一、Nifi 配置

1) nifi 基本配置

- 1. jdk 配置: nifi 新版本要求 jdk 版本必须 1.8 以上 目录: bin/nifi.env.sh export JAVA_HOME=/home/user/opt/jdk1.8.0 144 忽略此项配置,nifi将使用系统默认Java环境启动。 2. nifi运行内存配置 目录: conf/bootstarp.conf java.arg.2=-Xms512m java.arg.3=-Xmx20480m 根据服务器性能合理设置nifi内存 3. nifi日志配置 目录: conf/logback.xml 4. nifi端口配置 目录: conf/nifi.properties nifi.web.http.port=18083 默认为8080端口。 **5. nifi**基本命令 nifi.sh start 启动 nifi.sh stop 关闭 nifi.sh restart 重启
 - nifi.sh status 状态
- 以上基本配置完成,可单节点模式启动 nifi。

2) nifi 集群配置

配置 zookeeper
 目录: conf/zookeeper.properties
 clientPort=2181 zookeeper 端口

server.1=kube-master:2888:3888 server.2=kube-slave1:2888:3888 server.3=kube-slave2:2888:3888 zookeeper 服务器设置, nifi 集群有多少节点需要配置 server server.1 中的数字 1 为服务器的 ID,需要与 myid 文件中的 id 一致,下一步将配置 myid。

2. 配置 myid

在 nifi 目录下创建文件夹/state/zookeeper/并创建文件 myid



3. 配置 state-management.xml

目录: conf/state-management.xml

<cluster-provider>

<id>zk-provider</id>

<class>org.apache.nifi.controller.state.providers.zookeeper.ZooKeeperStateProvider</class> <property name="Connect String">kube-master:2182,kube-slave1:2182,kube-

slave2:2182,kube-slave3:2182</property>

<property name="Root Node">/nifi</property>

<property name="Session Timeout">10 seconds</property>

<property name="Access Control">Open</property></property>

</cluster-provider>

Connect String 配置为 nifi 节点的 ip 加 zookeeper 端口号,以逗号分隔

- 4. 配置 nifi 节点属性
 - 目录: conf/ nifi.properties
 - nifi.state.management.embedded.zookeeper.start=true
 - 使用内置 zookeeper

nifi.zookeeper.connect.string=kube-master:2182,kube-slave1:2182,kube-

slave2:2182, kube-slave3:2182

与上一步中的 connect string 一致。

下面配置需每个节点单独配,根据节点的 IP 相应配置,保持集群中节点使用的端口一致 nifi.cluster.is.node=true nifi.cluster.node.address=kube-slave1 nifi.cluster.node.protocol.port=9999 nifi.remote.input.host=kube-slave1 nifi.remote.input.secure=false nifi.remote.input.socket.port=9998 nifi.web.http.host=kube-slave1

配置完成后即可依次启动节点,集群将选举产生主节点。

二、Nifi 界面



1-8 为 nifi 提供的组件可以拖到下方的画布上使用。

- 1: processor : nifi 提供的多种处理器。
- 2: input port: flowfile 数据进入节点。
- **3:** output port flowfile 数据输出节点。

4: group 组,可以将多种处理器放入同一个组中,组与组之间可以通过 input port 和 output port 传输 flowfile 数据。

- 5: Remote process group: 远程进程组,可将 flowfile 发送至别的节点。
- 6: Funnel: 汇聚 flowfile 数据。
- 7: Template: 可拖放导入的模板配置。
- 8: Leble: 文字板
- 9-18 为 nifi 的基本状态栏
- 9: nifi 集群节点状态。9/9 表示有集群中有九个节点,并九个节点都在线。
- 10: nifi 当前运行线程数。
- 11: nifi 当前队列中待执行的 flowfile 的 size 及大小
- 12: nifi 启动的远程进程组。
- 13: nifi 未启动的远程进程组。
- 14: nifi 启动的处理器
- 15: nifi 未启动的处理器。
- 16: nifi 配置不可用的处理器
- 17: nifi 废弃不用的处理器。
- 18: nifi 界面最新刷新时间。
- 19: nifi 画布缩略图及操作按钮。
- 20: 当前位置。

