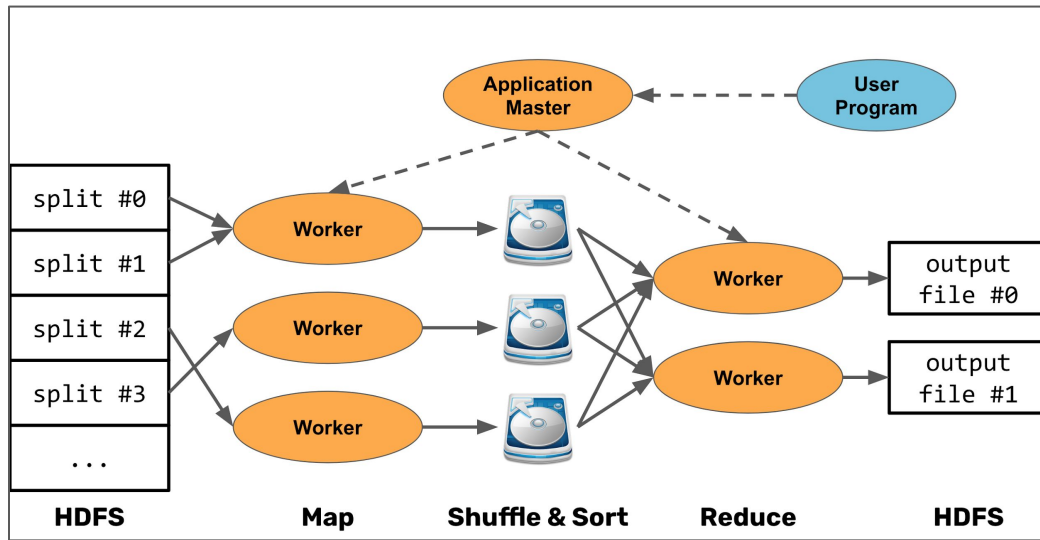
Б. HDFS

В. Local FS

Г. Где-то там



Где храним Shuffle & Sort?



A. RAM

Б. HDFS

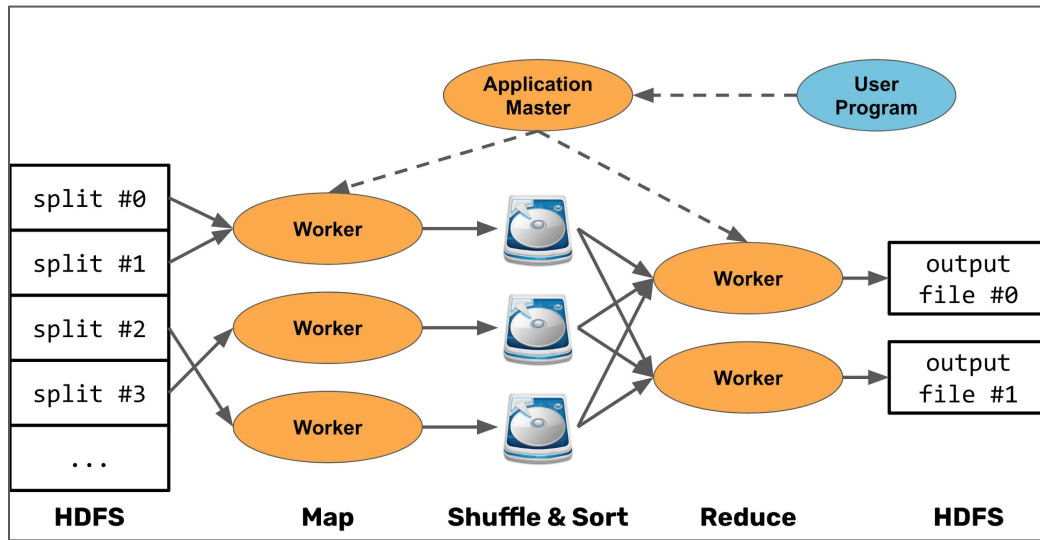
В. Local FS

Г. Где-то там





Где храним Shuffle & Sort?



