



菜鸟窝

菜鸟窝——程序猿的黄埔军校

7天入门大数据特训营

架构师和大数据专家手把手带你入门

主讲老师：lee

老师简介



jesson

复旦大学计算机硕士
10年以上研发经验

曾就职于**华为**、**百度**、**平安**等知名互联网公司。

近几年埋首大数据领域相关技术，对神经网络、机器学习有深入了解。热爱新技术，乐于学习与分享。

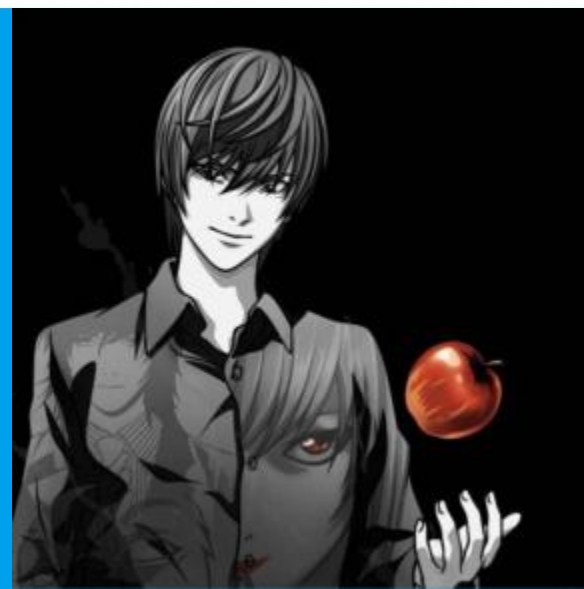


Brave

高级软件架构师
11年软件开发经验

曾任金蝶（中国）软件公司技术顾问，卫华集团高级需求分析师

曾主导开发多个大数据项目，精通Kafka、Spark等大数据相关技术



Lee

就职于**百度**
大数据开发岗位

5年研发经验，曾就职于BAT等一线互联网公司

精通Java开发，对大数据有深入的研究，喜欢前沿技术，热爱踩坑。



Franco

百度数据仓库工程师

曾任猎豹移动推荐算法工程师，newsrepublic海外推荐信息流推荐算法工程师

精通大数据算法推荐，有丰富的的大数据项目开发经验。



01 MapReduce相关定义

02 MapReduce流程详解

03 MapReduce简易案例

A background image showing a close-up of a laptop keyboard and a pen resting on a desk.

目录/CONTENTS



mapred-site.xml

//让MapReduce运行在Yarn上

```
<name>mapreduce.framework.name</name>
```

```
<value>yarn</value>
```



yarn-site.xml

//MapReduce运行服务

```
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
```

//resourcemanager默认启动机器

```
<name>yarn.resourcemanager.hostname</name>
```

```
<value>lee.cniao5.com</value>
```



MapReduce是一种计算模型，方便编程人员在不会分布式并行编程的情况下，将自己的程序运行在分布式系统上。作为一款计算框架，其表现形式就是有输入（input），通过MapReduce操作这个输入（input），根据定义好的计算模型，得到一个输出（output）。

