



北京航空航天大学  
BEIHANG UNIVERSITY

# <<流式实时分布式计算框架 Spark 的研究与应用>> 测试需求说明书



北京航空航天大学

2016-05

## 版本变更历史

版本	提交日期	主要编制人	审核人	版本说明
V1.0	2016/5/11	阳艳红、武一杰、 王铖成、于思民	于思民	初稿
V2	2016/5/19	阳艳红、武一杰、 王铖成、于思民	于思民	根据课堂评审以及 H/A 两组同学意见和建议,对 V1.0 版本进行修订。
V2.1	2016/5/19	阳艳红、武一杰、 王铖成、于思民	于思民	根据课堂评审以及 H/A 两组同学意见和建议,对 V1.0 版本进行修订。
V2.2	2016/5/20	阳艳红、武一杰、 王铖成、于思民	于思民	添加非功能性需求内容,添加集群部署测试模块,统一题注。

## 目 录

1. 范围.....	1
1.1 标识 .....	1
1.2 文档说明 .....	1
2. 测试计划.....	1
2.1 测试目标 .....	1
2.2 测试分工 .....	1
2.3 Spark 测试用例与需求用例对照表.....	2
3. 功能需求测试用例描述.....	4
3.1 数据编程模型测试 .....	4
3.1.1 RDD 的创建测试 .....	4
3.1.2 RDD 的转换测试 .....	5
3.1.3 RDD 的 Action 操作测试 .....	6
3.1.4 RDD 的缓存测试 .....	7
3.1.5 RDD 检查点测试 .....	8
3.2 数据存储测试 .....	9
3.2.1 启动 Master 测试 .....	9
3.2.2 注册 Slave 测试.....	10
3.2.3 Master 向 Slave 发送控制命令测试.....	11
3.2.4 Master 获取 Slave 状态测试.....	12
3.2.5 Slave 向 Master 发送状态更新测试.....	13
3.3 集群部署 .....	14
3.3.1 Local 模式 .....	14
3.3.2 Standalone 模式.....	15
3.3.3 Mesos 模式 .....	16
3.3.4 YARN 模式.....	17
3.4 作业调度 .....	19
3.4.1 用户作业提交测试.....	19
3.4.2 DAG 创建测试.....	20

3.4.3	Stage 划分测试.....	21
3.4.4	TaskSet 生成测试.....	21
3.4.5	Task 调度测试.....	22
4.	非功能需求测试用例描述.....	23
4.1.1	运行速度.....	23
4.1.2	容错性.....	24
4.1.3	安全性.....	26
5.	参考文献.....	27

## 图目表目

图 1 弹性分布式数据集用例图.....	4
图 2 数据存储模块用例图.....	9
表 1 测试分工.....	1
表 2 测试用例与需求用例对照表.....	2
表 3 创建 RDD 的测试用例 .....	4
表 4 转换 RDD 的测试用例 .....	5
表 5 RDD 的 Action 操作测试用例 .....	6
表 6 缓存 RDD 的测试用例 .....	7
表 7 创建 RDD 检查点的测试用例 .....	8
表 8 启动 Master 的测试用例 .....	10
表 9 注册 Slave 的测试用例.....	10
表 10 Master 向 Slave 发送控制命令的测试用例.....	11
表 11 Master 获取 Slave 状态的测试用例.....	12
表 12 Slave 向 Master 发送状态更新.....	13
表 13 Local 模式部署的测试用例 .....	14
表 14 Standalone 模式部署的测试用例.....	15
表 15 Mesos 模式部署的测试用例 .....	17
表 16 YARN 模式部署的测试用例.....	18
表 17 用户作业提交的测试用例.....	19
表 18 创建 DAG 的测试用例 .....	20
表 19 划分 Stage 的测试用例.....	21
表 20 生成 Taskset 的测试用例.....	22
表 21 Task 调度的测试用例 .....	23
表 22 运行速度的测试用例.....	24
表 23 容错性的测试用例.....	24
表 24 安全性的测试用例.....	26

# 1. 范围

## 1.1 标识

Spark 版本号: Spark1.2.1

版本: V2.2

## 1.2 文档说明

本文档用于对本组的实验项目进行测试.测试工作主要针对改进后的 spark 进行,测试用例设计完成后,对程序运行结果的准确性以及运行时所需的时间和空间进行详细的测试用例设计和说明。

# 2. 测试计划

## 2.1 测试目标

通过测试来发现 Spark1.2.1 版本所能提供给开发人员使用的功能是否符合《软件需求规格说明书》。在测试的过程中要尽可能的发现软件在现有阶段存在的问题，确保在软件交付客户使用之后出现的问题在可接受的范围之内。

## 2.2 测试分工

表 1 测试分工

成员	主要职责	测试用例
于思民	做数据编程模型相关测试需求，非功能性需求，汇总测试需求，编写测试需求规格说明书	RDD 的创建测试、RDD 的转换测试、RDD 的 Action 操作测试、RDD 的缓存测试、RDD 检查点测试、运行速度对比测试、