A. RAM

Б. HDFS

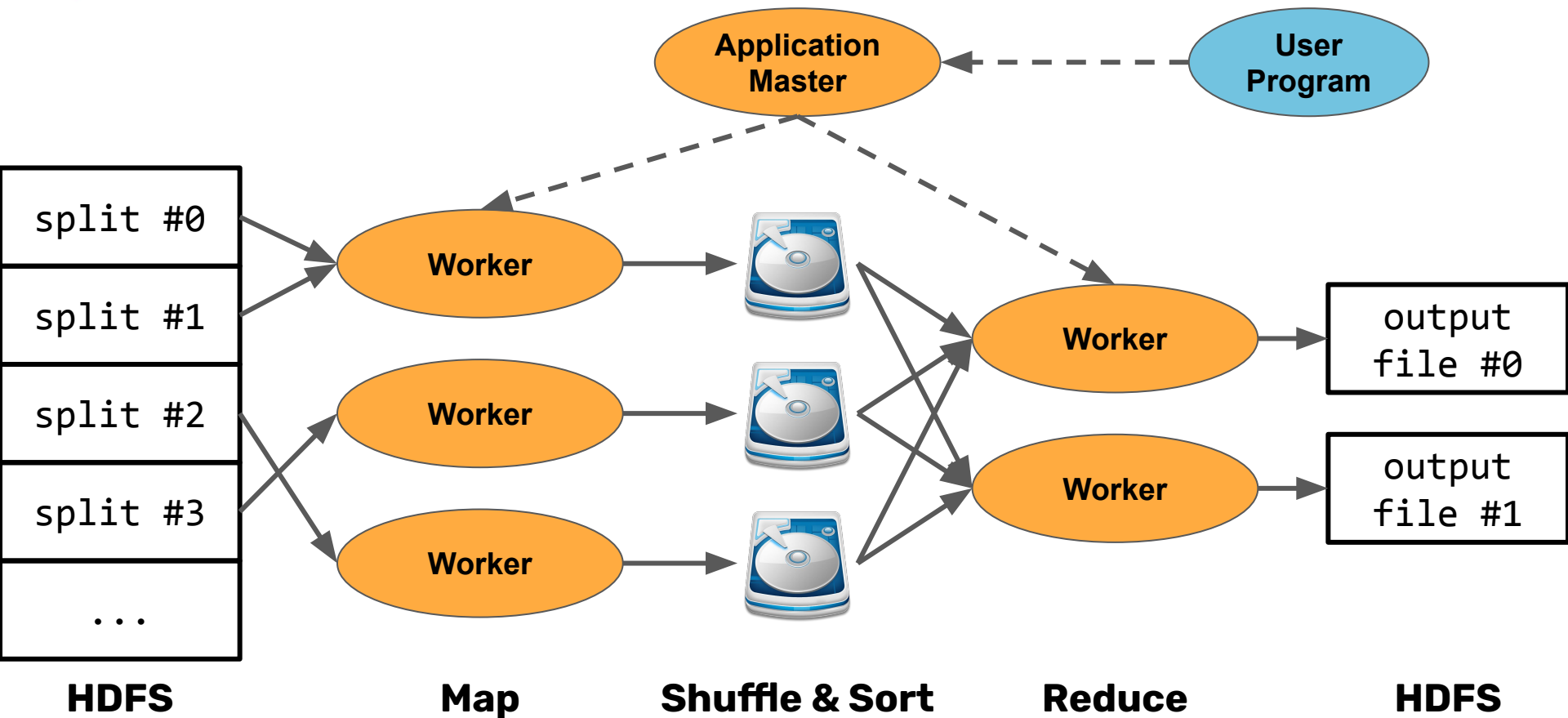
В. Local FS

Г. Где-то там



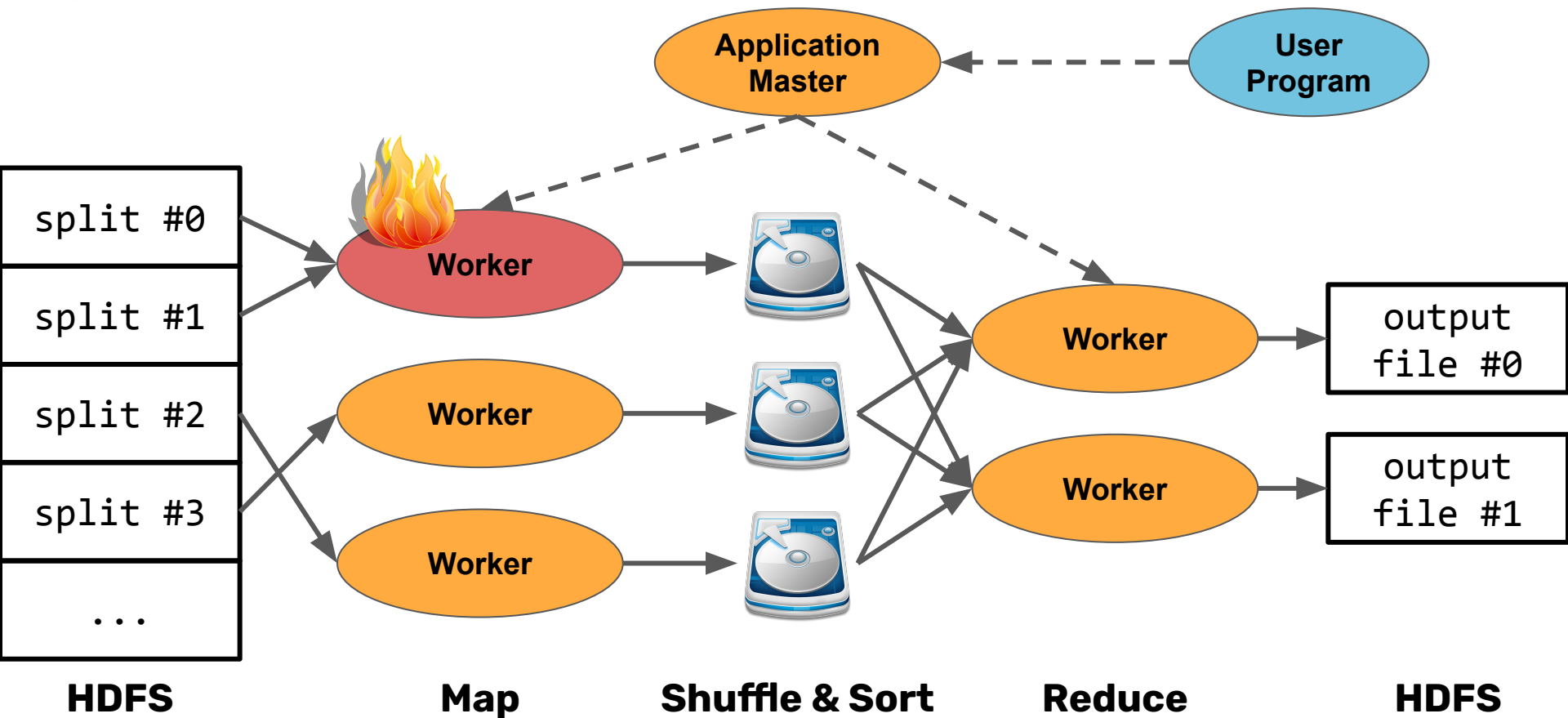


Fault Tolerance (защита от сбоев)



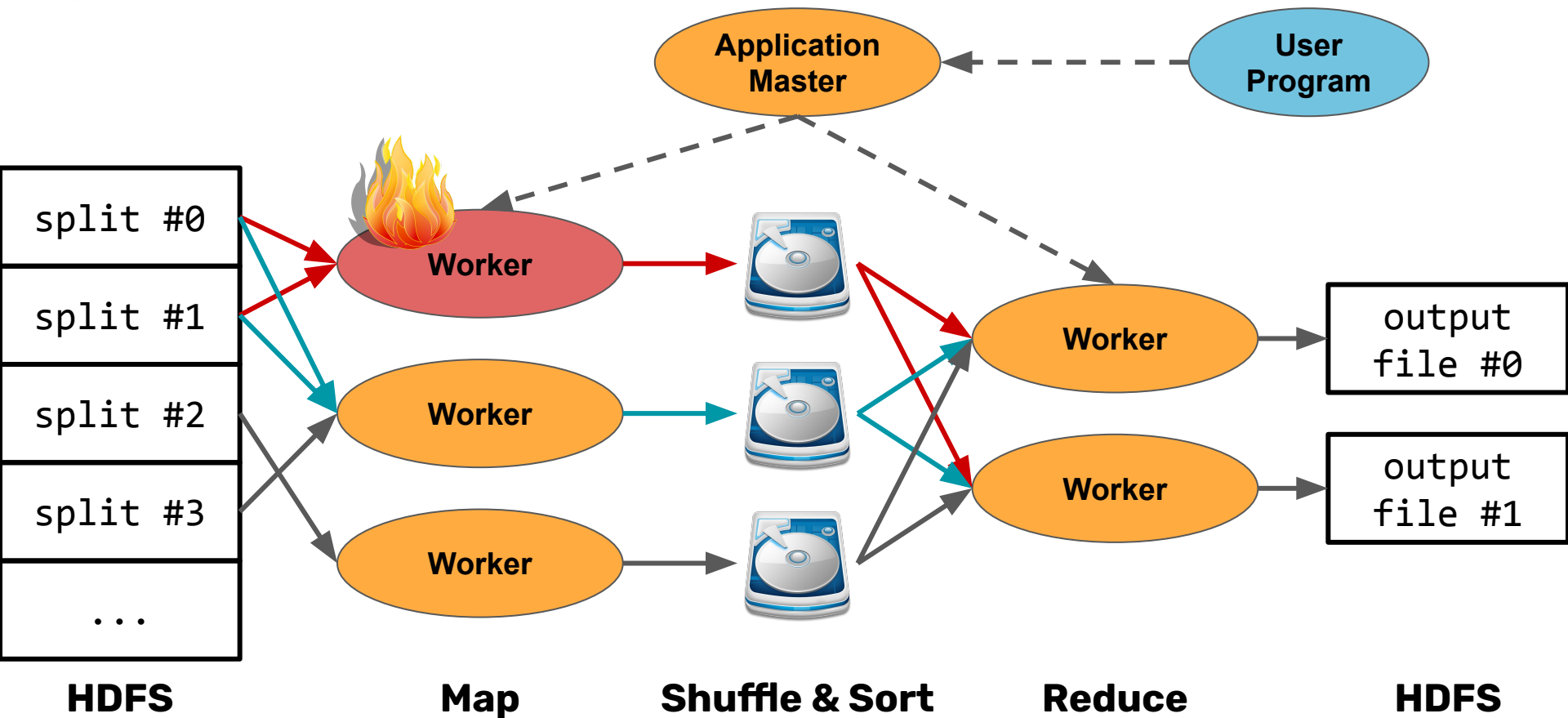


Fault Tolerance (защита от сбоев)



