

Integration with Kafka

– Integration with Kafka

Kafka is a highly available distributed messaging system for publishing and subscribing messages

NiFi provides capability to publish/subscribe messages to/from Kafka

Processor to publish messages to kafka → PublishKafka_{version}

Processor to subscribe messages from kafka → ConsumeKafka_{version}

PublishKafka_1_0, PublishKafka_2_0, PublishKafka_2_6

ConsumeKafka_1_0, ConsumeKafka_2_0, ConsumeKafka_2_6

NiFi as data producer



A screenshot of the NiFi interface showing the configuration for a "GenerateFlowFile" processor. The processor icon is a blue square with a white gear and square. The name "GenerateFlowFile" is displayed, along with its version "GenerateFlowFile 1.14.0" and source "org.apache.nifi - nifi-standard-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

A screenshot of the NiFi interface showing the configuration for a "PublishKafka_2_6" processor. The processor icon is a blue square with a white gear and square. The name "PublishKafka_2_6" is displayed, along with its version "PublishKafka_2_6 1.15.3" and source "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

– Publishing messages to Kafka

PublishKafka_2_6 Processor to send messages

	⚠ PublishKafka_2_6 PublishKafka_2_6 1.15.3 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

Required field	
Property	Value
Kafka Brokers	localhost:9092
Topic Name	No value set
Use Transactions	true
Transactional Id Prefix	No value set
Message Demarcator	No value set
Failure Strategy	Route to Failure
Delivery Guarantee	Best Effort
Attributes to Send as Headers (Regex)	No value set
Message Header Encoding	UTF-8
Security Protocol	PLAINTEXT
SASL Mechanism	GSSAPI
Kerberos Credentials Service	No value set

– Publishing messages to Kafka

PublishKafka_2_6 Processor to send messages

	⚠ PublishKafka_2_6 PublishKafka_2_6 1.15.3 org.apache.nifi - nifi-kafka-2-6-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Required field	
Property	Value
Kerberos User Service	?
Kerberos Service Name	?
Kerberos Principal	?
Kerberos Keytab	?
Username	?
Password	?
Token Auth	?
SSL Context Service	?
Kafka Key	?
Key Attribute Encoding	?
Max Request Size	?
Acknowledgment Wait Time	?

– Publishing messages to Kafka

PublishKafka_2_6 Processor
to send messages

 ⚠ PublishKafka_2_6	PublishKafka_2_6 1.15.3	
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

Required field

Property	Value
Kafka Brokers	ec2-34-235-87-105.compute-1.amazonaws.com:9093
Topic Name	test
Use Transactions	true
Transactional Id Prefix	No value set
Message Demarcator	No value set
Failure Strategy	Route to Failure
Delivery Guarantee	Guarantee Replicated Delivery
Attributes to Send as Headers (Regex)	No value set
Message Header Encoding	UTF-8
Security Protocol	SASL_PLAINTEXT
SASL Mechanism	PLAIN
Kerberos Credentials Service	No value set

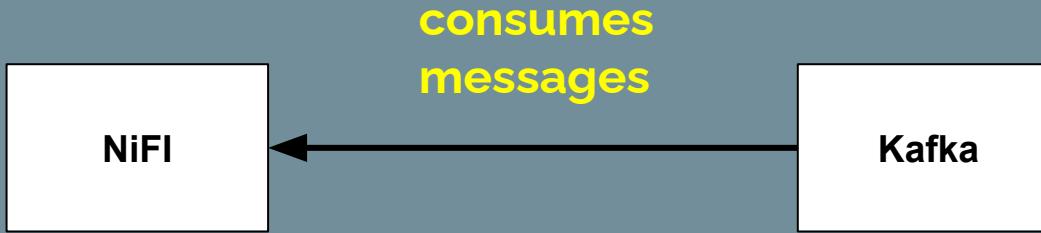
– Publishing messages to Kafka

PublishKafka_2_6 Processor
to send messages

	⚠ PublishKafka_2_6 PublishKafka_2_6 1.15.3 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

Required field	
Property	Value
Kerberos Service Name	?
Kerberos Principal	?
Kerberos Keytab	?
Username	?
Password	?
Token Auth	?
SSL Context Service	?
Kafka Key	?
Key Attribute Encoding	UTF-8 Encoded
Max Request Size	1 MB
Acknowledgment Wait Time	5 secs
Max Metadata Wait Time	5 sec

NiFi as data consumer



	⚠️ ConsumeKafka_2_6 ConsumeKafka_2_6 1.15.3 org.apache.nifi - nifi-kafka-2-6-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Consuming messages from Kafka

ConsumeKafka_2_6 Processor
to receive messages

	⚠️ ConsumeKafka_2_6 ConsumeKafka_2_6 1.15.3 org.apache.nifi - nifi-kafka-2-6-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Required field	
Property	Value
Kafka Brokers	localhost:9092
Topic Name(s)	No value set
Topic Name Format	names
Group ID	No value set
Commit Offsets	true
Max Uncommitted Time	1 secs
Honor Transactions	true
Message Demarcator	No value set
Separate By Key	false
Security Protocol	PLAINTEXT
SASL Mechanism	GSSAPI
Kerberos Credentials Service	No value set

– Consuming messages from Kafka

ConsumeKafka_2_6 Processor
to receive messages

	⚠️ ConsumeKafka_2_6 ConsumeKafka_2_6 1.15.3 org.apache.nifi - nifi-kafka-2-6-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Required field	
Property	Value
Kerberos User Service	?
Kerberos Service Name	?
Kerberos Principal	?
Kerberos Keytab	?
Username	?
Password	?
Token Auth	?
SSL Context Service	?
Key Attribute Encoding	UTF-8 Encoded
Offset Reset	latest
Message Header Encoding	?
Headers to Add as Attributes (Regex)	?

– Consuming messages from Kafka

ConsumeKafka_2_6 Processor
to receive messages

 ⚠	ConsumeKafka_2_6	ConsumeKafka_2_6 1.15.3
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

Required field	
Property	Value
Kafka Brokers	ec2-34-235-87-105.compute-1.amazonaws.com:9093
Topic Name(s)	test
Topic Name Format	names
Group ID	test-group
Commit Offsets	true
Max Uncommitted Time	1 secs
Honor Transactions	true
Message Demarcator	No value set
Separate By Key	false
Security Protocol	SASL_PLAINTEXT
SASL Mechanism	PLAIN
Kerberos Credentials Service	No value set

– Consuming messages from Kafka

ConsumeKafka_2_6 Processor
to receive messages

	⚠️ ConsumeKafka_2_6 ConsumeKafka_2_6 1.15.3 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

Required field	
Property	Value
Kerberos Service Name	?
Kerberos Principal	?
Kerberos Keytab	?
Username	?
Password	?
Token Auth	?
SSL Context Service	?
Key Attribute Encoding	?
Offset Reset	?
Message Header Encoding	?
Headers to Add as Attributes (Regex)	?
Max Poll Records	?

Integration with ElasticSearch

– Integration with ElasticSearch

Elasticsearch is a highly available indexing engine for providing unstructured searching of data

NiFi provides capability to index data to elasticsearch

Processor to connect to elasticsearch → PutElasticSearchHttp

Pushing data to ElasticSearch



GenerateFlowFile
GenerateFlowFile 1.14.0
org.apache.nifi - nifi-standard-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

PutElasticsearchHttp
PutElasticsearchHttp 1.15.3
org.apache.nifi - nifi-elasticsearch-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

– Pushing data to Elasticsearch

Processor to send data
To elasticsearch

 PutElasticsearchHttp	PutElasticsearchHttp 1.15.3	org.apache.nifi - nifi-elasticsearch-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

Required field	
Property	Value
Elasticsearch URL	ⓘ No value set
SSL Context Service	ⓘ No value set
Character Set	ⓘ UTF-8
Username	ⓘ No value set
Password	ⓘ No value set
Connection Timeout	ⓘ 5 secs
Response Timeout	ⓘ 15 secs
Proxy Configuration Service	ⓘ No value set
Proxy Host	ⓘ No value set
Proxy Port	ⓘ No value set
Proxy Username	ⓘ No value set
Proxy Password	ⓘ No value set

– Pushing data to Elasticsearch

Processor to send data
To elasticsearch

	PutElasticsearchHttp	PutElasticsearchHttp 1.15.3
		org.apache.nifi - nifi-elasticsearch-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

Required field	
Property	Value
Connection Timeout	5 secs
Response Timeout	15 secs
Proxy Configuration Service	?
Proxy Host	?
Proxy Port	?
Proxy Username	?
Proxy Password	?
Identifier Attribute	?
Index	No value set
Type	?
Batch Size	?
Index Operation	?

– Pushing data to Elasticsearch

Processor to send data
To elasticsearch

	PutElasticsearchHttp	PutElasticsearchHttp 1.15.3
		org.apache.nifi - nifi-elasticsearch-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

Required field	
Property	Value
Elasticsearch URL	http://ec2-54-90-91-197.compute-1.amazonaws.com:9200
SSL Context Service	No value set
Character Set	UTF-8
Username	No value set
Password	No value set
Connection Timeout	5 secs
Response Timeout	15 secs
Proxy Configuration Service	No value set
Proxy Host	No value set
Proxy Port	No value set
Proxy Username	No value set
Proxy Password	No value set

– Pushing data to Elasticsearch

Processor to send data
To elasticsearch

 A	PutElasticsearchHttp	PutElasticsearchHttp 1.15.3
		org.apache.nifi - nifi-elasticsearch-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

Required field	
Property	Value
Connection Timeout	5 secs
Response Timeout	15 secs
Proxy Configuration Service	No value set
Proxy Host	No value set
Proxy Port	No value set
Proxy Username	No value set
Proxy Password	No value set
Identifier Attribute	No value set
Index	test_dataflow
Type	_doc
Batch Size	100
Index Operation	index

Integration with AWS Services

– Integration with AWS Services

AWS Services - S3, SNS, SQS, Lambda

NiFi provides capability to integrate seamlessly with these AWS services

Processor to connect to S3 → **PutS3, FetchS3**

Processor to connect to SNS → **PutSNS**

Processor to connect to SQS → **PutSQS, GetSQS**

Processor to connect to Lambda → **PutLambda**

Pushing data to S3

Pushing data to S3



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000



	PutS3Object PutS3Object 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Pushing data to S3

Buckets (1) [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

Name	AWS Region	Access	Creation date
nifi-test--bucket	US East (N. Virginia) us-east-1	Bucket and objects not public	October 4, 2022, 13:26:30 (UTC+05:30)

[C](#) [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

– Pushing data to S3

Processor to send data
to S3

 PutS3Object	PutS3Object 1.15.3 org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Object Key	?(\${filename})
Bucket	?(No value set)
Content Type	?(No value set)
Content Disposition	?(No value set)
Cache Control	?(No value set)
Access Key ID	?(No value set)
Secret Access Key	?(No value set)
Credentials File	?(No value set)
AWS Credentials Provider service	?(No value set)
Object Tags Prefix	?(No value set)
Remove Tag Prefix	?(False)
Storage Class	?(Standard)

– Pushing data to S3

Processor to send data
to S3

	PutS3Object	PutS3Object 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Object Key	data.json
Bucket	nifi-test--bucket
Content Type	No value set
Content Disposition	No value set
Cache Control	No value set
Access Key ID	Sensitive value set
Secret Access Key	Sensitive value set
Credentials File	No value set
AWS Credentials Provider service	No value set
Object Tags Prefix	No value set
Remove Tag Prefix	False
Storage Class	Standard

Fetching data from S3

Fetching data from S3



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

	FetchS3Object FetchS3Object 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Fetching data from S3

Processor to fetch data from S3

 FetchS3Object	FetchS3Object 1.15.3	org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Bucket	?
Object Key	?
Region	?
Access Key ID	?
Secret Access Key	?
Credentials File	?
AWS Credentials Provider service	?
Communications Timeout	?
Version	?
SSL Context Service	?
Endpoint Override URL	?
Signer Override	?

Pushing data to SNS

Pushing data to SNS



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

	PutSNS PutSNS 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Pushing data to SNS

Processor to push data to SNS

 PutSNS	PutSNS 1.15.3 org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Amazon Resource Name (ARN)	?
ARN Type	?
E-mail Subject	?
Region	?
Access Key ID	?
Secret Access Key	?
Credentials File	?
AWS Credentials Provider service	?
Communications Timeout	?
Use JSON Structure	?
Character Set	?
Proxy Host	?