阳艳红	做集群部署模块相关测试需求，非功能性需求	Local 模式部署测试、Standalone 模式部署测试、Mesos 模式部署测试、YARN 模式部署测试、运行速度对比测试
武一杰	做数据存储模块相关测试需求，非功能性需求	启动 Master 测试、注册 Slave 测试、Master 向 Slave 发送控制命令测试、Master 向 Slave 获取状态测试、Slave 向 Master 发送状态更新测试、容错性测试
王铖成	做作业调度模块相关测试需求，非功能性需求	用户作业提交测试、DAG 创建测试、Stage 划分测试、TaskSet 声明测试、Task 调度测试、安全性测试

## 2.3 测试软硬件环境配置

表 2 测试软硬件环境配置

资源名称/类型	配置	备注
处理器（CPU）	Intel(R) Core(TM) i7-2670QM CPU @ 2.20GHZ	只使用 1 个核心
运行内存（Memory）	3072MB	
操作系统（OS）	CentOS7	
JDK	1.8.0_77	
Scala	2.11.2	
Hadoop	2.5.2	
Spark	1.2.1	

## 2.4 Spark 测试用例与需求用例对照表

表 3 测试用例与需求用例对照表

模块	需求用例	测试用例
数据编程模型	RDD 的创建	RDD 的创建测试
	RDD 的转换	RDD 的转换测试
	RDD 的 Action 操作	RDD 的 Action 操作测试
	RDD 的缓存	RDD 的缓存测试
	RDD 检查点	RDD 的检查点测试
数据存储	启动 Master	启动 Master 测试
	注册 Slave	注册 Slave 测试
	Master 向 Slave 发送控制命令	Master 向 Slave 发送控制命令测试
	Master 向 Slave 获取状态请求	Master 向 Slave 获取状态请求测试
	Slave 向 Master 发送状态更新	Slave 向 Master 发送状态更新测试
集群部署	Local 模式部署	Local 集群部署测试
	Standalone 模式部署	Standalone 集群部署测试
	Mesos 模式部署	Mesos 模式部署测试
	YARN 模式部署	YARN 模式部署测试
作业调度	用户作业提交	用户作业提交测试
	DAG 创建	DAG 创建测试
	STAGE 划分	STAGE 划分测试
	Taskset 生成	Taskset 生成测试
	Task 调度	Task 调度测试
非功能性	运行速度	运行速度对比测试
	容错性	容错性测试
	安全性	安全性测试



### 3. 功能需求测试用例描述

#### 3.1 数据编程模型测试

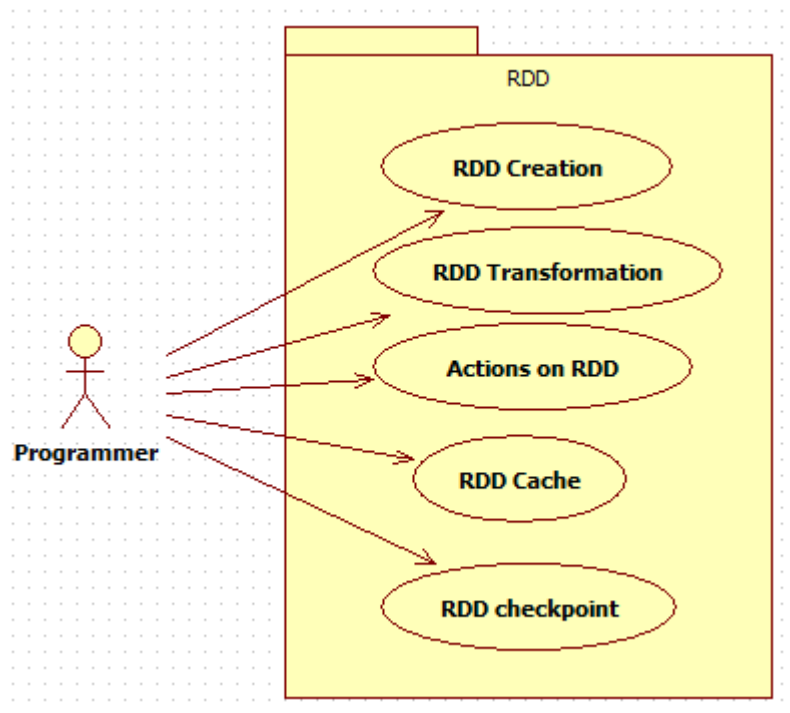


图 1 弹性分布式数据集用例图

##### 3.1.1 RDD 的创建测试

创建 RDD 是 Spark 的基础流程，所以设计 RDD 的创建测试用例的目的是测试所部属的 Spark 集群能否正确创建 RDD，从而测试 Spark 集群的部署效果和 RDD 的创建功能。如果结果正确创建，则返回相应创建成功的结果，可以进行下一步操作；否则返回创建失败的结果。

测试用例描述如下：

表 4 创建 RDD 的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	创建 RDD	编制人	于思民
用例 ID	TC001	时间	2016-05-19
相关用例	无		