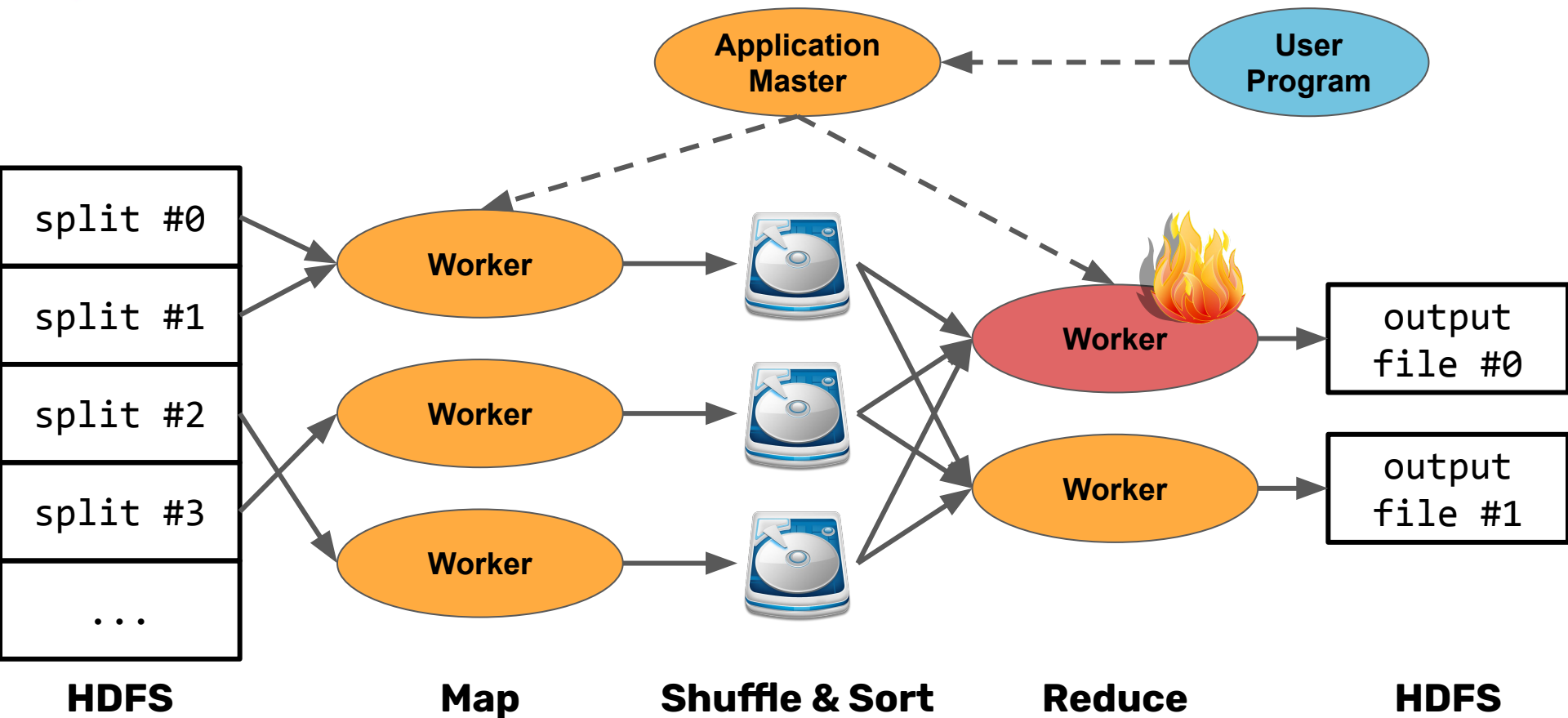


Fault Tolerance (защита от сбоев)



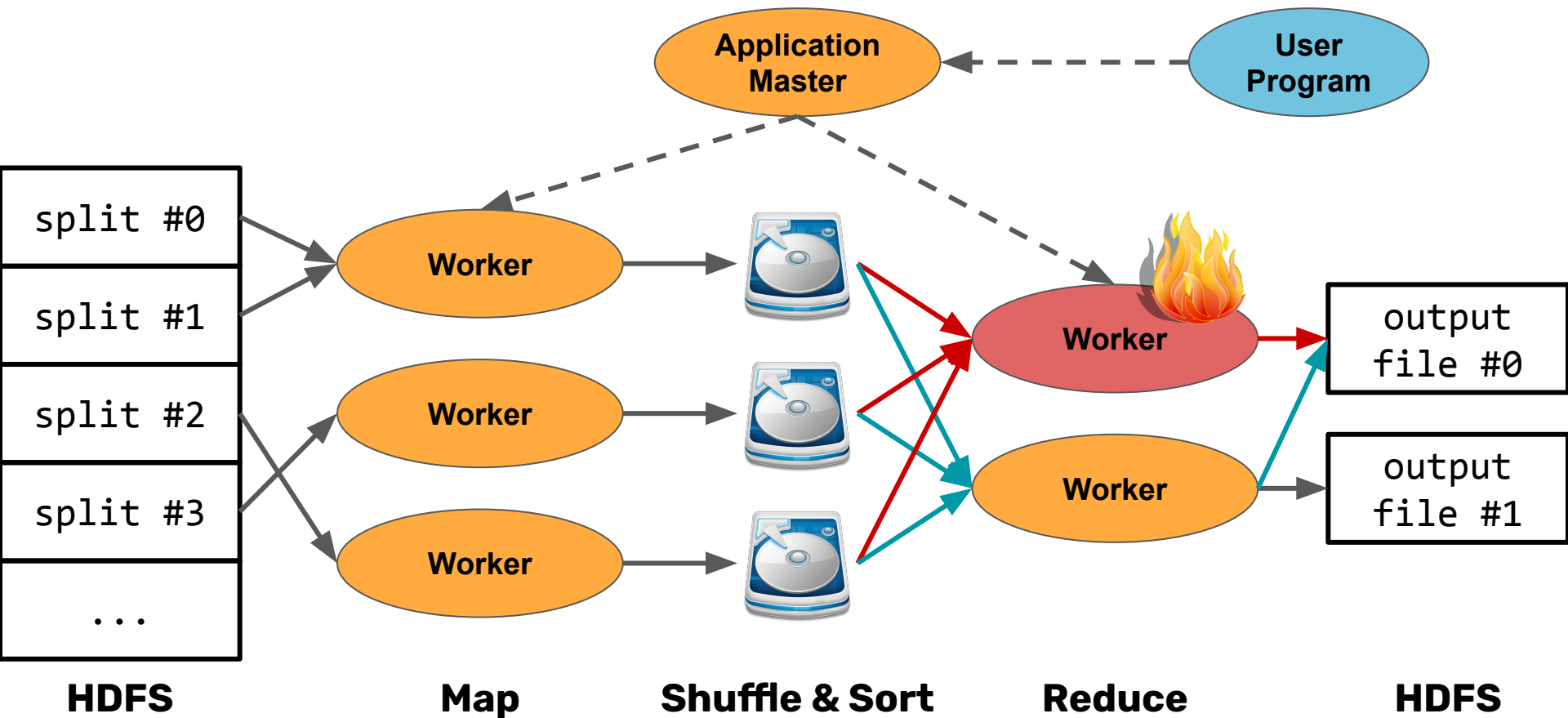


Fault Tolerance (защита от сбоев)



