Extra Lab: NiFi Enrichment

Background: NiFi Enrichment

In this lab, we will cover how to add processors to enrich events before sending them to the OpenSearch database.

Specifically, we will learn how to add GeoIP information to an event using the same schema as Beats processors.

Background: Record Paths

NiFi operates on JSON files using "Record Path" syntax, and is analogous to a file-path, with the root directory `/` being the root of the JSON object, and sub-keys being folder names. For example in the object:

```
{
  "source": {
    "ip": "192.168.1.1",
    "ip.keyword": "192 168 1 1",
    "geo": {
      "country_name": "Unknown"
    }
  }
}
```

the path to the `country_name` is: `/source/geo/country_name`, and the path to `ip.keyword` is `/source/ip.keyword`.

Exercise: NiFi Enrichment

Add a GeoEnrichIPRecord processor to enrich the destination IP of events from Packetbeat.

Set the following settings on the GeoEnrichIPRecord processor:

- MaxMind Database File: `/opt/nifi/nifi-current/state/GeoLite2-City.mmdb`
- City Record Path: `/destination/geo/city_name`
- Latitude Record Path: `/destination/geo/location/lat`
- Longitude Record Path: `/destination/geo/location/lon`
- Country Record Path: `/destination/geo/country_name`
- Country ISO Code Record Path: `/destination/geo/country_iso_code`
- Country Postal Code Record Path: `/destination/geo/postal_code`

Exercise: NiFi Enrichment (2)

Mini-Lab: find the correct `IP Address Record Path` for the destination IP by inspecting queued FlowFiles as shown above. Also set the Record Reader and Record Writer to appropriate values.

See next slide for the answer.

Exercise: NiFi Enrichment (3)

Mini-Lab Answer: `IP Address Record Path` should be set to

```
/destination/ip
```

in the properties of the GeoEnrichIPRecord processor. Set Record Reader to `JsonTreeReader` and Record Writer to `JsonRecordSetWriter`.

Exercise: NiFi Enrichment (4)

Hover over the GeoEnrichIPRecord processor to show the arrow icon, then drag it and **connect it to the PutElasticsearchRecord processor you configured for Packetbeat** to create a connection for the `found` relationship.

Now, hover over the GeoEnrichIPRecord processor and drag its arrow to the PutElasticsearchRecord processor *again*. This time, create the connection for the `not found` relationship.

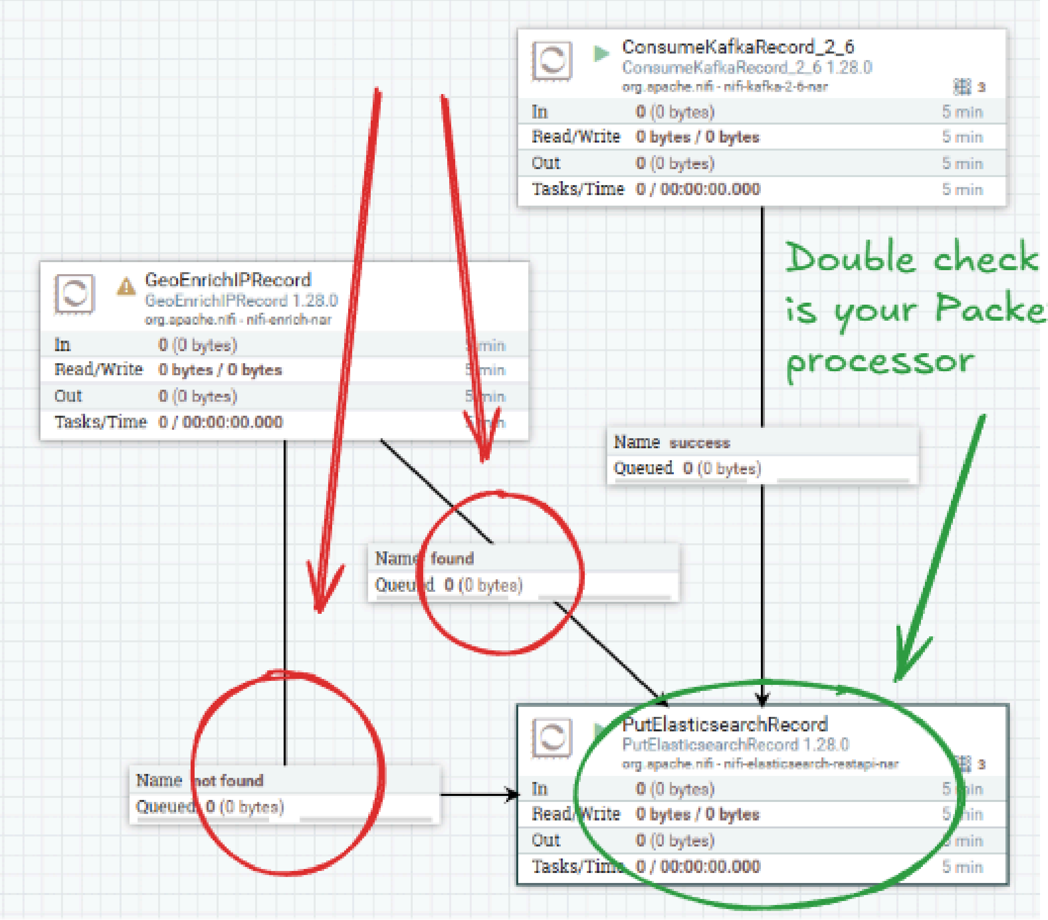
Finally, open the settings of the GeoEnrichIPRecord processor and terminate the `original` relationship.