功能特性	使用 Spark API 创建一个 RDD 对象		
测试目的	检测创建 RDD 创建功能的正确性		
预置条件	环境部署完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.1.1RDD 的创建部分		
测试数据	已存在的 RDD 对象，或者某个文件的 uri。		
测试人员	于思民		
操作步骤			
1. 启动 Master 节点； 2. 启动 Slave 节点； 3. 执行 spark-shell 命令进入命令行； 4. 使用 SparkContext 对象从某一文件 uri 读取数据； 5. 输出 RDD 对相关属性及数据，检查结果正确性。			
预测结果	RDD 对象各属性正确，RDD 数据正确。		

### 3.1.2 RDD 的转换测试

测试用例设计的目的是测试所创建的 RDD 能否正确转换。测试结果分为转换成功和转换失败两种。在 RDD 上调用代表某种数据转换的函数操作，操作完成后，测试成功将得到新的 RDD，失败则不能得到新的 RDD。

测试用例描述如下所示：

表 5 转换 RDD 的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	转换 RDD	编制人	于思民
用例 ID	TC002	时间	2016-05-19
相关用例	TC001		
功能特性	RDD 转换		
测试目的	测试环境部署效果以及 RDD 的转换功能		
预置条件	环境部署完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.1.2RDD 的		
测试数据	已存在的 RDD 对象，或者有效文件的 uri。		

测试人员	于思民
操作步骤	
1. 启动 Master 节点; 2. 启动 Slave 节点; 3. 执行 spark-shell 命令进入命令行; 4. 使用 SparkContext 对象从某一文件 uri 读取数据; 5. 对得到的 RDD 对象调用某种转换算子操作, 并得到新的 RDD 对象; 6. 输出新转换的 RDD 对象的属性及数据, 检查其结果的正确性	
预测结果	新转换的 RDD 对象各属性正确, RDD 数据结果符合转换算子操作。

### 3.1.3 RDD 的 Action 操作测试

在进行了 RDD 的转换操作后, 数据转换成不同的形式, 这时候开发人员可以对转换后的数据形式进行一些操作, 如存储到文件系统, 如提供给其他系统进行处理。此时的 Action 操作测试用例的目的是测试这些操作是否能正确地执行。操作正确执行则返回相应操作的结果, 如返回存储结果, 其他系统的接收结果, 否则操作失败。

测试用例描述如下:

表 6 RDD 的 Action 操作测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	RDD 的 action 操作	编制人	于思民
用例 ID	TC003	时间	2016-05-19
相关用例	TC001, TC002		
功能特性	Aciton		
测试目的	测试环境部署效果以及 RDD 的各种 Action 操作		
预置条件	环境部署完成	特殊规程说明	无

参考信息	<G-系统需求规格说明书-V3.2>中 3.1.3RDD 的操作部分
测试数据	已存在的 RDD 对象，或者某个文件的 uri。
测试人员	于思民
操作步骤	
1. 启动 Master 节点； 2. 启动 Slave 节点； 3. 执行 spark-shell 命令进入命令行； 4. 使用 SparkContext 对象从某一文件 uri 读取数据； 5. 对得到的 RDD 对象调用某种转换算子操作，并得到新的 RDD 对象； 6. 对 RDD 执行 Action 算子操作； 7. 将操作结果输出，分析输出结果的正确性。	
预测结果	输出结果符合 Action 算子操作。

### 3.1.4 RDD 的缓存测试

由于 RDD 可能会被用户应用程序重复使用，使用缓存可以提高效率。RDD 的缓存测试用例设计的目的是为了测试系统的 RDD 缓存功能是否正常开启。测试通过缓存 RDD 和对缓存读取来判断缓存功能。结果正常则能读取缓存，否则失败。

测试用例描述如下：

表 7 缓存 RDD 的测试用例

项目	Spark 研究与应用	版本	v1.0
功能需求	RDD 缓存	编制人	于思民
用例 ID	TC004	时间	2016-05-19
相关用例	TC001		
功能特性	缓存 RDD		
测试目的	环境部署效果以及 RDD 的缓存功能		
预置条件	环境部署完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.1.4RDD 的缓存部分		
测试数据	已存在的 RDD 对象，或者某个文件的 uri。		

测试人员	于思民
操作步骤	
1. 启动 Master 节点; 2. 启动 Slave 节点; 3. 执行 spark-shell 命令进入命令行; 4. 使用 SparkContext 对象从某一文件 uri 读取数据; 5. 将得到的 RDD 对象缓存起来; 6. 调用 API, 检测是否有该 RDD 对象的缓存。	
预测结果	内存中有该 RDD 对象的缓存

### 3.1.5 RDD 检查点测试

为避免缓存丢失而进行重复计算的开销, Spark 设计了检查点 (checkpoint) 机制。该测试用例设计的目的是测试 RDD 检查点是否能正确创建和使用。RDD 检查点创建和使用正常则返回成功的反馈, 否则返回失败的反馈消息。

测试用例的描述如下:

表 8 创建 RDD 检查点的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	RDD 检查点创建和使用	编制人	于思民
用例 ID	TC005	时间	2016-05-19
相关用例	TC001		
功能特性	RDD 检查点		
测试目的	测试环境部署效果以及 RDD 检查点设置		
预置条件	环境部署完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.1.6RDD 的检查点部分		
测试数据	已存在的 RDD 对象，或者某个文件的 uri。		
测试人员	于思民		
操作步骤			
1. 启动 Master 节点；			

