

Cloud Computing – Lecture 09

Data acquisition, migration and flow management

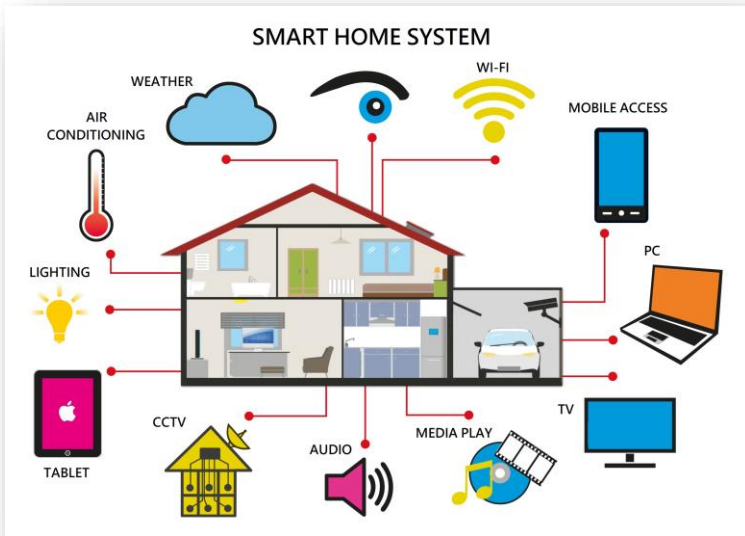
05 April 2022

Chinmaya Dehury

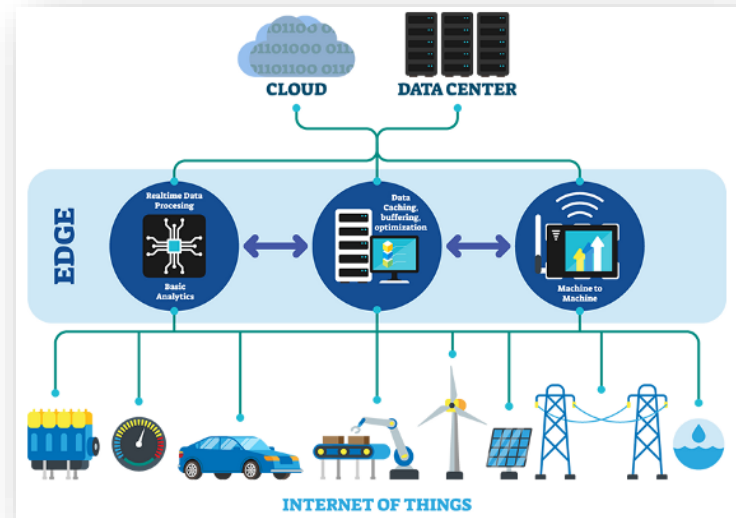
chinmaya.dehury@ut.ee

Outlines

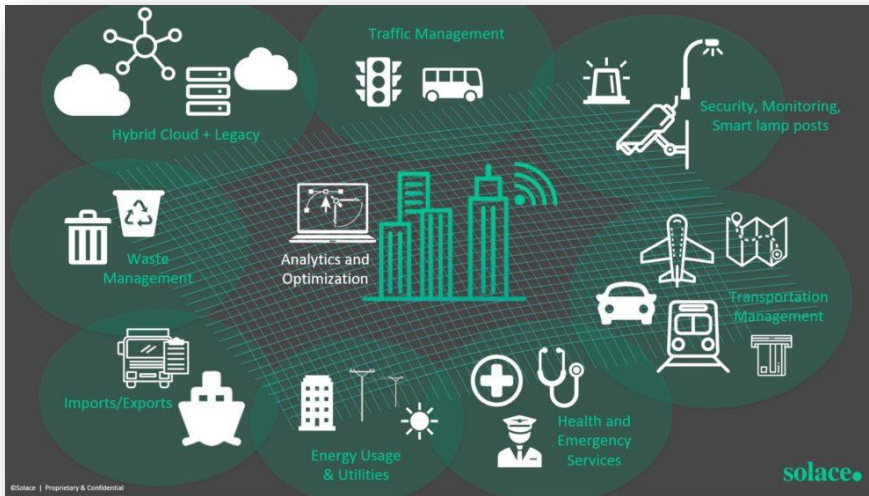
- Data Acquisition
- Data Migration
- Data Pipeline solutions
- AWS data pipeline
- Apache Nifi