Exercise: NiFi Enrichment (5)

If done correctly, your connection should look like the diagram on the right.

Now, stop both the ConsumeKafkaRecord_2_6 and PutElasticsearchRecord processors for Packetbeat.

Ensure that there are two distinct connections here



ConsumeKafkaRecord_2_6
ConsumeKafkaRecord_2_6 1.28.0
org.apache.nifi - nifi-kafka-2-6-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

GeoEnrichIPRecord
GeoEnrichIPRecord 1.28.0
org.apache.nifi - nifi-enrich-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

Name found
Queued 0 (0 bytes)

Name not found
Queued 0 (0 bytes)

PutElasticsearchRecord
PutElasticsearchRecord 1.28.0
org.apache.nifi - nifi-elasticsearch-restapi-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

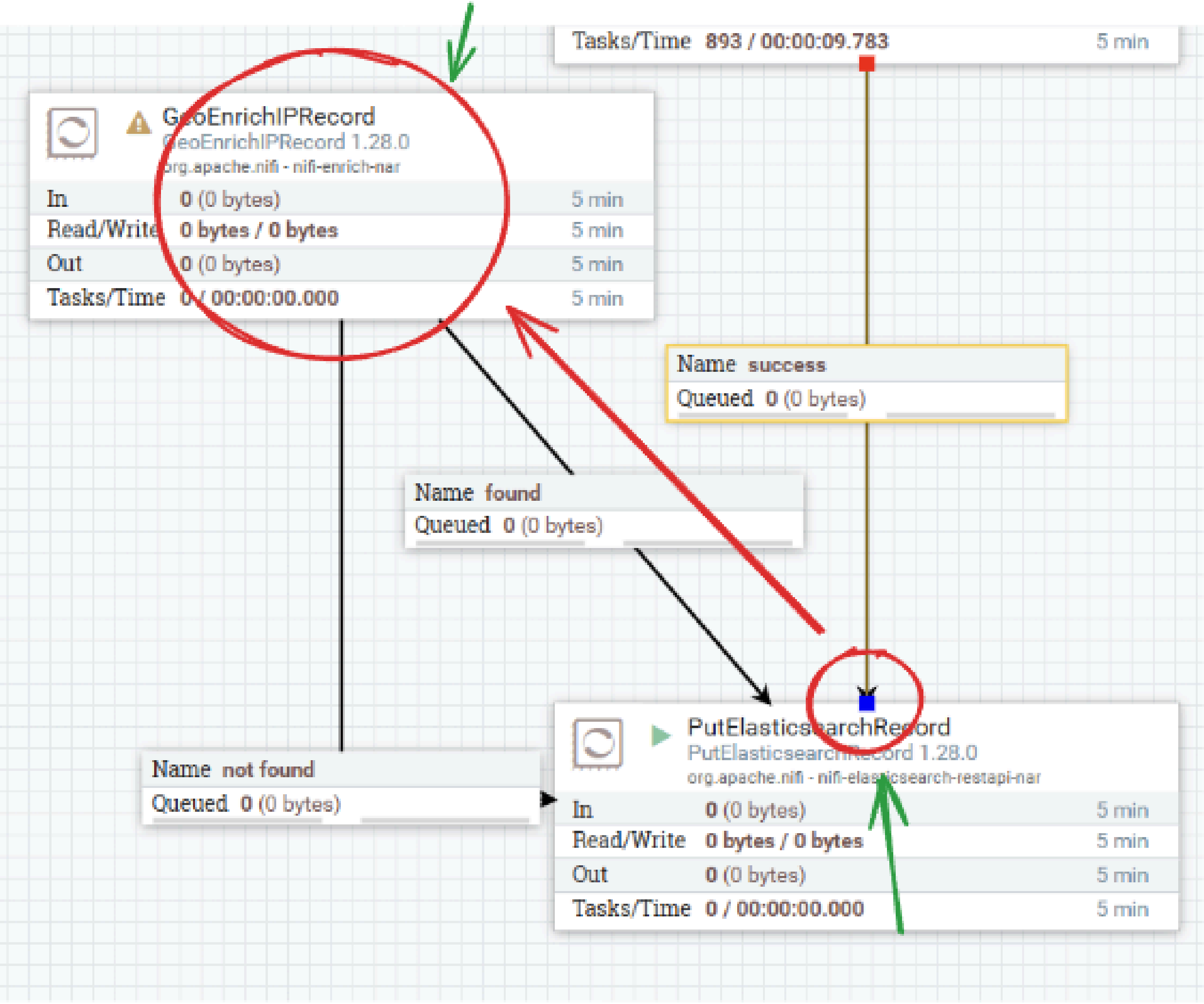
Double check that this is your Packetbeat processor

Exercise: NiFi Enrichment (6)

Click on the `success` relationship between ConsumeKafkaRecord_2_6 and PutElasticsearchRecord. A blue dot should appear in the arrowhead of the relationship.

Drag the blue dot and connect it to the GeoEnrichIPRecord processor.

to here



Tasks/Time 893 / 00:00:09.783 5 min

GeoEnrichIPRecord
GeoEnrichIPRecord 1.28.0
org.apache.nifi - nifi-enrich-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

Name found
Queued 0 (0 bytes)

Name not found
Queued 0 (0 bytes)

PutElasticsearchRecord
PutElasticsearchRecord 1.28.0
org.apache.nifi - nifi-elasticsearch-restapi-nar