– Pushing data to SNS

Processor to push data to SNS

 PutSNS	PutSNS 1.15.3	org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Amazon Resource Name (ARN)	arn:aws:sns:us-east-1:263423382682:nifi-test-topic
ARN Type	Topic ARN
E-mail Subject	No value set
Region	US East (N. Virginia)
Access Key ID	Sensitive value set
Secret Access Key	Sensitive value set
Credentials File	No value set
AWS Credentials Provider service	No value set
Communications Timeout	30 secs
Use JSON Structure	false
Character Set	UTF-8
Proxy Host	No value set

Pushing data to SQS

Pushing data to SQS



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000



	PutSQS PutSQS 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Pushing data to SQS

Processor to push data to SQS

 PutSQS	PutSQS 1.15.3	org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Queue URL	?
Access Key ID	?
Secret Access Key	?
Credentials File	?
AWS Credentials Provider service	?
Region	?
Delay	?
Communications Timeout	?
Endpoint Override URL	?
Proxy Host	?
Proxy Host Port	?

– Pushing data to SQS

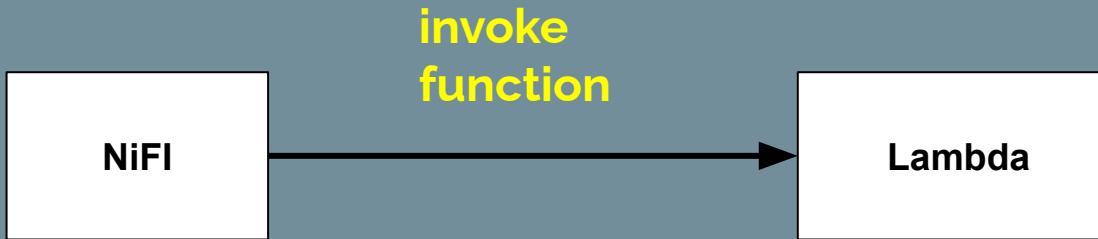
Processor to push data to SQS

 PutSQS	PutSQS 1.15.3	org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Queue URL	? https://sq.s.us-east-1.amazonaws.com/263423...
Access Key ID	? Sensitive value set
Secret Access Key	? Sensitive value set
Credentials File	? No value set
AWS Credentials Provider service	? No value set
Region	? US East (N. Virginia)
Delay	? 0 secs
Communications Timeout	? 30 secs
Endpoint Override URL	? No value set
Proxy Host	? No value set
Proxy Host Port	? No value set

Invoking AWS Lambda

Invoking AWS Lambda



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

	PutLambda PutLambda 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Invoking AWS Lambda

Processor to invoke aws
lambda

 A	PutLambda	PutLambda 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Amazon Lambda Name	?
Amazon Lambda Qualifier (version)	?
Region	?
Access Key ID	?
Secret Access Key	?
Credentials File	?
AWS Credentials Provider service	?
Communications Timeout	30 secs
Proxy Configuration Service	?
Proxy Host	?
Proxy Host Port	?
Proxy Username	?

– Invoking AWS Lambda

Processor to invoke aws
lambda

 PutLambda	PutLambda 1.15.3 org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Amazon Lambda Name	arn:aws:lambda:us-east-1:749039393955:function:lambda-test
Amazon Lambda Qualifier (version)	\$LATEST
Region	US East (N. Virginia)
Access Key ID	?
Secret Access Key	?
Credentials File	?
AWS Credentials Provider service	?
Communications Timeout	?
Proxy Configuration Service	?
Proxy Host	?
Proxy Host Port	?
Proxy Username	?

– Invoking AWS Lambda

FlowFile

DETAILS

ATTRIBUTES

Attribute Values

aws.lambda.result.log

START RequestId: 98b767b6-e152-496a-8ea3-0ad3f618cb20 Version: \$LATEST 2022-10-05T08:45:30.633Z 98b767b6-e152-496a-8ea3-0ad3f618cb20 INFO value1 = value1 2022-10-05T08:45:30.652Z 98b767b6-e152-496a-8ea3-0ad3f618cb20 INFO value2 = value2 2022-10-05T08:45:30.652Z 98b767b6-e152-496a-8ea3-0ad3f618cb20 INFO value3 = value3 END RequestId: 98b767b6-e152-496a-8ea3-0ad3f618cb20 REPORT RequestId: 98b767b6-e152-496a-8ea3-0ad3f618cb20 Duration: 60.95 ms Billed Duration: 61 ms Memory Size: 128 MB Max Memory Used: 57 MB

aws.lambda.result.payload

"value1 value2 value3"

aws.lambda.result.status.code

200

filename

28b2432a-b44f-45b1-8025-1cf40b2053ad

path

./

.....

Integration with NoSQL Database

– Integration with NoSQL Database

We will be using MongoDB and AWS DynamoDB as NoSQL databases

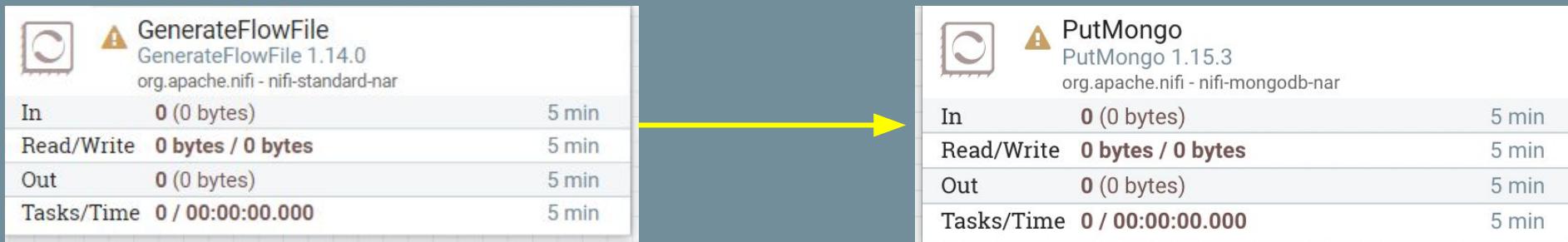
NiFi provides capability to store data to MongoDB and AWS DynamoDB

Processors to connect to MongoDB
→ GetMongo, PutMongo, DeleteMongo

Processors to connect to AWS DynamoDB
→ GetDynamoDB, PutDynamoDB

Inserting Data to MongoDB

Inserting data to MongoDB



– Inserting data to MongoDB

Processor to insert data to mongodb



PutMongo

PutMongo 1.15.3

org.apache.nifi - nifi-mongodb-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

Required field	
Property	Value
Client Service	?
Mongo URI	?
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
Mode	?
Upsert	?
Update Query Key	?
Update Query	?
Update Mode	?
Write Concern	?

– Inserting data to MongoDB

Processor to insert data to mongodb



PutMongo

PutMongo 1.15.3

org.apache.nifi - nifi-mongodb-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

Required field	
Property	Value
Client Service	?
Mongo URI	mongodb://nifi:password@ec2-34-234-88-176.compute-1.amazonaws.com:27017/test/nifi
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
Mode	?
Upsert	?
Update Query Key	?

```
1 mongodb://nifi:password@ec2-34-234-88-176.compute-1.amazonaws.com:27017
```

Updating Data in MongoDB

Updating data in MongoDB

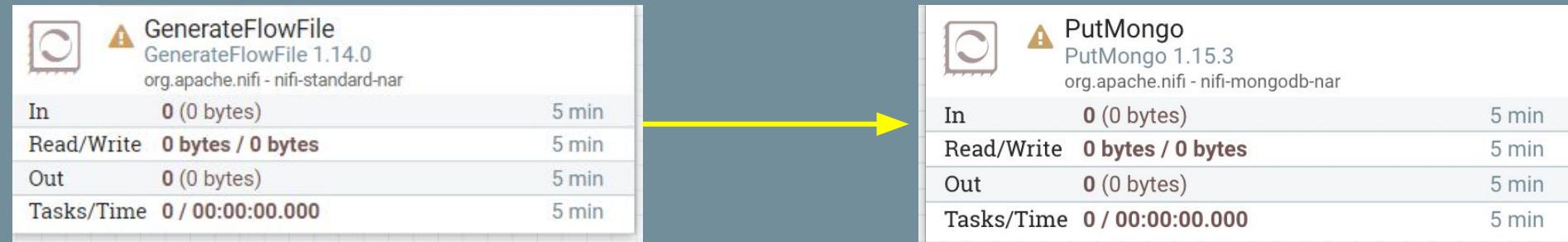
Source Json

```
{  
  "id": 6522,  
  "ident": "00A",  
  "type": "heliport",  
  "name": "Total Rf Heliport",  
  "latitude_deg": 40.0708,  
  "longitude_deg": -74.93360137939453,  
  "elevation_ft": 11,  
  "continent": "NA",  
  "iso_country": "US",  
  "iso_region": "US-PA",  
  "municipality": "Bensalem",  
  "scheduled_service": "yes",  
  "gps_code": "00A",  
  "iata_code": null,  
  "local_code": "00A",  
  "home_link": null,  
  "wikipedia_link": null,  
  "keywords": null  
}
```

Target Json

```
{  
  "id": 6522,  
  "ident": "00A",  
  "type": "small_airport",  
  "name": "Total Rf Heliport",  
  "latitude_deg": 40.0708,  
  "longitude_deg": -74.93360137939453,  
  "elevation_ft": 11,  
  "continent": "NA",  
  "iso_country": "US",  
  "iso_region": "US-PA",  
  "municipality": "Bensalem",  
  "scheduled_service": "yes",  
  "gps_code": "00A",  
  "iata_code": null,  
  "local_code": "00A",  
  "home_link": null,  
  "wikipedia_link": null,  
  "keywords": null  
}
```

Updating data in MongoDB



– Updating data in MongoDB

Processor to update data to mongodb



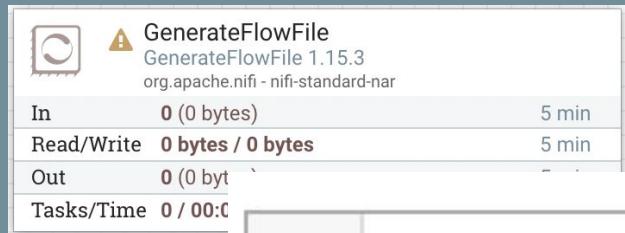
PutMongo
PutMongo 1.15.3
org.apache.nifi - nifi-mongodb-nar

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

Required field	
Property	Value
Client Service	?
Mongo URI	?
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
Mode	?
Upsert	?
Update Query Key	?
Update Query	?
Update Mode	?
Write Concern	?

– Updating data in MongoDB

GenerateFlowFile processor
to update the field



Required field	
Property	Value
File Size	0B
Batch Size	1
Data Format	Text
Unique FlowFiles	false
Custom Text	{ \$set: { "type" : "small_airport" } }

```
1 {  
2   $set: { "type" : "small_airport" }  
3 }
```

Updating data in MongoDB

Source Json

```
{  
  "id": 6522,  
  "ident": "00A",  
  "type": "heliport",  
  "name": "Total Rf Heliport",  
  "latitude_deg": 40.0708,  
  "longitude_deg": -74.93360137939453,  
  "elevation_ft": 11,  
  "continent": "NA",  
  "iso_country": "US",  
  "iso_region": "US-PA",  
  "municipality": "Bensalem",  
  "scheduled_service": "yes",  
  "gps_code": "00A",  
  "iata_code": null,  
  "local_code": "00A",  
  "home_link": null,  
  "wikipedia_link": null,  
  "keywords": null  
}
```

Target Json

```
{  
  "id": 6522,  
  "ident": "00A",  
  "type": "small_airport",  
  "name": "Total Rf Heliport",  
  "latitude_deg": 40.0708,  
  "longitude_deg": -74.93360137939453,  
  "elevation_ft": 11,  
  "continent": "NA",  
  "iso_country": "US",  
  "iso_region": "US-PA",  
  "municipality": "Bensalem",  
  "scheduled_service": "yes",  
  "gps_code": "00A",  
  "iata_code": null,  
  "local_code": "00A",  
  "home_link": null,  
  "wikipedia_link": null,  
  "keywords": null  
}
```

Querying Data from MongoDB

Getting data from MongoDB



– Querying data from MongoDB

Processor to query data from mongodb

	GetMongo	GetMongo 1.15.3
		org.apache.nifi - nifi-mongodb-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

Required field	
Property	Value
Client Service	?
Mongo URI	?
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
JSON Type	?
Pretty Print Results JSON	?
Character Set	?
Query	?
Query Output Attribute	?
Projection	?

– Querying data from MongoDB

Processor to query data from mongodb

	GetMongo	GetMongo 1.15.3
		org.apache.nifi - nifi-mongodb-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

Required field	
Property	Value
Client Service	?
Mongo URI	?
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
JSON Type	?
Pretty Print Results JSON	?
Character Set	?
Query	?
Query Output Attribute	?
Projection	?