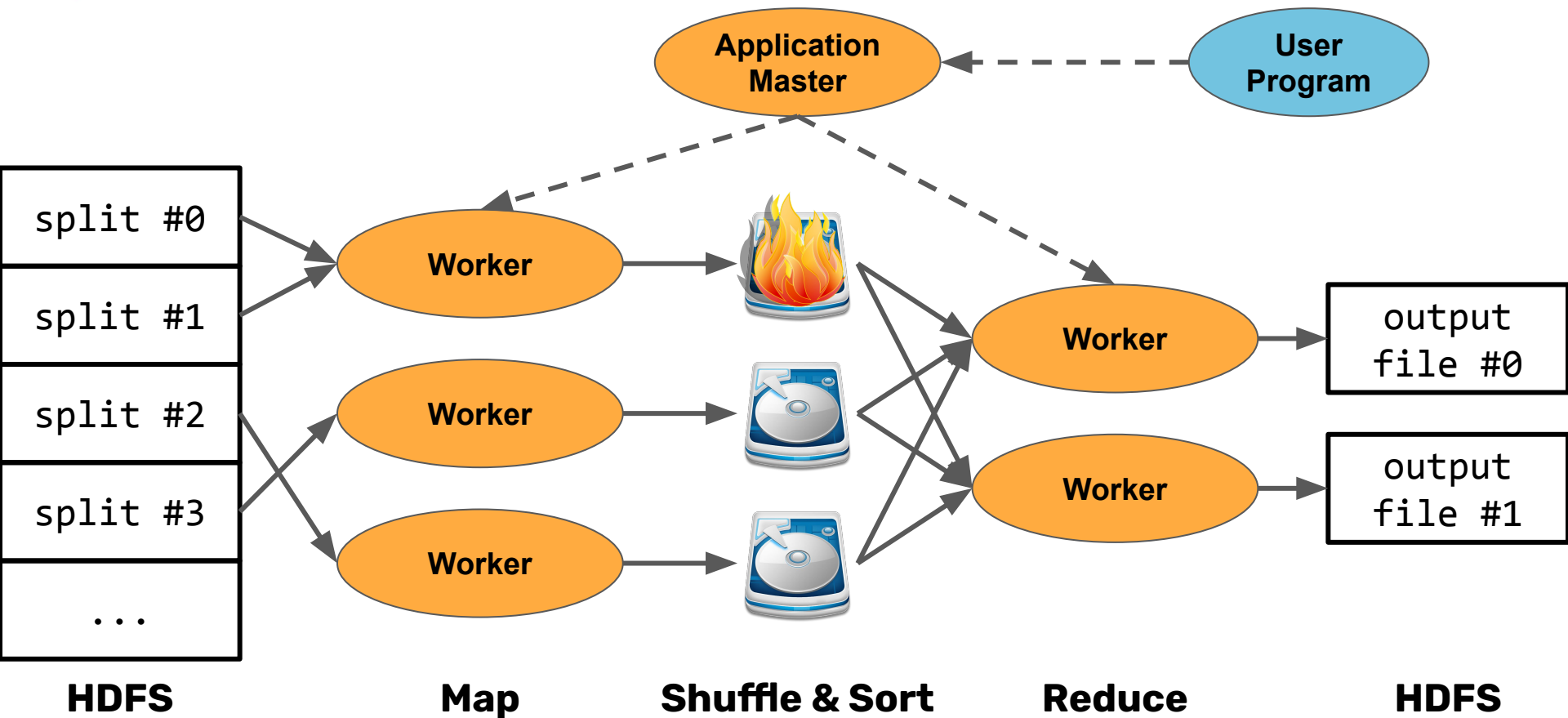


Fault Tolerance (защита от сбоев)



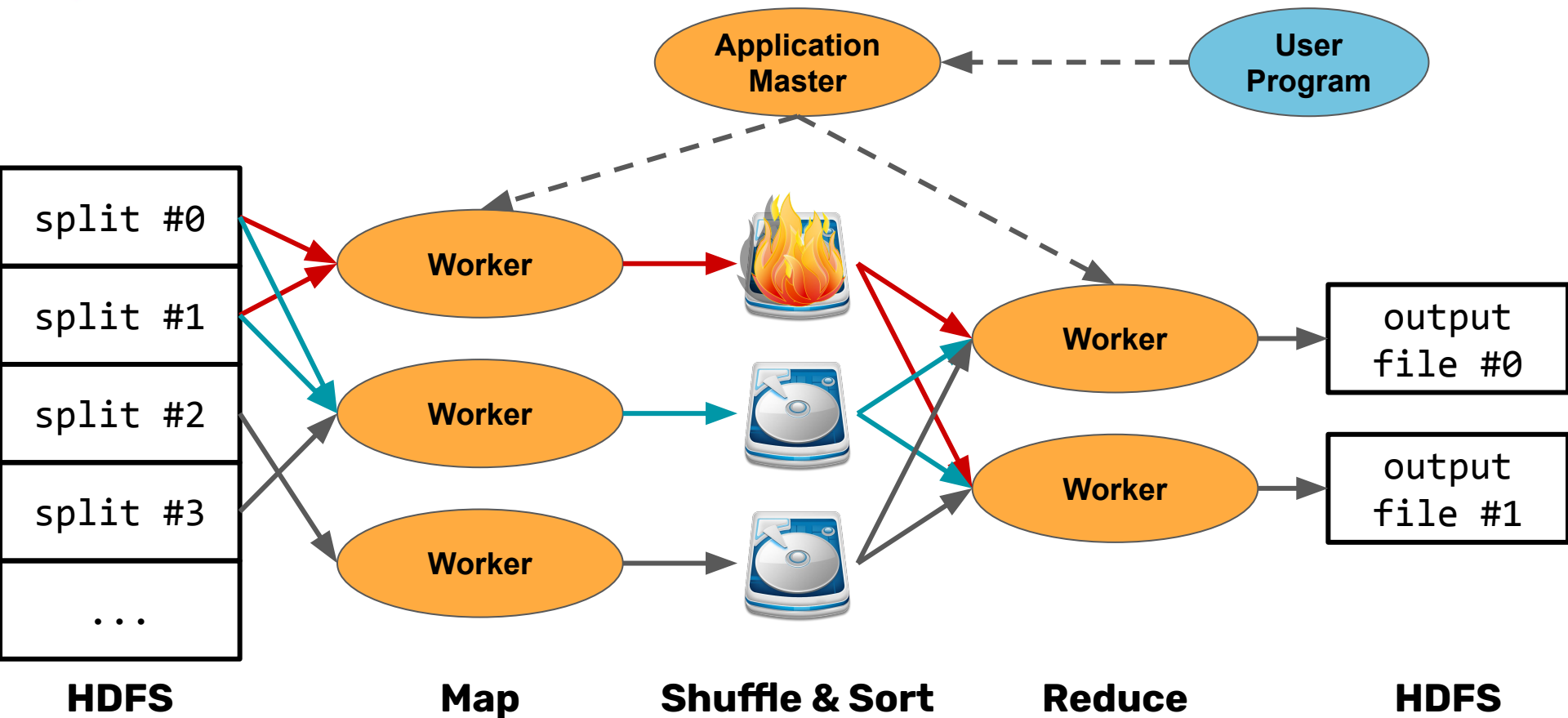


Fault Tolerance (защита от сбоев)



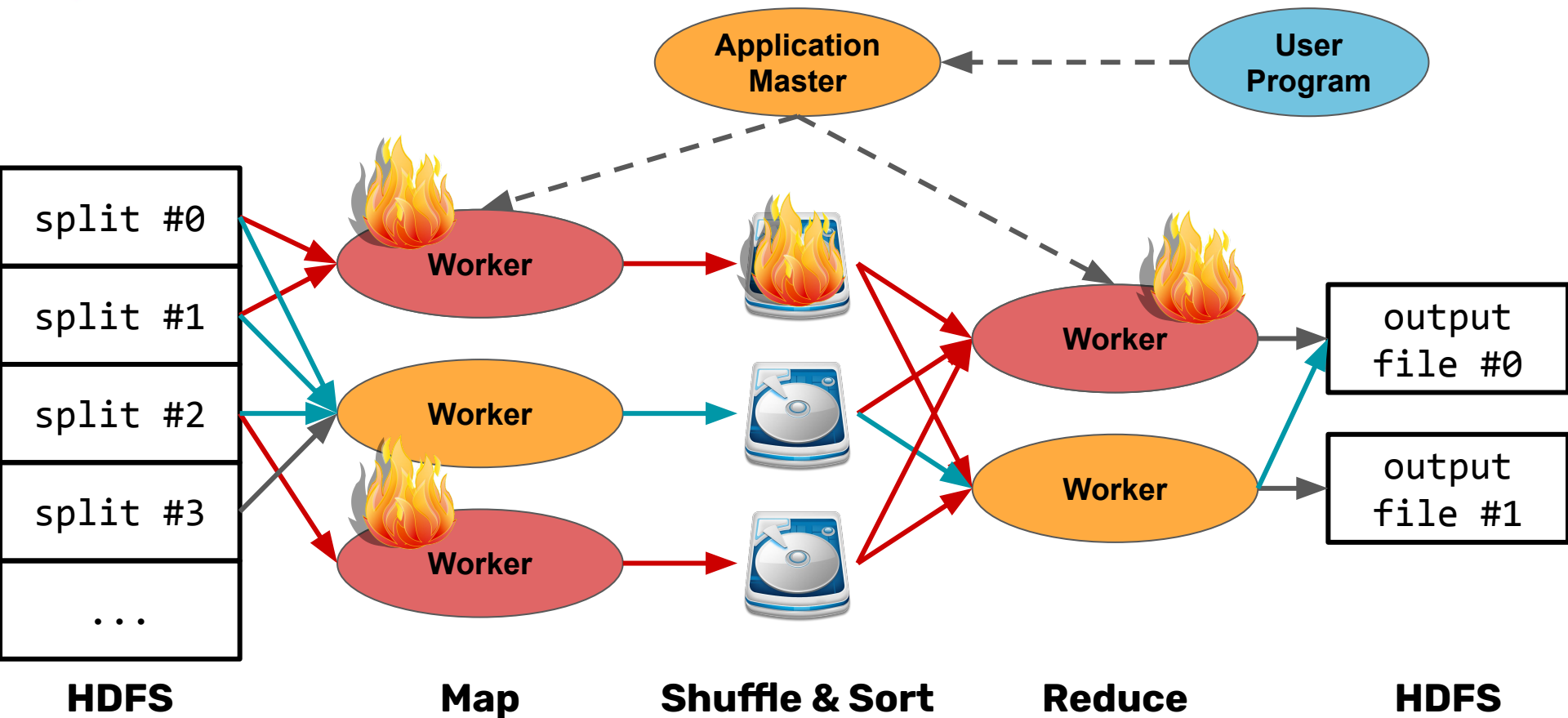


Fault Tolerance (защита от сбоев)





