

# 《Storm源码分析》

## 图书基本信息

书名：《Storm源码分析》

13位ISBN编号：9787115371261

出版时间：2014-11-1

作者：李明,王晓鹏

页数：472

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.com)

# 《Storm源码分析》

## 内容概要

本书从源代码的角度详细分析了Storm 的设计与实现，共分为三个部分，第一部分介绍了Storm 的基本原理以及Storm 集群系统的搭建方法，第二部分深入剖析了Storm 的底层架构，如Nimbus、Supervisor

、Worker 以及Task，第三部分系统讨论了Storm 如何实现可靠的消息传输，如Transaction Topology 以及Trident。

本书适用于程序员、架构师以及计算机专业的学生。

# 《Storm源码分析》

## 作者简介

李明

2001~2007年在哈尔滨工业大学完成本科以及硕士的学习。微软搜索技术部门资深研发工程师及高级开发经理，擅长Linux、Clojure、Java、C#等多种开发技术，长期致力于大数据、分布式系统的研究和应用，目前致力于实时性分布式处理系统的研究与开发。

王晓鹏

2004~2011年在北京邮电大学完成本科以及硕士的学习。微软搜索技术部门高级研发工程师，擅长Windows phonePhone、Silverlight、Clojure、Java、C#等多种开发技术，一直致力于大数据处理、分布式系统的研究和应用，目前致力于实时性分布式处理系统的研究与开发。

## 书籍目录

第1章 总体架构与代码结构	1
1.1 Storm的总体结构	1
1.2 Storm的元数据	3
1.2.1 元数据介绍	3
1.2.2 Storm怎么使用这些元数据	4
1.3 Storm的代码结构	7
1.3.1 Clojure代码	7
1.3.2 Java代码	8
1.3.3 Trident代码	9
1.3.4 其他代码	10
第2章 搭建Storm集群	11
2.1 搭建单机Storm集群	11
2.2 搭建多机Storm集群	14
2.2.1 设置环境	14
2.2.2 启动Storm集群	15
2.2.3 提交Topology	15
2.3 WordCountTopology介绍	15
2.3.1 RandomSentenceSpout	15
2.3.2 SplitSentence	16
2.3.3 WordCount	17
2.3.4 WordCountTopology构建	17
第3章 Storm编程基础	19
3.1 Fields定义	19
3.2 Tuple接口	20
3.3 常用声明接口	21
3.3.1 配置声明接口	22
3.3.2 输入声明接口	23
3.3.3 输出字段声明接口	24
3.3.4 组件声明接口	25
3.4 Spout输出收集器	25
3.4.1 ISpoutOutputCollector和SpoutOutputCollector	25
3.4.2 Executor中ISpoutOutputCollector的实现	27
3.5 Bolt输出收集器	28
3.5.1 IOutputCollector和OutputCollector	28
3.5.2 IBasicOutputCollector和BasicOutputCollector	31
3.5.3 BatchOutputCollector和BatchOutputCollectorImpl	32
3.5.4 Executor中的IOutputCollector实现	34
3.6 组件接口	35
3.7 Spout接口	35
3.7.1 ISpout	36
3.7.2 IRichSpout	38
3.8 Bolt接口	38
3.8.1 IBolt	38
3.8.2 IRichBolt	40
3.8.3 IBasicBolt	40
3.8.4 IBatchBolt	42
3.8.5 小结	45

3.9	Storm数据结构	46
3.9.1	GlobalStreamId	46
3.9.2	消息分组方式	46
3.9.3	StreamInfo	47
3.9.4	ShellComponent	47
3.9.5	ComponentObject	47
3.9.6	ComponentCommon	47
3.9.7	SpoutSpec	48
3.9.8	Bolt	48
3.9.9	StormTopology	49
3.9.10	TopologySummary	49
3.9.11	SupervisorSummary	49
3.9.12	ClusterSummary	50
3.9.13	BoltStats	50
3.9.14	SpoutStats	50
3.9.15	统计信息	50
3.9.16	DRPC	51
3.10	基本Topology构建器	52
3.10.1	TopologyBuilder	52
3.10.2	ConfigGetter	55
3.10.3	SpoutGetter和BoltGetter	55
3.10.4	一个简单例子	56
3.11	异常处理	57
第4章	基础函数和工具类	58
4.1	计时器	58
4.1.1	mk-timer	58
4.1.2	check-active!	60
4.1.3	schedule	60
4.1.4	schedule-recurring	60
4.1.5	cancel-timer	61
4.2	async-loop	61
4.3	event-manager	62
4.4	even-sampler	63
4.5	ZooKeeper工具类	64
4.5.1	mk-client	64
4.5.2	create-node	65
4.5.3	get-data	65
4.5.4	进程内启动ZooKeeper	66
4.6	LocalState	66
4.7	ClusterState	68
4.8	StormClusterState	69
第5章	通信机制	71
5.1	进程间通信	71
5.1.1	进程间通信协议	71
5.1.2	LocalCluster模式实现	72
5.1.3	分布式模式实现	73
5.1.4	协议使用	75
5.2	进程内通信	77
5.2.1	Disruptor Queue的使用	77