The table shows configuration properties for the GetMongo processor. The 'Mongo URI' (value: 'mongodb://nifi:password@ec2-34-234-88-176.c...'), 'Mongo Database Name' (value: 'test'), 'Mongo Collection Name' (value: 'nifi'), and 'Query' (value: '{ "id" : 6522 }') fields are highlighted with a green border.

Deleting Data from MongoDB

Deleting data from MongoDB



A screenshot of the NiFi interface showing the configuration for a "GenerateFlowFile" processor. The processor icon is a blue square with a white "G". The name is "GenerateFlowFile", version is "GenerateFlowFile 1.14.0", and the group is "org.apache.nifi - nifi-standard-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

A screenshot of the NiFi interface showing the configuration for a "DeleteMongo" processor. The processor icon is a blue square with a white "D". The name is "DeleteMongo", version is "DeleteMongo 1.15.3", and the group is "org.apache.nifi - nifi-mongodb-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

Deleting data from MongoDB

Processor to delete data to mongodb

	DeleteMongo DeleteMongo 1.15.3 org.apache.nifi - nifi-mongodb-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

Required field	
Property	Value
Client Service	?
Mongo URI	?
Mongo Database Name	?
Mongo Collection Name	?
SSL Context Service	?
Client Auth	?
Delete Mode	?
Fail When Nothing Is Deleted	?
Write Concern	?

The table shows configuration properties for a DeleteMongo processor. The properties listed are: Client Service (No value set), Mongo URI (mongodb://nifi:password@ec2-34-234-88-176.co..), Mongo Database Name (test), Mongo Collection Name (nifi), SSL Context Service (No value set), Client Auth (REQUIRED), Delete Mode (Delete One), Fail When Nothing Is Deleted (True), and Write Concern (ACKNOWLEDGED). The Mongo URI, Mongo Database Name, Mongo Collection Name, Delete Mode, and Write Concern fields are highlighted with a green border.

Integration with AWS DynamoDB

Inserting data to DynamoDB



A screenshot of the NiFi interface showing the configuration for a "GenerateFlowFile" processor. The processor icon is a blue square with a white "G". The title is "GenerateFlowFile".
Details:
- Version: GenerateFlowFile 1.14.0
- Group: org.apache.nifi - nifi-standard-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

A screenshot of the NiFi interface showing the configuration for a "PutDynamoDB" processor. The processor icon is a blue square with a white "P". The title is "PutDynamoDB".
Details:
- Version: PutDynamoDB 1.15.3
- Group: org.apache.nifi - nifi-aws-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

– Inserting data to DynamoDB

Processor to generate sample data

	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Required field	
Property	Value
File Size	0B
Batch Size	1
Data Format	Text
Unique FlowFiles	false
Custom Text	{ "id" : 6522, "ident" : "00A", "type" : "heliport", "name" : "Total ...
Character Set	UTF-8
Mime Type	No value set
dynamodb.item.hash.key.value	6522

– Inserting data to DynamoDB

DynamoDB > Items: nifi-test > Edit item

Form

JSON view

Edit item

You can add, remove, or edit the attributes of an item. You can nest attributes inside other attributes up to 32 levels deep. [Learn more](#)

Add new attribute ▾

Attributes

+ Attribute name

Value

Type

id - Partition key

6522

New

String

+ data

Insert a field ▾

Map

Remove

Cancel

Save changes

– Inserting data to DynamoDB

Processor to insert data to dynamodb

	PutDynamoDB	PutDynamoDB 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Table Name	?
Hash Key Name	?
Range Key Name	?
Hash Key Value	?
Range Key Value	?
Hash Key Value Type	?
Range Key Value Type	?
Json Document attribute	?
Character set of document	?
Batch items for each request (between 1 and 50)	?
Region	?

– Inserting data to DynamoDB

Processor to insert data to dynamodb

	PutDynamoDB	PutDynamoDB 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Region	US West (Oregon)
Access Key ID	No value set
Secret Access Key	No value set
Credentials File	No value set
AWS Credentials Provider service	No value set
Communications Timeout	30 secs
SSL Context Service	No value set
Proxy Configuration Service	No value set
Proxy Host	No value set
Proxy Host Port	No value set
Proxy Username	No value set
Proxy Password	No value set

– Inserting data to DynamoDB

Processor to insert data to dynamodb

	PutDynamoDB	PutDynamoDB 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Table Name	?
Table Name	nifi-test
Hash Key Name	?
Hash Key Name	id
Range Key Name	?
Range Key Name	No value set
Hash Key Value	?
Hash Key Value	`\${dynamodb.item.hash.key.value}`
Range Key Value	?
Range Key Value	`\${dynamodb.item.range.key.value}`
Hash Key Value Type	?
Hash Key Value Type	string
Range Key Value Type	?
Range Key Value Type	string
Json Document attribute	?
Json Document attribute	data
Character set of document	?
Character set of document	UTF-8
Batch items for each request (between 1 and 50)	?
Batch items for each request (between 1 and 50)	1
Region	?
Region	US East (N. Virginia)

– Inserting data to DynamoDB

Processor to insert data to dynamodb

	PutDynamoDB	PutDynamoDB 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Region	US East (N. Virginia)
Access Key ID	Sensitive value set
Secret Access Key	Sensitive value set
Credentials File	No value set
AWS Credentials Provider service	No value set
Communications Timeout	30 secs
SSL Context Service	No value set
Proxy Configuration Service	No value set
Proxy Host	No value set
Proxy Host Port	No value set
Proxy Username	No value set
Proxy Password	No value set

Querying Data from AWS DynamoDB

Querying data from DynamoDB



	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

	GetDynamoDB GetDynamoDB 1.15.3 org.apache.nifi - nifi-aws-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

– Querying data from DynamoDB

Processor to generate primary key to query

	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Required field	
Property	Value
File Size	0B
Batch Size	1
Data Format	Text
Unique FlowFiles	false
Custom Text	No value set
Character Set	UTF-8
Mime Type	No value set
dynamodb.item.hash.key.value	6522

– Querying data from DynamoDB

Processor to query data from dynamodb

	GetDynamoDB	GetDynamoDB 1.15.3
		org.apache.nifi - nifi-aws-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

Required field	
Property	Value
Table Name	?
nifi-test	
Hash Key Name	?
id	
Range Key Name	?
No value set	
Hash Key Value	?
\${dynamodb.item.hash.key.value}	
Range Key Value	?
\${dynamodb.item.range.key.value}	
Hash Key Value Type	?
string	
Range Key Value Type	?
string	
Json Document attribute	?
data	
Batch items for each request (between 1 and 50)	?
1	
Region	?
US East (N. Virginia)	
Access Key ID	?
Sensitive value set	
Secret Access Key	?
Sensitive value set	

NiFi Record Processing

NiFi Record Processing

Problem

Specialized Processors to handle different types

SplitText, SplitJson, SplitAvro,

EvaluateJsonPath, ExtractText,
RouteOnContent, etc

ConvertAvroToJson, ConvertJsonToAvro

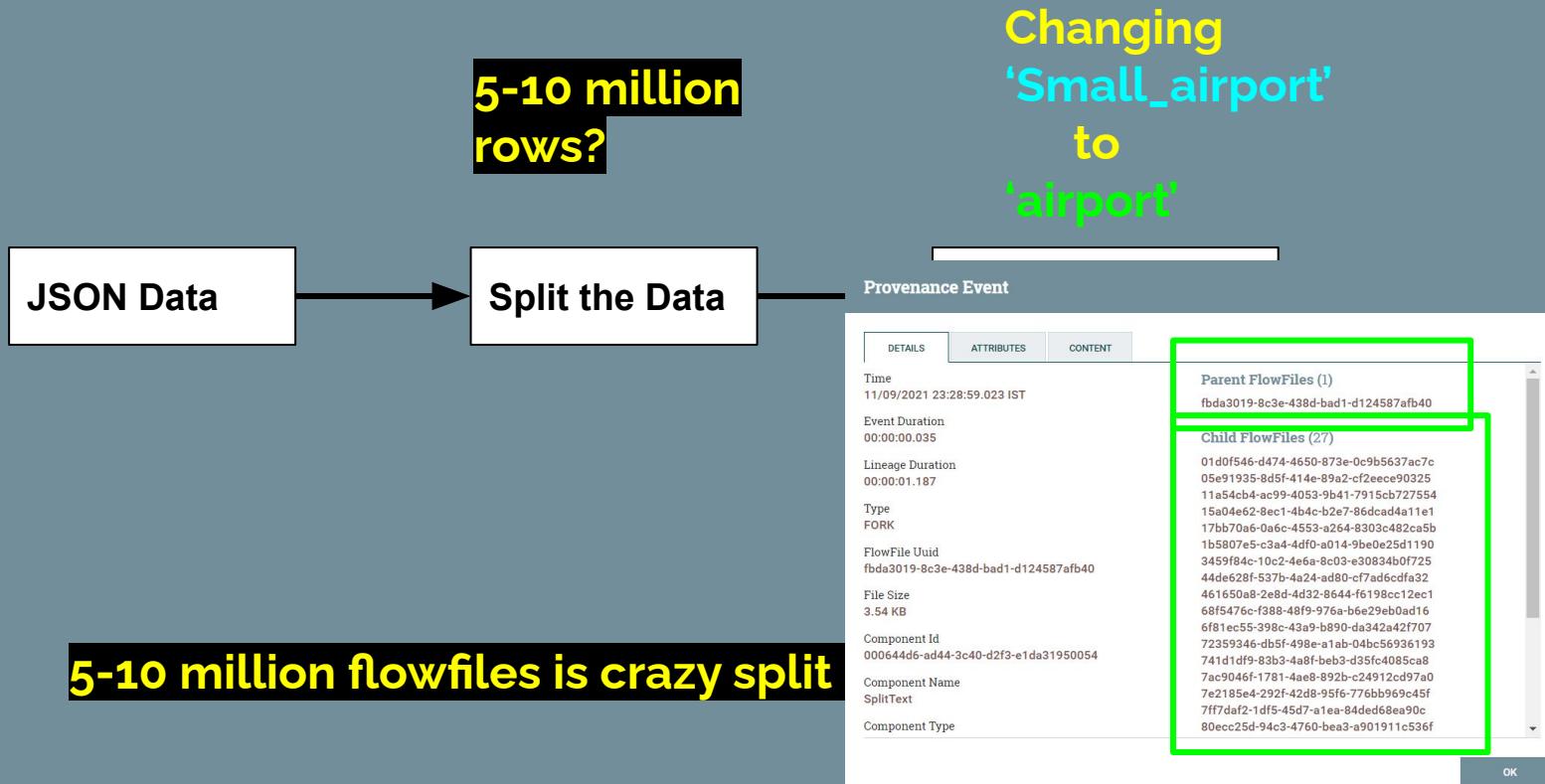
Sometimes missing conversions

No ConvertCSVToJson, ConvertXmlToJson, etc

Sometimes missing specific function for a data type

No EvaluateAvroPath, we have to use ConvertAvroToJson and then EvaluateJsonPath

NiFi Record Processing



Splitting, Transformation, Aggregating (JSON)

Data Generator

	■ GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

5-10 million rows?