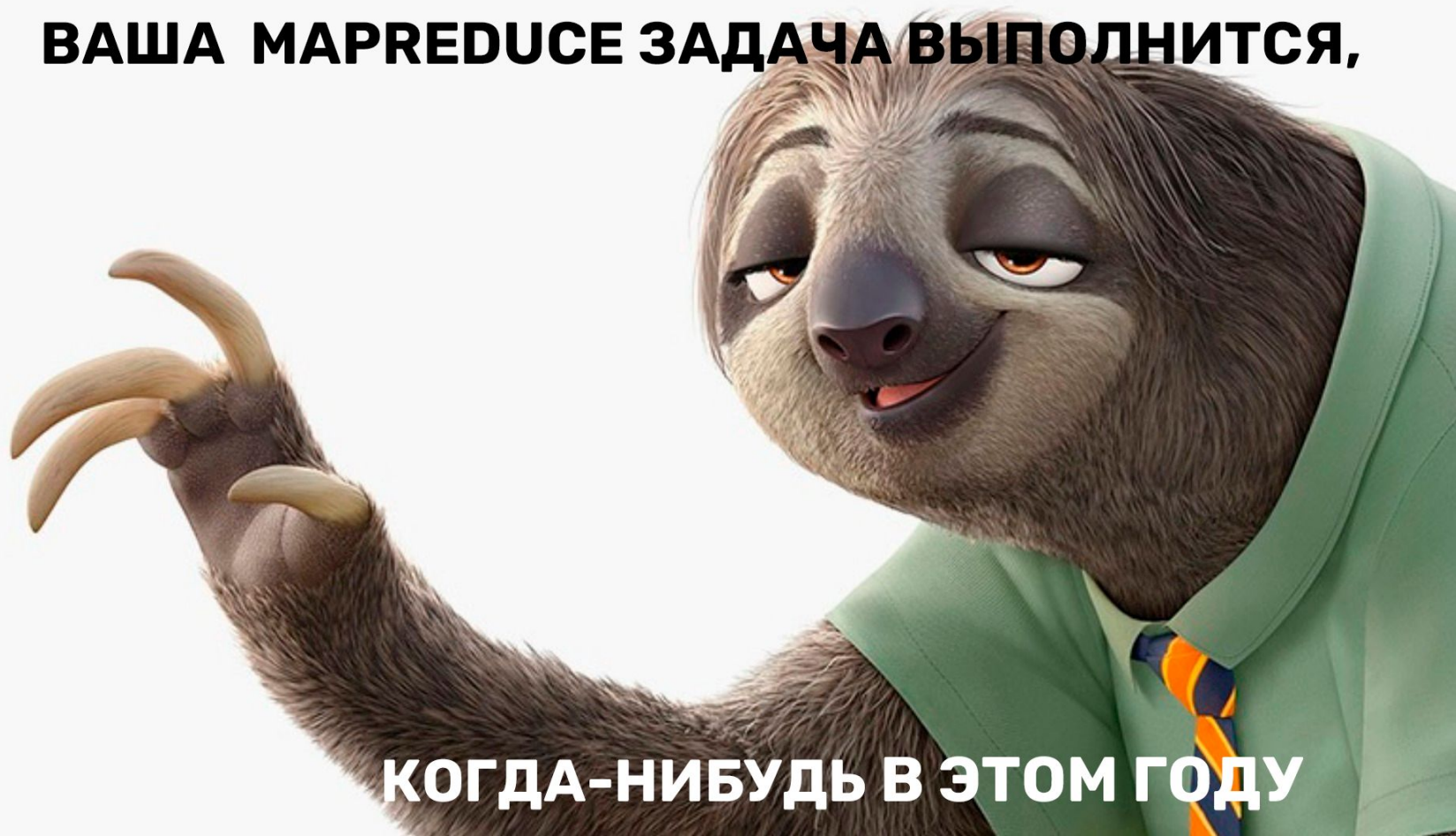
Fault Tolerance (защита от сбоев)