1) 总览界面

Summary 摘要包含处理器,GROUP 等一系列整体摘要信息。 Counters 计数器界面,需要配合计数器处理器(UpdateCounter)使用。 Bulletin Board nifi 通讯信息页面。 Date Provenance flowfile 数据总览。 Controller Settings 设置界面,可以配置线程数等一些设置。 Flow conf History nifi 配置历史。 Cluster nifi 集群模式时,可以查看集群状态,管理节点。

三、Nifi 使用

1) 模板

创建模板:在要创建模板的 group 中点击面板左侧的 create template 或者鼠标右键画布空白处。

🛦 NiFi	X 💩 N	iFi		× +											_	n	×
								240		0		1	141 O		_		
(←) → G @	0 🖌	10.254.103	171:1808	33/nifi/				88	··· ជ	9、捜索			III\ 🛩	ц	9	າ 🙂	=
火狐官方站点 登 最常访问	一 火狐官方	站点 📵 新手	上路 🎄 🛛	NiFi 🛅 常用	网址 🔟 京东商	城 📄 常用网址	JD 京东	商城 🕜		有道						[] 移动	版书签
nifi 🖸		ej 🔄	j jo		2	U											=[
🔛 0 🔳 20,530	6 / 10.06 ME	з (0	ا	▶ 32	63		<u>A</u> 9	🏌 5	C 1	5:54:35 CST					C	٦ 🗅
Ø Navigate	(Э — —															
0.0.111																	
Q Q []]																	
											PostH	TTP					
				C C Ge	etKafka etKafka 1.4.0						PostHT org.apac	TP 1.4.0 he.nifi - nifi-standard-n	ar				_
				In f	g.apache.nifi - nifi-kafka 0 (0 bytes)	-0-8-nar	5 min	Name su	iccess		In 60 (3	89.53 KB)		5 п	nin		
		_		Read/Write	0 bytes / 389.53 KE	1	5 min	Queued	0 (0 bytes)		Read/Write 389.5	3 KB / 0 bytes		5 m	nin 🖣	Name failu	are (
Operate	l	3		Out (60 (389.53 KB)		5 min				Tasks/Time 60/0	lytes) 10:00:05.114		5 m	nin	Queued 0	(U bytes)
NiFi Flow				Tasks/Time	60 / 00:00:00.037		5 min				Tusks, Thic boy c	0.00.03.114		511			
Process Group			C Refre	sh													
dd3a9/16-015t-1000-c8ed-e3862	9dc2605																
🍄 🤺 🕺 🕨 🔳	s 🖈		Contig	gure													
463/1	DELETE		variat	ules													
		_	 Start 														
			Stop														
			🚽 Uploa	d template													
			B Creat	e template													
NiFi Flow																	

下载模板:

															٥	×
(←) → 健 @ @	10.254.	.103.171:18083	3/nifi/				80	☆	Q. 搜索			111\	9 t	•	•	@ Ξ
▶ 火狐官方站点 🗘 最常访问 🗎 火狐	官方站点 📵) 新手上路 💧 N	liFi 🛅 常用网址	D 京东商城	🚞 常用网址	D 京东南城	ŧ 🛞		有道						日和	动版书签
nifil 🖸 🕹	⋳⋛	أما أما		}	:											Ξ
0 20,536 / 10.0	5 MB	. 0		▶ 32	63	A	9	× 5	C 15	:58:07 CST		23	Summ	ary		0
													Bulletin	n Board		
Ø Navigate												Ľ	Data P	rovena	nce	
												۶ 2	Contro Flow C	ller Set onfigur	tings ation Hi	story
												${\bf r}_{\rm s}$	Templa	ates		
			D GetKafk GetKafk org.apaci	fka ka 1.4.0 he.nifi-nifi-kafka-0-8-1 ytes)	har	5 min	Name su Queued	ccess 0 (0 bytes)		PostHTT PostHTTP org.apache. In 60 (383) Read/Write 383.68	P 1.4.0 nifi - nifi-standard-nar .68 KB) KB / 0 bytes	0	Help About	5 min 5 min -	Name	failure
🖕 Operate	8		Read/Write 0 byte Out 60 (3)	es / 383.68 KB 83.68 KB)		5 min 5 min				Out 0 (0 byte	es)			5 min 5 min	Queued	0 (0 bytes)
NIFLFIOW dd3a9716015F1000-68de e38629dc2605 P P			Tasks/Time 60 / 0	0:00:00.166		5 min										