Data Splitting

	■ SplitJSON SplitJSON 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Merge Data

	■ MergeContent MergeContent 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

5-10 million splits

Modify Data

	■ ReplaceText ReplaceText 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000



NiFi Record Processing

Solution

Introduce concept of “Record”

With Records, processors can accept any data formats without having to worrying about parsing and serialization logic

Big advantage is to keep the FlowFile larger, each with multiple records which results in far better performance

NiFi Record Processing

```
[  
  {  
    "id" : 6523,  
    "ident" : "00A",  
    "type" : "heliport",  
    "name" : "Total Rf Heliport",  
    "iso_region" : "Pennsylvania",  
    "country_name" : "United States"  
  },  
  
  {  
    "id" : 6524,  
    "ident" : "00B",  
    "type" : "small_airport",  
    "name" : "Total Rf Heliport",  
    "iso_region" : "Pennsylvania",  
    "country_name" : "United States"  
  },  
  
  {  
    "id" : 6525,  
    "ident" : "00C",  
    "type" : "medium_airport",  
    "name" : "Bay Area Airport",  
    "iso_region" : "Chicago",  
    "country_name" : "United States"  
  }]
```

→ Record

→ Record

→ Record

Processors based on Records

SplitRecord

ConvertRecord

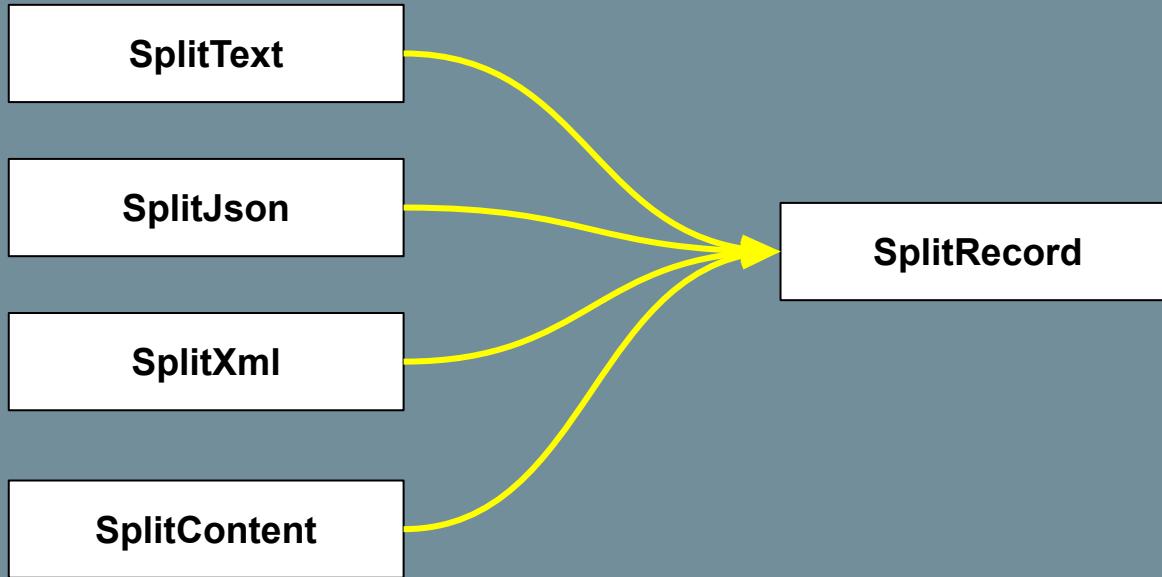
UpdateRecord

MergeRecord

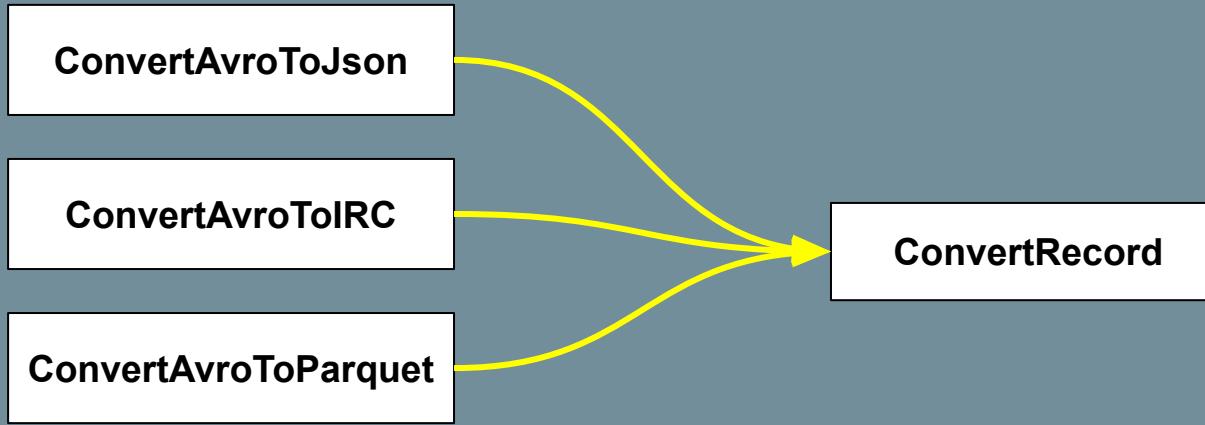
PartitionRecord

QueryRecord

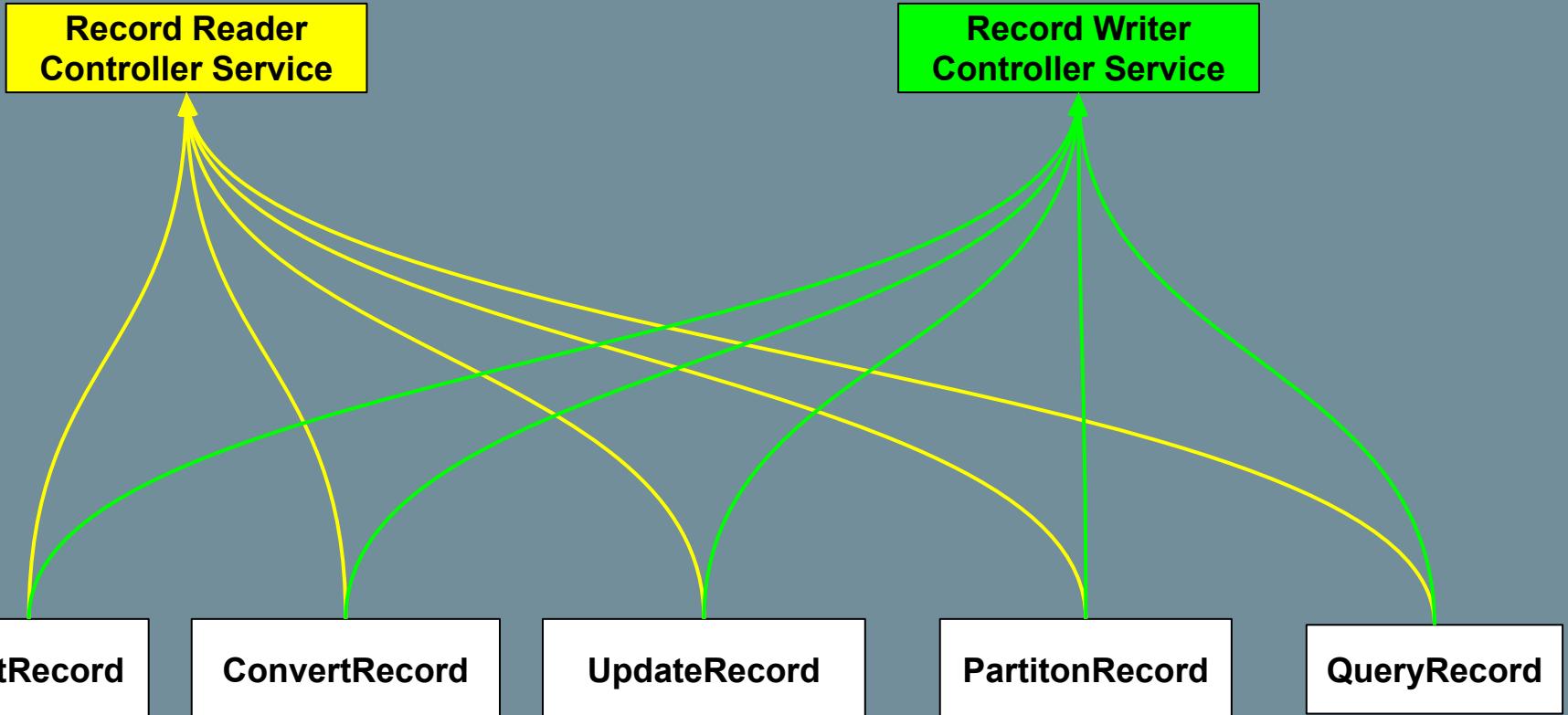
NiFi Record Processing



NiFi Record Processing



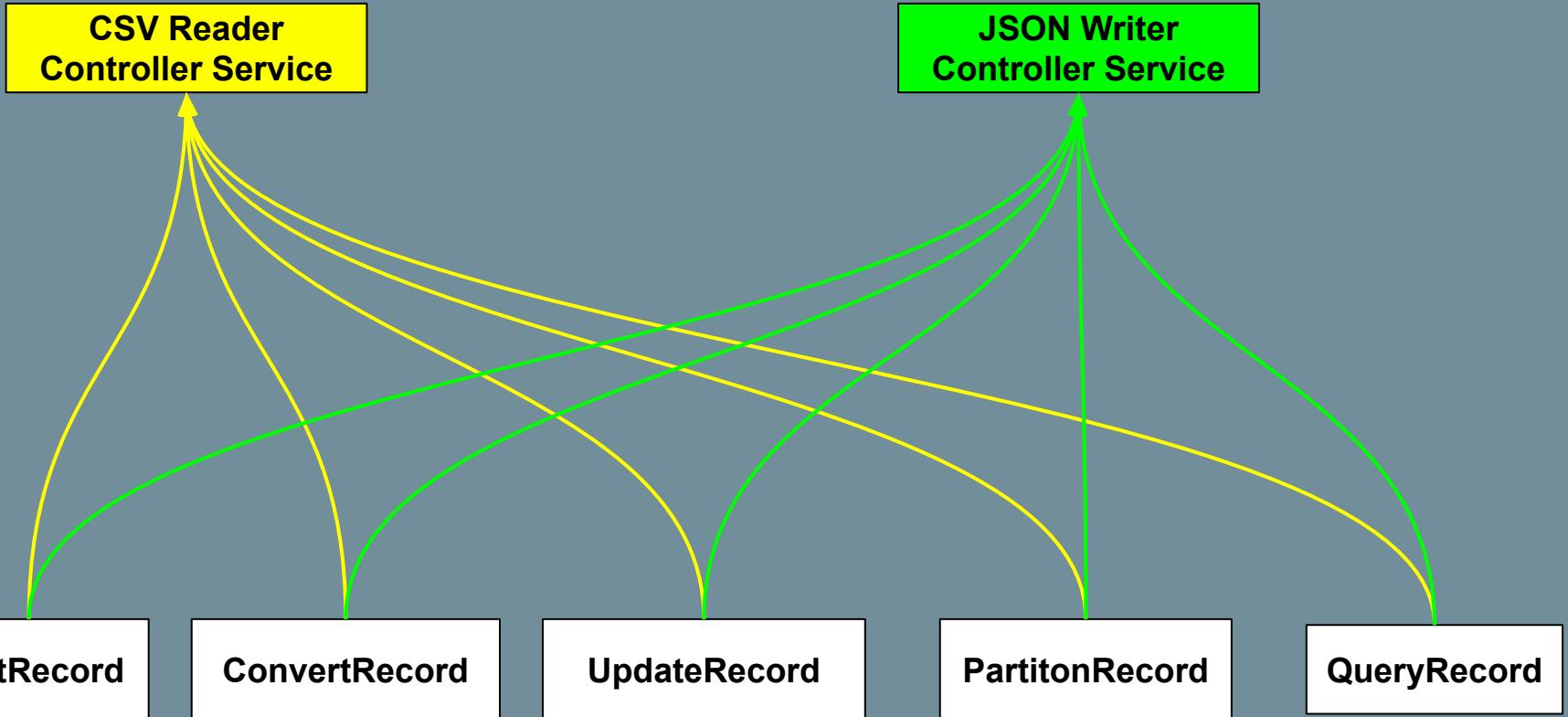
NiFi Record Processing



NiFi Record Processing

Record Readers	Record Writers
CSVReader	CSVRecordSetWriter
JsonPathReader JsonTreeReader	JsonRecordSetWriter
XMLReader	XMLRecordSetWriter
ParquetReader	ParquetRecordSetWriter
AvroReader	AvroRecordSetWriter
	FreeFormTextRecordSetWriter
GrokReader	

NiFi Record Processing



NiFi Record Processing

-

Add Controller Service

Record Reader Controller Service

Source	Type	Version	Tags
cache	AvroReader	1.14.0	comma, reader, record, values, ...
connection	CSVReader	1.14.0	comma, reader, csv, record, val...
credentials	GrokReader	1.14.0	logstash, regex, reader, grok, lo...
distributed	JsonPathReader	1.14.0	parser, reader, record, jsonpath...
enrich	JsonTreeReader	1.14.0	parser, reader, record, tree, json
invoke	ParquetReader	1.14.0	reader, record, parse, row, parq...
javascript	ReaderLookup	1.14.0	lookup, reader, record, parse, r...
jruby	ScriptedReader	1.14.0	luaj, python, jython, js, invoke, ...
js	Syslog5424Reader	1.14.0	reader, logfiles, record, syslog ...
jython	SyslogReader	1.14.0	reader, logfiles, record, syslog, ...
key	WindowsEventLogReader	1.14.0	parser, log, xml, reader, record, ...

AvroReader 1.14.0 org.apache.nifi - nifi-record-serialization-services-nar

Parses Avro data and returns each Avro record as an separate Record object. The Avro data may contain the schema itself, or the schema can be externalized and accessed by one of the methods offered by the 'Schema Access Strategy' property.

