



Using the FLaNK Stack for Edge AI (Apache Flink, Apache NiFi, Apache Kafka, Apache Kudu)

APACHECON @HOME
Sept. 29th – Oct. 2nd

2020

Join Me Virtually @ Future of Data - Princeton on Youtube & Zoom



<https://www.meetup.com/futureofdata-princeton/>
<https://github.com/tspannhw>

From Big Data to AI to Streaming to Containers to
Cloud to Analytics to Cloud Storage to Fast Data to
Machine Learning to Microservices to ...



@PaasDev



FLaNK Stack

Tim Spann
Cloudera
@PaasDev

 APACHECON
@Home
Sept. 29-Oct. 1, 2020

SPEAKER



My Talk List

29/9/2020, 1:35 PM EDT - Incrementally Streaming RDBMS Data to Your Data Lake
Automagically with **John Kuchmek**

[29/9/2020, 2:15 PM EDT - Using the Mm FLaNK Stack for Edge AI](#)

29/9/2020, 3:35 PM EDT - Real-Time Stock Processing With Apache NiFi, Apache Flink and
Apache Kafka with **Pierre Villard**

30/9/2020, 1:35 PM EDT - Apache Deep Learning 301 with **Ian Brooks, PhD.**

30/9/2020, 2:15 PM EDT - Edge to AI: Analytics from Edge to Cloud with Efficient Movement of
Machine Data with **Paul Vidal**

30/9/2020, 2:55 PM EDT - Utilizing Apache NiFi and MiNiFi for EdgeAI IoT at Scale with
Sunile Manjee

Edge AI with FLaNK

- FLaNK Overview
- Edge AI Agent
- Ingesting Edge Data Events
- Stream Processing
- Event Storage
- Event Queries, Reporting, Analytics and Applications

FLaNK Stack Overview

The FLaNK stack is a team of Apache Open Source software that can be used to rapidly build streaming event pipelines that span use cases from Cloud Ingest, I(I)oT, Log Processing, Cybersecurity Analytics and more.



The FLaNK stack consists of Apache Flink, Apache NiFi and Apache Kafka at it's core. With Apache Kudu as the second option for the K and a preferred data store.



APACHECON NA

2020

FLaNK Stack Diagram

<https://github.com/tspannhw/MmFLaNK>



APACHECON NA

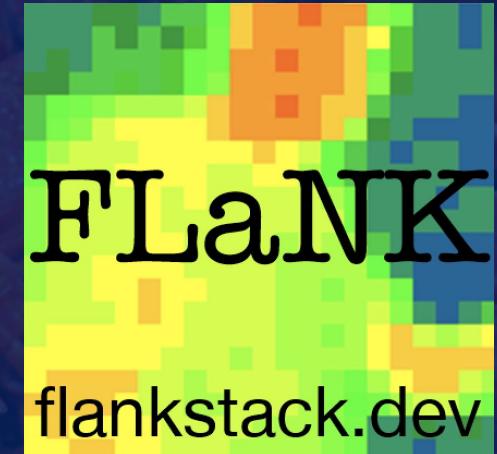
2020

<https://www.datainmotion.dev/2019/11/introducing-mm-flank-apache-flink-stack.html>

FLaNK Stack Data Science

Apache MXNet and DJL.AI as a preferred option for running Deep Learning as part of data flows.

We can run Deep Learning on edge devices, in NiFi flows, in Flink apps and in NiFi connected cloud machine learning services like Cloudera Machine Learning.