Img src: <http://visioforce.com/smarthome.html>



Img src: <https://medium.com/dataseries/a-primer-on-edge-computing-3ef550c3d84e>



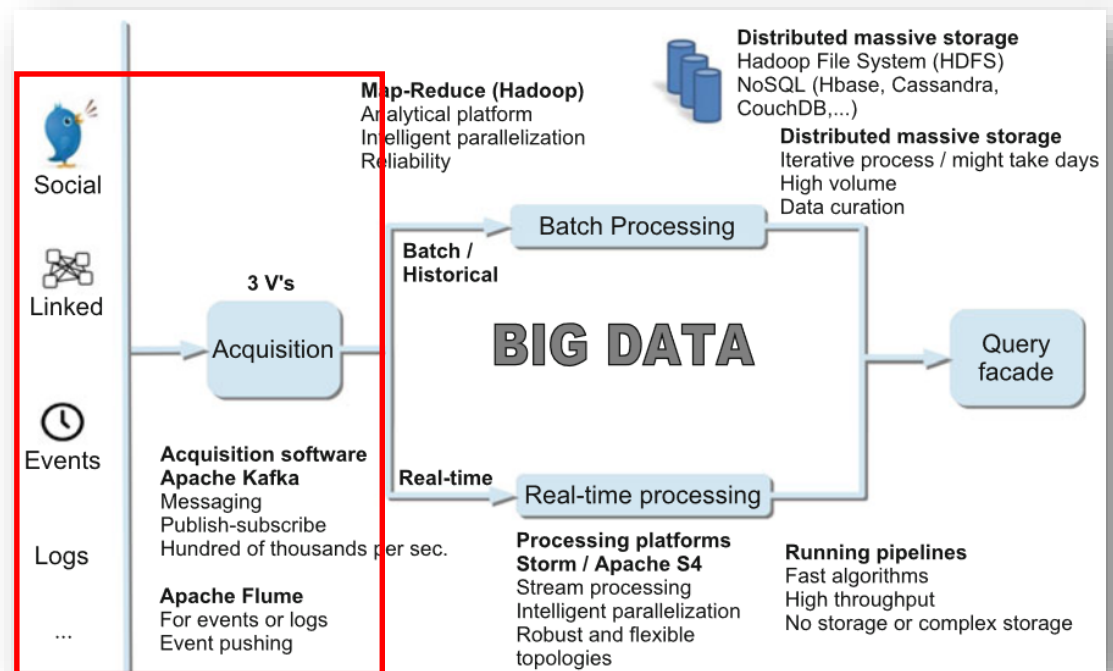
Img: <https://solace.com/blog/smart-city-data-management/>



Img: <https://www.pcmag.com/picks/the-best-cloud-storage-and-file-sharing-services>

Data Acquisition

- Process of gathering, filtering, and cleaning data
- Difficult to find complete set of required data in one place
- Data sources:
 - Social medias
 - IoT
 - Events
 - Logs
 - Linked



Img src: https://link.springer.com/content/pdf/10.1007/978-3-319-21569-3_4.pdf

Data Acquisition

- Data can be:
 - Text
 - Audio
 - Video
- 5V's of the data:
 - Volume (size of the data)
 - Velocity (how fast the data is generated?)
 - Variety (Structured, Semi- Structured, Unstructured data)
 - Veracity (messy, quality, and accuracy?)
 - Value

Data Migration

- transferring data from one computer storage system to another. e.g. :
 - Transferring images from your smart phone to your laptop
 - Transferring data from old laptop to new one
 - Transferring data from Google Drive to Dropbox
 - Sending data from CCTV to cloud storage
 - Sending sensors' data to cloud storage
- process of *selecting, preparing, extracting, transforming* data and *transferring*
- Usually thousands of data sources are involved
- Generated data are of small size
- Higher frequency of data generation

Data Migration

Challenges & Risks

- Data Loss
 - At source
 - At intermediate devices
 - Over network
 - At target
- Knowing data source
 - Can you identify duplicate, missing data, erroneous data
- Data validation
 - Validation at source
 - Merged data validation
 - Tools validation
 - Integration validation, etc

Data Migration

Challenges & Risks

- Compatibility issues:
 - Storage Compatibility (e.g. S3 <-> DynamoDB, local harddisk <-> cloud storage)
 - Application compatibility (e.g. old excel file with Excel 2019)
 - Platform compatibility (e.g. from on-premise to cloud),
 - Cloud compatibility (e.g. AWS <-> Azure, UT's Openstack <-> AWS)

Data Migration

2 Broad categories

- Online:
 - Migrating data without disrupting other applications
 - e.g. live VM migration
- Offline
 - This migration approach would invite disruption to user and application
 - E.g. data migration during scheduled maintenance, backup and restore purpose.

Challenges in choosing data migration method:

- Impact on downtime:
 - Estimating downtime
- Risk while migrating data online
- Emergency decision and rollback plan