CANCEL ADD

NiFi Record Processing

-

Add Controller Service

Source Displaying 8 of 100

writer|

Type	Version	Tags
AvroRecordSetWriter	1.14.0	result, set, record, serializer, rec...
CSVRecordSetWriter	1.14.0	result, set, tab, tsv, csv, record, ...
FreeFormTextRecordSetWriter	1.14.0	serialize, expression, freeform, ...
JsonRecordSetWriter	1.14.0	serialize, record, json, resultset, ...
ParquetRecordSetWriter	1.14.0	result, set, record, serializer, rec...
RecordSetWriterLookup	1.14.0	result, lookup, set, record, serial...
ScriptedRecordSetWriter	1.14.0	luaj, python, jython, js, invoke, s...
XMLRecordSetWriter	1.14.0	serialize, xml, record, resultset, ...

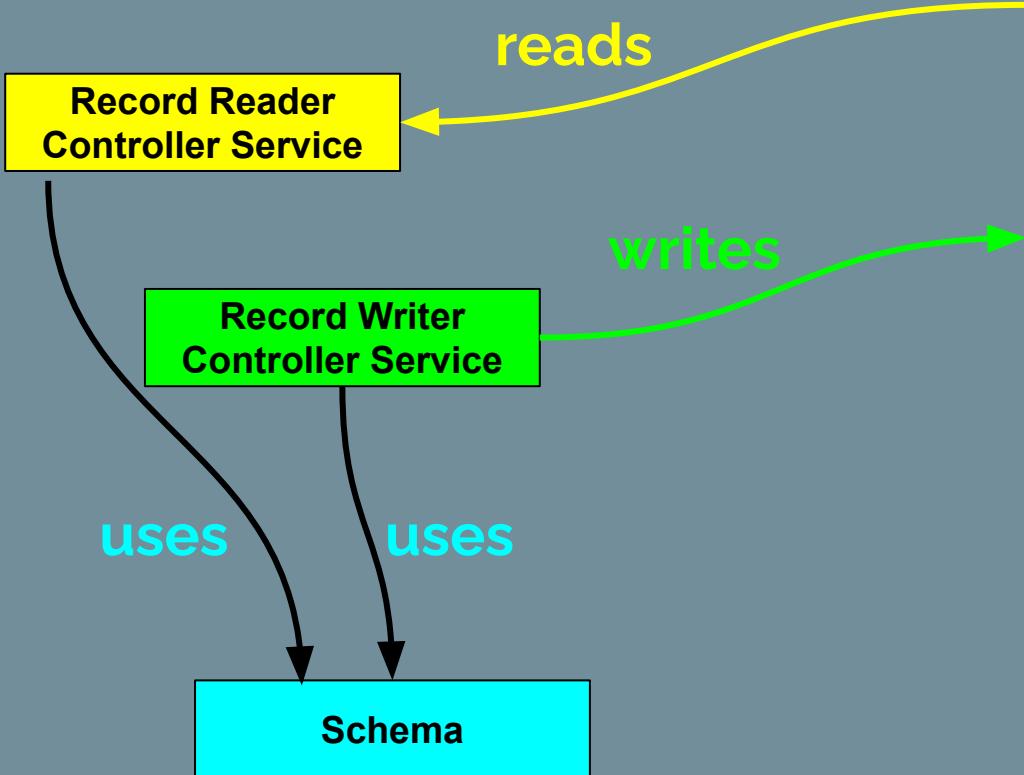
Record Writer Controller Service

AvroRecordSetWriter 1.14.0 org.apache.nifi - nifi-record-serialization-services-nar

Writes the contents of a RecordSet in Binary Avro format.

CANCEL ADD

NiFi Record Processing



```
[  
  {  
    "id" : 6523,  
    "ident" : "00A",  
    "type" : "heliport",  
    "name" : "Total Rf Heliport",  
    "iso_region" : "Pennsylvania",  
    "country_name" : "United States"  
  },  
  {  
    "id" : 6524,  
    "ident" : "00B",  
    "type" : "small_airport",  
    "name" : "Total Rf Heliport",  
    "iso_region" : "Pennsylvania",  
    "country_name" : "United States"  
  },  
  {  
    "id" : 6525,  
    "ident" : "00C",  
    "type" : "medium_airport",  
    "name" : "Bay Area Airport",  
    "iso_region" : "Chicago",  
    "country_name" : "United States"  
  }]  
]
```

NiFi Record Processing

Schema (**Avro Schema Definition**)

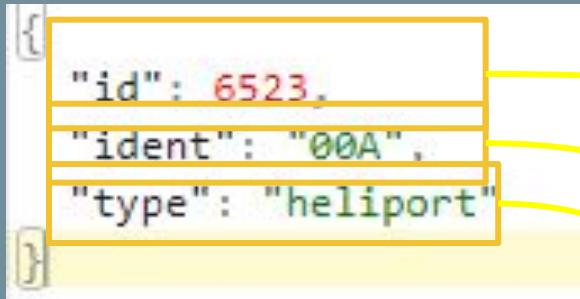
To interpret data as Record, Readers/Writers rely on Schema (**Avro Schema Definition**)

Avro schema is written in **JSON Format**

Avro schema is used to serialize data into compact **binary format** which can be deserialized easily

NiFi Record Processing

Json Record



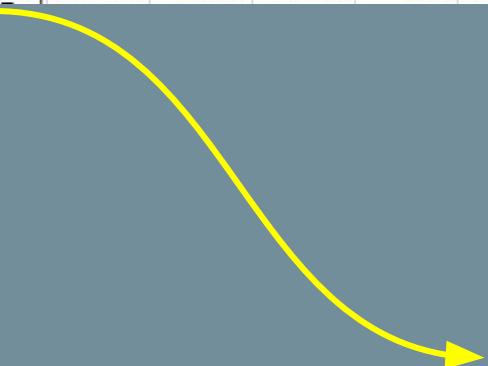
Schema

```
postgres=# CREATE TABLE EMPLOYEE(
    ID INT PRIMARY KEY      NOT NULL,
    NAME          VARCHAR(50) NOT NULL,
    COMPANY       VARCHAR(50) NOT NULL
);
CREATE TABLE
```

NiFi Record Processing

CSV to JSON Conversion

id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	iata_code	local_code	home_link	wikipedia_link	keywords
6523	00A	heliport	Total Rf Heliport	40.0708	-74.9336	11	NA	US	US-PA	Bensalem	no	00A		00A			
323361	00AA	small_airport	Aero B Rai	38.70402	-101.474	3435	NA	US	US-KS	Leoti	no	00AA		00AA			
6524	00AK	small_airport	Lowell Field	59.94773	-151.693	450	NA	US	US-AK	Anchor Point	no	00AK		00AK			
6525	00AL	small_airport	Epps Airpark	34.8648	-86.7703	820	NA	US	US-AL	Harvest	no	00AL		00AL			
6526	00AR	closed	Newport Landing	35.6087	-91.2549	237	NA	US	US-AR	Newport	no					00AR	
322127	00AS	small_airport	Fulton Air	34.9428	-97.818	1100	NA	US	US-OK	Alex	no	00AS		00AS			



```
{  
  "id" : 6523,  
  "ident" : "00A",  
  "type" : "heliport",  
  "name" : "Total Rf Heliport",  
  "latitude_deg" : 40.0708,  
  "longitude_deg" : -74.93360137939453,  
  "elevation_ft" : 11,  
  "continent" : "NA",  
  "iso_country" : "US",  
  "iso_region" : "US-PA",  
  "municipality" : "Bensalem",  
  "scheduled_service" : "no",  
  "gps_code" : "00A",  
  "iata_code" : null,  
  "local_code" : "00A",  
  "home_link" : null,  
  "wikipedia_link" : null,  
  "keywords" : null  
}
```

NiFi Record Processing

CSV to Json convertor

	ConvertRecord	ConvertRecord 1.14.0
org.apache.nifi - nifi-standard-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

Configure Processor

Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Include Zero Record FlowFiles	true

+

CANCEL APPLY

Updating Record Field

Configure Processor

Add Controller Service

Requires Controller Service
RecordReaderFactory 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

- AvroReader 1.14.0
- AvroReader 1.14.0
- CSVReader 1.14.0**
- GrokReader 1.14.0
- JsonPathReader 1.14.0
- JsonTreeReader 1.14.0
- ParquetReader 1.14.0
- ReaderLookup 1.14.0
- ScriptedReader 1.14.0
- Syslog5424Reader 1.14.0
- SyslogReader 1.14.0

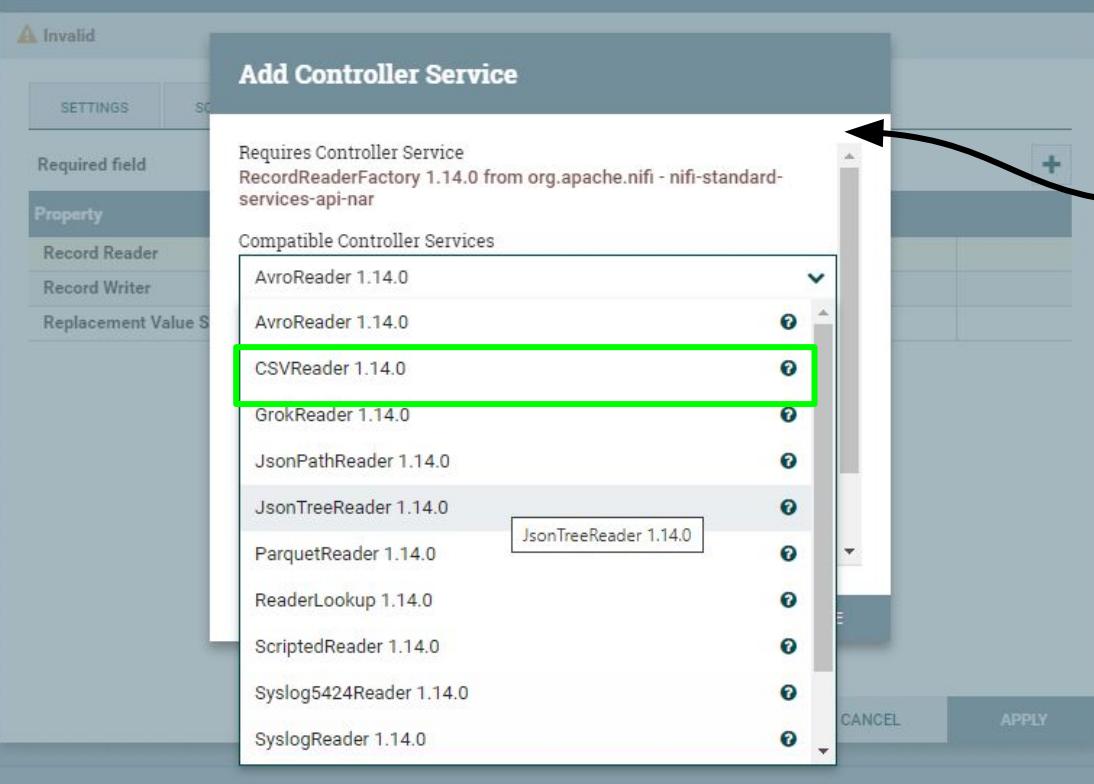
Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Replacement Value Strategy	Literal Value

CANCEL APPLY



Configure Processor

Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Replacement Value Strategy	Literal Value

CANCEL APPLY

Updating Record Field

Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property

Record Reader
Record Writer
Replacement Value Strategy

/continent

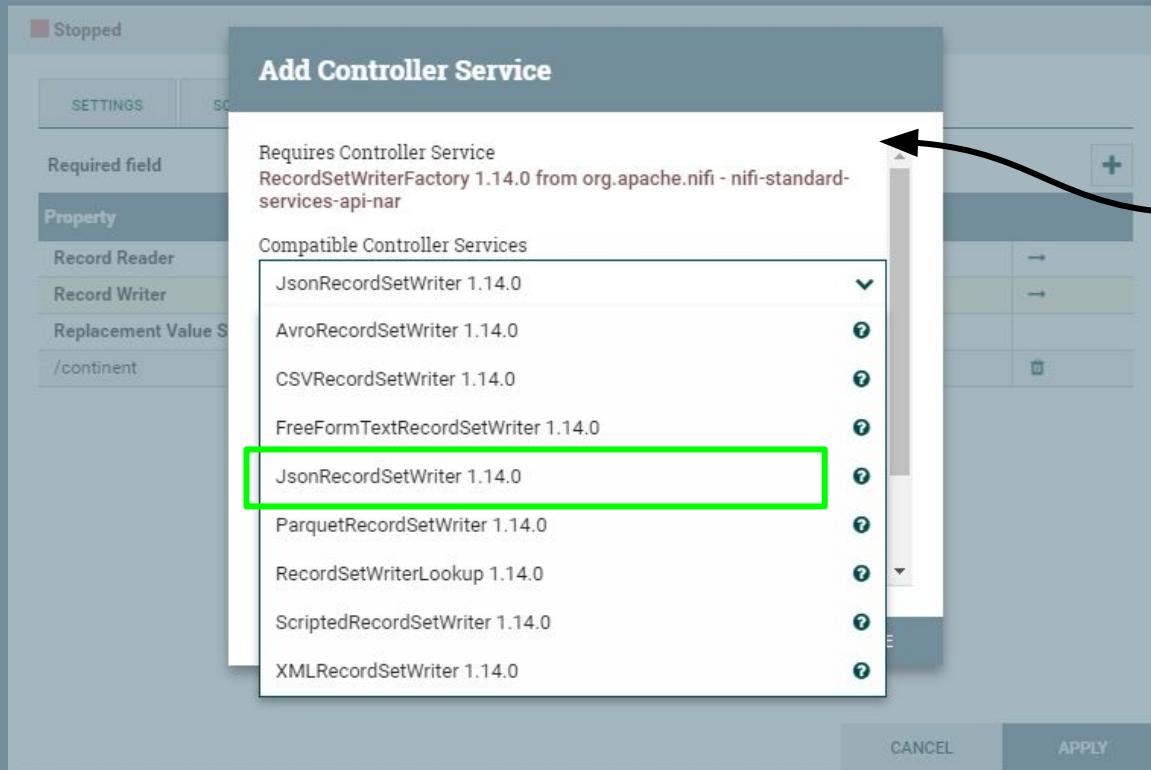
Add Controller Service

Requires Controller Service
RecordSetWriterFactory 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