2. 启动 Slave 节点;
3. 执行 spark-shell 命令进入命令行;
4. 使用 SparkContext 对象从某一文件 uri 读取数据;
5. 为得到的 RDD 对象创建检查点;
6. 调用 API, 检测是否有该 RDD 对象的检查点, 及检查点的相关属性 (时间、大小等) 的正确性

预测结果

检查点的相关属性与该 RDD 对象相一致

表 5 RDD 检查点的测试用例

## 3.2 数据存储测试

数据存储模块的用例图如下图所示。

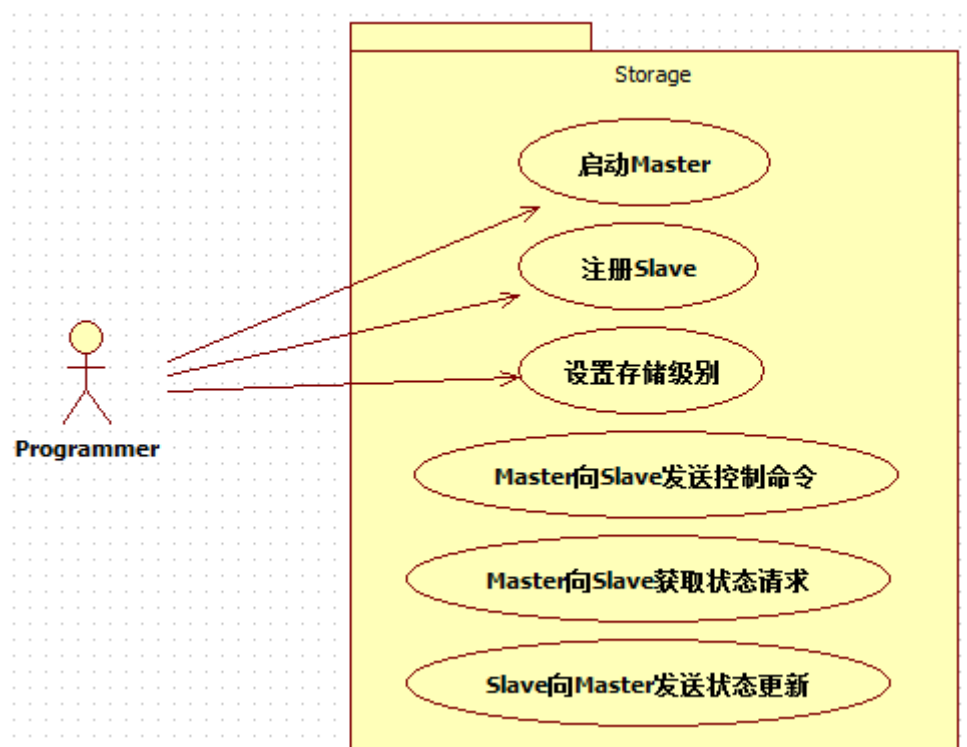


图 2 数据存储模块用例图

### 3.2.1 启动 Master 测试

Spark 集群启动时, 同时也要启动主从结构的数据存储节点 (进程)。该测试用例设计的目的是检查 Master 节点是否能正确启动。正确启动则返回成功结

果，否则报告启动错误。

测试用例描述如下：

表 9 启动 Master 的测试用例

项目	Spark 研究与应 用	版本	v2.0
功能需求	Master 启动	编制人	武一杰
用例 ID	TC006	时间	2016-05-19
相关用例	无		
功能特性	Master 的启动		
测试目的	确保 Master 能够成功启动		
预置条件	RDD 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.2.1 启动 Master 部分		
测试数据	无		
测试人员	武一杰		
操作步骤			
1.设置 Master 节点配置信息			
2.发送启动 Master 命令			
3.读取 Master 启动信息			
预测结果	Master 启动信息显示 Master 节点正常。		

### 3.2.2 注册 Slave 测试

在启动主节点后，Slave 节点会启动，并向 Master 节点注册，成为其从节点。本测试用例设计的目的是测试 Slave 节点是否正确向 Master 节点注册，即 Slave 的配置是否正确。Slave 注册成功则返回的 Slave 启动信息中显示 Slave 启动成功，否则失败。

测试用例描述如下：

表 10 注册 Slave 的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	Slave 注册	编制人	武一杰

用例 ID	TC007	时间	2016-05-19
相关用例	TC006		
功能特性	Slave 功能		
测试目的	测试 Slave 的配置功能		
预置条件	RDD 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.2.2 注册 Slave 部分		
测试数据	无		
测试人员	武一杰		
操作步骤			
1.读取 Master 节点信息			
2.启动 Master 节点			
3.读取 Slave 节点信息			
4.Slave 向 Master 发送状态更新信息,进行注册			
5.Master 获取 Slave 节点的启动信息			
预测结果	Master 能够读取到 Slave 节点的注册信息（启动信息）		