Data Migration

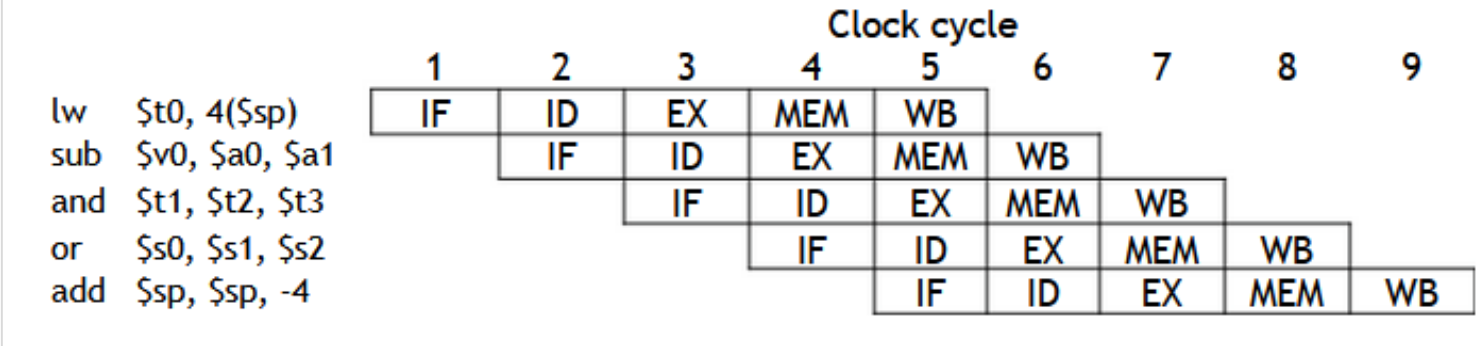
Factors to consider:

- Type of workload:
 - Databases, virtual machines (VMs), Backups, etc
- Amount of data
 - Imagine migrating some Petabytes of data online
 - Imagine migrating few GBs of data in offline mode
- Speed to completion:
 - For online migrations: amount of data
 - For offline migrations: shipping time

Then What is Data Pipeline ?

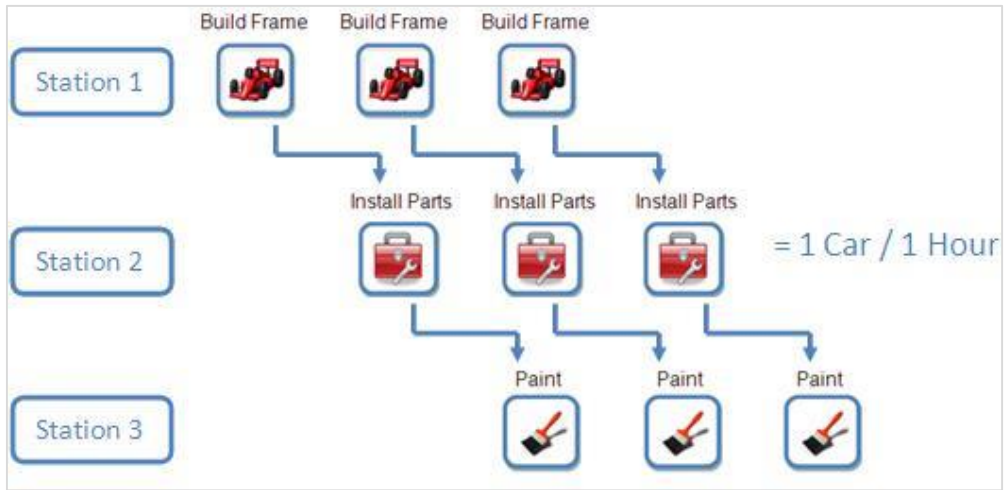
Data Pipeline

Pipeline approach for computer instruction execution:



<https://slideplayer.com/slide/8207220/>

Assembly-line in automobile industry:




<http://www.ni.com/cms/images/devzone/tut/final.JPG>



Data Pipeline

Pipeline approach in logistic:

Logistics Information:

International Shipping Company	Tracking Number	Remarks	Details
菜鸟超级经济Global	S00000090969004		<p>2019.11.26 19:37 (GMT-7): Departed country of origin 2019.11.26 14:37 (GMT-7): Shipment accepted by airline 2019.11.26 14:37 (GMT-7): Shipment left country of origin warehouse 2019.11.26 04:01 (GMT-7): Shipment at country of origin warehouse 2019.11.26 03:49 (GMT-7): Shipment dispatched</p> <p>Refresh</p> <div style="border: 1px solid orange; padding: 5px;"><p> Tracking information is available within 5-10 days. You can track your order here 菜鸟超级经济 Global.</p></div>

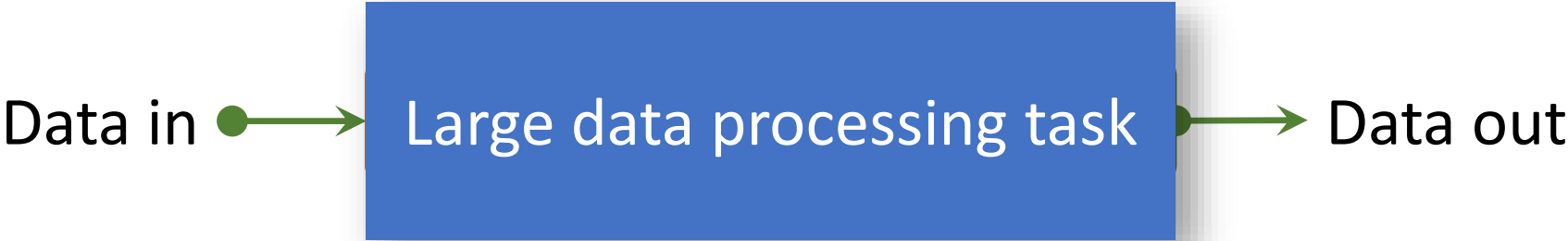
[View Delivery Detail](#)



Data Pipeline

Pipeline approach for handling the data acquisition, migration and its flow.

