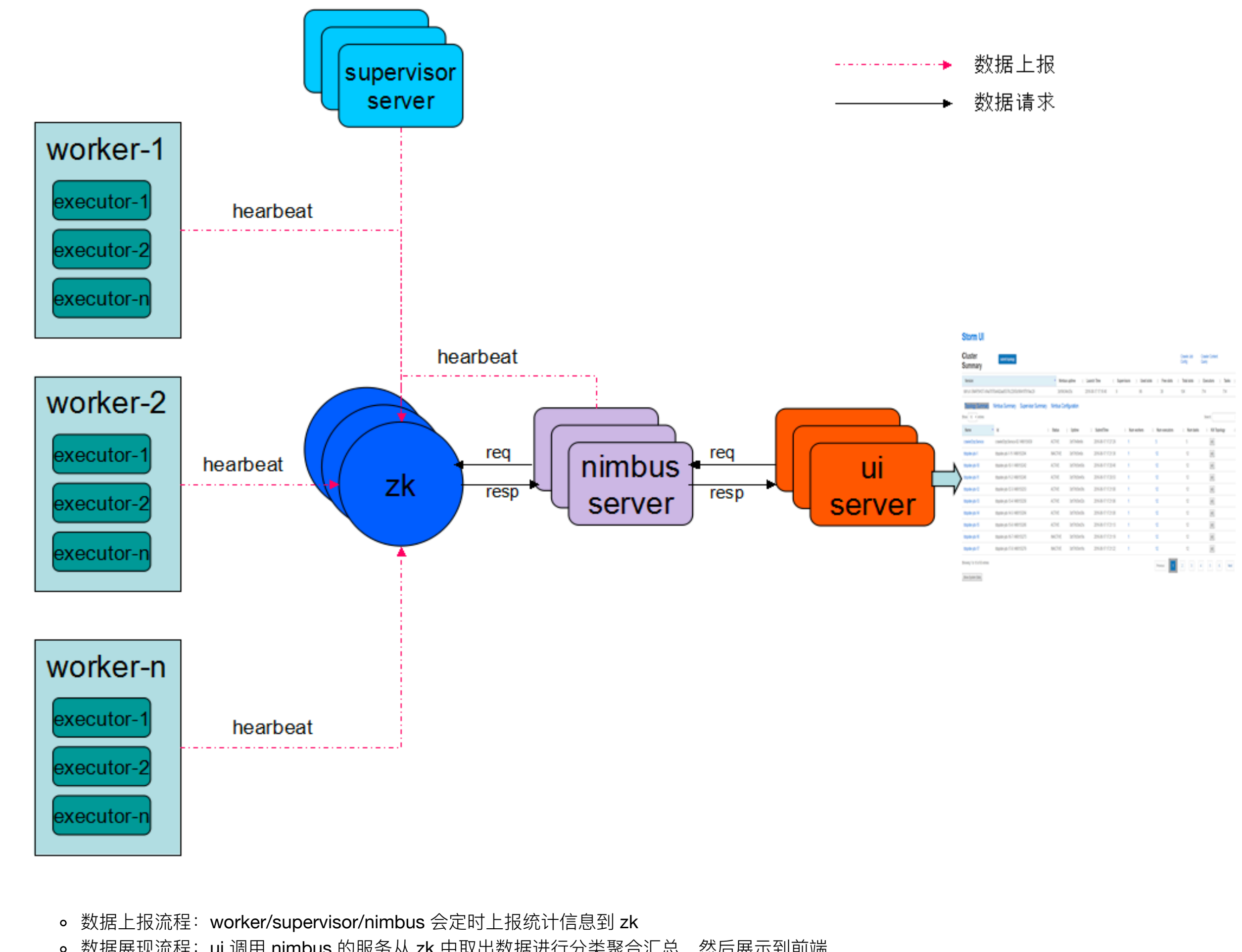


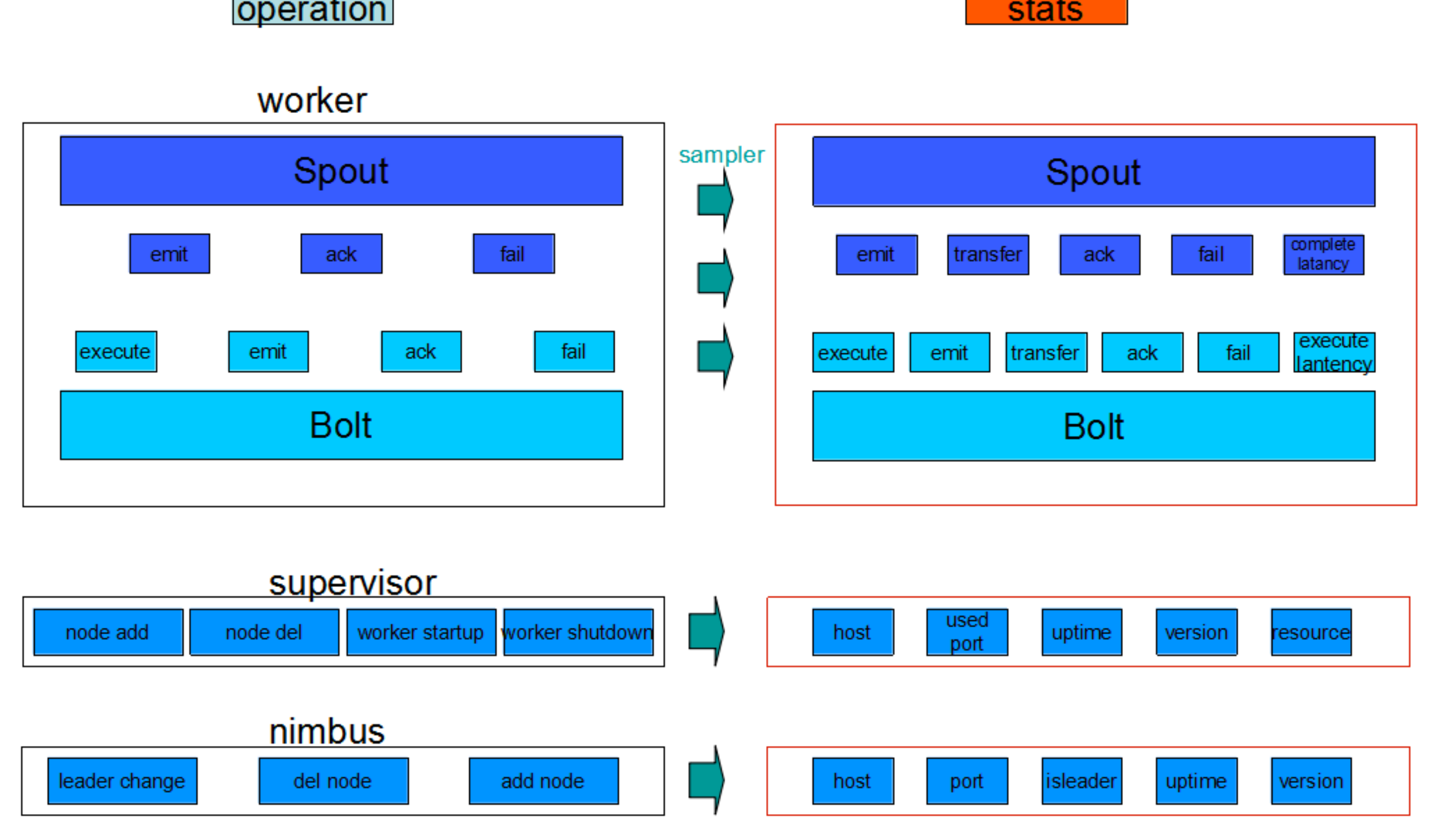
众所周知，storm 已经是业界主流的流时处理框架，Storm 被广泛应用于实时分析，在线机器学习，持续计算、分布式远程调用等领域。当前无论是内部还是外部论坛介绍原理的文档都比较多，但主要都是从运行机制和原理方面的介绍，在 UI 方面的介绍甚少，今天我试着向大家介绍一下 storm ui，一方面可以让大家了解一下 storm 的机制，另外也可以让大家更好的使用好 storm ui 协助大家自助解决使用过程中遇到的问题。

## 一. 数据上报及展示流程



## 二. 数据类型

通常我们要真正理解一个事物，通常都会从来龙去脉进行解剖；理解 storm ui 也是，想理解 storm 有什么数据，那么先去理解 storm 有哪些动作就事半功倍了，以下是 storm 中 worker/supervisor/nimbus 的基本操作及对应的数据类型，左边为操作，右边为数据。



- spout:**
  - emit: 向下游提交 tuple，对应的统计数据为 emit/transfer，emit=emit 执行数量，transfer=emit 执行数量 \* 下游 bolt 个数
  - ack: 成功流转之后的操作，对应的统计数据为 ack/complete-latency，ack=成功流转次数 complete-latency=成功流转平均时间
  - fail: 失败流转之后的操作 对应的统计数据为 fail，fail=失败流转次数
- bolt:**
  - execute: 接收到上游数据之后的执行操作 对应的统计数据为 execute/execute-latency，execute=执行次数 execute-latency=平均执行时间
  - emit: 与 spout.emit 相同
  - ack: bolt 完成 execute 主动发起的 ack 操作，对应的数据为 ack，ack=执行次数