5.2.2	DisruptorQueue的Clojure处理器	80
第6章 Nimbus 81		
6.1	Nimbus服务接口定义	81
6.2	Nimbus相关的数据结构	83
6.2.1	Java 数据结构	83
6.2.2	Clojure数据结构	84
6.3	Nimbus中的线程介绍	86
6.3.1	mk-assignments	87
6.3.2	do-cleanup	89
6.3.3	clean-inbox	90
6.4	Topology状态转移	90
6.4.1	transition-name!	90
6.4.2	transition!	91
6.4.3	state-transitions	92
6.5	启动Nimbus服务	96
6.5.1	launch-server !	96
6.5.2	service-handler	97
6.6	关闭Nimbus服务	99
6.7	主要服务方法	99
6.7.1	submitTopology	99
6.7.2	kill、rebalance、activate、deactivate 方法	101
6.7.3	文件上传与下载	102
6.7.4	获取UI所需的信息	104
6.7.3	获取Topology	106
6.7.6	获取Storm配置项	107
6.8	主要辅助方法	107
6.8.1	system-topology!	107
6.8.2	normalize-topology	112
6.8.3	compute-new-topology->executor->node+port	114
6.8.4	compute-executors	117
第7章 Scheduler 119		
7.1	IScheduler接口	119
7.2	EvenScheduler	120
7.2.1	schedule-topologies-evenly	120
7.2.2	schedule-topology	121
7.2.3	get-alive-assigned-node+port->executors	122
7.2.4	sort-slots	123
7.3	DefaultScheduler	124
7.3.1	default-schedule	124
7.3.2	slots-can-reassign	126
7.3.3	bad-slots	126
7.4	IsolationScheduler	127
7.5	调度示例	131
7.5.1	EvenScheduler和DefaultScheduler	131
7.5.2	IsolationScheduler	134
第8章 Scheduler 137		
8.1	与Supervisor相关的数据结构	137
8.1.1	standalone-supervisor	137

8.1.2	Supervisor的数据	138
8.1.3	本地存储数据	139
8.2	Supervisor中的线程	140
8.2.1	计时器线程	140
8.2.2	同步Nimbus任务的线程	140
8.2.3	管理Worker进程的线程	143
8.3	启动Supervisor	145
8.4	关闭Supervisor	147
8.5	重要方法介绍	147
8.5.1	launch-worker	147
8.5.2	read-allocated-workers	150
8.5.3	wait-for-worker-launch	151
8.5.4	shutdown-worker	152
8.5.5	download-storm-code	152
第9章	Worker	155
9.1	Worker中的数据	155
9.2	Worker中的计时器	157
9.2.1	Worker的心跳	157
9.2.2	Executor的心跳	158
9.2.3	Worker中对ZMQ连接的维护	159
9.2.4	从ZooKeeper获取Topology的活跃情况	161
9.2.5	小结	162
9.3	创建Worker	163
9.4	关闭Worker	164
9.5	重要辅助方法介绍	165
9.5.1	Worker中的接收函数	166
9.5.2	Worker中的发送函数	167
9.5.3	获取属于Worker的Executor	169
9.5.4	创建Executor的接收消息队列和查找表	169
9.5.5	下载Topology的配置项以及代码	170
9.6	小结	171
第10章	Executor	172
10.1	Executor的数据	172
10.2	Executor的输入和输出	174
10.2.1	Executor的输入及处理	174
10.2.2	Executor的输出及发送	175
10.3	Spout类型的Executor	176
10.3.1	准备消息循环的数据	176
10.3.2	Spout输入处理函数	178
10.3.3	Spout消息发送函数	180
10.3.4	Spout对象的初始化	181
10.3.5	消息循环	182
10.4	Bolt类型的Executor	184
10.4.1	准备消息循环的数据	184
10.4.2	Bolt输入处理函数	184
10.4.3	Bolt的消息发送函数	185
10.4.4	Bolt对象的初始化	185
10.4.5	消息循环	186
10.5	创建Executor	187