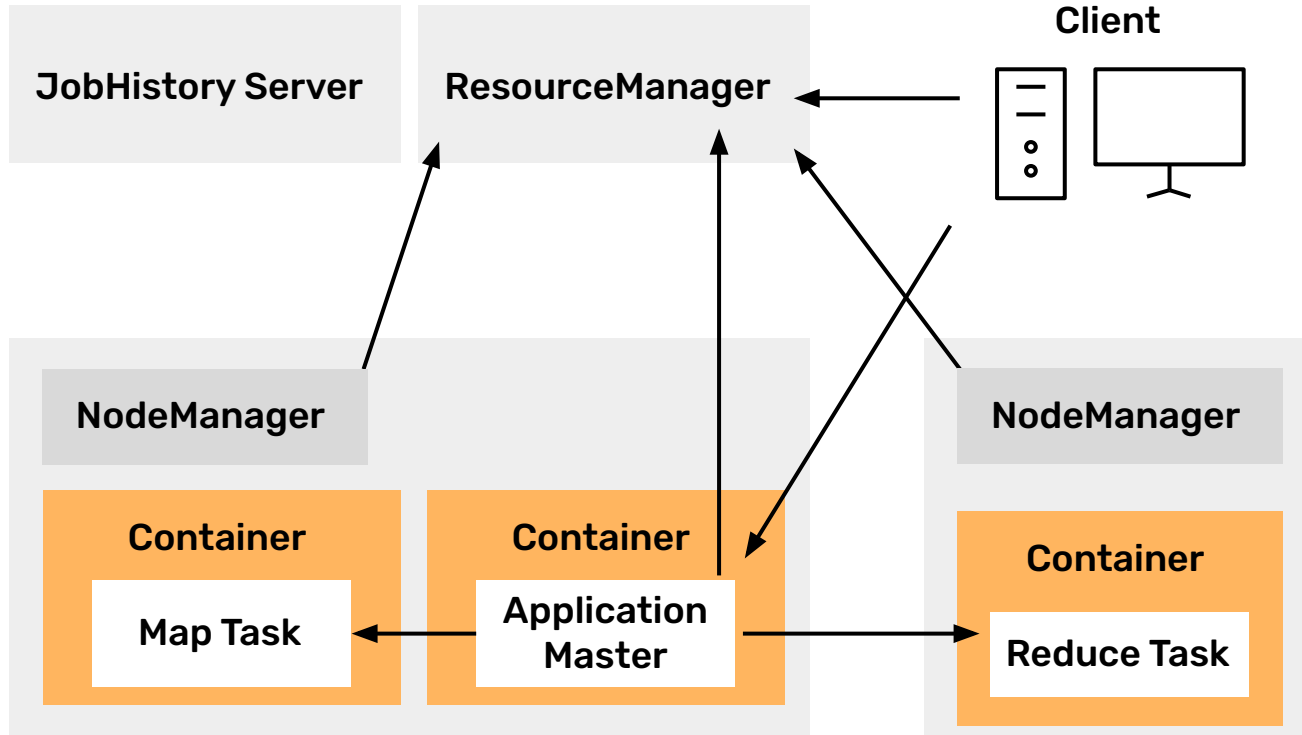


BIGDATA

ВАША MAPREDUCE ЗАДАЧА ВЫПОЛНИТСЯ,



КОГДА-НИБУДЬ В ЭТОМ ГОДУ

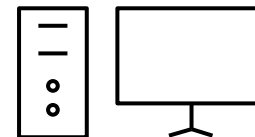




JobHistory Server

ResourceManager

Client



NodeManager

Container

Map Task

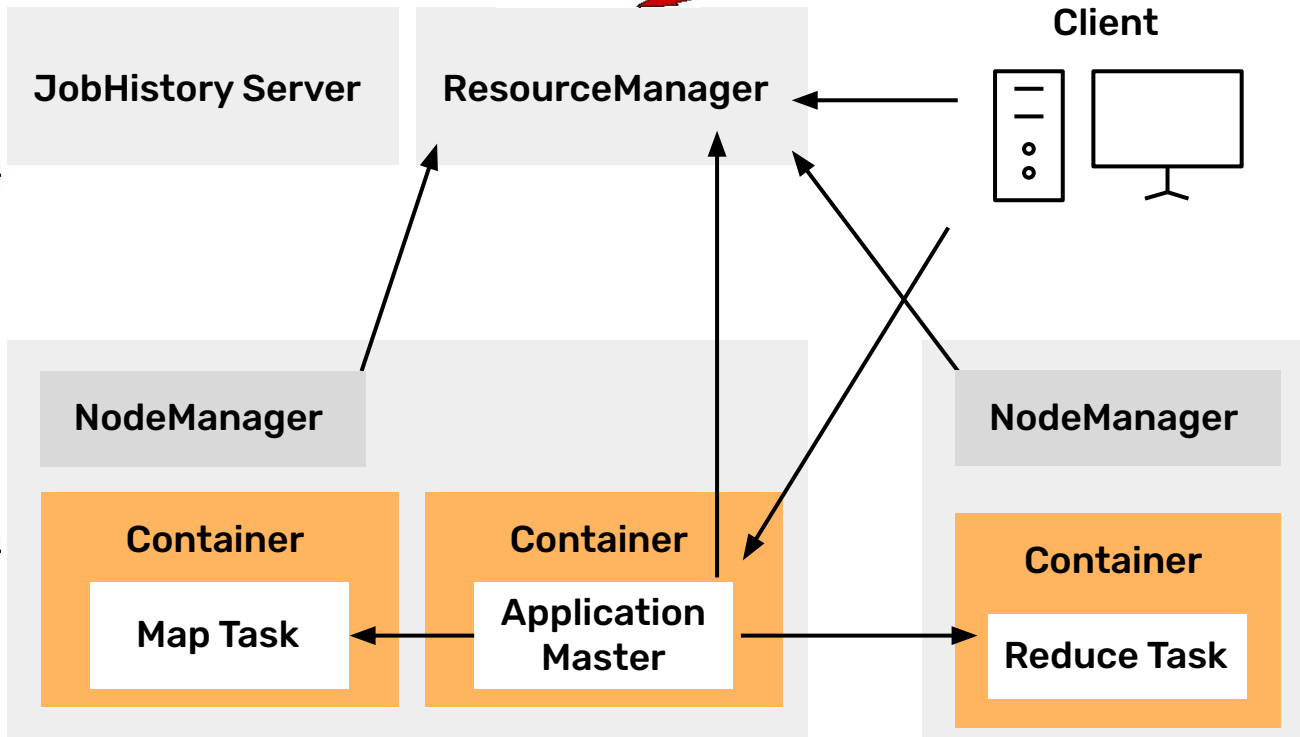
Container

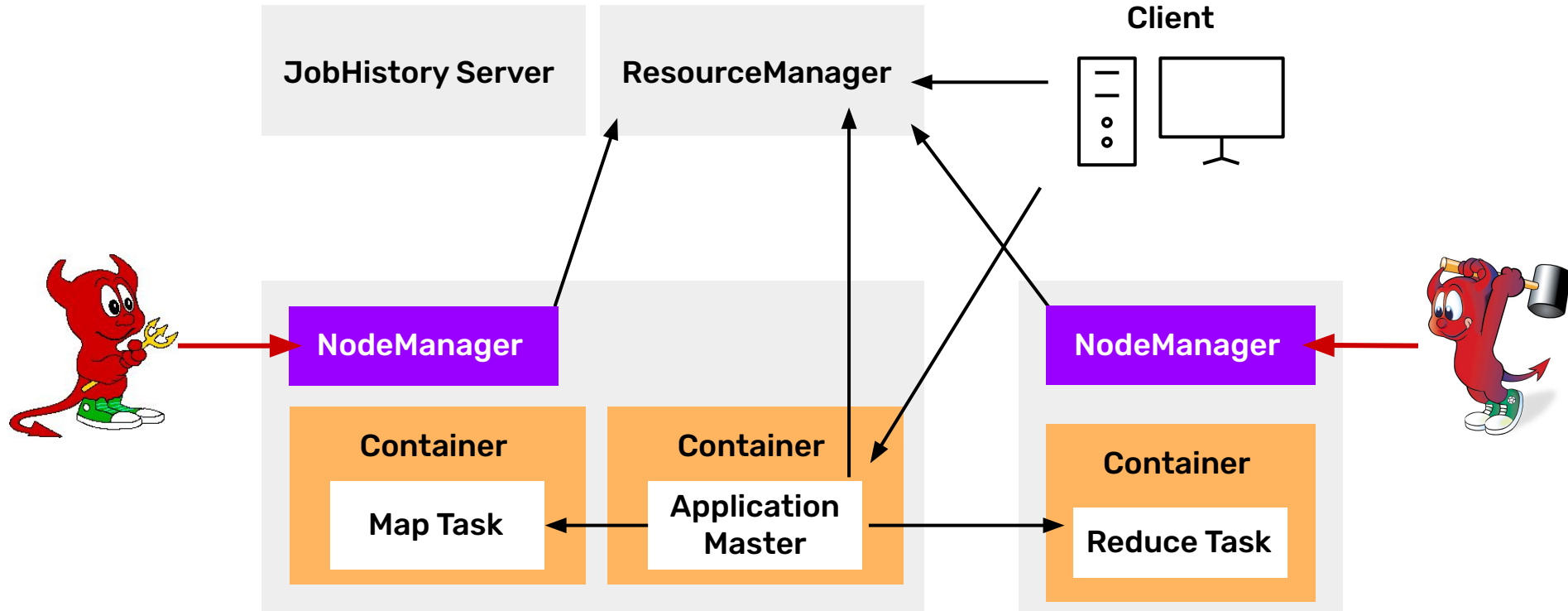
**Application
Master**

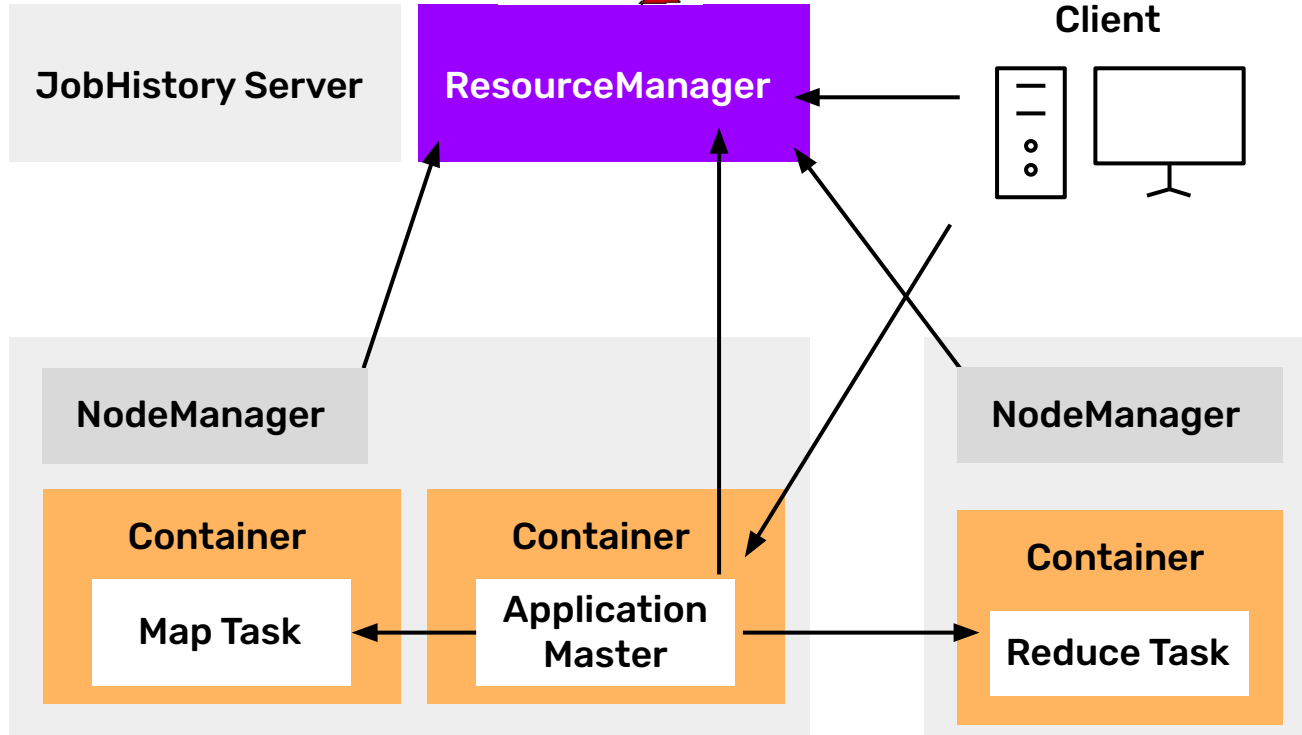
NodeManager

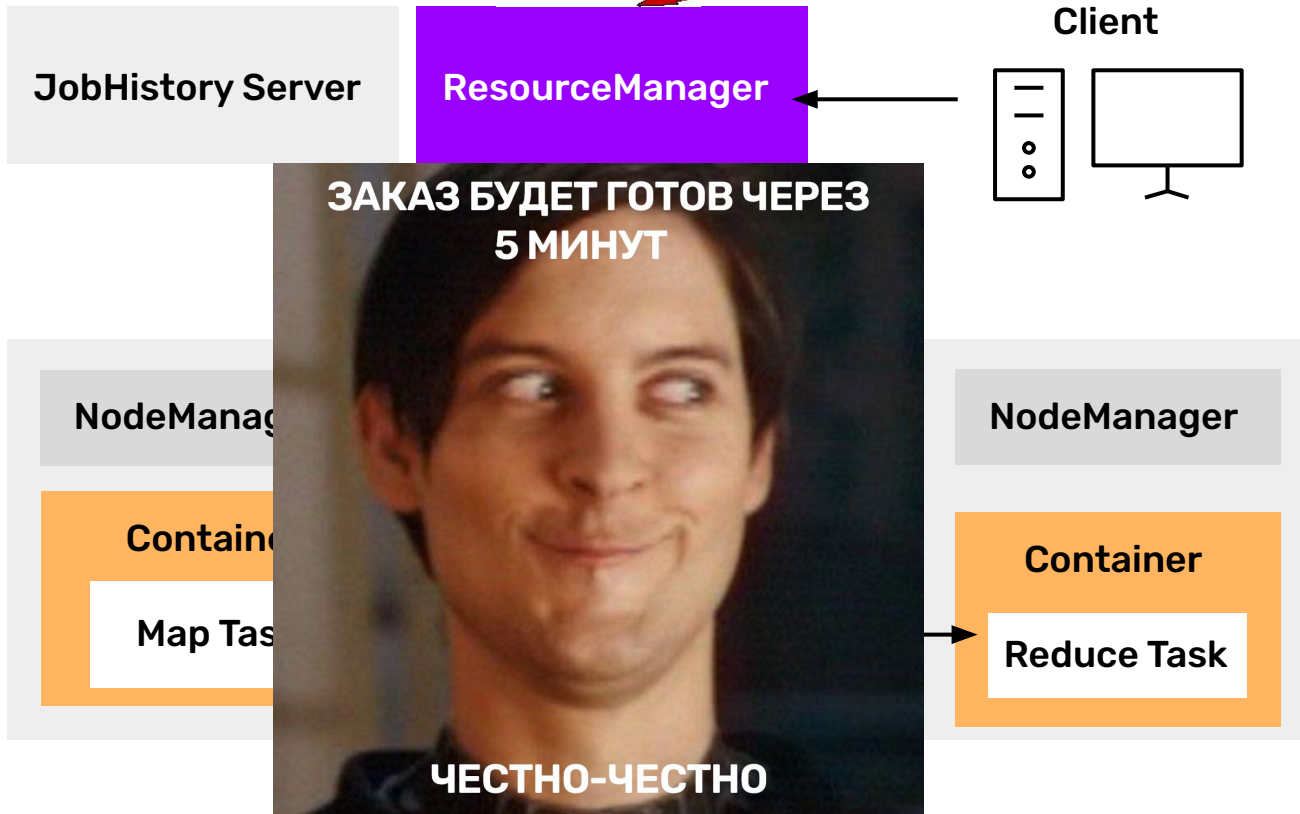
Container

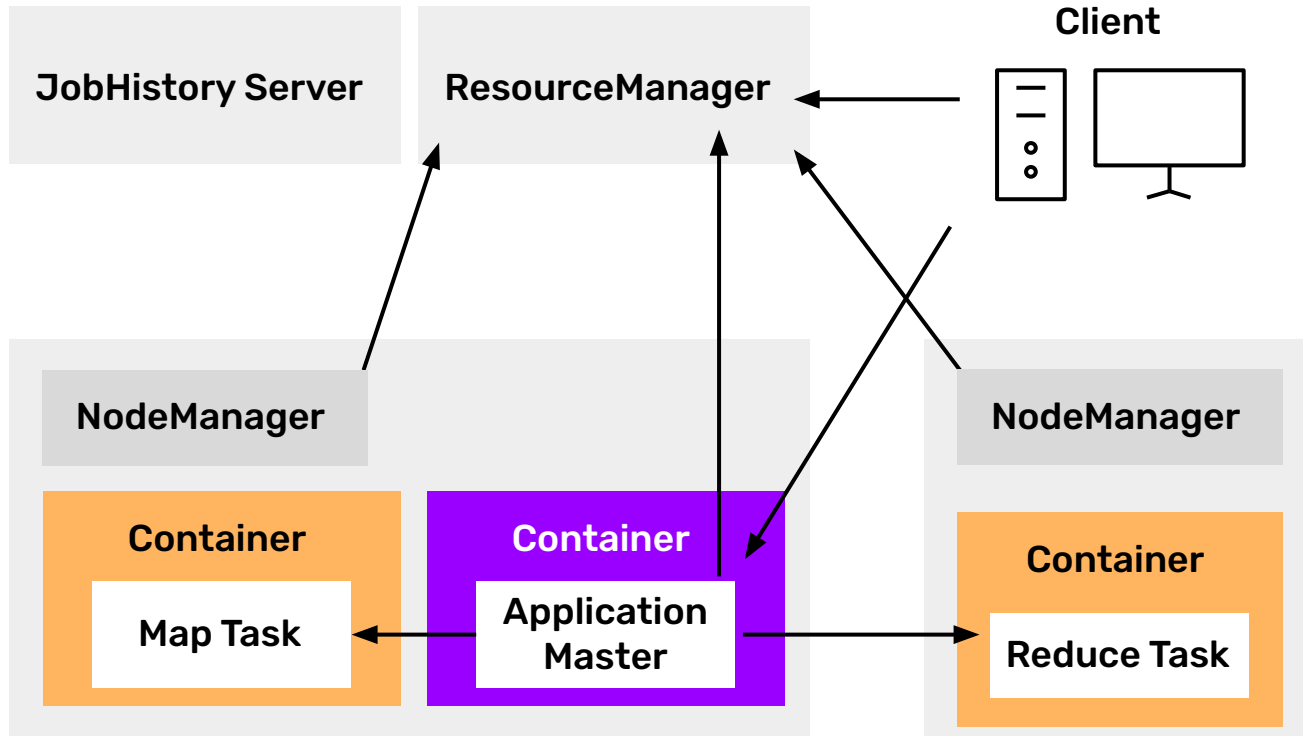
Reduce Task

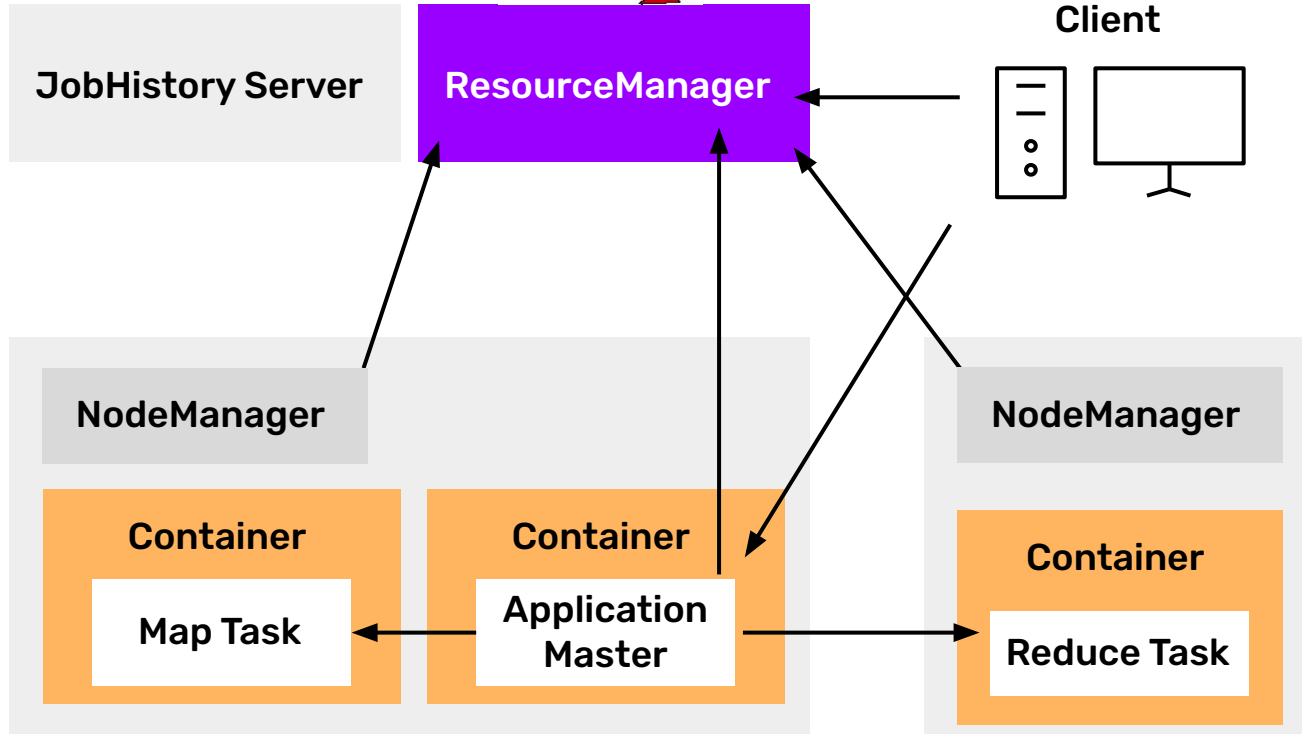










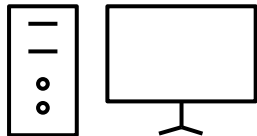