通常数据的统计会对系统的性能造成一定的影响，那么 storm 中为了平衡这种影响，采取了抽样的方式进行上报数据；storm 中把此值默认为 0.05，即每执行 20 次则执行一次数据累加(value = 20)，当然我们也可以修改抽样的配置值为 1，那么就是逐条上报了。

- supervisor**
  - node add: 增加 supervisor 节点，对应的会向 zk 注册一条 supervisorinfo 信息，包含：host all-slots used-slots uptime version resource
  - node del: 删除 supervisor 节点，那么会从 zk 中删除一条 supervisorinfo 信息
  - worker startup: 启动了一个 worker，那么 supervisorinfo 的 used-slots(1) resource 会发生变化
  - worker shutdown: 关闭一个 worker，那么 supervisorinfo 的 used-slots(-1) resource 会发生变化
- nimbus**
  - node add: 增加一个 nimbus 节点，那么会向 zk 注册一条 nimbus-summary 信息，包含：host port isleader uptime version
  - node del: 删除一个 nimbus 节点，那么会从 zk 中删除一条 nimbus-summary 信息
  - leader change: leader 发生变更，那么对应的 nimbus-summary 中的 isleader 发生变化

## 三. 页面展现结构

在上述的第二点中已经说明了基本操作及对应的数据类型，那么在页面中就可以很好的对应了，先从整体视图来看看我们 storm ui 有什么信息，此 storm ui 为我们数平自己开发维护的 jstorm，一方面用 tab 页把 topology summary/nimbus summary/supervisor summary/nimbus configuration 分类展示，另外加了很多快捷链接入口用于查看关注的内容，以及把常用的 kill topology 操作移动到首页

- 首页-初始化

### Storm UI

Cluster Summary

submit topology

Crawler Job Config Crawler Content Query

Version	Nimbus uptime	Launch Time	Supervisors	Used slots	Free slots	Total slots	Executors	Tasks
tdh1u1-SNAPSHOT-d4a78793e6d2aa85576c22950c0841f7819ac29	3d22h29m53s	2016-06-17 17:16:48	9	66	38	104	714	714

Topology Summary Nimbus Summary Supervisor Summary Nimbus Configuration

Show 10 entries

Name	ID	Status	Uptime	SubmitTime	Num workers	Num executors	Num tasks	Kill Topology
crawlerDrpcService	crawlerDrpcService-62-1466155639	ACTIVE	5d0h10m29s	2016-06-17 17:27:25	1	5	5	[kill]
tdspider-job-1	tdspider-job-1-11-1466155294	ACTIVE	5d0h16m16s	2016-06-17 17:21:38	1	12	12	[kill]
tdspider-job-10	tdspider-job-10-1-1466155242	ACTIVE	5d0h17m10s	2016-06-17 17:20:44	1	12	12	[kill]
tdspider-job-11	tdspider-job-11-2-1466155248	ACTIVE	5d0h17m10s	2016-06-17 17:20:49	1	12	12	[kill]
tdspider-job-12	tdspider-job-12-3-1466155253	ACTIVE	5d0h16m59s	2016-06-17 17:20:55	1	12	12	[kill]
tdspider-job-13	tdspider-job-13-4-1466155258	ACTIVE	5d0h16m52s	2016-06-17 17:21:02	1	12	12	[kill]
tdspider-job-14	tdspider-job-14-5-1466155264	ACTIVE	5d0h16m49s	2016-06-17 17:21:05	1	12	12	[kill]
tdspider-job-15	tdspider-job-15-6-1466155268	ACTIVE	5d0h16m44s	2016-06-17 17:21:10	1	12	12	[kill]
tdspider-job-16	tdspider-job-16-7-1466155273	ACTIVE	5d0h16m36s	2016-06-17 17:21:18	1	12	12	[kill]
tdspider-job-17	tdspider-job-17-8-1466155279	ACTIVE	5d0h16m33s	2016-06-17 17:21:21	1	12	12	[kill]

Showing 1 to 10 of 61 entries

Previous 1 2 3 4 5 6 7 Next

Show System Stats

cluster summary: 展示集群的整体信息，包含版本、在线时间、启动时间、管理的 supervisor 个数、总的节点数、已用节点数、空闲节点数、在运行的 executor/task 数

topology summary: 展示所有正在运行的 topology 信息，包含名称、id、状态、在线时间、提交时间、占用的 worker 数、拥有的 executor 数、拥有的 task 数、快捷关闭按钮