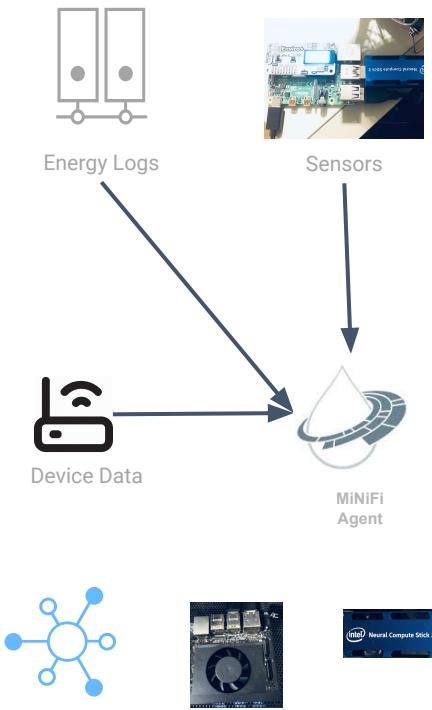
APACHECON NA

2020

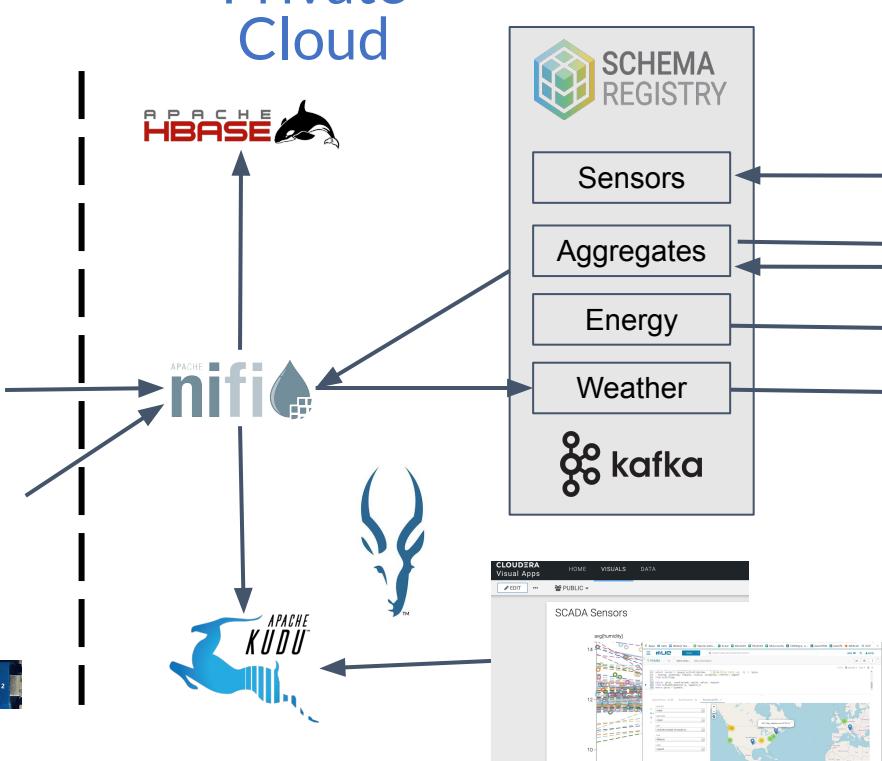
Edge AI to Cloud Streaming Pipeline



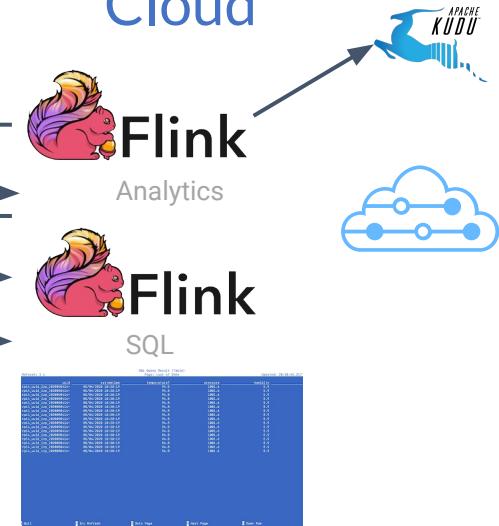
Edge



Private
Cloud



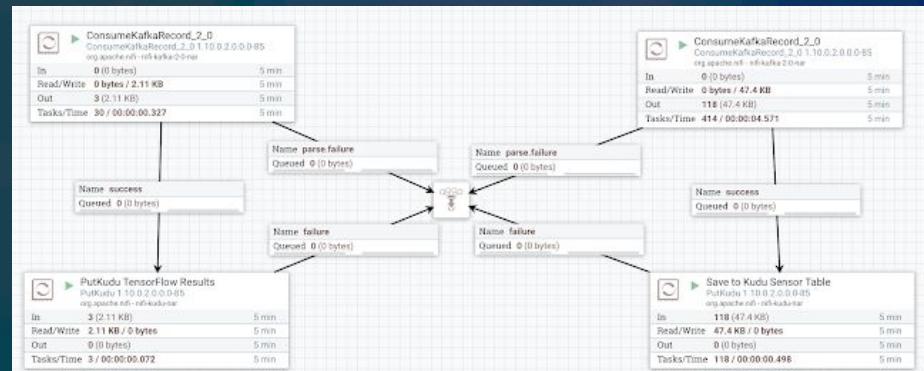
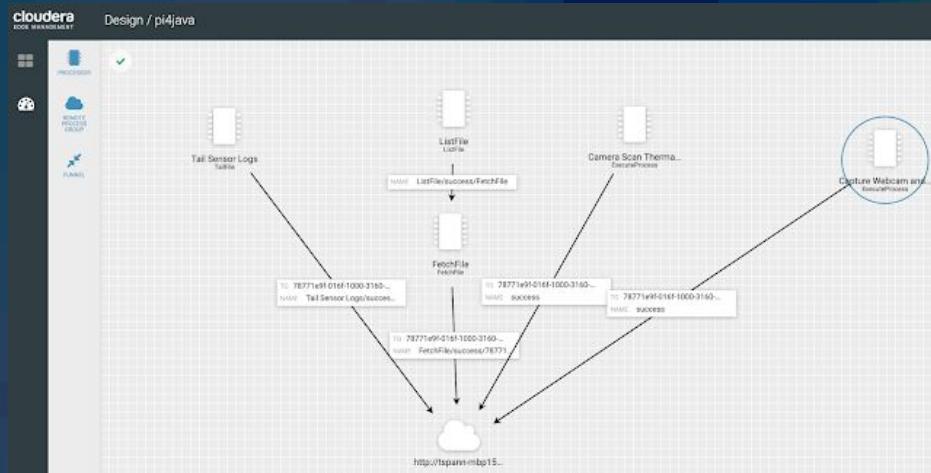
Multi-Public
Cloud



Edge AI Connected Applications

<https://www.datainmotion.dev/2020/06/the-rise-of-mega-edge-flink.html>

<https://www.datainmotion.dev/2020/01/cloudera-edge2ai-minifi-java-agent-with.html>



APACHECON NA

2020

Edge

Edge Devices

NVIDIA Jetson Xavier NX

NVIDIA Jetson Nano
RPI 4 NCC2 Enviro+
USB Web Cameras
Enviro+ Sensors



Edge Software

Apache NiFi / MiNiFi
Java and C++ Agents
Python 3
DetectNet
GoogleNet
MQTT
HTTP
JDK 8



APACHECON NA

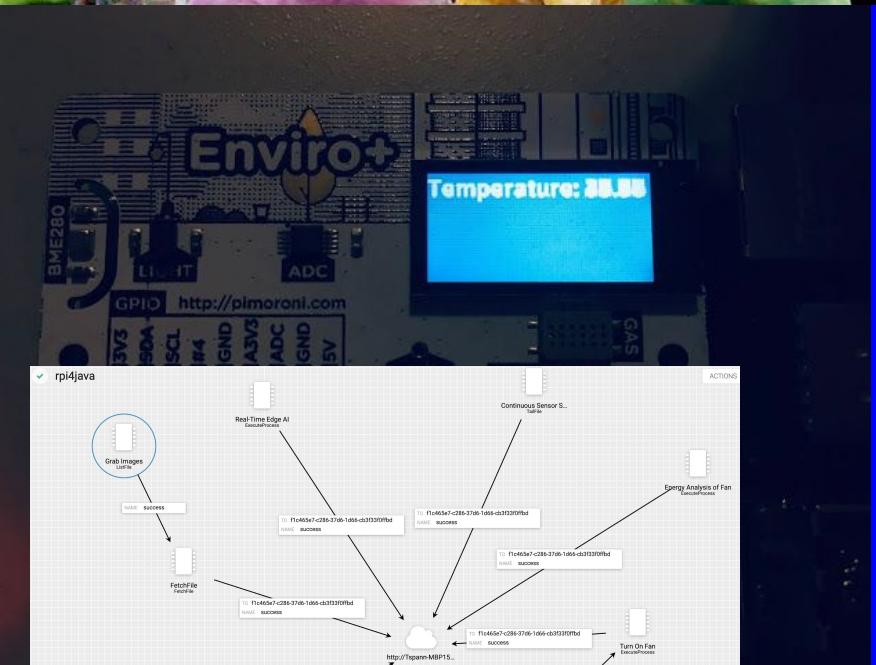
2020

<https://www.datainmotion.dev/2020/06/unboxing-most-amazing-edge-ai-device.html>

Raspberry Pi with Enviro+ Sensors

MiNiFi Java Agent

- Reads Sensor Logs
- OpenVino NCC2 AI
- Reads Images
- Sends to NiFi Gateway



Apache NiFi Gateway processors, validates, transforms, cleans, routes and streams events for additional processing through Apache Kafka topics.

APACHECON NA

2020

Sensor Output

```
{"uuid": "rpi4_uuid_jfx_20200826203733", "amplitude100": 1.2, "amplitude500": 0.6, "amplitude1000": 0.3, "lownoise": 0.6, "midnoise": 0.2, "highnoise": 0.2, "amps": 0.3, "ipaddress": "192.168.1.76", "host": "rp4", "host_name": "rp4", "macaddress": "6e:37:12:08:63:e1", "systemtime": "08/26/2020 16:37:34", "endtime": "1598474254.75", "runtime": "28179.03", "starttime": "08/26/2020 08:47:54", "cpu": 48.3, "cpu_temp": "72.0", "diskusage": "40219.3 MB", "memory": 24.3, "id": "20200826203733_28ce9520-6832-4f80-b17d-f36c21fd8fc9", "temperature": "47.2", "adjtemp": "35.8", "adjtempf": "76.4", "temperaturef": "97.0", "pressure": 1010.0, "humidity": 8.3, "lux": 67.4, "proximity": 0, "oxidising": 77.9, "reducing": 184.6, "nh3": 144.7, "gasKO": "Oxidising: 77913.04 Ohms\nReducing: 184625.00 Ohms\nNH3: 144651.47 Ohms"}  
APACHECON NA  
2020
```



Sensor List

BME280 - temperature, pressure, humidity sensor
LTR-559 - light and proximity sensor
MICS6814 - analog gas sensor
ADS1015 ADC
MEMS - microphone
0.96-inch, 160 x 80 color LCD

APACHECON NA

2020

NVIDIA Xavier NX with MiNiFi + Python 3

MiNiFi Java Agent

- Runs Google Net on Three Cameras
- Sends Images
- Sends classification and device data to NiFi Gateway

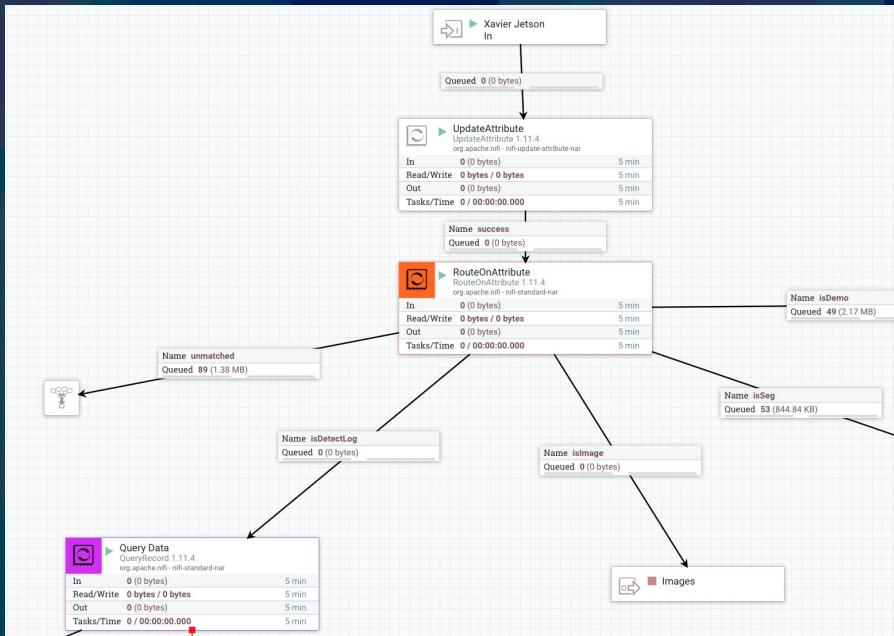
Apache NiFi Gateway processors, validates, transforms, cleans, routes and streams events for additional processing through Apache Kafka topics.