**YARN Timeline Server
(JobHistory Server)**

ResourceManager

Client



NodeManager

Container

Map Task

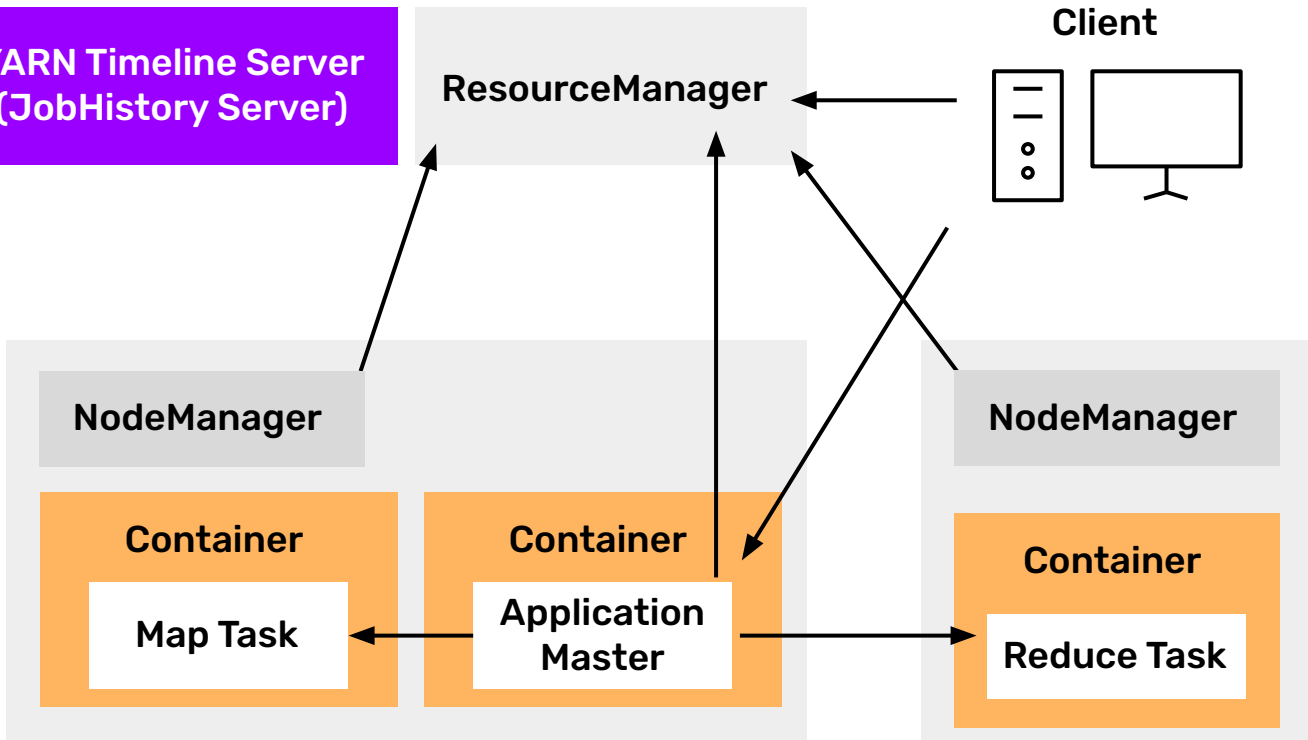
Container

**Application
Master**

NodeManager

Container

Reduce Task







Q&A

Как зная топологию кластера
оптимизировать MapReduce?



```
$ yarn jar map_reduce_example.jar
```

```
...
```

```
INFO mapreduce.Job: Counters: 30
```

```
...
```

```
    Job Counters
```

```
        Launched map tasks=2
```

```
        Data-local map tasks=2
```

```
        Total time spent by all maps in occupied slots (ms)=27360
```

```
...
```



```
$ yarn jar map_reduce_example.jar
```

```
...
```

```
INFO mapreduce.Job: Counters: 30
```

```
...
```

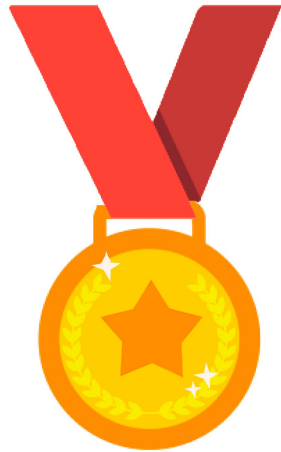
```
    Job Counters
```

```
        Launched map tasks=2
```

```
        Data-local map tasks=2
```

```
        Total time spent by all maps in occupied slots (ms)=27360
```

```
...
```





**BIGDATA
TEAM**

Резюме



Теперь вы можете:

- Объяснить каким образом достигается Fault Tolerance



Теперь вы можете:

- ▶ Объяснить каким образом достигается Fault Tolerance
- ▶ Рассказать друзьям о базовых оптимизациях вычислений MapReduce