- 首页-supervisor summary: 展示所有的 supervisor 节点，包含 id、主机名称、在线时间、启动时间、总的 worker 节点数、已经使用的 worker 节点数

Topology Summary Nimbus Summary Supervisor Summary Nimbus Configuration

Show 10 entries

ID	Host	Uptime	Launch Time	Slots	Used slots
02537491-cdb7-48cd-934f-afeb6f6100b3	10-209-16-217	5d0h38m23s	2016-06-17 17:16:42	24	16
174eb584-633c-442f-3a0f-359b1d0b1484	10-208-154-103	5d0h37m18s	2016-06-17 17:17:47	8	5
16a6067-d7f7-468b-99a8-ae48bea21473	10-208-154-32	5d0h37m12s	2016-06-17 17:17:53	8	5
4bf4f880-1437-4745-b730-519bac7ae895	172-27-28-192	5d0h37m10s	2016-06-17 17:17:55	4	3
5598d313-4ae0-487c-0891-6a13c57400b9	10-219-5-22	5d0h37m30s	2016-06-17 17:17:35	24	15
56875588-a5d0-4ab4-b42a-c34f3d402f6	10-224-151-207	5d0h37m28s	2016-06-17 17:17:37	16	11
9a6c8dcb-0fa1-44d5-b653-bc0da33bb1ad	10-208-154-225	5d0h37m17s	2016-06-17 17:17:48	8	5
9a9ebc49-4412-4aa2-888e-b4383c35afe1	10-149-28-18	5d0h37m20s	2016-06-17 17:17:45	4	3
ba189889-1493-4020-b0fc-2c3a0a6700c1	10-208-154-208	5d0h37m16s	2016-06-17 17:17:49	8	5

Showing 1 to 9 of 9 entries

Show System Stats

supervisor summary: 展示所有的 supervisor 节点，包含 id、主机名称、在线时间、启动时间、总的 worker 节点数、已经使用的 worker 节点数

- 首页-nimbus summary: 展示所有的 nimbus 节点，包含主机名称、端口号、是否 leader、版本号、在线时间、启动时间

Topology Summary Nimbus Summary Supervisor Summary Nimbus Configuration

Show 10 entries

ID	Port	Status	Version	Uptime	Launch Time
10-209-16-217	6027	Leader	tdh1u1-SNAPSHOT	5d0h37m59s	2016-06-17 17:16:38
10-219-5-22	6027	Not a Leader	tdh1u1-SNAPSHOT	5d0h37m6s	2016-06-17 17:17:31
10-224-151-207	6027	Not a Leader	tdh1u1-SNAPSHOT	5d0h37m3s	2016-06-17 17:17:34

Showing 1 to 3 of 3 entries

Show System Stats

nimbus summary: 展示所有的 nimbus 节点，包含主机名称、端口号、是否 leader、版本号、在线时间、启动时间

- 首页-nimbus configuration: 把配置信息以列表展现出来

Topology Summary Nimbus Summary Supervisor Summary Nimbus Configuration

Show 10 entries

Key	Value
backpressure-disruptor-high-watermark	0.9
backpressure-disruptor-low-watermark	0.4
client-blobstore-class	org.apache.storm.blobstore.NimbusBlobStore
crawler-jobs-dir	/data/home/crawler/deploy/tdspider-current/jobs/
dev-zookeeper-path	/tmp/dev-storm-zookeeper
drpc-authorizer-acl-filename	drpc-auth-acl.yaml
drpc-authorizer-acl-strict	false
drpc-childopts	-Xmx768m
drpc-http-creds-plugin	org.apache.storm.security.auth.DefaultHttpCredentialsPlugin
drpc-http-port	3774

Showing 1 to 10 of 213 entries

Previous 1 2 3 4 5 ... 22 Next

Show System Stats

nimbus configuration: 把配置信息以列表展现出来

### 5. topology 详情页

Storm UI

Topology Summary

Show 1 to 1 of 1 entries

Name	ID	Status	Uptime	SubmitTime	Num workers	Num executors	Num tasks
tdspider-job-1	tdspider-job-1-11-1466155294	ACTIVE	4d23h16m34s	2016-06-17 17:21:38	1	12	12