- JsonRecordSetWriter 1.14.0
- AvroRecordSetWriter 1.14.0
- CSVRecordSetWriter 1.14.0
- FreeFormTextRecordSetWriter 1.14.0
- JsonRecordSetWriter 1.14.0
- ParquetRecordSetWriter 1.14.0
- RecordSetWriterLookup 1.14.0
- ScriptedRecordSetWriter 1.14.0
- XMLRecordSetWriter 1.14.0

CANCEL APPLY



Configure Processor

Invalid

NiFi Record Processing

CSV Reader
Controller
Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

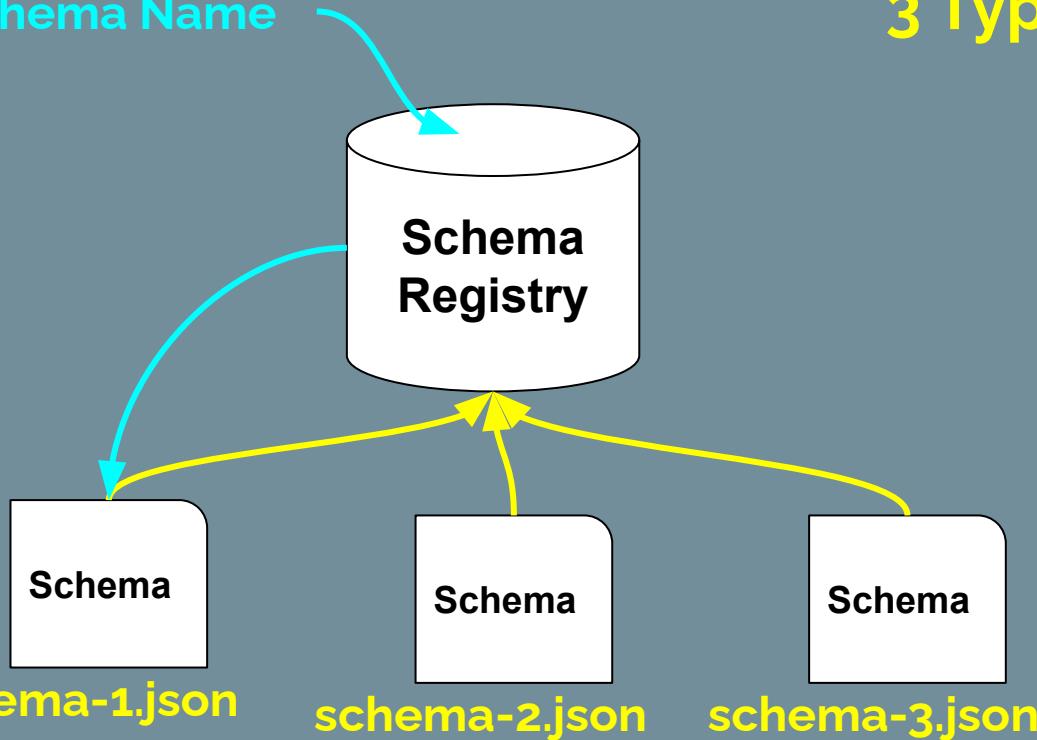
Property	Value
Schema Access Strategy	Infer Schema
CSV Parser	Apache Commons CSV
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
CSV Format	Custom Format
Value Separator	,
Record Separator	\n
Treat First Line as Header	false
Ignore CSV Header Column Names	false
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set

CANCEL APPLY

NiFi Record Processing

Schema Access Strategy

Schema Name



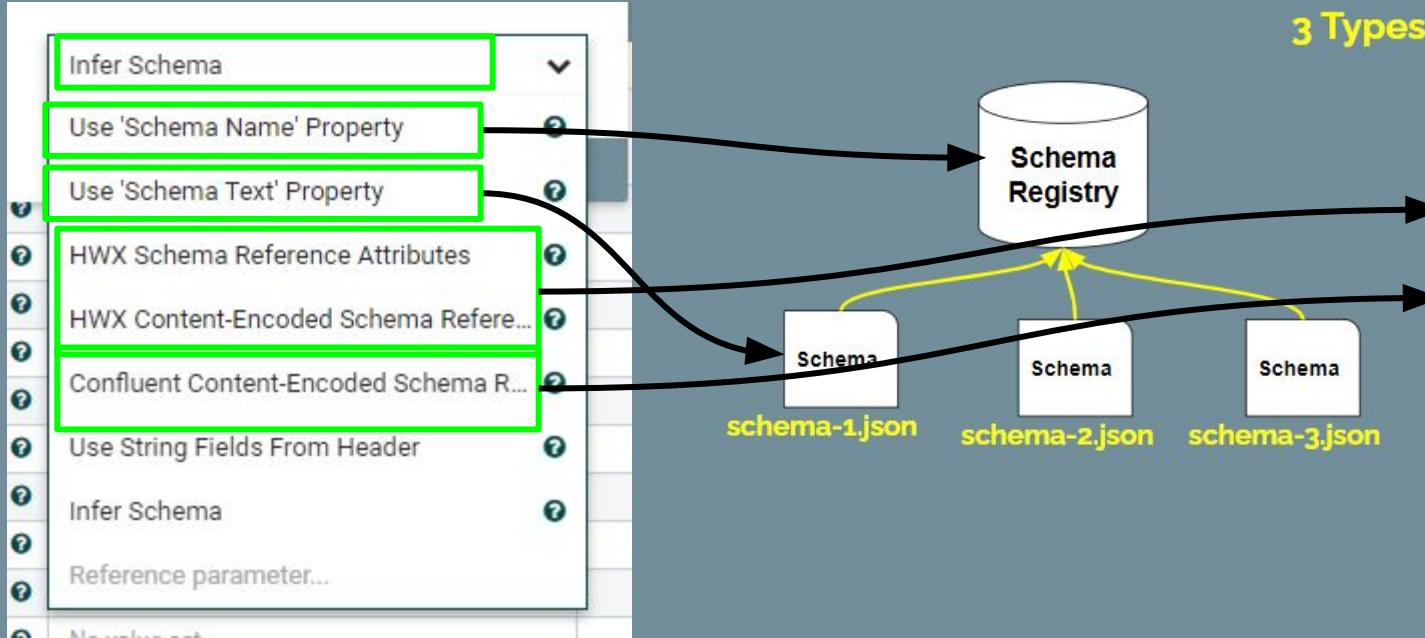
3 Types of Schema Registry

Avro
Schema Registry

Hortonworks Schema Registry

Confluent Schema Registry

NiFi Record Processing



3 Types of Schema Registry

Avro
Schema Registry

Hortonworks
Schema Registry

Confluent
Schema Registry

NiFi Record Processing

JSON Writer Controller Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

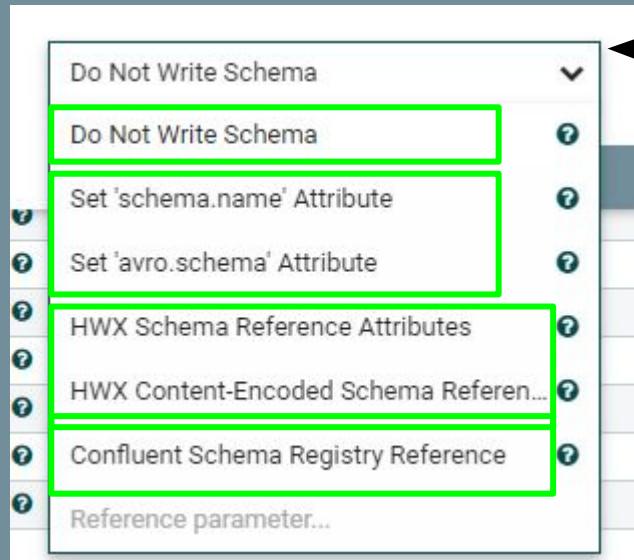
Required field

Property	Value
Schema Write Strategy	Do Not Write Schema
Schema Cache	No value set
Schema Access Strategy	Inherit Record Schema
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
Pretty Print JSON	false
Suppress Null Values	Never Suppress
Output Grouping	Array
Compression Format	none

+ CANCEL APPLY

NiFi Record Processing

Schema Write Strategy



Configure Controller Service		
SETTINGS	PROPERTIES	COMMENTS
Required field		
Property	Value	
Schema Write Strategy	<input checked="" type="radio"/> Do Not Write Schema	
Schema Cache	<input type="radio"/> No value set	
Schema Access Strategy	<input type="radio"/> Inherit Record Schema	
Date Format	<input type="radio"/> No value set	
Time Format	<input type="radio"/> No value set	
Timestamp Format	<input type="radio"/> No value set	
Pretty Print JSON	<input type="radio"/> false	
Suppress Null Values	<input type="radio"/> Never Suppress	
Output Grouping	<input type="radio"/> Array	
Compression Format	<input type="radio"/> none	

NiFi Record Processing

JSON Writer Controller Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

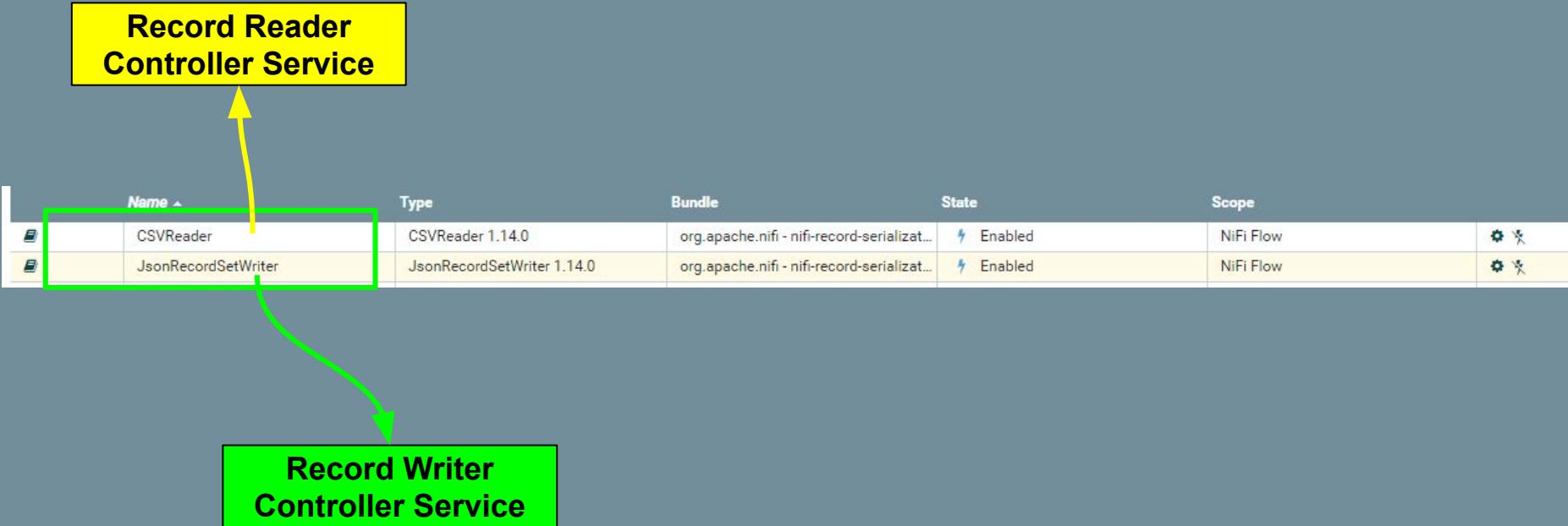
Required field

Property	Value
Schema Write Strategy	Do Not Write Schema
Schema Cache	No value set
Schema Access Strategy	Inherit Record Schema
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
Pretty Print JSON	false
Suppress Null Values	Never Suppress
Output Grouping	Array
Compression Format	none