10.6	辅助函数介绍	188
10.6.1	组件的Grouper函数	188
10.6.2	带流量控制的错误报告方法	193
10.6.3	触发系统Ticks	194
10.7	小结	196
第11章	Task	198
11.1	Task的上下文对象	198
11.1.1	TopologyContext	198
11.1.2	GeneralTopologyContext	199
11.1.3	WorkerTopologyContext	200
11.1.4	TopologyContext	201
11.2	创建Task数据	202
11.3	mk-tasks-fn函数	204
11.4	send-unanchored	205
11.5	创建Task	206
11.6	Storm中传输的消息以及序列化	206
第12章	Storm的Ack框架	208
12.1	Acker Bolt的实现分析	209
12.2	启动消息跟踪	211
12.3	消息跟踪	212
12.4	Ack机制的例子	214
第13章	系统运行统计	216
13.1	基础数据结构以及更新算法	216
13.1.1	滑动窗口的数据结构	216
13.1.2	滑动窗口的回调函数	220
13.1.3	滑动窗口集合的类型	221
13.2	Storm中的统计信息	222
13.2.1	Stats中定义的统计类别	222
13.2.2	运行统计的更新	223
13.2.3	运行统计的更新时间点	223
13.2.4	获取统计数据	228
13.3	运行统计的Thrift结构	229
第14章	系统运行统计的另一种实现	231
14.1	内置统计信息的计算	231
14.1.1	MultiCountMetric	232
14.1.2	MultiReducedMetric	233
14.2	内置统计类型	234
14.2.1	Spout类型的内置统计	235
14.2.2	Bolt类型的内置统计	235
14.3	统计触发消息	235
14.3.1	注册统计信息	236
14.3.2	触发消息的产生与发送	237
14.3.3	处理统计触发消息	238
14.4	运行统计收集节点	239
14.5	SystemBolt	241
第15章	事务Topology的实现	243
15.1	事务Topology的实现概述	243
15.1.1	事务Topology的类型	244
15.1.2	事务Topology的类关系	245



15.2	ITransactionalSpout接口	246
15.3	协调Spout节点的执行器	248
15.3.1	ZooKeeper客户端工具	248
15.3.2	协调Spout的执行器	255
15.3.3	消息发送Bolt的执行器	261
15.4	CoordinatedBolt的实现分析	264
15.4.1	TrackingInfo	264
15.4.2	CoordinatedOutput-Collector	265
15.4.3	CoordinatedBolt中的消息类型	267
15.4.4	成员变量以及主要方法分析	267
15.5	分区的事务类型	271
15.5.1	分区的事务Spout接口	271
15.5.2	分区的事务Spout的执行器	273
15.6	分区的模糊事务Spout	277
15.6.1	分区的模糊事务Spout的接口	277
15.6.2	模糊的事务Spout执行器	278
15.7	事务Topology的构建器	281
15.7.1	构建器的构造函数及成员变量	281
15.7.2	设置Bolt对象	283
15.7.3	构建Topology	284
15.7.4	输入流声明器	286
第16章	事务Topology示例	288
16.1	例子代码	288
16.1.1	分区的事务Spout	288
16.1.2	局部计数Bolt的实现	291
16.1.3	全局计数Bolt的实现	292
16.2	构建Topology	293
16.3	事务处理示例	295
第17章	Trident的Spout节点	298
17.1	ITridentSpout接口	298
17.1.1	BatchCoordinator接口	299
17.1.2	TridentSpoutCoordinator	300
17.1.3	MasterBatchCoordinator	301
17.1.4	消息发送节点接口	306
17.1.5	消息发送接口的执行器	306
17.2	适配IRichSpout接口	307
17.3	适配IBatchSpout接口	311
17.4	Trident中分区的Spout类型	311
17.4.1	分区Spout接口	311
17.4.2	分区Spout的执行器	313
17.5	模糊事务类型的Spout节点	316
17.5.1	模糊事务类型的Spout接口	317
17.5.2	模糊事务类型Spout的执行器	317
17.6	构建Spout节点	320
17.6.1	TridentTopology的newStream调用	320
17.6.2	TridentTopology中新DRPCStream调用	321
第18章	Trident的存储	322
18.1	存储的基本接口	322

18.2	MapState接口的实现	323
18.2.1	非事务类型的存储	324
18.2.2	事务类型的存储	325
18.2.3	模糊事务类型存储	327
18.3	值的序列化方法	329
18.4	数据更新接口	330
18.4.1	CombinerValueUpdater	330
18.4.2	ReducerValueUpdater	331
18.5	存储更新接口	331
18.5.1	ReducerAggStateUpdater	332
18.5.2	MapReducerAggStateUpdater	332
18.5.3	BaseStateUpdater	334
18.6	创建存储对象	334
第19章	Trident消息	336
19.1	ValuePointer	336
19.2	Factory接口及其实现	337
19.2.1	ProjectionFactory	338
19.2.2	FreshOutputFactory	339
19.2.3	OperationOutputFactory	339
19.2.4	RootFactory	341
19.3	消息工厂的例子	342
19.4	TridentTupleView	342
19.5	ComboList	343
第20章	Trident操作与处理节点	346
20.1	操作的基本接口	346
20.2	Aggregator实现	347
20.2.1	GroupedAggregator	348
20.2.2	ChainedAggregatorImpl	350
20.2.3	SingleEmitAggregator	353
20.3	用户接口及其实现	355
20.3.1	ReducerAggregator接口及其实现	355
20.3.2	CombinerAggregator接口及其实现	356
20.4	所有处理节点的上下文	357
20.4.1	单个处理节点的上下文	358
20.4.2	操作执行的上下文	359
20.5	Trident的输出收集器	359
20.5.1	FreshCollector	359
20.5.2	CaptureCollector	360
20.5.3	GroupCollector	360
20.5.4	AppendCollector	361
20.5.5	AddIdCollector	361
20.6	Trident的处理节点	362
20.6.1	TridentProcessor接口	363
20.6.2	PartitionPersistProcessor	363
20.6.3	StateQueryProcessor	365
20.7	聚集器的执行	367
第21章	Trident流的基本操作	370
21.1	流的成员变量和基础方法	370
21.1.1	流的成员变量	370