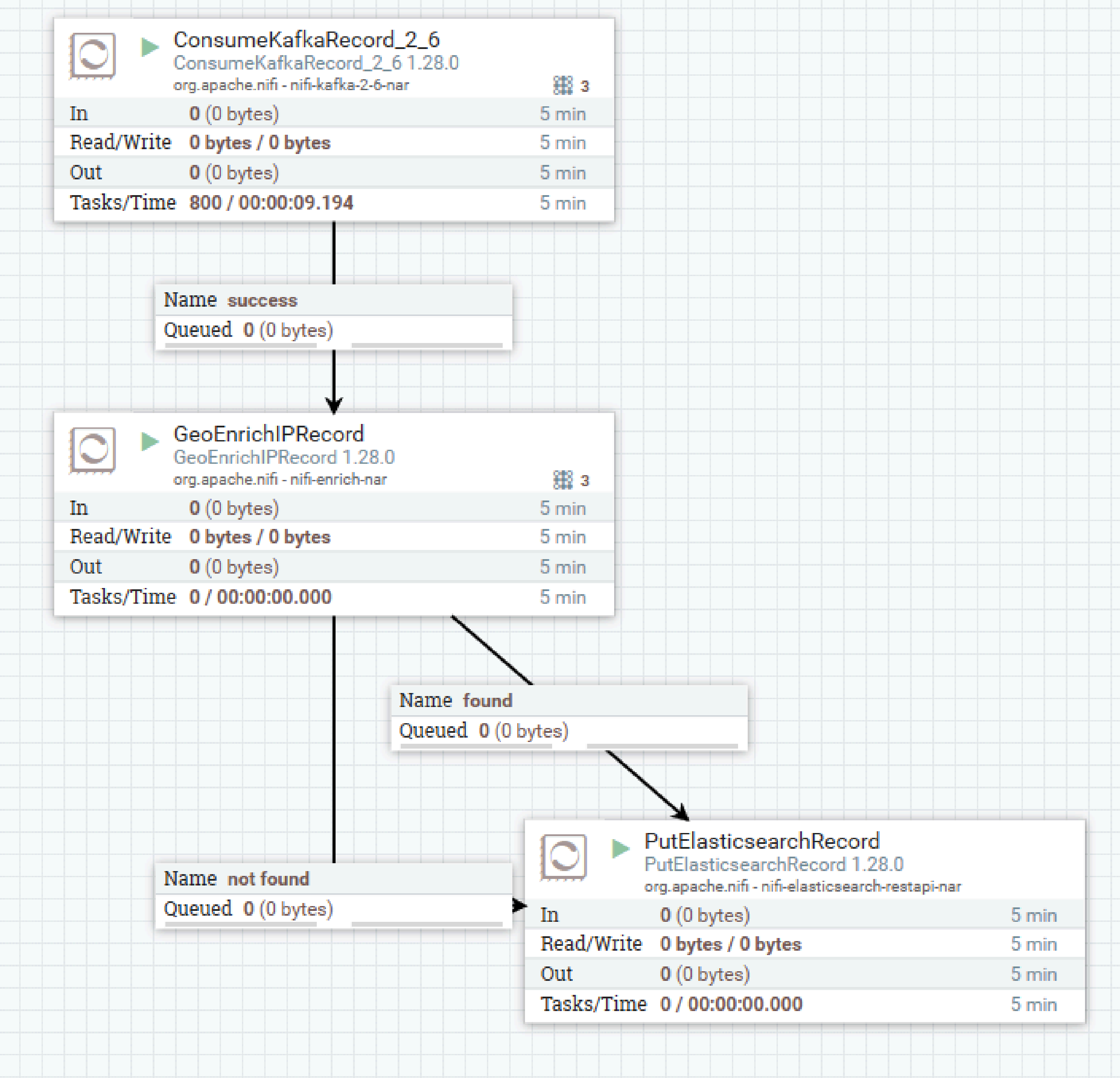
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Drag the blue dot from here...

Exercise: NiFi Enrichment (7)

Finally, start all the processors. The final diagram should look like the image on the right.

Notice that almost all events are being sent to the `not found` relationship. This is normal as most of the traffic in your Kubernetes cluster consists mainly of connections between containers with private IP addresses.



ConsumeKafkaRecord_2_6
ConsumeKafkaRecord_2_6 1.28.0
org.apache.nifi - nifi-kafka-2-6-nar
3

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	800 / 00:00:09.194	5 min

Name success
Queued 0 (0 bytes)

GeoEnrichIPRecord
GeoEnrichIPRecord 1.28.0
org.apache.nifi - nifi-enrich-nar
3

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name found
Queued 0 (0 bytes)

Name not found
Queued 0 (0 bytes)

PutElasticsearchRecord
PutElasticsearchRecord 1.28.0
org.apache.nifi - nifi-elasticsearch-restapi-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Background: NiFi Enrichment

By separating the different data sources (i.e., Beats outputs) into different topics in Kafka, and creating separate pipelines per topic, we were able to trivially apply specialized processing for a particular type of data (e.g., Packetbeat).

If all the events were sent to a single pipeline, it would be much more difficult to integrate enrichment for only one type of data.

Next Steps

Congratulations!

You have successfully completed the following:

- Completed a deployment of a GeoIP enrichment processor
- Learned how to modify an existing transformation to insert a new processor
- Applied Record Paths to specify fields in JSON FlowFiles

What's Next?

You can go back to the main session here.