

CLOUDERA

# Cracking the Nut, Solving Edge AI with Apache Tools and Frameworks

Timothy Spann

Principal DataFlow Field Engineer

Cloudera

@PaasDev



# Tim SPANN

<https://github.com/tspannhw>

<https://www.datainmotion.dev/>

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



# Welcome to Future of Data - Princeton - Virtual

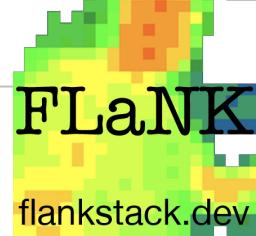


<https://www.meetup.com/futureofdata-princeton/>

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 for Cloud Data Engineers - Edge AI

Multiple users, frameworks, languages, clouds, data sources & clusters



CLOUD DATA ENGINEER

- Experience in ETL/ELT
- Coding skills in Python or Java
- Knowledge of database query languages such as SQL
- Experience with Streaming
- Knowledge of Cloud Tools



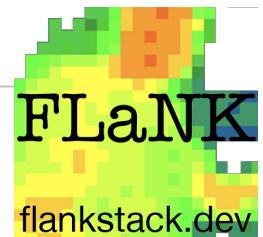
CAT

- Expert in ETL (Eating, Ties and Laziness)
- Edge Camera Interaction
- Typical User
- No Coding Skills
- Can use NiFi
- Questions your cloud spend



AI / Deep Learning / ML / DS

- Can run in Apache NiFi
- Can run in Kafka Streams
- Can run in Apache Flink
- Can run in MiNiFi Agents



# Apache Tools and Frameworks Used



# Apache MXNet Native Processor through DJL.AI for Apache NiFi



#workshop

11:30 AM =====

Deep Learning Class Label: person  
File: cc0a469f-c108-42c7-95c6-10e5fda95006.person.png  
Probability: 0.96  
UUID: 32ef65a3-0650-42cd-965c-ba25597eb1ad  
Rank: 1

Bounding Box (Height/Width, X,Y)  
0.74 / 0.69  
0.27, 0.25

Image (Height/Width, X,Y)  
480 / 640  
0, 0

=====

11:30 AM tspann 371bdb8f-35bc-4a2a-919c-bdeb609b726c.person.png

```
=====
private void runMockUserDeploy() {
    testRunner.setWorkflowExpressionUsage(false);
    testRunner.run();
    testRunner.assertValid();
}

testRunner.assertAllFlowFilesTransferred(DeepLearningProcessor.REL_SUCCESS);
List<MockFlowFile> successfuls = testRunner.getFlowFilesForRelationship(DeepLearningProcessor.REL_SUCCESS);

for (MockFlowFile mockFile : successfuls) {
    assertEquals("car", mockFile.getAttribute("category"));
    assertEquals("1.0", mockFile.getAttribute("probability"));

    System.out.println("MockFlowFile: " + mockFile);
    Map<String, String> attributes = mockFile.getAttributes();
    for (String attribute : attributes.keySet()) {
        System.out.println("Attribute: " + attribute);
    }
}

@test
public void testProcessor() throws Exception {
    java.io.File resourcesDirectory = new java.io.File(System.out.print(resourcesDirectory.getAbsolutePath()));

    testRunner.setProperty(DeepLearningProcessor.MXNET_MODEL_PATH, testRunner.setProvenv(DeepLearningProcessor.DATASET));
    DeepLearningProcessorTest testProcessor = new DeepLearningProcessorTest();
    testProcessor.setProcessor(DeepLearningProcessor.REL_SUCCESS);
    testProcessor.start();
    testProcessor.process();
    testProcessor.stop();

    TestRunner.Result result = testRunner.getResult();
    assertEquals("Test passed: 1 of 1 test - 4 o 018 ms", result.toString());
    assertEquals("Size:176328", result.toString());
    assertEquals("Attribute:boundingbox_height_1 = 0.35", result.toString());
    assertEquals("Attribute:boundingbox_width_1 = 0.25", result.toString());
    assertEquals("Attribute:image_min_x = 0", result.toString());
    assertEquals("Attribute:class_1 = car", result.toString());
    assertEquals("Attribute:rank_3 = 1", result.toString());
    assertEquals("Attribute:uuid = e9993c52-f5ab-4849-8876-a25796714984", result.toString());
    assertEquals("Attribute:boundingbox_width_1 = 0.24", result.toString());
}
=====
```