Topology actions

Actions: Overview (Balance) Debug (Stop Debug) Change Log Level

Topology stats

Window	Emitted	Transferred	Execute latency (ms)	Acked	Failed
5d0h10m	60	60	0.000	0	0
5d0h5m	1020	1020	319.349	2164	5362.486
5d0h0m	41015	41015	319.334	116817	3200.562
1d0h0m	167965	167965	30918.770	220	126720
All time	1101340	1101340	24420.764	3700	785940

Spouts (All time)

ID	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error
formerSpout	1	1	789660	789660	24420.764	3700	789940			

Showing 1 to 1 of 1 entries

Bolts (All time)

ID	Executors	Tasks	Emitted	Transferred	Capacity (last 10s)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error
fetcherParseBolt	8	8	311880	311880	1.082	2560.612	767720	2560.743	696980	58180			
persistBolt	1	1	0	0	0.000	0.589	311640	0.583	311640	0			

Showing 1 to 2 of 2 entries

Topology Visualization

Topology Visualization

Streams

Topology Configuration

Key	Value
auth-info	nu11
backpressure-disruptor-high-watermark	0.9

Topology summary: 与首页对应的 topology summary 对应。

topology actions: 用于 topology 的一些操作：激活、注销、重分配、debug/stop debug、修改日志级别。

topology stats: 按时间窗口展现统计数据：十分钟表示近十分钟的统计，3 小时表示近 3 小时的统计等。

spouts: spout 组件的统计数据，数据含义在第二点的数据类型已有说明。

bolts: bolt 组件的统计数据，数据含义在第二点的数据类型已有说明。

topology visualization: 展示 topology 的拓扑结构图及数据传输情况。

topology configuration: 展示 topology 的配置信息。

### 6. component 详情页

Storm UI

Component summary

Show 1 to 1 of 1 entries

ID	Topology	Executors	Tasks	Debug
fetcherParseBolt	tdspider-job-1	8	8	events

Component actions

Debug (Stop Debug)

Bolt stats

Window	Emitted	Transferred	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
5d0h10m	40	40	1977.600	100	7713.000	120	0
5d0h5m	1020	1020	30840	2164	5362.486	2169	40
5d0h0m	41015	41015	319.334	116817	3200.562	104021	8840
All time	311880	311880	2562.191	768020	2562.795	697280	58180

Showing 1 to 4 of 4 entries

Input stats (All time)

Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
formerSpout	former_border	2562.259	767700	2562.795	697280	58180

Showing 1 to 1 of 1 entries

Output stats (All time)

Stream	Emitted	Transferred
fetcher_parse_result	311880	311880

Showing 1 to 1 of 1 entries

Executors

ID	Uptime	Host	Port	Emitted	Transferred	Capacity (last 10s)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
[10-10]	4d23h16m12s	10-209-16-217	6701	flex	33620	0.000	2500.971	82360	2568.503	74440	6380
[5-5]	4d23h16m12s	10-209-16-217	6701	flex	33620	0.000	2500.971	82360	2568.503	74440	6380
[5-5]	4d23h16m12s	10-209-16-217	6701	flex	34800	0.000	2525.293	85080	2467.425	75660	7140
[5-5]	4d23h16m12s	10-209-16-217	6701	flex	40660	0.000	2684.891	103260	2685.749	93700	7720
[5-6]	4d23h16m12s	10-209-16-217	6701	flex	30840	0.000	2242.881	86020	2411.414	79680	6900
[5-7]	4d23h16m12s	10-209-16-217	6701	flex	38840	0.330	2568.460	107080	2793.065	96420	19120
[5-8]	4d23h16m12s	10-209-16-217	6701	flex	37160	0.000	2468.804	90280	2535.196	81320	7560
[5-9]	4d23h16m12s	10-209-16-217	6701	flex	43000	0.000	2615.599	95500	2602.002	88540	5720

Showing 1 to 8 of 8 entries

Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)

Actions

Stack Resident Worker Jmap

Errors

component summary: 展示 component 的 id、executor 个数、task 个数、debug 日志。

component actions: 提供 component 的 debug/stop debug 操作。

spout/bolt stats: 按时间窗口展示数据，数据含义在第二点的数据类型已有说明。

input stats: 输入流数据，包含上游 component、stream、执行时间、执行次数、bolt 正常流转的处理时间、ack 数、失败数。

output stats: 输出流数据，包含 stream、emit 个数、transfer 个数。

executors: executor 数据，包含 id、在线时间、所在主机及端口、emit 个数、transfer 个数、执行时间、执行个数、成功流转直接、ack 个数、失败个数。

profile and debugging: 用于诊断 worker 的运行情况，执行 jstack/jmap 等操作对 jvm 进行检测输出到文件，供用户下载打开查看 jvm 的运行情况。