APACHECON NA

2020

NVIDIA Xavier NX Processing



APACHECON NA
2020

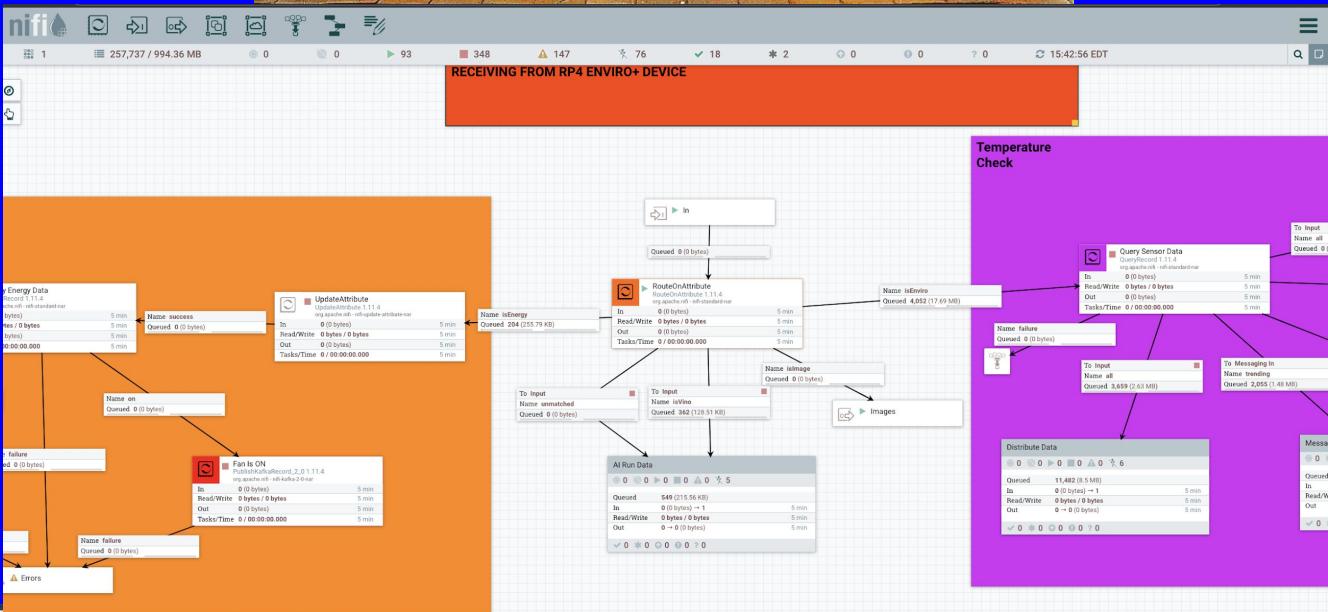
Flink SQL Insert into Kafka

```
INSERT INTO global_sensor_events  
SELECT scada.uuid, scada.systemtime, scada.temperatureref,  
scada.pressure, scada.humidity, scada.lux, scada.proximity,  
scada.oxidising, scada.reducing, scada.nh3, scada.gasko,  
energy.`current`, energy.voltage, energy.`power`,  
energy.`total`, energy.fanstatus  
FROM energy, scada  
WHERE scada.systemtime = energy.systemtime;
```