### 3.2.3 Master 向 Slave 发送控制命令测试

Master 节点可以向 Slave 发送控制信息，如删除数据块、获取 RDD 相关的 RDD、广播变量相关的数据等。Master 向 Slave 发送正确的操作命令请求之后，Slave 返回相应的处理信息则表示成功，否则失败。

用例描述如下：

表 11 Master 向 Slave 发送控制命令的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	Master 向 Slave 发送控制命令	编制人	于思民
用例 ID	TC008	时间	2016-05-19
相关用例	TC006, TC007		
功能特性	发送控制 Slave 命令		
测试目的	测试 Master 与 Slave 的连接状态与控制功能		



预置条件	RDD 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.2.3Master 向 Slave 发送控制命令部分		
测试数据	无		
测试人员	于思民		
操作步骤			
1.读取 Master 节点信息 2.启动 Master 节点 3.读取 Slave 节点信息 4.Slave 向 Master 的特定端口发送状态更新信息,进行注册 5.Master 获取 Slave 节点的启动信息 6.Master 向 Slave 发送控制命令，并监听 Slave 的状态更新			
预测结果	Slave 的状态更新结果与控制命令相一致		

表 8 Master 向 Slave 发送控制命令的测试用例

### 3.2.4 Master 获取 Slave 状态测试

Master 有时还需要获取数据块的状态等信息，这时 Master 需要向 Slave 发送获取状态的请求。获取结果成功则 Slave 向 Master 发送状态信息，Master 获取到 Slave 发送过来的状态信息；否则 Master 未收到信息，失败。

用例描述如下所示：

表 12 Master 获取 Slave 状态的测试用例

项目	Spark 研究与应用	版本	v1.0
功能需求	Master 获取 Slave 状态	编制人	于思民
用例 ID	TC009	时间	2016-05-19
相关用例	TC006, TC007		
功能特性	获取 Slave 状态		
测试目的	测试从 Slave 状态信息获取		
预置条件	RDD 测试完成	特殊规程说明	无

参考信息	<G-系统需求规格说明书-V3.2>中 3.2.4Master 向 Slave 获取状态请求部分
测试数据	无
测试人员	于思民
操作步骤	
1.读取 Master 节点信息 2.启动 Master 节点 3.读取 Slave 节点信息 4.Slave 向 Master 的特定端口发送状态更新信息,进行注册 5.Master 获取 Slave 节点的启动信息，并获取 Slave 状态	
预测结果	能够获取 Slave 的状态，与 Slave 的状态一致。

### 3.2.5 Slave 向 Master 发送状态更新测试

Slave 管理和维护的数据发生变更时，需要向 Master 发送状态更新，与 Master 保持同步。Slave 向 Master 发送状态更新成功则收到 Master 提供的更新后的 Slave 信息，同步成功。否则同步失败。

测试用例描述如下：

表 13 Slave 向 Master 发送状态更新

项目	Spark 研究与应用	版本	v2.0
功能需求	Slave 向 Master 发送状态更新	编制人	于思民
用例 ID	TC010	时间	2016-05-19
相关用例	TC006, TC007		
功能特性	状态更新测试		
测试目的	Slave 向 Master 发送状态更新		
预置条件	环境部署完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.2.5Slave 向 Master 发送状态更新部分		
测试数据	无		

测试人员	于思民
操作步骤	
1.读取 Master 节点信息 2.启动 Master 节点 3.读取 Slave 节点信息 4.Slave 向 Master 的特定端口发送状态更新信息,进行注册 5.Master 监听特定端口, 获取状态更新信息	
预测结果	Master 获取到状态更新信息

### 3.3 集群部署

Spark 是一个快速数据处理引擎或者平台。Spark 需要庞大的集群节点来支撑其快速的数据处理。由于已经存在比较成熟的集群部署框架, Spark 可以使用这些已有的集群, 在这些集群上运行 Spark 作业。

Spark 为适应已有的集群, 对各类集群框架进行适配。Spark 集群有以下部署方式: Local、Standalone、Hadoop YARN、Mesos、EC2 等方式。不同的部署方式就有不同的作业调度模式。

#### 3.3.1 Local 模式

Spark 可以部署在单机节点上, 也就是说, Local 部署模式下, 没有集群的概念, 所有的工作节点都以进程或者线程的方式出现。

Local 模式部署的测试用例如下:

表 14 Local 模式部署的测试用例

项目	Spark 研究与应 用	版本	v2.0
功能需求	Local 模式部署 的测试	编制人	阳艳红
用例 ID	TC011	时间	2016-05-19
相关用例	无		

功能特性	Local 模式部署		
测试目的	检测 Local 模式部署的正确性		
预置条件	无	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.3.1Local 模式		
测试数据	无		
测试人员	阳艳红		
操作步骤			
1.设置 Spark 的配置信息，其中将 Spark 集群的部署模式配置为 Local; 2.启动 Master 节点; 3.启动 Slave 节点; 4.执行 JConsole 可执行程序，查看已经启动的 java 线程 ; 5.使用 Spark 提供的应用程序例子，在此模式下提交作业运行;			
预测结果	JConsole 显示的线程数，符合设置的 Master 和 Slave 节点数;  应用程序跑出的结果正确。		