+ CANCEL APPLY

Convert CSV to JSON

NiFi Record Processing



NiFi Record Processing

Record Reader
Controller Service

Configure Controller Service

SETTINGS

PROPERTIES

COMMENTS

Required field

Property	Value
Schema Access Strategy	Infer Schema
CSV Parser	Apache Commons CSV
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
CSV Format	Custom Format
Value Separator	,
Record Separator	\n
Treat First Line as Header	false
Ignore CSV Header Column Names	false
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set

CANCEL

APPLY

Json Record

```
{  
  "id": 6523,  
  "ident": "00A",  
  "type": "heliport"  
}
```

NiFi Record Processing

Record Writer Controller Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Write Strategy	Do Not Write Schema
Schema Cache	No value set
Schema Access Strategy	Inherit Record Schema
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
Pretty Print JSON	false
Suppress Null Values	Never Suppress
Output Grouping	Array
Compression Format	none

CANCEL APPLY

Record Reader Controller Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Access Strategy	<input checked="" type="radio"/> Infer Schema <input type="radio"/> Apache Commons CSV
CSV Parser	No value set
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
CSV Format	Custom Format
Value Separator	,
Record Separator	\n
Treat First Line as Header	false
Ignore CSV Header Column Names	false
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set

CANCEL APPLY



Converting CSV to JSON

Converting xml to JSON

	ConvertRecord	ConvertRecord 1.14.0
		org.apache.nifi - nifi-standard-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

Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Record Reader	CSVReader
Record Writer	JsonRecordSetWriter
Include Zero Record FlowFiles	true

CANCEL APPLY

Converting CSV to JSON

Generate CSV Data

	■ GenerateFlowFile	GenerateFlowFile 1.14.0	org.apache.nifi - nifi-standard-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	



Converts CSV to JSON

	⚠ ConvertRecord	ConvertRecord 1.14.0	org.apache.nifi - nifi-standard-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	



Logs JSON Data

	⚠ LogMessage	LogMessage 1.14.0	org.apache.nifi - nifi-standard-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	

– Getting Schema from Record Writer

JSON Writer
Controller
Service

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Write Strategy	Set 'avro.schema' Attribute
Schema Cache	No value set
Schema Access Strategy	Inherit Record Schema
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
Pretty Print JSON	false
Suppress Null Values	Never Suppress
Output Grouping	Array
Compression Format	none

CANCEL APPLY

Getting Schema from Record Writer

FlowFile

DETAILS ATTRIBUTES

Attribute Values

```
avro.schema
{"type": "record", "name": "nifiRecord", "namespace": "org.apache.nifi", "fields": [{"name": "id", "type": ["null", "int"]}, {"name": "ident", "type": ["null", "string"]}, {"name": "type", "type": ["null", "string"]}, {"name": "name", "type": ["null", "string"]}, {"name": "latitude_deg", "type": ["null", "float"]}, {"name": "longitude_deg", "type": ["null", "double"]}, {"name": "elevation_ft", "type": ["null", "int"]}, {"name": "continent", "type": ["null", "string"]}, {"name": "iso_country", "type": ["null", "string"]}, {"name": "iso_region", "type": ["null", "string"]}, {"name": "municipality", "type": ["null", "string"]}, {"name": "scheduled_service", "type": ["null", "string"]}, {"name": "gps_code", "type": ["null", "string"]}, {"name": "iata_code", "type": ["null", "string"]}, {"name": "local_code", "type": ["null", "string"]}, {"name": "home_link", "type": ["null", "string"]}, {"name": "wikipedia_link", "type": ["null", "string"]}, {"name": "keywords", "type": ["null", "string"]}]}
```

filename
532f8693-dc6d-4861-ba8b-544a0745623f

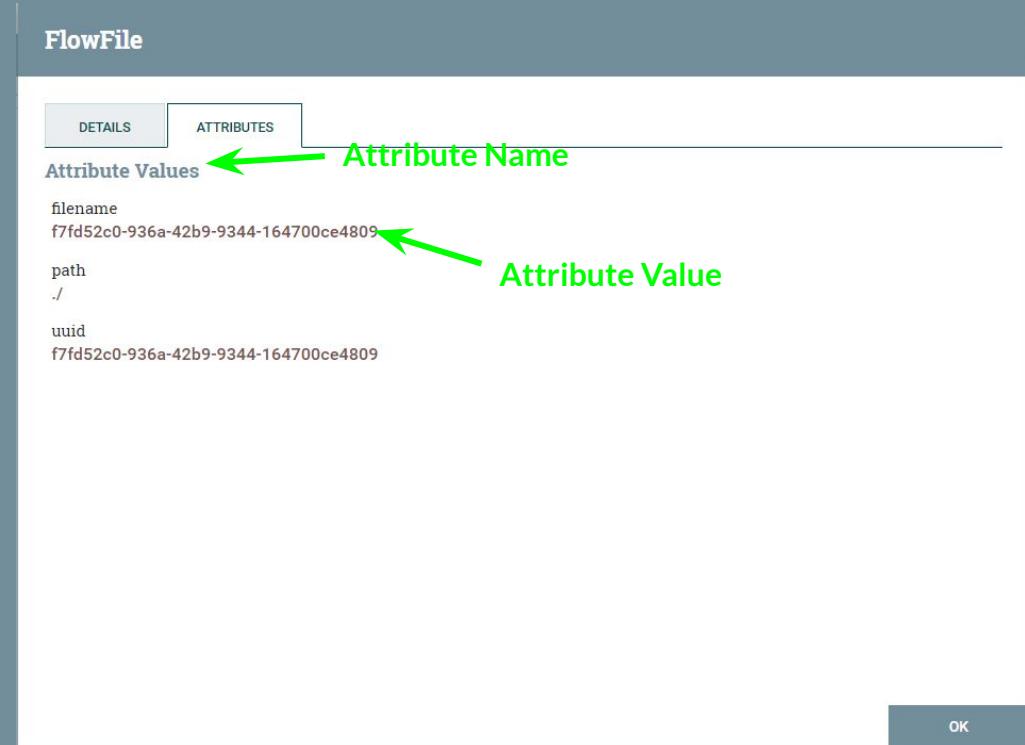
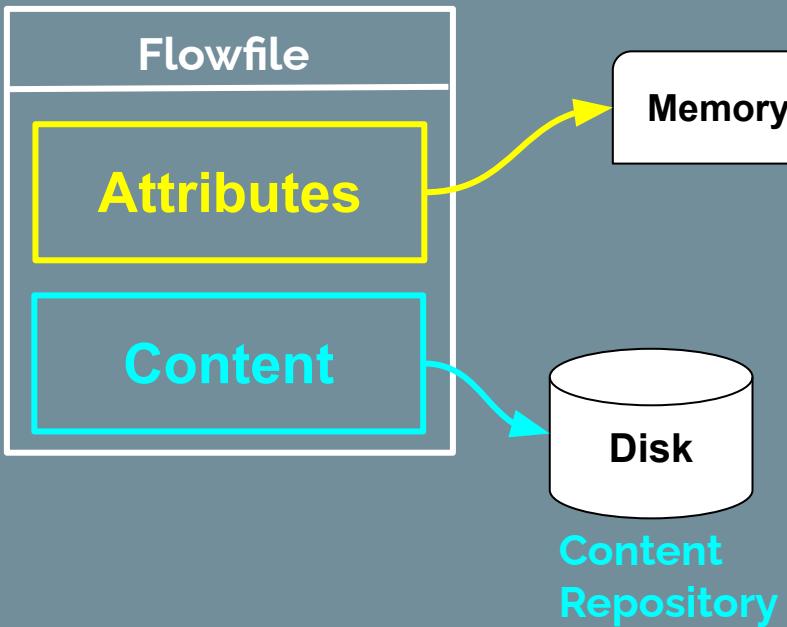
mime.type
application/json

path
./

record_count

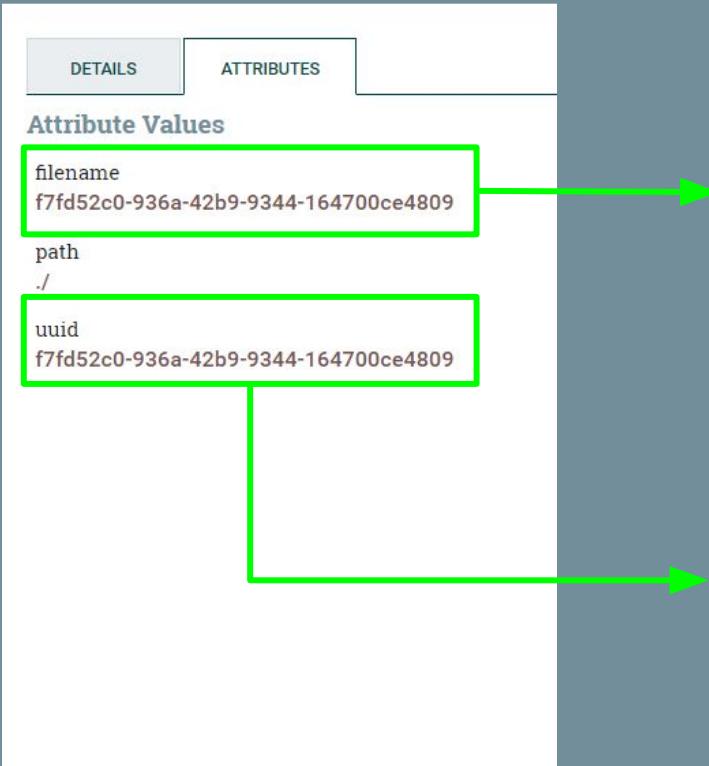
OK

Recap : NiFi Expression Language



– NiFi Expression Language

FlowFile



Attribute Name
\${ filename }
Expression Syntax '\${ [] }'

\${ uuid }

General Syntax
\${ **AttributeName** }

– NiFi Expression Language

General Syntax

`${AttributeName}`

`${ name }`



FlowFile

DETAILS	ATTRIBUTES
Attribute Values	
filename	907061b8-a642-4ed9-97f0-d3ff214cbd46
name	John
path	./
uuid	907061b8-a642-4ed9-97f0-d3ff214cbd46