APACHECON NA

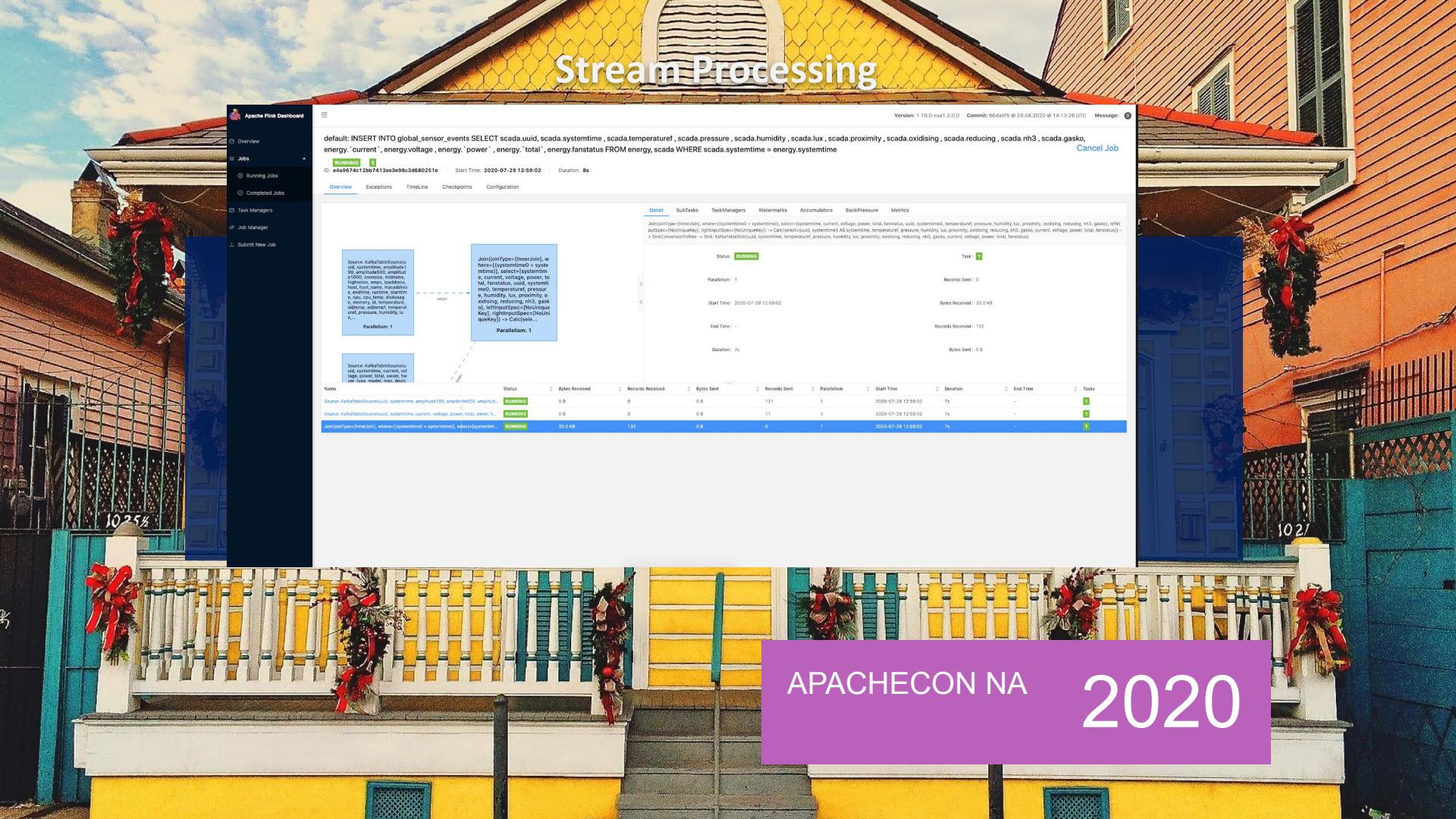
2020

Ingesting Edge Data Events



APACHECON NA
2020

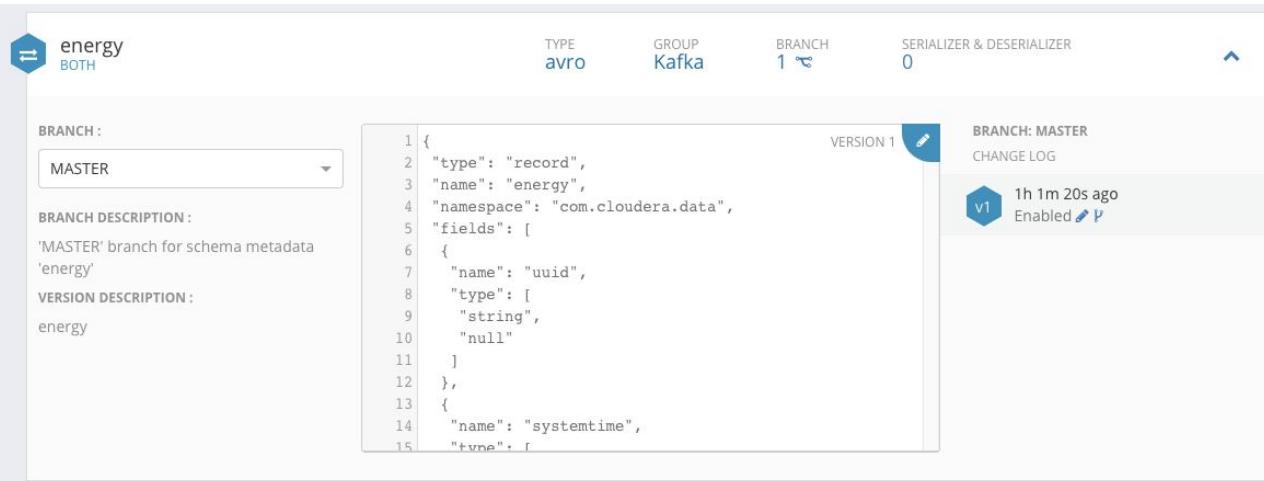
Stream Processing



APACHECON NA
2020

One Schema To Rule Them All

```
oppable".false,"state":"CANCELED","start-time":1599576455857,"end-time":1599576486876,"duration":31019,"now":1599576486888,"timestamps":  
57,"CANCELLING":1599576486855,"FAILED":0,"RESTARTING":0,"FINISHED":0,"RUNNING":1599576455905),"vertices":[{"id":"cbc357ccb763df2852fee8c4fc7d55f2","name":"Source: KafkaTableSource(upc, originstore, updatedate, longdescription,  
gistry.default_database.itemprice, source: [KafkaTableSource(upc, originstore, updatedate, longdescription, itemdescription, brandname, displayimage, price, msrp, tpr)]}, fields=[upc, originstore, updatedate, longdescription,  
Sink: SQL Client Stream Collect Sink,"parallelism":1,"status":"CANCELED","start-time":1599576455941,"end-time":1599576486876,"duration":30935,"tasks":  
SCHEDULED":0,"CREATED":0,"DEPLOYING":0),"metrics":{"read-bytes":0,"read-bytes-complete":true,"write-bytes":0,"write-bytes-complete":true,"read-records":0,"read-records-complete":true,"write-records":0,"write-records-complete":true}}],"state:  
D":0,"CREATED":0,"DEPLOYING":0),"plan": {"jid": "7c01884b74ff981a896307c4a06f2b15","name": "default: select * from itemprice", "nodes":  
y": "", "description": "Source: KafkaTableSource(upc, originstore, updatedate, longdescription, itemdescription, brandname, displayimage, price, msrp, tpr) -> SourceConversion(table=[registry.default_database.itemprice, source:  
me, displayimage, price, msrp, tpr]), fields=[upc, originstore, updatedate, longdescription, itemdescription, brandname, displayimage, price, msrp, tpr]) -> SinkConversionToTuple2 -> Sink: SQL Client Stream Collect
```

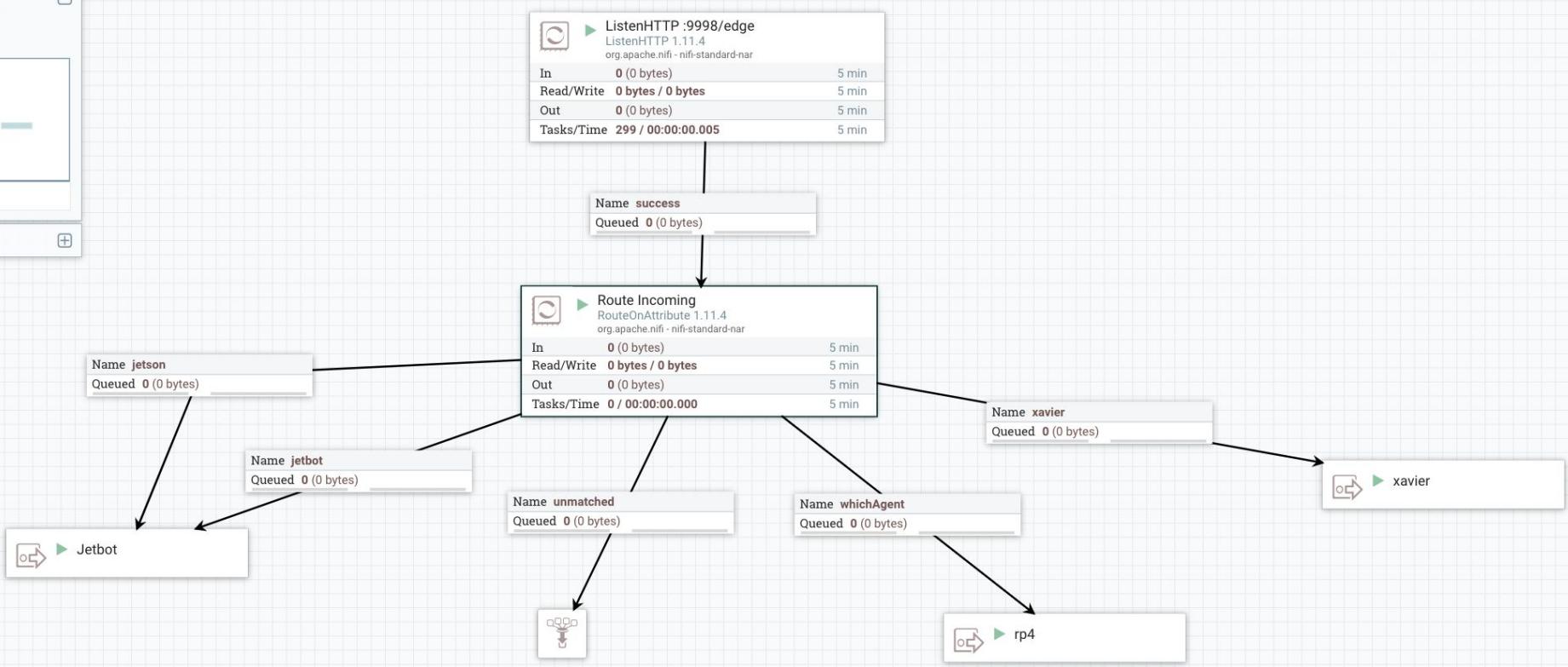


The screenshot shows the Confluent Schema Registry interface. The schema is named 'energy' and is defined as 'BOTH'. It is grouped under 'Kafka' and has a single branch labeled 'MASTER'. The schema version is 1. The code editor contains the following Avro schema:

```
1 {  
2   "type": "record",  
3   "name": "energy",  
4   "namespace": "com.cloudera.data",  
5   "fields": [  
6     {  
7       "name": "uuid",  
8       "type": [  
9         "string",  
10        "null"  
11      ]  
12    },  
13    {  
14      "name": "systemtime",  
15      "type": "long"  
16    }  
17  ]  
18}  
19}
```

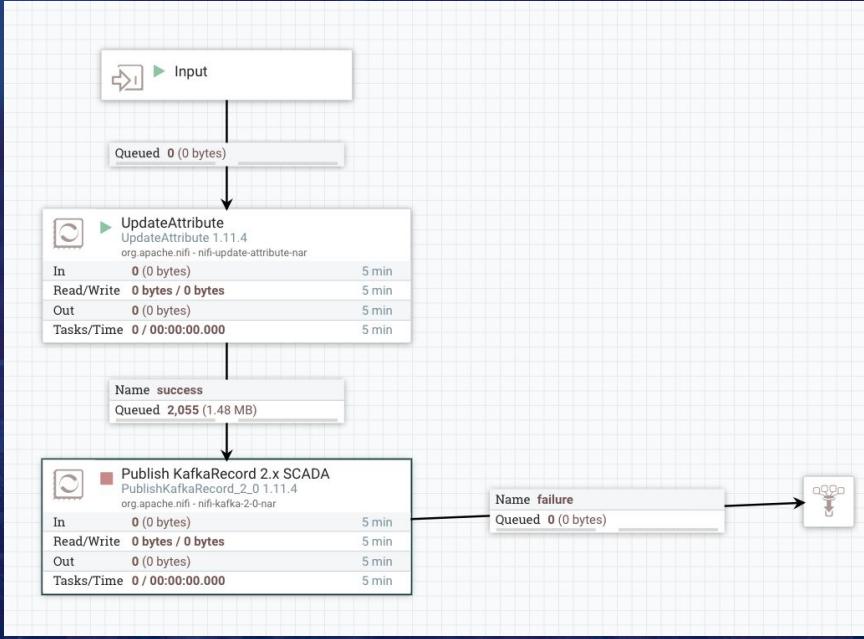
On the right, there is a 'VERSION 1' button with a pencil icon, a 'CHANGE LOG' section showing 'v1 1h 1m 20s ago Enabled', and a 'BRANCH: MASTER' label.