Data Pipeline (DP)



Data Pipeline (DP)

- A system for moving data from one system to another.
- Encompasses ETL as a subsystem
- Transformation of data is optional
- May process data in real-time or in batch manner

Data Pipeline properties

1. Low Event Latency:

- query recent event data within mins/secs

2. Scalability

- Able to scale to billions of data points

3. Interactive Querying

- support both long-running batch queries and smaller interactive queries

4. Versioning

5. Monitoring

6. Testing



Types of data pipeline solutions

1. Batch:

- Suitable for large-volume of data
- Move in a regular time interval

2. Real-time:

- Move and process data in a real-time

3. Cloud native

4. Open source

5. Proprietary Solution

Types of data pipeline solutions

Solution type	Solutions
Batch	Apache Spark, Astera Centerprise, Hevo Data,
Real-time	Apache Kafka, Apache Spark, Astera Centerprise, Hevo Data,
Cloud-Native	AWS Data pipeline, Hevo Data, Blendo, Confluent
Open-source	Apache Spark, Apache Kafka, Apache Nifi
Proprietary Solution	Astera Centerprise, Hevo Data

Data Pipeline Technologies

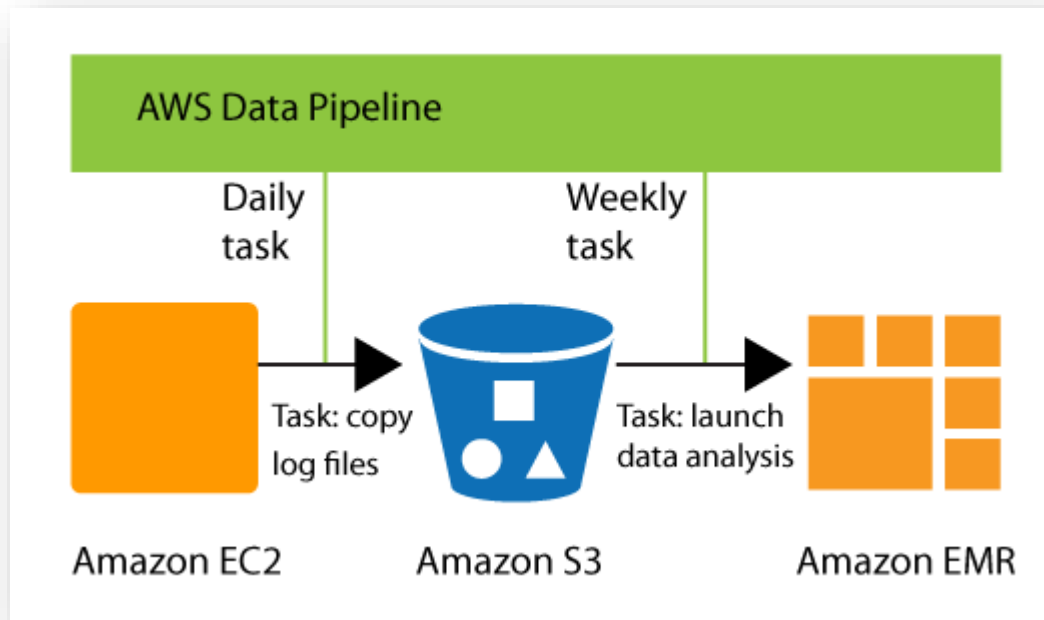
1. Amazon Data pipeline
2. Apache Nifi

Data Pipeline Technologies

1. Amazon Data pipeline
2. Apache Nifi



Amazon Data Pipeline



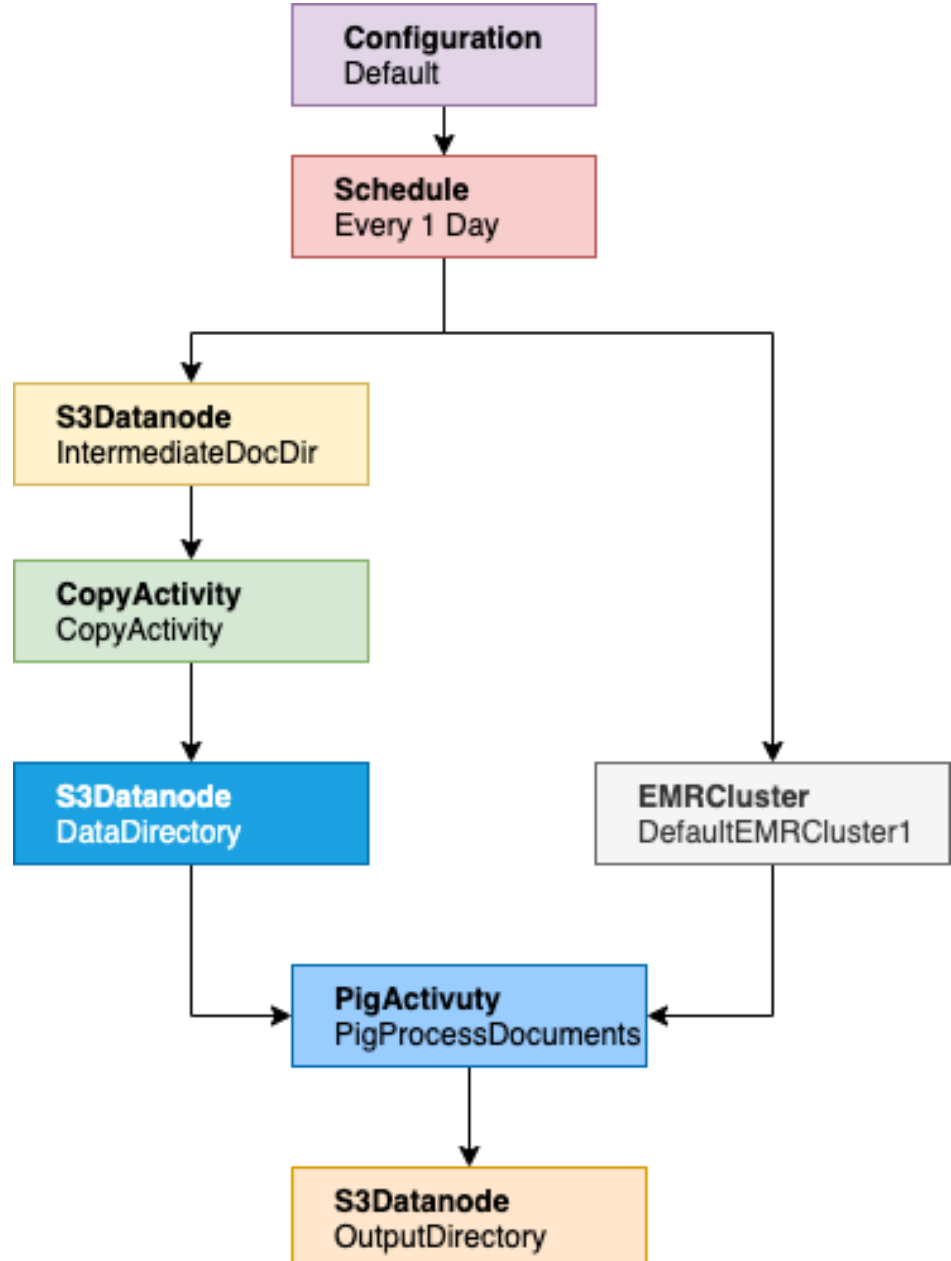
Img src: <https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/images/dp-how-dp-works-v2.png>

- A web service for reliable process and movement of data
- Focus is on AWS compute and storage services