### 3.3.2 Standalone 模式

对于 Local 模式下，主节点和工作节点都以进程的方式存在于单一节点。这些都不是分布式的集群，或者只能称之为伪分布式的进群部署。Spark 还应提供真正的分布式部署模式，即 Spark Standalone。

在 Spark Standalone 模式下，Spark 集群有 Master 节点和 Worker 节点构成，用户程序通过与 Master 节点交互，申请所需的资源，Worker 节点负责具体 Executor 的启动运行。

对 Standalone 模式部署的测试，主要是检测 Standalone 模式集群启动后是否运行正常。

测试用例如下：

表 15 Standalone 模式部署的测试用例

项目	Spark 研究与应	版本	v2.0
----	------------	----	------

	用		
功能需求	Standalone 模式部署的测试	编制人	阳艳红
用例 ID	TC012	时间	2016-05-19
相关用例	无		
功能特性	Standalone 模式部署		
测试目的	检测 Standalone 模式部署的正确性		
预置条件	无	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.3.2Standalone 模式		
测试数据	无		
测试人员	阳艳红		
操作步骤			
1.设置 Spark 的配置信息，其中将 Spark 集群的部署模式配置为 Standalone； 2.启动 Master 节点； 3.启动 Slave 节点； 4.执行 JPS 命令，查看已经启动的 java 进程； 5.使用 Spark 提供的应用程序例子，在此模式下提交作业运行；			
预测结果	JPS 显示的进程数，符合设置的 Master 和 Slave 节点数； 应用程序跑出的结果正确。		

### 3.3.3 Mesos 模式

Apache Mesos 采用了 Master/Slave 的架构。主要由四部分组成：Mesos Master、Mesos Slave、Framework（Mesos Application）和 Executor。

为使得 Spark 的计算可以运行在 Apache Mesos 之上。Spark 还要对 Apache Mesos 进行适配，即 Spark 应有一套 Mesos 部署模式。

对 Mesos 模式部署的测试，主要是检测 Spark 与 Mesos 集群适配启动后是否运行正常。

测试用例如下：

表 16 Mesos 模式部署的测试用例

项目	Spark 研究与应 用	版本	v2.0
功能需求	Mesos 模式部 署的测试	编制人	阳艳红
用例 ID	TC013	时间	2016-05-19
相关用例	无		
功能特性	Mesos 模式部署		
测试目的	检测 Mesos 模式部署的正确性		
预置条件	Mesos 集群搭 建好	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.3.3Mesos 模式		
测试数据	无		
测试人员	阳艳红		
操作步骤			
1.设置 Spark 的配置信息，将集群部署模式配置为 Mesos（与 Mesos 适配的相关配置信息）； 2.启动 Mesos 集群； 3.启动 Master 节点； 4.启动 Slave 节点； 5.执行 JPS 命令，查看已经启动的 java 进程； 6.在 Mesos 各节点执行 JPS 命令； 7.使用 Spark 提供的应用程序例子，在此模式下提交作业运行；			
预测结果	JPS 显示的进程，符合设置的 Master 和 Slave 节点数； JPS 显示的进程，符合 Mesos 正常运行的应有进程； 应用程序跑出的结果正确。		

### 3.3.4 YARN 模式

Spark 也为 Hadoop 集群适配了 YARN 部署模式。

测试用例如下：

表 17 YARN 模式部署的测试用例

项目	Spark 研究与应 用	版本	v2.0
功能需求	YARN 模式部 署的测试	编制人	阳艳红
用例 ID	TC014	时间	2016-05-19
相关用例	无		
功能特性	YARN 模式部署		
测试目的	检测 YARN 模式部署的正确性		
预置条件	YARN 集群搭 建好	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.3.4YARN 模式		
测试数据	无		
测试人员	阳艳红		
操作步骤			
1.设置 Spark 的配置信息，将集群部署模式配置为 YARN（与 YARN 适配的相关配置信息）； 2.启动 Hadoop 集群； 3.启动 Master 节点； 4.启动 Slave 节点； 5.执行 JPS 命令，查看已经启动的 java 进程； 6.在 Hadoop 各节点执行 JPS 命令； 7.使用 Spark 提供的应用程序例子，在此模式下提交作业运行；			
预测结果	JPS 显示的进程，符合设置的 Master 和 Slave 节点数； JPS 显示的进程，符合 Hadoop 正常运行的应有进程； 应用程序跑出的结果正确。		

### 3.4 作业调度

#### 3.4.1 用户作业提交测试

该测试用例设计的目的是测试用户编写的应用程序能否正确提交，生成 JOB 并通过 DAGScheduler 提交到事件处理器。测试中提交完作业后的结果包括提交成功和提交失败。从事件处理器中获取事件处理结果显示处理成功则表示提交成功，否则失败。