## Attribute Values

boundingbox\_height\_1

0.99

No value set

boundingbox\_width\_1

0.90

No value set

boundingbox\_x\_1

0.09

No value set

boundingbox\_y\_1

0.01

No value set

class\_1

tmonitor

No value set

filename

2020-08-26\_1330.jpg.tmonitor.png

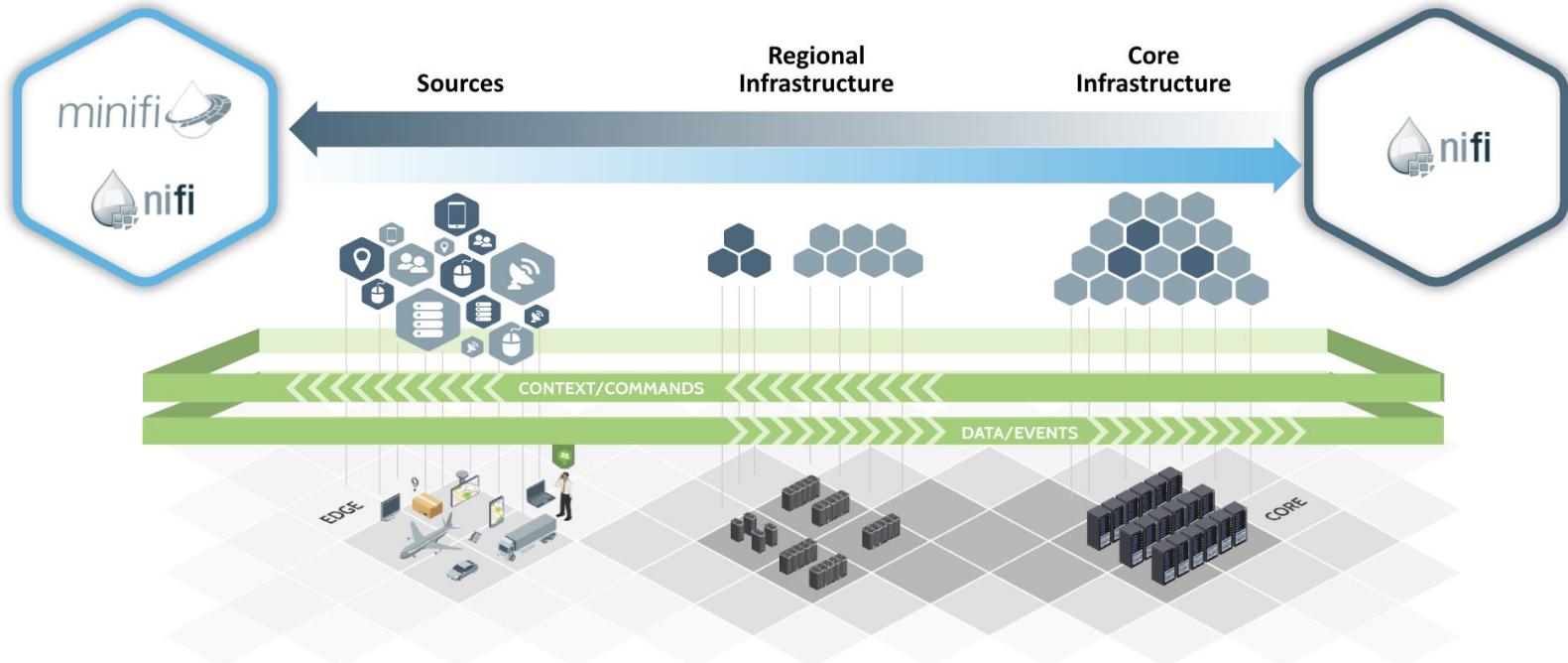
2020-08-26\_1330.jpg (previous)