APACHECON NA
2020



APACHECON NA

2020



APACHECON NA

2020

Kafka Messages

Topics / scada

Cluster: OneNodeCluster

METRICS DATA EXPLORER CONFIGS LATENCY

DESERIALIZER: Keys: String Values: Avro

VALUE SCHEMA NAME:

scada

VALUE SCHEMA VERSIONS:

1

Show schema text

TO OFFSET

24784

FROM OFFSET

24769

0

4130

8260

12390

16520

20650



Partition 0

Offset ▲	Timestamp	Key	Value
24769	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_zfb_20200910235738	{"uuid": "rpi4_uuid_zfb_20200910235738", "systemtime": "09/10/2020 19:57:39", "temperaturef": "92.3", "pressure": 1017.2, "humidity": 9.6, "lux": 14.5... show more
24770	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_avl_20200910235725	{"uuid": "rpi4_uuid_avl_20200910235725", "systemtime": "09/10/2020 19:57:26", "temperaturef": "91.9", "pressure": 1017.2, "humidity": 9.5, "lux": 14.0... show more
24771	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_lpp_20200910235744	{"uuid": "rpi4_uuid_lpp_20200910235744", "systemtime": "09/10/2020 19:57:45", "temperaturef": "92.3", "pressure": 1017.2, "humidity": 9.4, "lux": 14.5... show more
24772	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_uqx_20200910235757	{"uuid": "rpi4_uuid_uqx_20200910235757", "systemtime": "09/10/2020 19:57:58", "temperaturef": "90.7", "pressure": 1017.1, "humidity": 10.0, "lux": 14.... show more
24773	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_raw_20200910235800	{"uuid": "rpi4_uuid_raw_20200910235800", "systemtime": "09/10/2020 19:58:01", "temperaturef": "90.8", "pressure": 1017.2, "humidity": 10.1, "lux": 14.... show more
24774	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_gcm_20200910235806	{"uuid": "rpi4_uuid_gcm_20200910235806", "systemtime": "09/10/2020 19:58:07", "temperaturef": "89.9", "pressure": 1017.2, "humidity": 10.5, "lux": 14.... show more
24775	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_evr_20200910235735	{"uuid": "rpi4_uuid_evr_20200910235735", "systemtime": "09/10/2020 19:57:36", "temperaturef": "91.9", "pressure": 1017.1, "humidity": 9.6, "lux": 13.6... show more
24776	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_lsf_20200910235728	{"uuid": "rpi4_uuid_lsf_20200910235728", "systemtime": "09/10/2020 19:57:29", "temperaturef": "92.3", "pressure": 1017.2, "humidity": 9.7, "lux": 2.3... show more
24777	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_ilv_20200910235722	{"uuid": "rpi4_uuid_ilv_20200910235722", "systemtime": "09/10/2020 19:57:23", "temperaturef": "91.4", "pressure": 1017.1, "humidity": 9.7, "lux": 14.9... show more
24778	Fri, Sep 18 2020, 17:05:49	rpi4_uuid_uug_20200910235741	{"uuid": "rpi4_uuid_uug_20200910235741", "systemtime": "09/10/2020 19:57:42", "temperaturef": "92.1", "pressure": 1017.2, "humidity": 9.5, "lux": 14.5... show more

1 - 10 of 15 < >

APACHECON NA
2020

Kafka Connect HDFS Sink

Connect Cluster / Connector Profile

Cluster: OneNodeCluster

scada

Pause Resume Restart Delete New Connector

[Connector Profile](#) [Connector Settings](#)

Connector Profile

CLASSNAME

com.cloudera.dim.kafka.connect.hdfs.HdfsSinkConnector

ASSIGNED WORKER

10.0.1.65:38083

STATUS
RUNNING

TOTAL TASKS
1

RUNNING TASKS
1

FAILED TASKS
0

PAUSED TASKS
0

Tasks

Search by host

Status **Worker ID: ↓** Task ID Put Batch Avg Time Sink Record Send Rate Partition Count

Status	Worker ID	Task ID	Put Batch Avg Time	Sink Record Send Rate	Partition Count
	10.0.1.65	0	0	0	1

Running Ratio

36.3% RUNNING

63.7% PAUSED

Offset Commits

100%

SUCCESSFUL COMMITS

0%

FAILED COMMITS

Additional Sink Record Metrics

SINK RECORD LAG MAX
NA

Additional Sink Task Metrics

BATCH SIZE MAX
0 mbs

BATCH SIZE AVERAGE
0 mbs

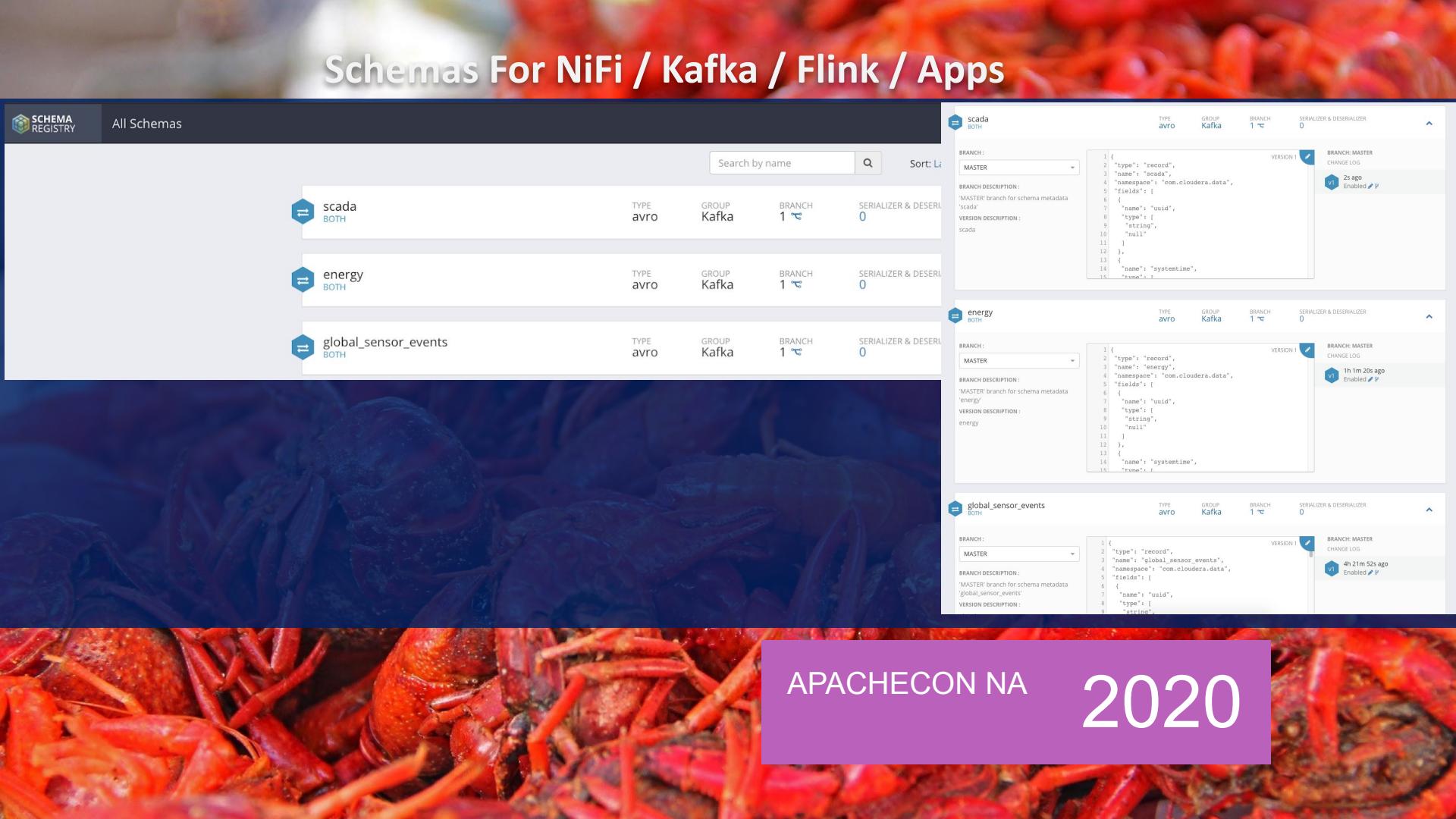
OFFSET COMMIT MAX TIME
0 ms

OFFSET COMMIT AVERAGE TIME
0 ms

APACHECON NA

2020

Schemas For NiFi / Kafka / Flink / Apps



The background of the slide features a close-up photograph of several bright red, cooked crayfish tails and claws.