测试用例的描述如下：

表 18 用户作业提交的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	用户作业提交	编制人	王铖成
用例 ID	TC015	时间	2016-05-19
相关用例	无		
功能特性	用户作业提交		
测试目的	测试用户作业提交功能		
预置条件	Master 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.4.1 用户提交作业部分		
测试数据	搜狗 2012 年 8 月的用户搜索数据		
测试人员	王铖成		
操作步骤			
1.用户向 Spark 集群提交作业（搜狗月度前 100 热门关键词检索）			
2.作业输出到 DAGScheduler			
3.生成 Job			
4.DAGScheduler 将作业输出到事件处理器			
5.从事件处理器获取事件处理结果			
预测结果	检索得到前 100 热门关键词正确		



### 3.4.2 DAG 创建测试

用户作业（Job）在 Spark 集群中执行的过程是通过原始数据集的 RDD 转换过程后执行 RDD Action 操作，并得到最终的计算结果的过程。在 RDD 转换的过程中，为了使运行在 RDD 上任务最大化并行操作，Spark 将这些 RDD 划分阶段，形成一个有向无环图的 RDD 转化流程。DAG 创建是为用户应用程序生成 RDD 转换的有向无环图。该 DAG 创建测试用例设计的目的是测试 DAG 是否能正确创建。创建之后可以查看 DAG 的信息，创建成功则该 DAG 生成，否则失败。

测试用例描述如下：

表 19 创建 DAG 的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	DAG 创建	编制人	王铖成
用例 ID	TC016	时间	2016-05-19
相关用例	TC015		
功能特性	计算功能		
测试目的	DAG 创建测试		
预置条件	Master 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.4.2DAG 创建部分		
测试数据	无		
测试人员	王铖成		
操作步骤			
1.获取计算任务			
2.读取 RDD 信息			
3.将 RDD 转化成 DAG			
4.查看生成的 DAG			
预测结果	生成的 DAG 信息符合计算任务的依赖关系。		

### 3.4.3 Stage 划分测试

为了提高作业执行的并行化效率，Spark 根据 RDD 转换的关系，即依赖关系，将 RDD 进行不同阶段的划分。该测试用例的设计目的是测试 STAGE 能否正确划分并转换为相应的 DAG。划分成功并转换则结果信息中显示相应的 STAGE 信息，否则失败。

测试用例的描述如下：

表 20 划分 Stage 的测试用例

项目	Spark 研究与应用	结果版本	v2.0
功能需求	Stage 划分	编制人	王铖成
用例 ID	TC017	时间	2016-05-19
相关用例	TC015		
功能特性	Stage		
测试目的	测试阶段划分功能		
预置条件	Master 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.4.3Stage 划分部分		
测试数据	无		
测试人员	王铖成		
操作步骤			
1.获取触发 Action 操作的 RDD1 2.获取该 RDD 所依赖的父 RDD2 3.将与 RDD2 与依赖关系的分为同一 STAGE 4.将生成的 STAGE 转化为 DAG 5.查看结果			
预测结果	结果现实的阶段划分与实际向一致。		

### 3.4.4 TaskSet 生成测试

在划分完 Stage 阶段后，DAGScheduler 要取得需要计算的 Partition，在判断哪些 Partition 需要计算后，就会为每个 Partition 生成 Task，然后将这些 Task 封

装成 TaskSet，即任务集，最后将任务集提交给 TaskScheduler。该测试用例设计的目的是测试 TaskSet 能否正确生成并提交。结果成功则在获取的提交信息中有提交成功的信息，否则失败。

测试用例的描述如下：

表 21 生成 Taskset 的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	Taskset 生成	编制人	王铖成
用例 ID	TC018	时间	2016-05-19
相关用例	TC015		
功能特性	TaskSet		
测试目的	测试 TaskSet 的生成功能		
预置条件	Master 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.4.4TaskSet 生成部分		
测试数据			
测试人员	于思民,王铖成,阳艳红,武一杰		
操作步骤			
1.判断需要进行计算 Partion			
2.依次为每个 Partition 创建对应的 Task			
3.将这些 Task 封装成对应的 Task			
4.向 TaskScheduler 提交任务			
5.查看提交结果			
预测结果	生成的 TaskSet 与配置信息相符合。		

### 3.4.5 Task 调度测试

任务的调度操作是由 TaskScheduler 进行，TaskScheduler 只是 Spark 提供的任务调度的抽象，具体的任务操作可以由 YAEN/Mesos 等第三方资源管理器进行管理。该测试用例的设计目的是测试 Task 能否正确调度并发送到 Executor 执行。如果 Task 被正确调度并提交到 Executor 执行，则 Executor 返回相应的处理结果，调度成功；否则没有处理结果，调度失败。

测试用例的描述如下：

表 22 Task 调度的测试用例

项目	Spark 研究与应用	版本	v2.0
功能需求	Task 调度测试	编制人	王铖成
用例 ID	TC019	时间	2016-05-19
相关用例	TC015		
功能特性	任务调度		
测试目的	测试任务调度功能		
预置条件	Master 测试完成	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 3.4.5Task 调度部分		
测试数据			
测试人员	王铖成		
操作步骤			
1.提交任务 2.分配资源 3.发送到执行模块 4.执行作业 5.查看结果			
预测结果	任务调度结果符合进群调度策略。		