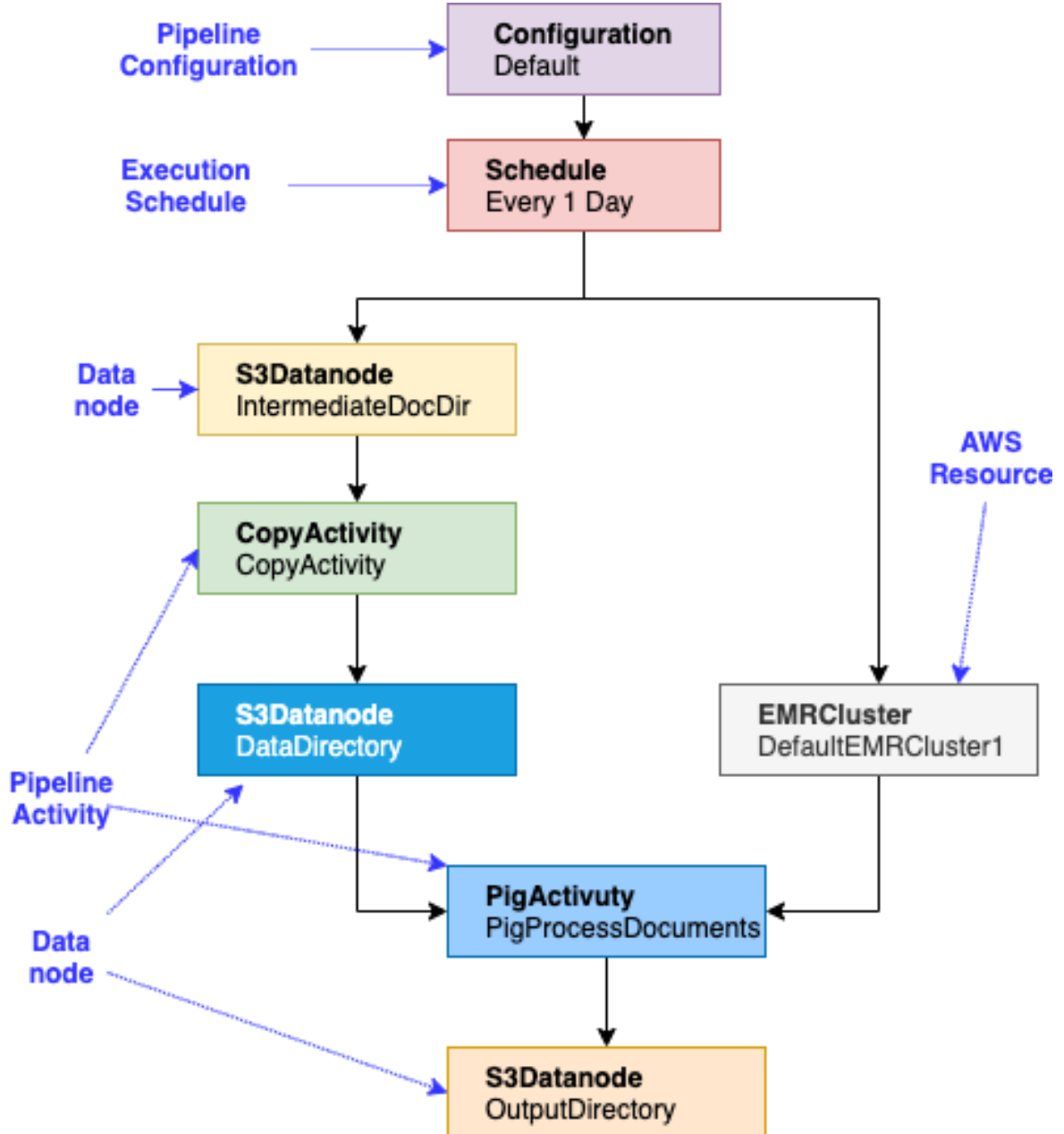
Amazon Data Pipeline

- AWS services such as
 - **Storage services:** Amazon S3, Amazon RDS, Amazon DynamoDB, Amazon Redshift
 - **Compute services:** Amazon EC2, Amazon EMR
- Data processing workloads can be
 - fault tolerant
 - repeatable
 - highly available