This processor uses the DJL.AI Java Interface

<https://github.com/tspannhw/nifi-djl-processor>

<https://dev.to/tspannhw/easy-deep-learning-in-apache-nifi-with-djl-2d79>

# What is Apache NiFi and MiNiFi used for?



---

# Demo

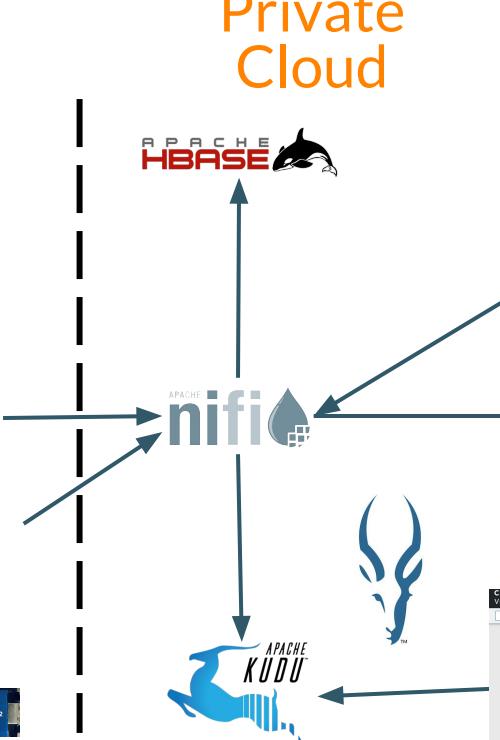
# Edge AI to Cloud Streaming Pipeline



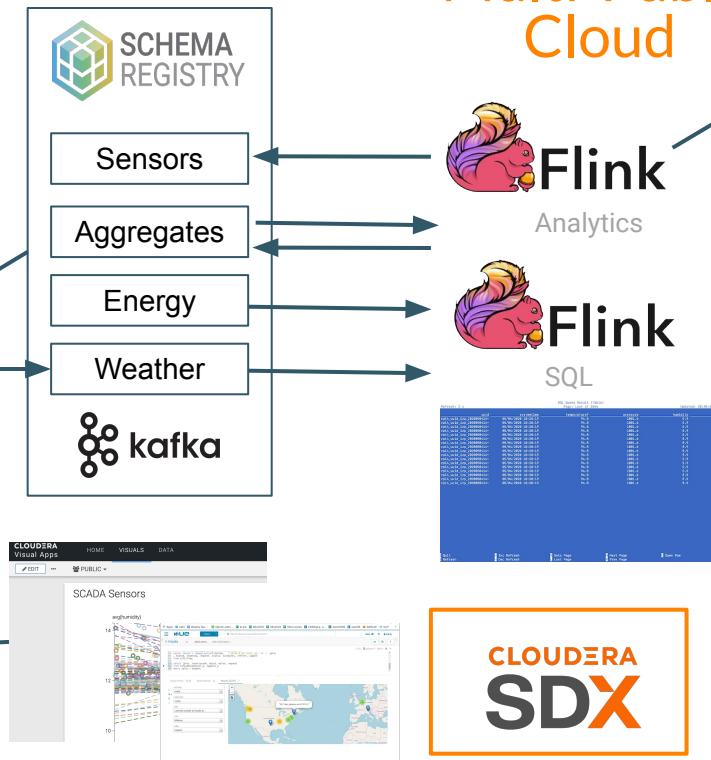
Edge



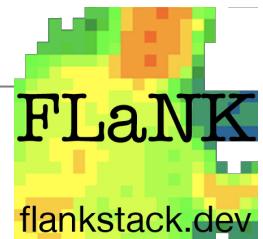
Private  
Cloud



Multi-Public  
Cloud



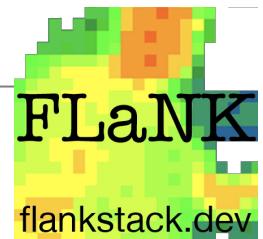
Centralized Schema Repository  
SCHEMA REGISTRY



## SHOW ME THE DATA

```
{"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", "temperatureref": "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"}
```





