

Spark企业级大数据项目实战 第2课

DATAGURU专业数据分析社区

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

- Dataguru (炼数成金) 是专业数据分析网站 , 提供教育 , 媒体 , 内容 , 社区 , 出版 , 数据分析业务等服务。我们的课程采用新兴的互联网教育形式 , 独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围 , 重竞争压力的特点 , 同时又发挥互联网的威力打破时空限制 , 把天南地北志同道合的朋友组织在一起交流学习 , 使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本 , 直线下降至百元范围 , 造福大众。我们的目标是 : 低成本传播高价值知识 , 构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情 , 请看我们的培训网站 <http://edu.dataguru.cn>

- 在虚拟机安装CentOS
- Hadoop、Spark、Hive等软件安装
- 开发工具（IDEA）安装调试
- Maven工程构建
- Spark源码编译
- Spark作业提交

■ 软件

- 虚拟机：VMware® Workstation 12 Pro
- 操作系统：CentOS-6.6-x86_64

■ 步骤

- 操作系统安装
- 网络设置
- 虚拟机和宿主机（windows主机）相互无缝复制、粘贴
- Yum源设置

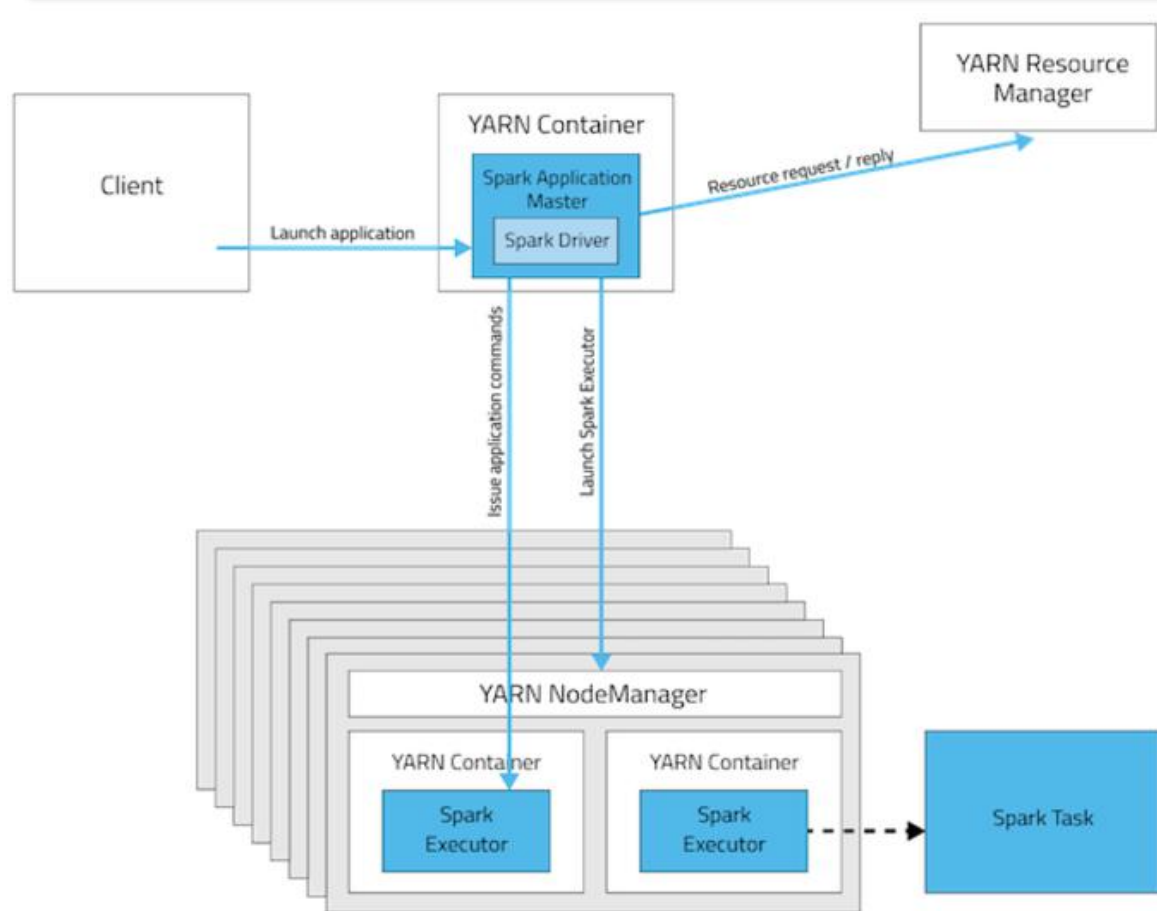
■ 软件：

- Hadoop
- Hive
- MySQL
- Spark

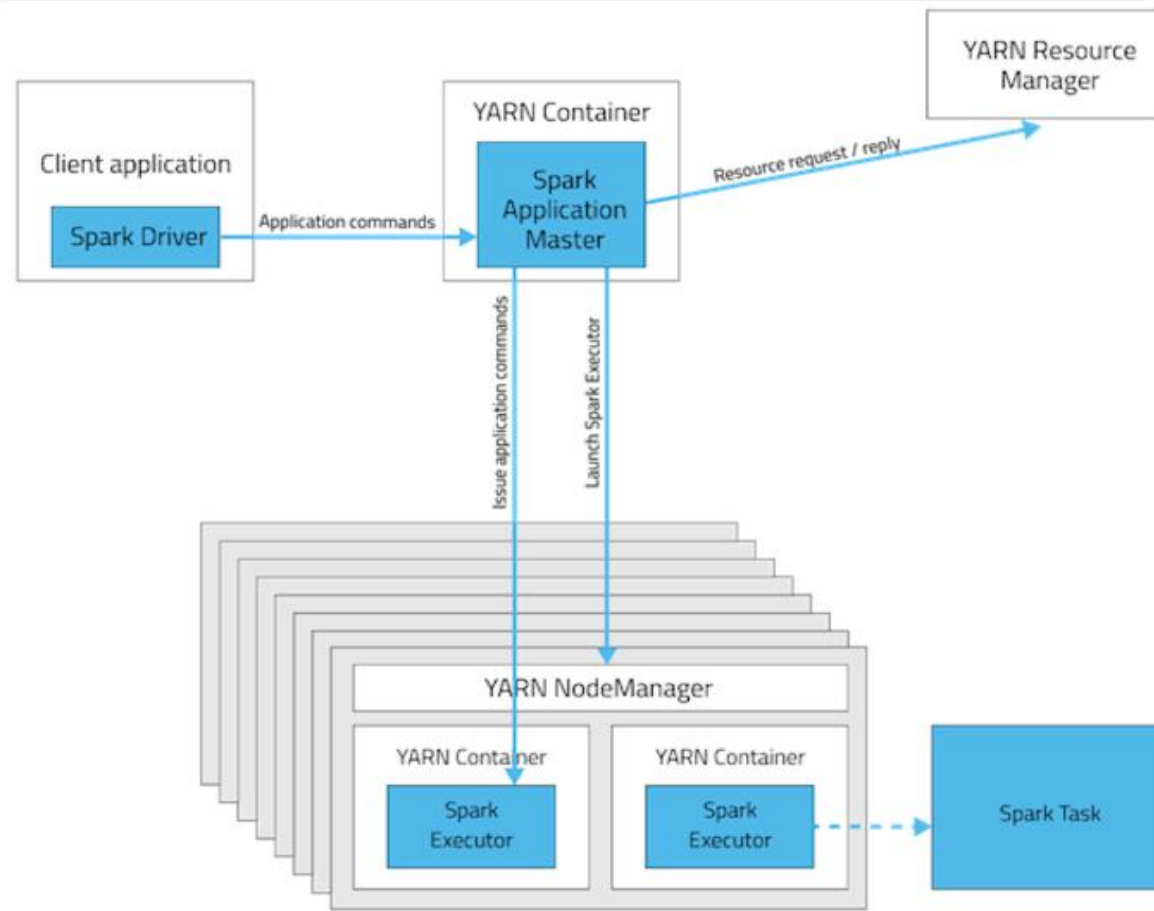
....

Spark源码编译

Spark作业提交



Yarn Cluster



Yarn Client

yarn client:

```
spark-submit --master yarn --class com.test.Test ./Test-1.0-SNAPSHOT.jar
```

```
spark-submit --master yarn --deploy-mode client --class com.test.Test ./Test-1.0-SNAPSHOT.jar
```

```
spark-submit --master yarn-client --class com.test.Test ./Test-1.0-SNAPSHOT.jar
```

yarn cluster:

```
spark-submit --master yarn --deploy-mode cluster --class com.test.Test ./Test-1.0-SNAPSHOT.jar
```

关于**YARN**作业提交的其他参数， 会在**spark**调优部分详细讲解

Thanks

FAQ时间