Amazon Data Pipeline : An Example



Amazon Data Pipeline : An Example



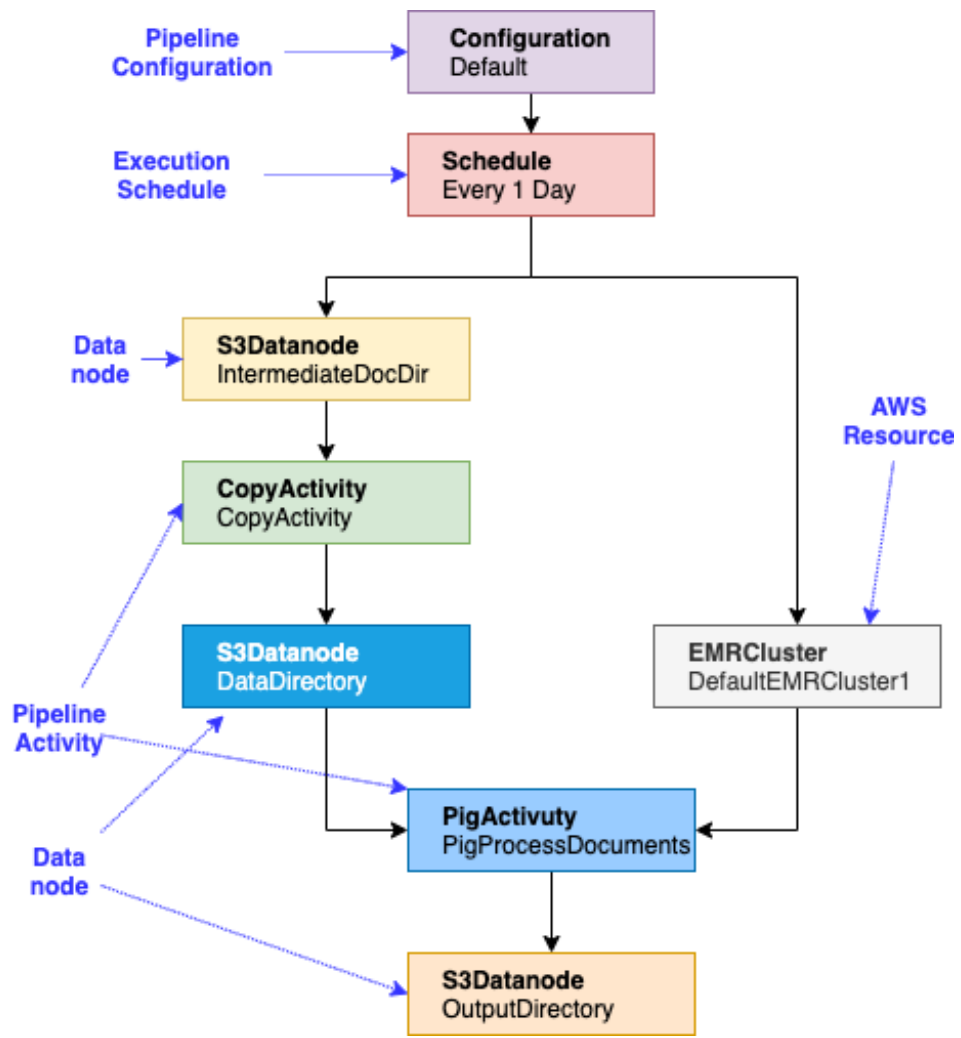
Amazon Data Pipeline – Components

1. Major components

- I. DataNodes
- II. Activities

2. Additional components

- I. Schedules
- II. Preconditions
- III. Resources



Amazon Data Pipeline – Major components

1. Major components

I. DataNodes: It specifies the name, location, and format of the data sources such as Amazon S3, Dynamo DB, etc.

i. DynamoDBDataNode

ii. SqlDataNode

iii. RedshiftDataNode

iv. S3DataNode

v. SqlDataNode



DynamoDBDataNode

MySqlDataNode

RedshiftDataNode

S3DataNode

SqlDataNode

II. Activities: Activities are the actions that perform the SQL Queries on the databases, transforms the data from one data source to another data source.

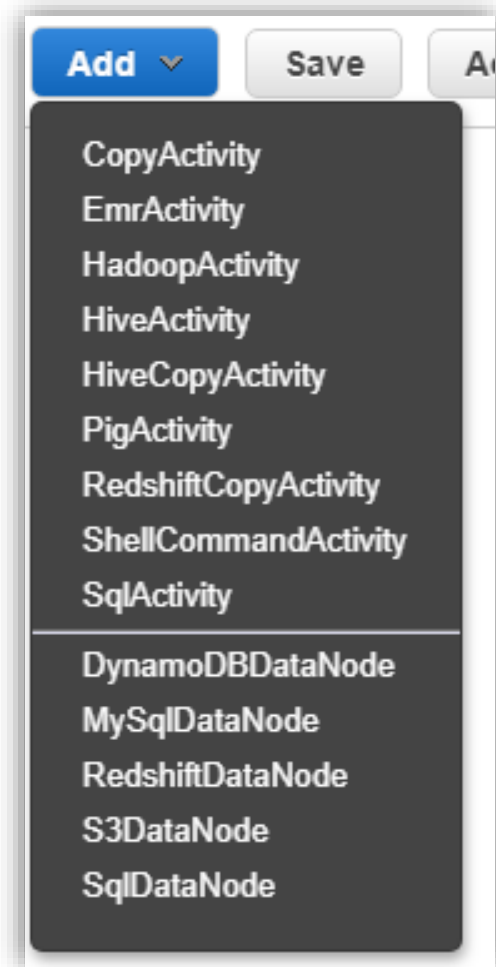
Amazon Data Pipeline – Major components

1. Major components

I. DataNodes

II. Activities

- i. CopyActivity
- ii. EmrActivity
- iii. HadoopActivity
- iv. HiveActivity
- v. HiveCopyActivity
- vi. PigActivity
- vii. RedshiftCopyActivity
- viii. ShellCommandActivity
- ix. SqlActivity



Amazon Data Pipeline - Additional components

1. Major components

I. DataNodes

II. Activities

2. Additional components

I. Schedules: Schedule defines the timing of a scheduled event, such as when an activity runs.

Schedule

i You can run your pipeline once or specify a schedule. [More](#)