## WHERE DID THAT DATA COME FROM?

BME280 - temperature, pressure, humidity sensor

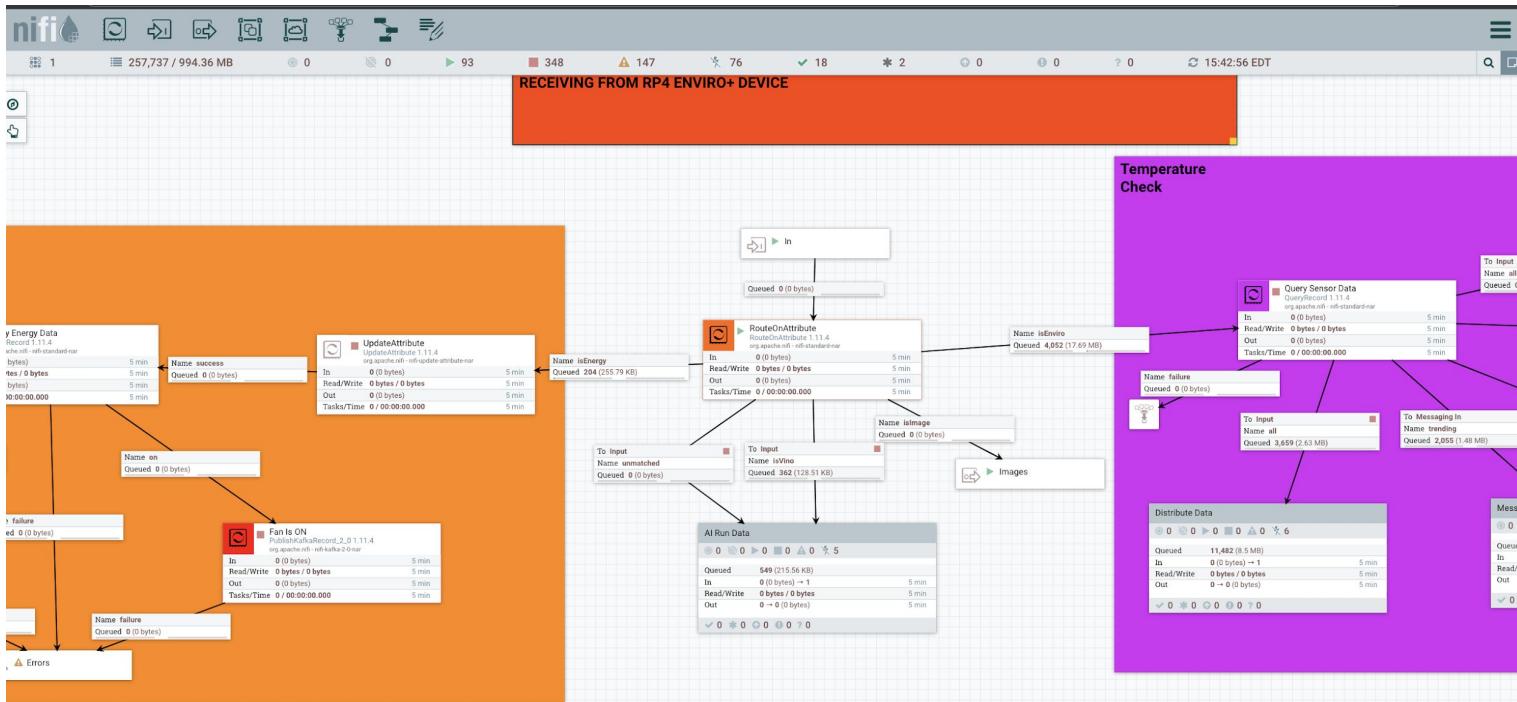
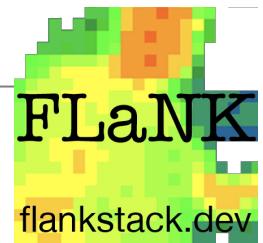
LTR-559 - light and proximity sensor