SCHEMA REGISTRY

All Schemas

Search by name Sort: List

SCHEMA	TYPE	GROUP	BRANCH	SERIALIZER & DESERIALIZER
scada BOTH	avro	Kafka	1 ↗ 0	
energy BOTH	avro	Kafka	1 ↗ 0	
global_sensor_events BOTH	avro	Kafka	1 ↗ 0	

scada (Branch: MASTER)

BRANCH DESCRIPTION: 'MASTER' branch for schema metadata 'scada'

VERSION DESCRIPTION: scada

```
1 {  
2   "type": "record",  
3   "name": "scada",  
4   "namespace": "com.cloudera.data",  
5   "fields": [  
6     {"name": "uuid",  
7      "type": [  
8        "string",  
9        "null"  
10       ]  
11     }  
12   ],  
13   {"name": "systemtime",  
14     "+some": f  
15   }
```

energy (Branch: MASTER)

BRANCH DESCRIPTION: 'MASTER' branch for schema metadata 'energy'

VERSION DESCRIPTION: energy

```
1 {  
2   "type": "record",  
3   "name": "energy",  
4   "namespace": "com.cloudera.data",  
5   "fields": [  
6     {"name": "uuid",  
7       "type": [  
8         "string",  
9         "null"  
10        ]  
11     },  
12   ],  
13   {"name": "systemtime",  
14     "+some": f  
15   }
```

global_sensor_events (Branch: MASTER)

BRANCH DESCRIPTION: 'MASTER' branch for schema metadata 'global_sensor_events'

VERSION DESCRIPTION:

```
1 {  
2   "type": "record",  
3   "name": "global_sensor_events",  
4   "namespace": "com.cloudera.data",  
5   "fields": [  
6     {"name": "uuid",  
7       "type": [  
8         "string",  
9         "null"  
10        ]  
11     }  
12   ],  
13   {"name": "systemtime",  
14     "+some": f  
15   }
```

APACHECON NA

2020

CLI: flink list

```
20/07/31 15:55:46 INFO configuration.GlobalConfiguration: Loading configuration property: execution.job-list
20/07/31 15:55:46 INFO configuration.GlobalConfiguration: Loading configuration property: historyserver.wet
20/07/31 15:55:46 INFO configuration.GlobalConfiguration: Loading configuration property: historyserver.wet
20/07/31 15:55:46 INFO cli.FlinkYarnSessionCli: Found Yarn properties file under /tmp/.flink.yarn-properties
20/07/31 15:55:46 INFO modules.HadoopModule: Hadoop user set to root (auth:SIMPLE), credentials check status: OK
20/07/31 15:55:46 INFO modules.JaasModule: Jaas file will be created as /tmp/jaas-2656189102676203006.conf
20/07/31 15:55:46 INFO cli.CliFrontend: Running 'list' command.
20/07/31 15:55:47 INFO cli.CliFrontend: Waiting for response...
Waiting for response...
20/07/31 15:55:47 INFO cli.CliFrontend: Successfully retrieved list of jobs
----- Running/Restarting Jobs -----
31.07.2020 15:55:35 : d255d8b5ef4c2cc4bf34593bf3bb817 : default: INSERT INTO global_sensor_events
SELECT
    scada.uuid,
    scada.systemtime ,
    scada.temperaturef ,
    scada.pressure ,
    scada.humidity ,
    scada.lux ,
    scada.proximity ,
    scada.oxidising ,
    scada.reducing ,
    scada.nh3 ,
    scada.gasko,
    energy.`current`,
    energy.voltage ,
    energy.`power` ,
    energy.total ,
    energy.fanstatus

FROM energy,
     scada
WHERE
    scada.systemtime = energy.systemtime (RUNNING)
-----
```

CLI: flink sql

```
20/03/18 18:54:49 INFO cli.CliClient: SQL Cli command history file set to /tmp/flink-sql.history
```



BETA

```
[FLINK] [SQL] [CLUSTER] [TABLE]
```

Welcome! Enter 'HELP;' to list all available commands. 'QUIT;' to exit.

Flink SQL> |

APACHECON NA

2020



Global Flink Dashboard

Apache Flink Dashboard

Version: 1.10.0-csa1.2.0.0 Commit: 664a5f5 @ 29.04.2020 @ 14:13:26 UTC Message: 0

default: INSERT INTO global_sensor_events SELECT scada.uuid, scada.systemtime , scada.temperatureref , scada.pressure , scada.humidity , scada.lux , scada.proximity , scada.oxidising , scada.reducing , scada.nh3 , scada.gasko , energy.`current` , energy.voltage , energy.`power` , energy.`total` , energy.fanstatus FROM energy, scada WHERE scada.systemtime = energy.systemtime

[Cancel Job](#)

RUNNING 3
ID: e4a9674c12bb7413ee3e98c3d680251e Start Time: 2020-07-29 12:59:02 Duration: 8s

[Overview](#) [Exceptions](#) [TimeLine](#) [Checkpoints](#) [Configuration](#)

Detail SubTasks TaskManagers Watermarks Accumulators BackPressure Metrics

Join(joinType=[InnerJoin], where=[(systemtime0 = systemtime)], select=[systemtime, current, voltage, power, total, fanstatus, uid, systemtime0, temperature, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko], leftInputSpec=[NoUniqueKey], rightInputSpec=[NoUniqueKey]) -> Calc(select=[uid, systemtime0 AS systemtime, temperature, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko, current, voltage, power, total, fanstatus]) -> SinkConversionToRow -> Sink: KafkaTableSink(uid, systemtime, temperature, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko, current, voltage, power, total, fanstatus)

Status: **RUNNING** Task: 1

Parallelism: 1 Records Sent: 0

Start Time: 2020-07-29 12:59:02 Bytes Received: 20.0 KB

End Time: - Records Received: 132

Duration: 7s Bytes Sent: 0 B

Source: KafkaTableSource[uid, systemtime, amplitude100, amplitude500, amplitude1000, lowrange, midrange, highrange, amps, ipaddress, host, host_name, macaddresses s, energy, current, voltage, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko, current, voltage, power, total, fanstatus, memory_id, temperature, adjtemp, adjtemp, temperature, pressure, humidity, lu x, ...] Parallelism: 1

Join(joinType=[InnerJoin], where=[(systemtime0 = systemtime)], select=[systemtime, current, voltage, power, to tal, fanstatus, uid, systemtime0, temperature, pressur e, humidity, lux, proximity, o xidising, reducing, nh3, gask o], leftInputSpec=[NoUnique Key], rightInputSpec=[NoUni queKey]) -> Calc(select...

Source: KafkaTableSource[uid, systemtime, current, vol tage, power, total, swver, hw ver, tvos, model, mac, devic e] Parallelism: 1

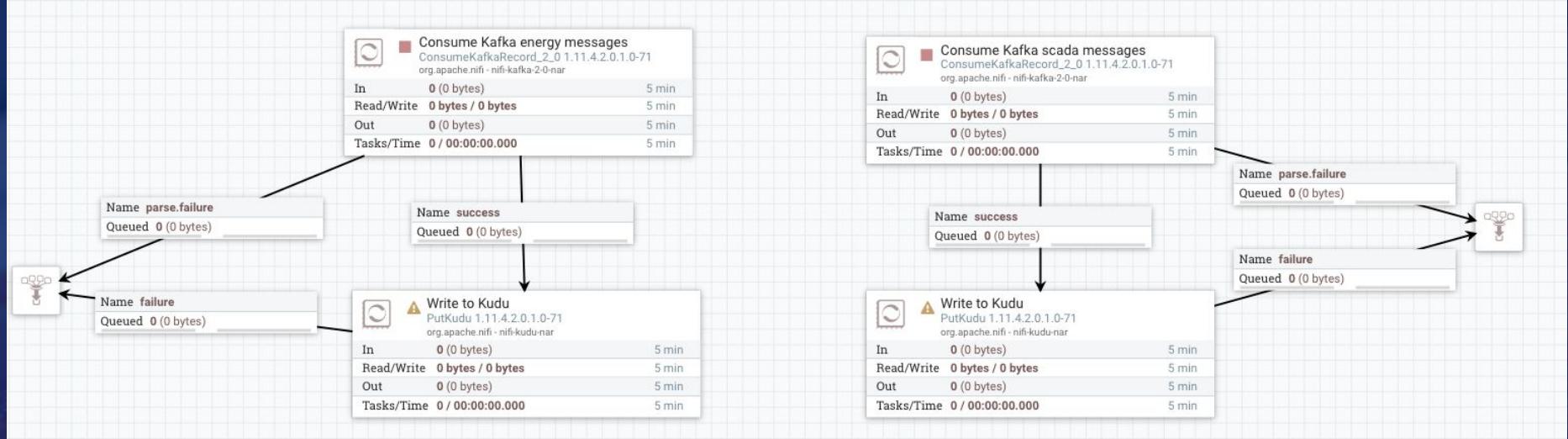
Name Status Bytes Received Records Received Bytes Sent Records Sent Parallelism Start Time Duration End Time Tasks

Name	Status	Bytes Received	Records Received	Bytes Sent	Records Sent	Parallelism	Start Time	Duration	End Time	Tasks
Source: KafkaTableSource[uid, systemtime, amplitude100, amplitude500, amplitude1000, lowrange, midrange, highrange, amps, ipaddress, host, host_name, macaddresses s, energy, current, voltage, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko, current, voltage, power, total, fanstatus, memory_id, temperature, adjtemp, adjtemp, temperature, pressure, humidity, lu x, ...]	RUNNING	0 B	0	0 B	121	1	2020-07-29 12:59:02	7s	-	1
Source: KafkaTableSource[uid, systemtime, current, voltage, power, total, swver, h...	RUNNING	0 B	0	0 B	11	1	2020-07-29 12:59:02	7s	-	1
Join(joinType=[InnerJoin], where=[(systemtime0 = systemtime)], select=[systemtim...	RUNNING	20.0 KB	132	0 B	0	1	2020-07-29 12:59:02	7s	-	1