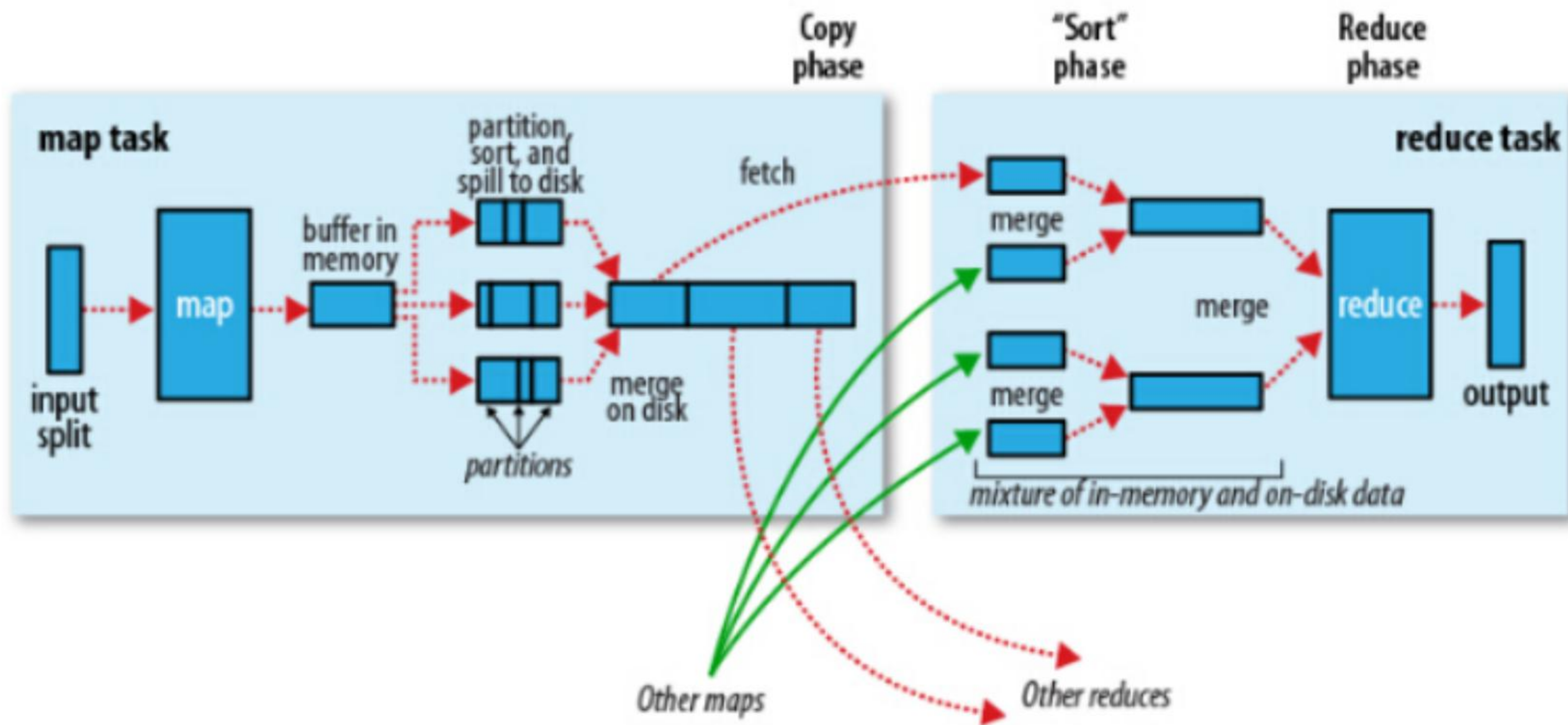


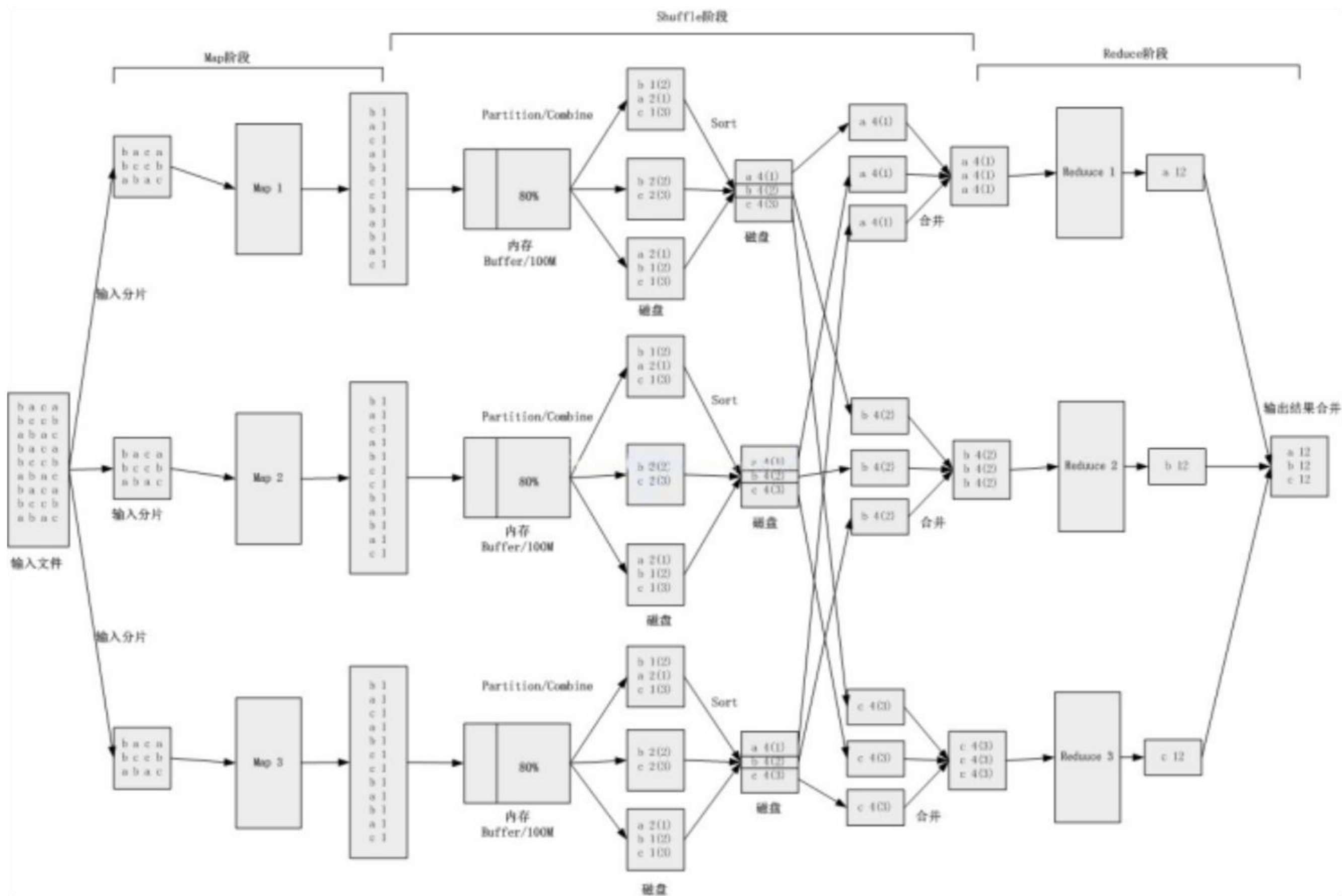
其中心思想为“**分而治之**”，致使MapReduce极其擅长处理大数据。



分/而/治/之

- Mapper负责“分”
将复杂任务分解成“简单任务”
- Reducer负责“合”
map阶段的结果汇总





三种基本模式：



开发工具：Eclipse+插件



实现系统案例Wordcount



MapReduce代码模板

大数据实战就业班

226个知识点&3大商业项目实战，全能班型，学完即用



+ 2位10年经验架构师
+ 2位百度大数据大神

● 扫码关注大数据技能
顺利加薪 / 升职 / 跳槽



现在报名**菜鸟窝大数据实战就业班**
限量送在售价 **¥ 1898**的《**java web工程师**》

报名咨询还有减免优惠：
windy老师微信：phoebe2016524