使用模板:选择界面上分的 template 拖放至画布,选择要使用的模板。



Nifi 的模板会保存组中的处理器配置及 CONTROLLER SERVERS。例如数据库连接, 但是不会保存密码。

2) CONTROLLER SERVERS

Nifi 中的控制器,例如数据库连接,XML 读取,JSON 读取器。

Nifi 中子 group 可以继承父 group 中的控制器。建议在根画布及 new flow 中配置控制器。 每个控制器都是一个实例,如数据库连接,交给同一个控制器管理更为节省资源及方便。 在空白处右键点击 CONFIGURE:

下图为一个 DBCPConnectionPool 数据库连接控制器的配置说明:

¢	NiFi	🗙 🎄 Nifi	× +									-	٥	×
(-) → C' @	0 🔏 10.25	4.103.171:18083/nifi/		题 … 合	Q. 搜索			lii\	P 1	4 1	•	5 🙆	≡
6	」火狐官方站点 ♀	最常访问 🛅 火狐官方站点 🌘	🖢 新手上路 🎄 NiFi 🛅 常用网址 🔟 京东	商城	🖿 常用网址 🔟 京东商城 🕣	有道] 移动	版书签 🛾
	NiFi Flow C	configuration	Configure Controller Servic	е									×	
0	GENERAL	CONTROLLER SERVICES	SETTINGS PROPERTIES Required field	CON	IMENTS		+						4	
	_		Property		Value									
		Name 🔺	Database Connection URL	0	jdbc:oracle:thin:@//10.250.90.200:1521/orcl									
	2	JTWFCL	Database Driver Class Name	0	oracle.jdbc.driver.OracleDriver			w				\$ 1	0	
	2	MYNJ	Database Driver Location(s)	0	/home/user/opt/jdbc/ojdbc7.jar			ow.				• *		
	2	MYNJ-ORACLE	Database User	0	JTWFCL			w				• 1	0	
		ZNMH	Password	0	Sensitive value set			w				\$ 1	Ċ.	
0	2	data_asset_manager	Max Wait Time	0	500 millis			w				• *	0	
5		mysql	Max Total Connections	0	8			w				\$ 1	Ċ.	es)
			Validation query	0	No value set									
d														
	C Last upda	ted: 16:09:41 CST		_		CANCEL		rocessors and	l servi	ces of	this	Proce	as Group	

依次为:

- 1. 数据库连接的 URL,类似于 JDBC 中的 URL。
- 2. 数据库驱动类名
- 3. 数据库驱动类目录 (jar 包需要放置在 nifi 可访问到的地址,集群需要每个节点都上传 jar 包)。
- 4. 数据库用户名
- 5. 密码
- 6. 最大等待时间
- 7. 最大连接数
- 8. 测试 SQL。

3) PROCESSOR

Nifi 处理器:



- 1: 处理器名字
- 2: 处理器的 ID
- 3: 处理器类型
- 4: 处理器所属 nar 包
- 5: 处理器输出日志级别

6: 处理器输出的关系,配置处理器时所有关系需要处理:拖放给下一处理器,或者选择终止关系(配置处理器时选中要自动终止的关系)。

- 7: 处理器基础设置
- 8: 处理器定时设置
- 9: 处理器属性
- 10: 备注

(←) → C 企 ③ Ø 10.254.103.171:18083/nfif/?processGroupId=f7c2c139-c520-3bcc-71 ○ Ø 總定 ○ Ø 總定	
🖹 火狐官方站点 💩 最常访问 🖹 火狐官方站点 🧶 新手上路 🐁 Nifi 📄 常用网址 🔟 东东南城 📄 常用网址 🚺 东东南城 🛞 有道] 移动版书签
nifi 🖉 🖓 🖬 📴 🎬 🏲 形	
🗄 0 🗏 20,536 / 10.06 MB Processor Details	Q 🗆
Navigate SETTINGS SCHEDULING PROPERTIES COMMENTS	
Q Q [] : Scheduling Strategy • Timer driven Run Duration • 00:00:00.000	
Concurrent Tasks Concurrent Tasks Concu	
Execution •	
Primary node only 4	
🖏 operate 😑	
GenerateTableFetch	
D305724-1042-306-26601e72607f	
	nar 5 mir
Read/Write 1.58 KB / 0 bytes	5 min
Out 0 (0 bytes)	5 min
Tasks/Time 1/00:00:00.101	5 min
ок	

- 1: 处理器定时任务类型 1.timer 2.cron
- 2: 单个节点处理器并发数,如图单节点并发数为1,集群为9个节点时并发数为9
- 3: 根据定时任务类型不同配置不同,图中为600秒执行一次

4: 当 nifi 为集群模式时:可以配置处理器为主节点运行或所有节点运行。一般获取数据的 处理器使用主节点运行,防止重复消费。

🎄 NiFi 🛛 🗙 🎄 NiFi	× +		- o ×
(←) → C' ŵ () <u>∅</u> 10.25	4.103.171:18083/nifi/?processGroupId=f7c2c139-c	520-3bcc-71 器 ···· 合 🔍 搜索	III\ 🗭 🖪 🕤 🤝 🚍
▶ 火狐官方站点 🗘 最常访问 📄 火狐官方站点 🌘	😼 新手上路 💧 NiFi 🛅 常用网址 🔟 京东商城 🛅 常用	网址 🔟 京东商城 🕣 有道	□移动版书签
nifil 🖸 🕹 🖙			
🏦 0 🗮 20,536 / 10.06 MB	Processor Details		۹ 🗖
🙆 Navigate 🖂	SETTINGS SCHEDULING PROPERTIES	COMMENTS	
Q Q []!	Paguinad Gald		
	Requirea fiela		
	Property	Value	
	Database Connection Pooling Service 0	→ Lux	
	Database Type 🛛 🔞	Oracle	
	Table Name 0	USER_ZNMH.T_ZNMH_ALIPAYZMXY	
	Columns to Return 0	IDNUMBER,CUSTOMERID,CREATETIME,UPDATETIME,SY	
	Maximum-value Columns 0	CREATETIME	
ර Operate	Max Wait Time 🛛 🔞	0 seconds	
	Partition Size 0	1000	
GenerateTableFetch	Additional WHERE clause 0	No value set	
35/05174-1a04-39-05-26071			PostHTTP 0 orgaschaft (TP 1.0. orgaschaft (TP 1.0.) Strikt In 0.0 tytes) 5 min Read/Write 0.9 tytes) 5 min Out 0.0 (0.9 tytes) 5 min Tasks/Time 0./ 00.00.000 5 min
	****	ок	
NiFi Flow » 芝麻信用存放customerId表同步			

处理器属性:某些属性需配合 CONTROLLER SERVERS 使用,如图中的数据库连接。

依次配置项为:

- 1: 数据库连接
- 2: 数据库类型
- 3: 表名
- 4: 需要返回的字段
- 5: 最大值列,可依据此列完成增量同步
- 6: 最大等待时间 0S 表示一直等待。
- 7: 分页条数, 1000 表示每条 SQL 查询出 1000 条记录。
- **8:** 查询时的 where 条件。

GenerateTableFetch 拥有 state 属性:记录 Maximum-value Columns 的值。



- 1: 处理器五分钟内接受到的 flowfile 数据。
- 2: 处理器五分钟读取和写的数据大小
- 3: 处理器五分钟输出的数据
- 4: 处理器五分钟内的执行次数和时间。
- 5: 处理器当前使用的线程数。如图为一个线程在运行。
- 6:处理器的 SUCCESS 关系输出值 ExecuteSQL 处理器。其中 failure 关系为自动终止。

点击处理器左键可以查看 Data Provenance

Data Provenance 为一天内通过处理器的数据记录。

\rightarrow	C' 🏠 🛛 🖉	10.254.103.171:18083/ni	ifi/?processGroupId=f7c2c139-c520-3	3bcc-71 🗱 ••• 🕻	Q. 搜索	lii\ 🗭 🛱	•	ð (
瓜官方	站点 🗘 最常访问 📄 火狐官方	游点 🥹 新手上路 💧 NiFi	🛅 常用网址 🔟 京东商城 🛅 常用网址	JD 京东商城 🕣	有道			日移
£1		a—a a—a	n90n 📮 💻 ,					
liFi	Data Provenance							
isplay Idest	ring 130 of 130 event available: 12/20/2017 16:	:44:08 CST			Showing the ev	ents that match the specified qu	Jery. Cle	ar sea
Filter	by cor	mponent name 🛛 🗸					_	
	Date/Time 🗸	Туре	FlowFile Uuid	Size	Component Name	Component Type		
0	12/21/2017 16:39:56.443 CST	CREATE	01ec2f91-b82b-46f1-a24d-3a5c	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 16:29:56.271 CST	CREATE	b3b048f2-30ab-48b2-b1e5-0d23	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 16:22:31.706 CST	CREATE	11f3d469-d1ae-44b7-91c5-ae87	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 16:11:57.365 CST	CREATE	95886cc5-3d95-4555-b8ef-9b65	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 16:01:56.975 CST	CREATE	fa8e1d3d-7398-49d1-874f-52b9	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,‰ →
0	12/21/2017 15:51:56.437 CST	CREATE	3ff4fcc5-0fd7-47fb-8586-da78d	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 15:41:55.909 CST	CREATE	fd20a595-bb9c-4bc4-a639-9624	365 bytes	GenerateTableFetch	GenerateTableFetch	i	"հ→
0	12/21/2017 15:31:55.227 CST	CREATE	99f60701-97b0-4820-88bd-38f4	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,թ.→
0	12/21/2017 15:21:54.859 CST	CREATE	41c067e8-bff0-4064-8aae-3928	365 bytes	GenerateTableFetch	GenerateTableFetch	i	"հ→
0	12/21/2017 15:11:53.927 CST	CREATE	bcc47780-b057-40c4-acca-aa7c	365 bytes	GenerateTableFetch	GenerateTableFetch	1	,‰→
0	12/21/2017 15:01:53.466 CST	CREATE	0f112f47-002c-425c-a521-b548	365 bytes	GenerateTableFetch	GenerateTableFetch	i	"հ→
0	12/21/2017 14:51:53.272 CST	CREATE	2e284959-254b-4ff9-9ec7-42b3	365 bytes	GenerateTableFetch	GenerateTableFetch	1	‰→
0	12/21/2017 14:41:52.867 CST	CREATE	03eade73-39e4-436e-8dd4-ac57	365 bytes	GenerateTableFetch	GenerateTableFetch		"հ→
A	10/01/0017 11/01 51 000 007	005475	5 004557 7000 4 50 0 10 040	0001	0 ITUEII	0		. .
a .	ast updated: 16:44:16 CST							