APACHECON NA

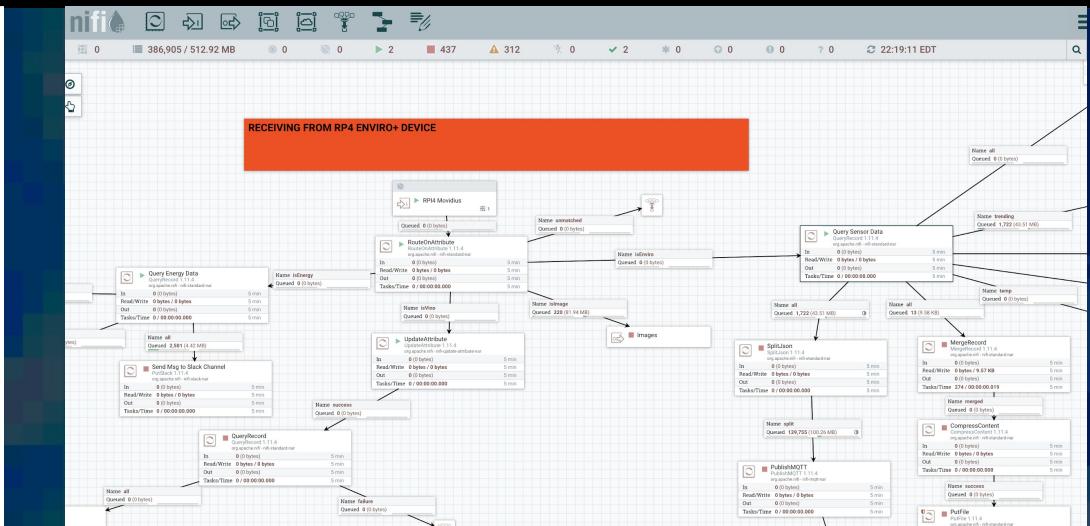
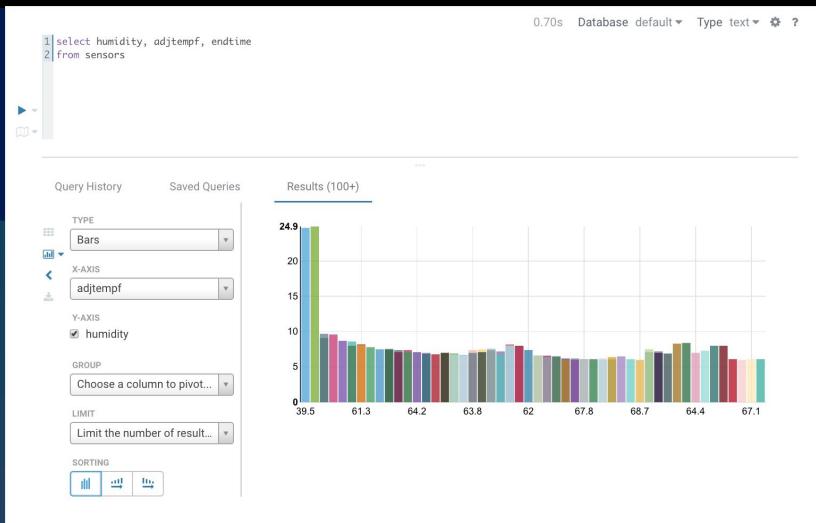
2020

Event Storage: Kafka to Kudu DataFlow



APACHECON NA
2020

Hydrate the Sensor Data Lake



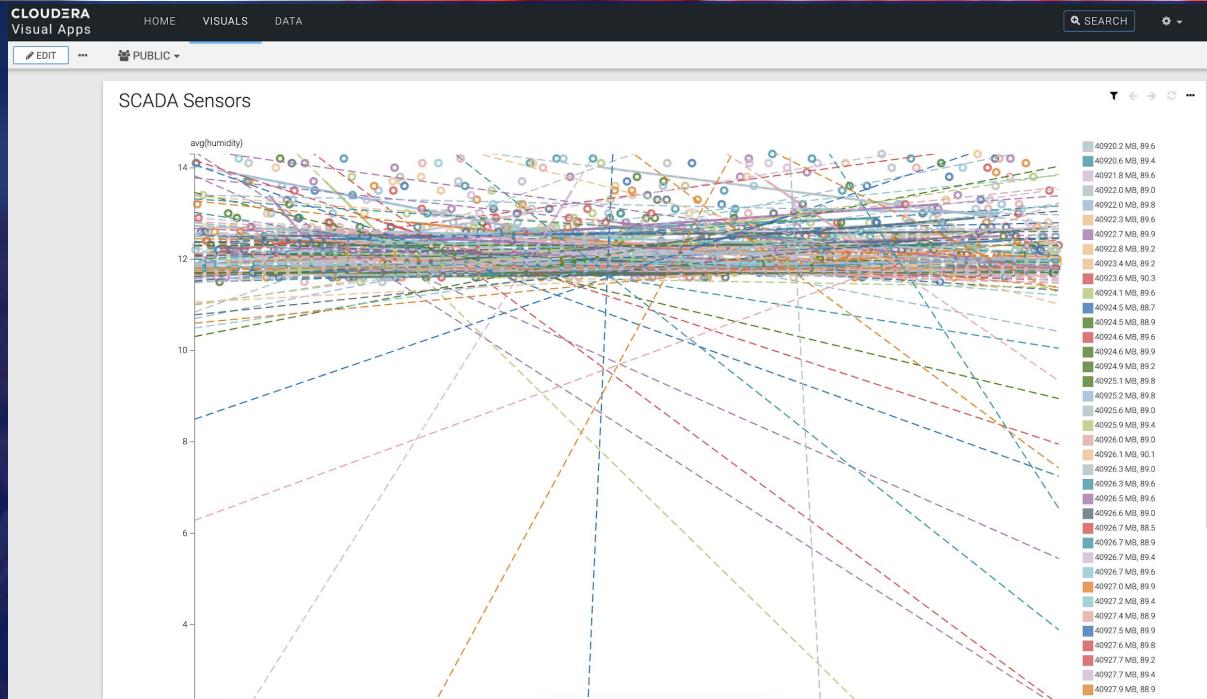
flankstack.dev

flankstack.dev

APACHECON NA

2020

Event Queries, Reporting, Analytics and Applications



APACHECON NA

2020



FLINK

FLaNK

flankstack.dev

APACHECON NA

2020

Flink Jobs

Schema Write Strategy

HWX Content-Encoded Schema Reference

Schema Cache

No value set

Schema Protocol Version

1

Schema Access Strategy

Inherit Record Schema

Schema Registry

HortonworksSchemaRegistry

Schema Name

\$schema.name

Schema Version

No value set

Schema Branch

No value set

Schema Text

\$avro.schema

Compression Format

NONE

Cache Size

1000

Encoder Pool Size

32

Running/Restarting Jobs

31.07.2020 15:55:35 : d255d8b5ef4c2cc4bf34593bf3b817 : default: INSERT INTO global_sensor_events

```

SELECT
    scada.uuid,
    scada.systemtime ,
    scada.temperaturef ,
    scada.pressure ,
    scada.humidity ,
    scada.lux ,
    scada.proximity ,
    scada.oxidising ,
    scada.reducing ,
    scada.nh3 ,
    scada.gasko ,
    energy.current ,
    energy.voltage ,
    energy.power ,
    energy.total ,
    energy.fanstatus
FROM energy,
     scada
WHERE
    scada.systemtime = energy.systemtime (RUNNING)

```

Job Overview

ID: 69044507424071310ffbd459c2770ed Start Time: 2020-09-23 14:53:59 Duration: 1m 6s

Overview Exceptions Timeline Checkpoints Configuration

Detail

Parallel: 1 Records Set: 0

Start Time: 2020-09-23 14:54:06 End Time: - Duration: 1m 6s

Records Received: 0/144000 Records Sent: 0/144000 Bytes Received: 0/432000 Bytes Sent: 0/432000

Subtasks

TaskManagers Watermarks Accumulators Backpressure Metrics

Task 1

Source: KafkaTopicSource[scada, systemtime, current, voltage, power, total, fanstatus] >> CalcJobs[calc, systemtime, current, voltage, power, total, fanstatus] >> SinkKafkaTopicSink[global_sensor_events, systemtime, current, temperaturef, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko]

Task 2

Source: KafkaTopicSource[energy, systemtime, current, voltage, power, total, fanstatus] >> CalcJobs[calc, systemtime, current, voltage, power, total, fanstatus] >> SinkKafkaTopicSink[global_sensor_events, systemtime, current, temperaturef, pressure, humidity, lux, proximity, oxidising, reducing, nh3, gasko]

SQL Query Result (Table)