21.1.2	流节点名字	370
21.1.3	流的映射检查	372
21.1.4	添加节点	372
21.2	流映射操作	373
21.3	流的分组操作	374
21.4	流的逐行操作	374
21.5	流的分区操作	374
21.6	流的单聚集器聚集操作	376
21.7	流的多聚集器聚集操作	377
21.7.1	ChainedAggregatorDeclarer	377
21.7.2	分区上的局部聚集操作	379
21.7.3	全局聚集操作	379
21.7.4	含有多个聚集器的partitionAggregate操作	381
21.8	流的聚集操作	382
21.9	流的分区写入操作	383
21.10	查询操作	384
21.11	流的全局写入操作	384
21.12	流的操作与有向图构建	384
21.13	分组流	385
21.13.1	成员变量	385
21.13.2	逐行操作	385
21.13.3	分组流的分区聚集操作	386
21.13.4	查询操作	386
21.13.5	聚集操作	386
21.13.6	写入操作	387
21.14	利用流操作来构建Topology的例子	388
第22章	Trident中流的交互操作	392
22.1	基本接口	392
22.2	JoinerMultiReducer	393
22.2.1	成员变量及构造函数	393
22.2.2	execute方法	395
22.2.3	complete方法	397
22.3	GroupedMultiReducerExecutor	397
22.4	MultiReducerProcessor	399
22.5	连接操作	401
22.6	流合并操作	403
第23章	Trident中的Bolt节点	404
23.1	SubTopologyBolt	404
23.1.1	输入准备	404
23.1.2	成员变量	405
23.1.3	主要方法	406
23.2	Trident中的Bolt执行器	409
23.2.1	ITridentBatchBolt接口	410
23.2.2	TrackedBatch	410
23.2.3	定制的输出收集器	412
23.2.4	消息类型	414
23.2.5	数据成员分析	414
23.2.6	主要成员方法分析	416
第24章	Trident的执行优化	420

24.1	节点类型	420
24.1.1	基本节点类型	420
24.1.2	Spout节点	422
24.1.3	处理节点	422
24.1.4	分区节点	423
24.2	执行优化算法	426
24.2.1	节点组	426
24.2.2	节点组的合并算法	427
24.2.3	处理节点组中的分区节点	431
24.2.4	节点组以不同的方式收听相同流	431
24.2.5	执行优化后的节点组	434
24.2.6	计算节点组的并行度	434
第25章	Trident与DRPC	437
25.1	DRPC服务器	438
25.1.1	DRPC服务器的成员变量	438
25.1.2	DRPC用户接口及其实现	439
25.1.3	DRPC Topology端接口及其实现	440
25.1.4	启动DRPC服务器	441
25.2	DRPC的客户端	442
25.3	DRPC中Spout节点	443
25.4	DRPC Spout的执行器	446
25.5	completeDRPC操作	449
25.6	返回DRPC结果	451
第26章	Trident的Topology构建器	453
26.1	基本工具函数	453
26.1.1	committerBatches	453
26.1.2	fleshOutputStreamBatchIds	453
26.1.3	getOutputStreamBatchGroups	454
26.2	TridentTopologyBuilder	455
26.2.1	成员变量	455
26.2.2	设置Spout节点	456
26.2.3	设置Bolt节点	458
26.3	一个例子	460
第27章	多语言	462
27.1	ShellProcess	462
27.2	ShellBolt	464
27.2.1	成员变量	464
27.2.2	读写线程	465
27.3	ShellSpout	467
第28章	Storm中的配置项	469

# 《Storm源码分析》

## 精彩短评

- 1、这本书通篇粘贴代码，没有整体框架，没有逻辑梳理和示例，真是浪费钱浪费时间，还不如去看源码注释。
- 2、纯粹粘贴代码
- 3、好书！！！！

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu111.com](http://www.tushu111.com)