## 4. 非功能需求测试用例描述

非功能需求测试用例的设计主要是考察 Spark 在执行过程中的运行速度、执行的容错性、安全性等非功能需求是否满足软件系统规格需求。执行前面设计的一系列测试用例从中总结出相应的信息并进行记录。

### 4.1.1 运行速度

这里用实现相同的用户程序、用户数据分别提交到 Hadoop 集群和 Spark 集群运行，统计二者的运行时长。运行速度的测试用例如下。

表 23 运行速度的测试用例

项目	Spark 研究与应 用	版本	v2.0
非功能性需求	运行速度	编制人	阳艳红
用例 ID	TC020	时间	2016-05-19
相关用例	无		
非功能特性	运行速度		
测试目的	与 Hadoop 对比测试 Spark 的运行速度		
预置条件	无	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 5.1 运行速度		
测试数据	搜狗 2012 年 8 月的用户搜索数据		
测试人员	阳艳红		
操作步骤			
1.设置 Spark 的配置信息，将集群部署模式配置为 YARN（与 YARN 适配的相关配置信息）； 2.启动 Hadoop 集群； 3.启动 Master 节点； 4.启动 Slave 节点； 5.将用户应用程序（搜狗月度前 100 热门关键词检索）分别提交到 Hadoop 集群和 Spark 集群执行。 注：两个集群运行作业的内存与 CPU 配置信息要一样。 6.统计作业运行时间，并对比。			
预测结果	Spark 的作业运行时长比 Hadoop 的作业运行时长短，比 Hadoop 快了 10-30 倍。		

#### 4.1.2 容错性

表 24 容错性的测用例

项目	Spark 研究与应 用	版本	v2.0
----	-----------------	----	------

非功能性需求	容错性	编制人	于思民	
用例 ID	TC021	时间	2016-05-19	
相关用例	无			
非功能特性	容错性			
测试目的	检测 Spark 的容错性			
预置条件	无	特殊规程说明	无	
参考信息	<G-系统需求规格说明书-V3.2>中 5.1 运行速度			
测试数据	搜狗 2012 年 8 月的用户搜索数据			
测试人员	于思民			
测试过程描述				
编号	步骤	预期结果	评价标准	测试结论
1	1. 设置 Spark 的配置信息, 并将其中某项配置设置为异常值。 2. 启动 Master、Slave 节点 3. 观察各节点是否正常启动	1. 节点服务启动	预期与实际相一致	无
2	1.设置 Spark 的配置信息, 将集群部署模式配置为 YARN（与 YARN 适配的相关配置信息）; 2.启动 Hadoop 集群; 3.启动 Master 节点; 4.启动 Slave 节点; 5.提交 Spark 提供的例子作业; 6.将 Hadoop 集群的某个节点关闭	作业正常运行	预期与实际相一致	无

	7.查看作业执行结果			
3	1.设置 Spark 的配置信息，将集群部署模式配置为 YARN（与 YARN 适配的相关配置信息）； 2.启动 Hadoop 集群； 3.启动 Master 节点； 4.启动 Slave 节点； 5.提交 Spark 提供的例子作业； 6.将 Hadoop 集群中执行任务的节点关闭 6.查看作业执行结果	作业正常运行，作业运行时间变长	预期与实际相一致	无

#### 4.1.3 安全性

表 25 安全性的测试用例

项目	Spark 研究与应 用	版本	v2.0
非功能性需求	安全性	编制人	武一杰
用例 ID	TC022	时间	2016-05-19
相关用例	无		
非功能特性	容错性		
测试目的	检测 Spark 的容错性		
预置条件	无	特殊规程说明	无
参考信息	<G-系统需求规格说明书-V3.2>中 5.1 运行速度		
测试数据	搜狗 2012 年 8 月的用户搜索数据		
测试人员	武一杰		
测试过程描述			

编号	步骤	预期结果	评价标准	测试结论
1	1. 设置 Spark 的配置信息，将 <code>spark.ui.filters</code> 参数设置为用户 A，启用 <code>javax.servlet.filters</code> 进行 UI 安全设置。 2. 启动 Master、Slave 节点 3. 用户 B 登录到 Spark，并查看用户 A 的 UI 界面	用户 B 无法获取 UI 界面数据	预期与实际相一致	无
2	1. 设置 Spark 的配置信息； 2. 启动 Master、Slave 节点； 3. 用户 A 提交用户 B 某目录下的作业（A 用户与 B 用户不属于相同的组）到 Spark 集群	Spark 拒绝作业的执行	预期与实际相一致	无

## 5. 参考文献

1. Spark 官方指导文档翻译：《spark-programming-guide-zh-ch.pdf》
2. Spark 官方开发者文档翻译：《spark-developer-guide.pdf》
3. 网络 Spark 运维实战文档：《spark-operation-maintenance-management.pdf》
4. 网络 Apache Spark 设计与实现文档：《SparkInternals.pdf》



- 
5. 张安站 《Spark 技术内幕：深入解析 Spark 内核架构设计与实现原理》，机械工业出版社. 2015
  6. 耿嘉安《深入理解 SPARK：核心思想与源码分析》，机械工业出版社. 2016
  7. 《G-系统需求规格说明书-V3.2》 2016.5.12