Refresh: 1 s Updated: 18:56:50.126

uuid	systemtime	temperaturef	pressure	humidity	lux	proximity	oxidising	redu
rpi4_uuid_sxk_2020092219~	09/22/2020 15:12:06	73.9	1012.3	15.0	25.3	0	3.7	1
rpi4_uuid_izc_2020092318~	09/23/2020 14:43:08	87.6	1007.9	13.2	25.3	0	10.2	1
rpi4_uuid_ntz_2020092318~	09/23/2020 14:45:27	89.6	1008.0	12.3	25.6	0	14.7	2
rpi4_uuid_afu_2020092318~	09/23/2020 14:53:30	91.9	1007.8	11.0	25.3	0	27.4	2

Flink SQL Output

Running Job List

Job Name	Start Time	Duration	End Time	Tasks	Status	Cluster
default: INSERT INTO global_sensor_events SELECT scada.uuid, scada.systemtime , scada.temperatureref , scada.pressure , scada.humidity , scada.lux , scada.proximity , scada.oxidising , scada.reducing , scada.nh3 , scada.gasko, energy.`current` , energy.voltage , energy.`power` , energy.`total` , energy.fanstatus FROM energy, scada WHERE scada.systemtime = energy.systemtime	2020-09-23 14:53:59	1m 48s	-	3 / 3	RUNNING	application_1600703208565_0003

uuid	systemtime	current	voltage	power	total	fanstatus	diskusage
20200921202335_221e27cc~-	09/21/2020 16:23:37	0	121	0	0	ON	40213.4
20200921202149_6c0f816d~-	09/21/2020 16:21:50	0	121	0	0	ON	40204.7
20200921202938_cfe6cd3d~-	09/21/2020 16:29:39	0	121	0	0	ON	40211.2
20200921202558_edd007fc~-	09/21/2020 16:25:59	0	121	0	0	ON	40195.0
20200921202963_2023d63b~-	09/21/2020 16:29:04	0	121	0	0	ON	40238.3
20200921202634_9bca4d67~-	09/21/2020 16:26:35	0	121	0	0	ON	40166.0
20200921202411_ee2abac3~-	09/21/2020 16:24:12	0	121	0	0	ON	40184.0
20200921202224_14ccae1~-	09/21/2020 16:22:26	--	--	--	--	--	--
uuid	systemtime	temperature	pressure	humidity	lux	proximity	oxidising
20200921202744_1b493658~-	09/21/2020 16:27:49	74.8	1812.2	14.8	25.3	0	4.5
20200921202522_87f9853e~-	09/21/2020 16:25:24	75.0	1812.2	14.7	25.3	0	4.5
20200921202824_81cce633~-	09/21/2020 16:28:29	75.0	1812.2	14.6	25.3	0	4.5
20200921202446_10d94542~-	09/21/2020 16:24:47	75.0	1812.2	14.7	25.3	0	4.5
20200921202300_18fc9044~-	09/21/2020 16:23:01	75.2	1812.3	14.7	25.3	0	4.5
20200921202299_8258a97e~-	09/21/2020 16:22:59	75.4	1812.3	14.3	25.3	0	4.7
20200921202084_6ab338c98~-	09/21/2020 16:30:47	75.5	1812.3	14.6	25.3	0	4.8
202009212020841_9a8adfd9~-	09/21/2020 16:20:41	75.5	1812.3	14.3	25.3	0	4.8
20200921202709_5d360728~-	09/21/2020 16:27:10	75.7	1812.2	14.0	25.3	0	4.9
20200922191057_efb97bddd~-	09/22/2020 15:10:57	75.9	1812.2	13.9	25.3	1	4.9
20200922191206_3a13096f~-	09/22/2020 15:12:06	76.1	1807.9	19.5	25.3	0	1.1
20200923183753_8258a97e~-	09/23/2020 14:37:53	76.1	1807.9	19.3	25.3	0	1.3
20200923183827_3e2f12dd~-	09/23/2020 14:38:28	76.4	1807.9	19.0	25.3	0	1.6
20200923183937_2de0fbf3~-	09/23/2020 14:39:37	76.7	1807.9	18.7	25.3	0	1.8
20200923184012_cbacc07~-	09/23/2020 14:40:12	77.0	1807.9	18.6	25.3	0	2.0
20200923184047_b6573a3e~-	09/23/2020 14:40:47	78.2	1807.9	18.2	25.3	0	2.2
20200923183967_2939fe5f~-	09/23/2020 14:39:03	78.4	1807.9	18.0	25.3	0	2.4
20200923183972_d2e0fbf3~-	09/23/2020 14:39:37	78.8	1807.9	18.0	25.3	0	2.7
20200923184012_cbacc07~-	09/23/2020 14:40:12	78.8	1807.9	18.0	25.3	0	2.8
20200923184047_b6573a3e~-	09/23/2020 14:40:47	78.8	1807.9	18.0	25.3	0	3.0
20200923183918_07e0fbf3~-	09/23/2020 14:39:01	78.8	1807.9	17.9	24.4	0	3.2
20200923183920_2028092318~-	09/23/2020 14:39:04	79.0	1807.9	17.8	25.3	0	3.5
20200923183921_2028092318~-	09/23/2020 14:39:07	79.0	1807.9	17.8	25.3	0	3.6
20200923183922_2028092318~-	09/23/2020 14:39:10	79.0	1807.9	17.8	25.3	0	3.8
20200923183923_2028092318~-	09/23/2020 14:39:14	79.0	1807.9	17.8	25.3	0	3.9
20200923183924_2028092318~-	09/23/2020 14:39:17	79.0	1807.9	17.8	25.3	0	4.0
20200923183925_2028092318~-	09/23/2020 14:39:20	79.0	1807.9	17.7	25.3	0	4.2
20200923183926_2028092318~-	09/23/2020 14:39:23	79.1	1807.9	17.6	25.3	0	4.3
20200923183927_2028092318~-	09/23/2020 14:39:26	79.3	1807.9	17.6	25.3	0	4.4
20200923183928_2028092318~-	09/23/2020 14:39:29	79.5	1807.9	17.6	25.3	0	4.5
20200923183929_2028092318~-	09/23/2020 14:39:32	79.7	1807.9	17.5	25.3	0	4.6
20200923183930_2028092318~-	09/23/2020 14:39:35	79.9	1807.9	17.5	25.3	0	4.7
20200923183931_2028092318~-	09/23/2020 14:39:36	79.9	1807.9	17.3	25.3	0	4.9
20200923183932_2028092318~-	09/23/2020 14:39:39	79.9	1807.9	17.3	25.3	0	5.1
20200923183933_2028092318~-	09/23/2020 14:39:42	79.9	1807.9	17.2	25.3	0	5.1
20200923183934_2028092318~-	09/23/2020 14:39:45	79.9	1807.9	17.1	25.3	0	5.1
20200923183935_2028092318~-	09/23/2020 14:39:48	80.0	1808.0	17.0	25.3	0	5.1
20200923183936_2028092318~-	09/23/2020 14:39:51	80.2	1808.0	17.0	25.3	0	5.2
20200923183937_2028092318~-	09/23/2020 14:39:55	80.4	1808.0	16.9	25.3	0	5.3

References

- <https://www.datainmotion.dev/2019/08/rapid-iot-development-with-cloudera.html>
- <https://www.datainmotion.dev/2019/09/powering-edge-ai-for-sensor-reading.html>
- <https://www.datainmotion.dev/2019/05/dataworks-summit-dc-2019-report.html>
- <https://www.datainmotion.dev/2019/03/using-raspberry-pi-3b-with-apache-nifi.html>
- <https://www.datainmotion.dev/2020/06/unboxing-most-amazing-edge-ai-device.html>
- <https://www.datainmotion.dev/2019/08/edge-processing-with-jetson-nano-part-3.html>
- <https://www.datainmotion.dev/2020/04/predicting-sensor-readings-with-time.html>
- <https://www.datainmotion.dev/2020/05/time-series-analysis-dataflow.html>



Demos / Source / DevOps / Scripts

The code, build scripts, schemas, table DDL, Flink SQL, Kafka Connect configuration, NiFi flows, HBase tables, Kudu tables, Hive tables, HDFS directories, alerts, images, HTML, docs, links and all the goodies are here. Please fork and contribute.

<https://github.com/tspannhw/ApacheConAtHome2020>

<https://github.com/tspannhw/minifi-xaviernx>

<https://github.com/tspannhw/minifi-jetson-nano>

<https://github.com/tspannhw/minifi-enviroplus>

<https://github.com/tspannhw/meetup-sensors/blob/main/flink-sql/democdf.sh>

<https://github.com/tspannhw/FlinkSQLDemo>

APACHECON NA

2020



CONNECT!

FLaNK

flankstack.dev

APACHECON NA

2020