MICS6814 - analog gas sensor

ADS1015 ADC

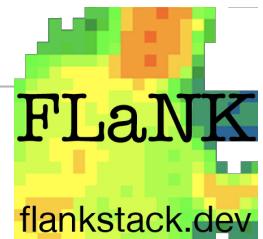
MEMS - microphone

0.96-inch, 160 x 80 color LCD



---

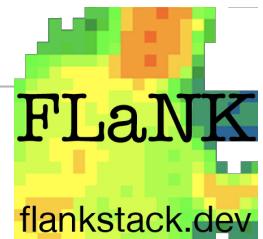
Learn More



## DEMO SOURCE CODE

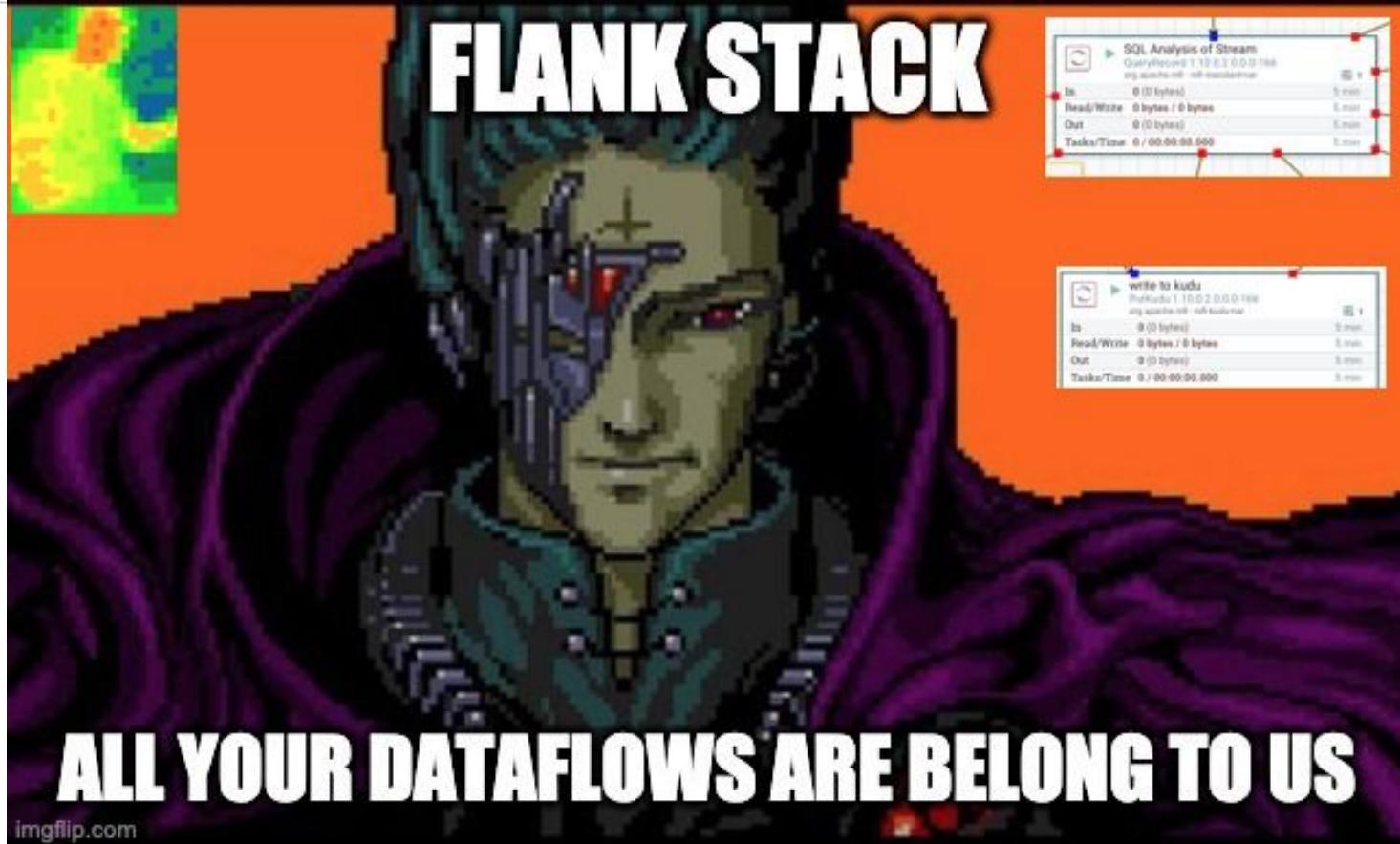
- <https://github.com/tspannhw/FlinkForwardGlobal2020>
- <https://github.com/tspannhw/ApacheConAtHome2020>
- <https://github.com/tspannhw/minifi-xaviernx>
- <https://github.com/tspannhw/minifi-jetson-nano>
- <https://github.com/tspannhw/minifi-enviroplus>
- <https://github.com/tspannhw/EverythingApacheNiFi>
- <https://github.com/tspannhw/CloudDemo2021>
- <https://github.com/tspannhw/FlinkSQLWithCatalogsDemo>

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.



## DEEPER CONTENT

- <https://www.datainmotion.dev/2020/10/running-flink-sql-against-kafka-using.html>
- <https://www.datainmotion.dev/2020/10/top-25-use-cases-of-cloudera-flow.html>
- <https://github.com/tspannhw/EverythingApacheNiFi>
- <https://github.com/tspannhw/CloudDemo2021>
- <https://github.com/tspannhw/StreamingSQLExamples>

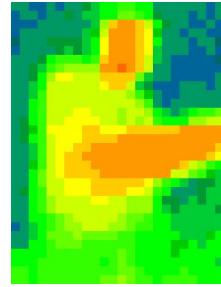


# Upcoming Events

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

May 19





TH<sup>DATA</sup>N Y<sup>ML</sup> U<sup>ML</sup>