– Why Expression Language

Language to access FlowFile Attributes

Provides wide variety of functions

In-memory processing, no disk access

Makes processing faster

NiFi Record Path Language

– Why Record Path Language

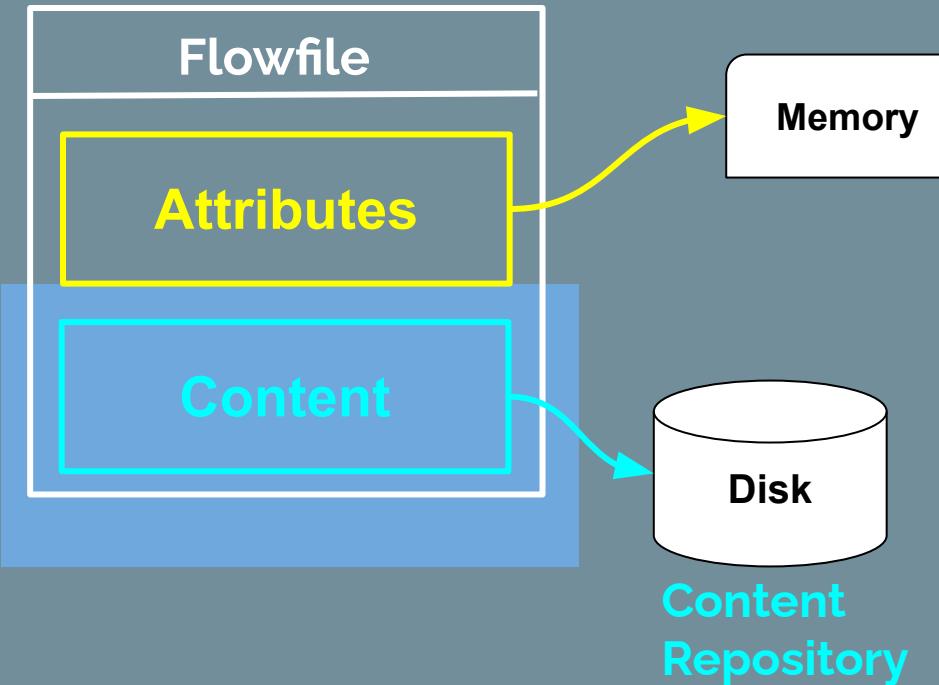
Language to access FlowFile Content

Provides wide variety of functions

Since data is read as stream of records, any data size could be handled

Makes processing faster, no more splitting and replacetext

NiFi Record Path Language

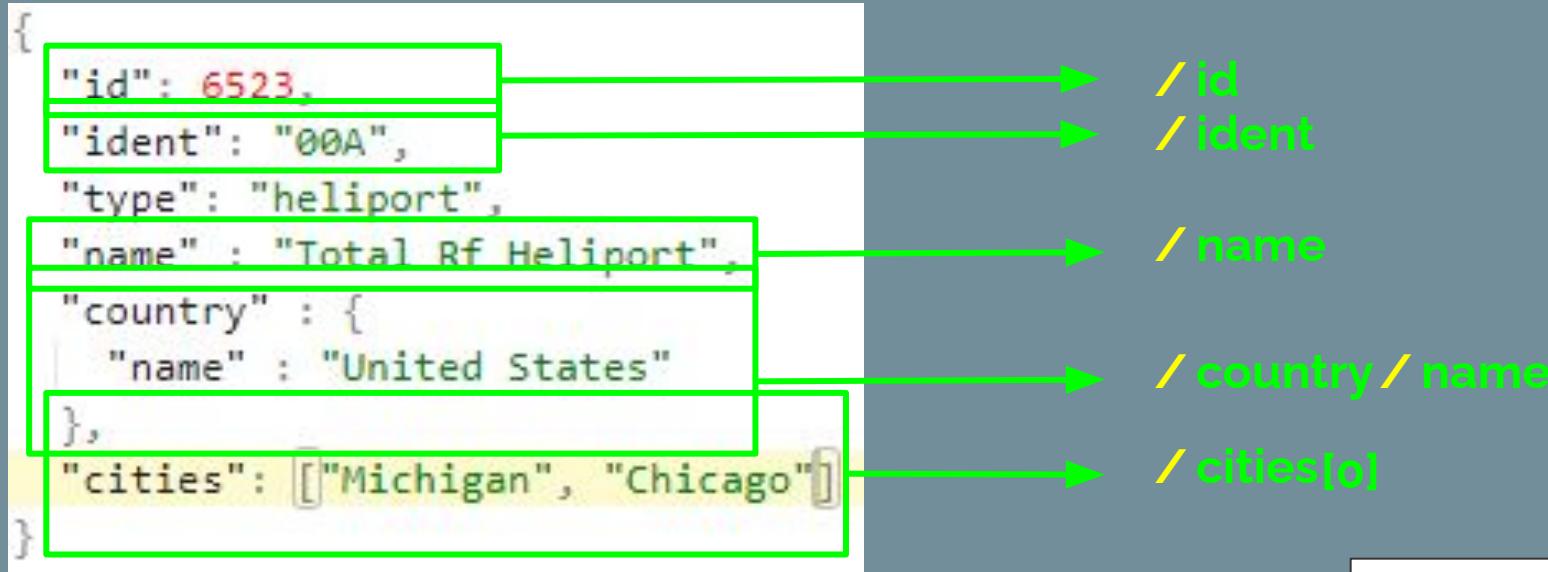


Json Record

```
{  
  "id": 6523,  
  "ident": "00A",  
  "type": "heliport",  
  "name" : "Total Rf Heliport",  
  "country" : {  
    "name" : "United States"  
  },  
  "cities": ["Michigan", "Chicago"]  
}
```

NiFi Record Path Language

Json Record



General Syntax

/ **FieldName**

NiFi Record Path Language - Standalone Functions

Record Path Expression	Result
/ type	Heliport
toUpperCase(/ type)	HELIPORT
toLowerCase(/ type)	heliport
replace(/ type , 'port', 'copter')	Helicopter
concat(/ type , ' in', ' New Zealand')	Heliport in New Zealand

Function Name

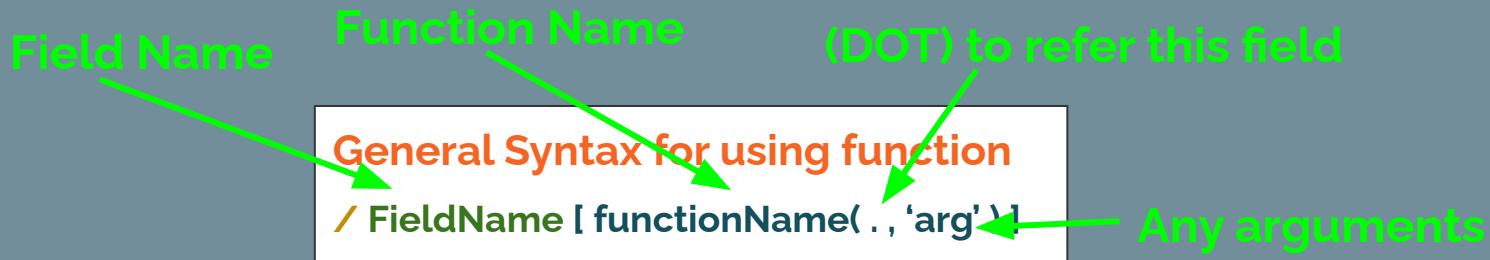
Field Name

General Syntax for using function

functionName(/ FieldName)

– NiFi Record Path Language - Filter Functions

Record Path Expression	Result
/ type	Heliport
/ type[contains(. , 'port')]	true
/ type[isEmpty(.)]	false
/ type [startsWith(. , 'Heli')]	true



Updating Record

Updating Record Field

Source JSON Record

```
{  
  "id" : 6523,  
  "ident" : "00A",  
  "type" : "heliport",  
  "name" : "Total Rf Heliport",  
  "latitude_deg" : 40.0708,  
  "longitude_deg" : -74.93360137939453,  
  "elevation_ft" : 11,  
  "continent" : "NA",  
  "iso_country" : "US",  
  "iso_region" : "US-PA",  
  "municipality" : "Bensalem",  
  "scheduled_service" : "no",  
  "gps_code" : "00A",  
  "iata_code" : null,  
  "local_code" : "00A",  
  "home_link" : null,  
  "wikipedia_link" : null,  
  "keywords" : null  
}
```

Target JSON Record

```
{  
  "id" : 6523,  
  "ident" : "00A",  
  "type" : "heliport",  
  "name" : "Total Rf Heliport",  
  "latitude_deg" : 40.0708,  
  "longitude_deg" : -74.93360137939453,  
  "elevation_ft" : 11,  
  "continent" : "North America",  
  "iso_country" : "US",  
  "iso_region" : "US-PA",  
  "municipality" : "Bensalem",  
  "scheduled_service" : "no",  
  "gps_code" : "00A",  
  "iata_code" : null,  
  "local_code" : "00A",  
  "home_link" : null,  
  "wikipedia_link" : null,  
  "keywords" : null  
}
```

Updating Record Field

UpdateRecord Processor
to update fields

	UpdateRecord	UpdateRecord 1.14.0
		org.apache.nifi - nifi-standard-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

Configure Processor

Invalid

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Replacement Value Strategy	Literal Value

CANCEL **APPLY**

Updating Record Field

Configure Processor

Add Controller Service

Requires Controller Service
RecordReaderFactory 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

- AvroReader 1.14.0
- AvroReader 1.14.0
- CSVReader 1.14.0
- GrokReader 1.14.0
- JsonPathReader 1.14.0
- JsonTreeReader 1.14.0** (highlighted with a green box)
- ParquetReader 1.14.0
- ReaderLookup 1.14.0
- ScriptedReader 1.14.0
- Syslog5424Reader 1.14.0
- SyslogReader 1.14.0

IconTreeReader 1.14.0

CANCEL APPLY

Configure Processor

Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Replacement Value Strategy	Literal Value

CANCEL APPLY

Updating Record Field

Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property

Record Reader
Record Writer
Replacement Value Strategy

/continent

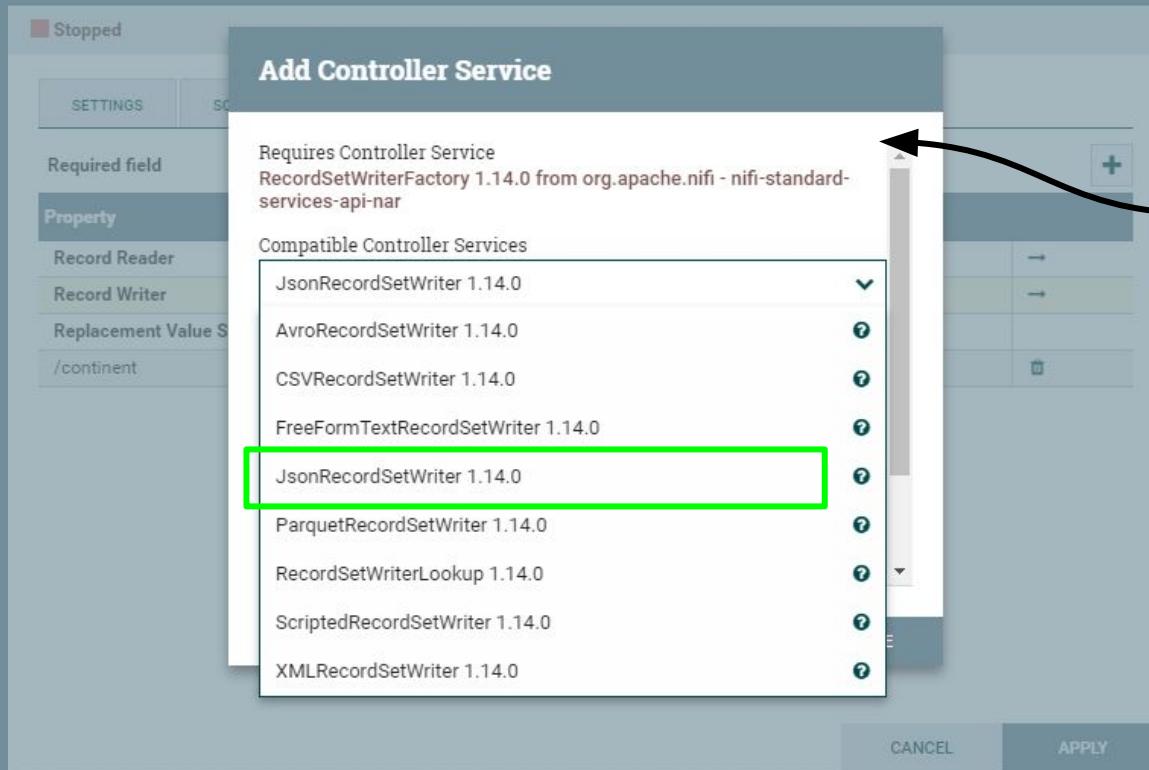
Add Controller Service

Requires Controller Service
RecordSetWriterFactory 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

- JsonRecordSetWriter 1.14.0
- AvroRecordSetWriter 1.14.0
- CSVRecordSetWriter 1.14.0
- FreeFormTextRecordSetWriter 1.14.0
- JsonRecordSetWriter 1.14.0
- ParquetRecordSetWriter 1.14.0
- RecordSetWriterLookup 1.14.0
- ScriptedRecordSetWriter 1.14.0
- XMLRecordSetWriter 1.14.0

CANCEL APPLY



Configure Processor

Invalid

Updating Record Field

UpdateRecord Processor
to update fields

	UpdateRecord	UpdateRecord 1.14.0
		org.apache.nifi - nifi-standard-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

Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

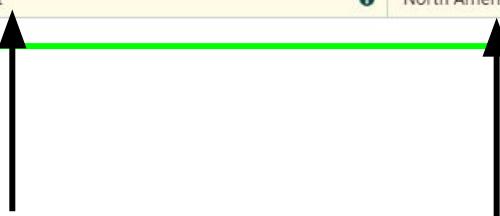
Required field

+ 

Property	Value
Record Reader	JsonTreeReader
Record Writer	JsonRecordSetWriter
Replacement Value Strategy	Literal Value
/continent	North America

Field Name Property Value

CANCEL APPLY



Updating Record Field

Using Record Path Strategy

Configure Processor

Stopped

Required field

Property	Value
Record Reader	JsonTreeReader
Record Writer	JsonRecordSetWriter
Replacement Value Strategy	Record Path Value
/continent	replace(/continent, 'NA', 'North America')

replace (/continent, 'NA', 'North America')

CANCEL APPLY

Updating Record Field

Generate JSON Data

	■ GenerateFlowFile	GenerateFlowFile 1.14.0	org.apache.nifi - nifi-standard-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	



Updates JSON Data

	⚠ UpdateRecord	UpdateRecord 1.14.0	org.apache.nifi - nifi-standard-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	



Logs JSON Data

	⚠ LogMessage	LogMessage 1.14.0	org.apache.nifi - nifi-standard-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	