Run on pipeline activation
 on a schedule

Run every

Starting on pipeline activation
 UTC (Current time is 12:11 UTC)
YYYY-MM-DD HH:MM

Ending never
 after occurrence(s)
 UTC (Current time is 12:11 UTC)
YYYY-MM-DD HH:MM

Amazon Data Pipeline - Additional components

2. Additional components

I. Schedules

II. **Preconditions:** A condition that must be true before an activity can run. E.g., check if the data is present on the source before attempting to run CopyActivity.

A. System-managed Precondition:

- a) **DynamoDBDataExists**
- b) **DynamoDBTableExists**
- c) **S3KeyExists**, etc..

B. User-managed precondition

- a) **Exists:** Checks whether a data node exists.
- b) **ShellCommandPrecondition:** Unix/Linux shell command that can be run as a precondition

Amazon Data Pipeline - Additional components

2. Additional components

I. Schedules

II. Preconditions

III. **Resources**: refer to the computational resource that performs the work that a pipeline activity specified

I. **Ec2Resource**: An EC2 instance

II. **EmrCluster**: An Amazon EMR cluster

Trying Amazon Data Pipeline

If your AWS account is less than 12 months old, you are eligible to use the free tier. ([url](#))

Other commercial data pipeline solutions

Microsoft Azure Data Factory:

<https://docs.microsoft.com/en-us/azure/data-factory/>

Google Cloud Dataflow:

<https://cloud.google.com/dataflow>

IBM InfoSphere Virtual Data pipeline:

<https://www.ibm.com/products/ibm-infosphere-virtual-data-pipeline>



Data Pipeline Technologies

1. Amazon Data pipeline

2. Apache Nifi



Apache Nifi Data Pipeline

- Open-source, under the Apache Software Foundation
- Automates and manages the flow of data between systems
- Web-based User Interface for creating, monitoring, & controlling data flows.
- Clients [\[src\]](#):
 - Micron: Semiconductor Manufacturing
 - Payoff: Financial Wellness (fintech)
 - Slovak: Telekom Telecommunications
 - Looker: SaaS & Analytics Software
 - Hastings Group: Insurance
 - and many more....
- Latest version 1.15.3 (as on April 2022)

Apache Nifi Data Pipeline

Key Features

Flow Management:

- Data Buffering
- Prioritized Queuing
- Guaranteed Delivery

Ease of Use:

- Flow Templates
- Data Provenance
- Fine-grained history



Apache Nifi Data Pipeline

Key Features

Security

- System to System
- User to System
- Multi-tenant Authorization

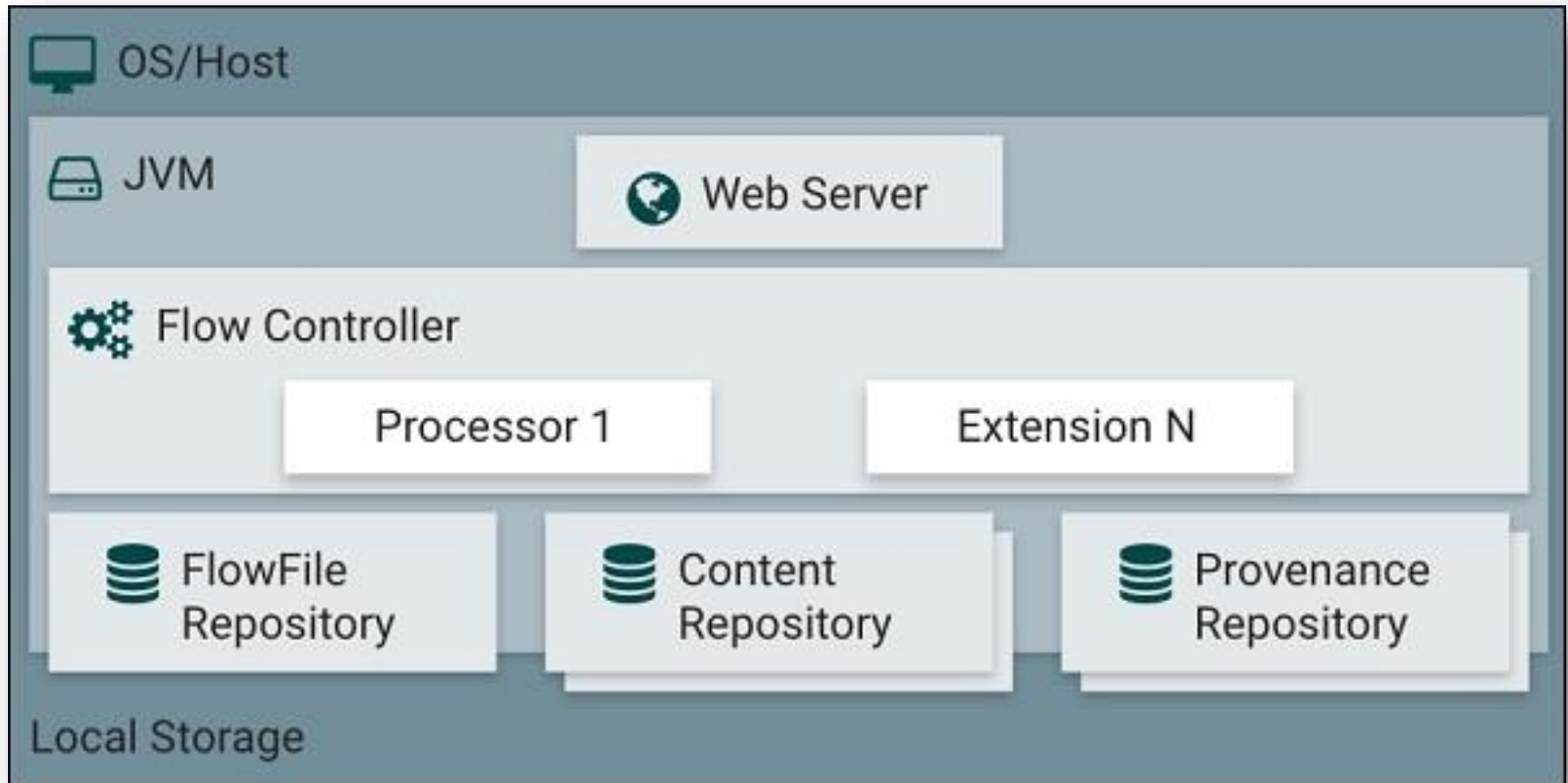
Extensible Architecture

- Extension (e.g. having custom processor)
- Site-to-Site Communication Protocol



Apache Nifi Data Pipeline

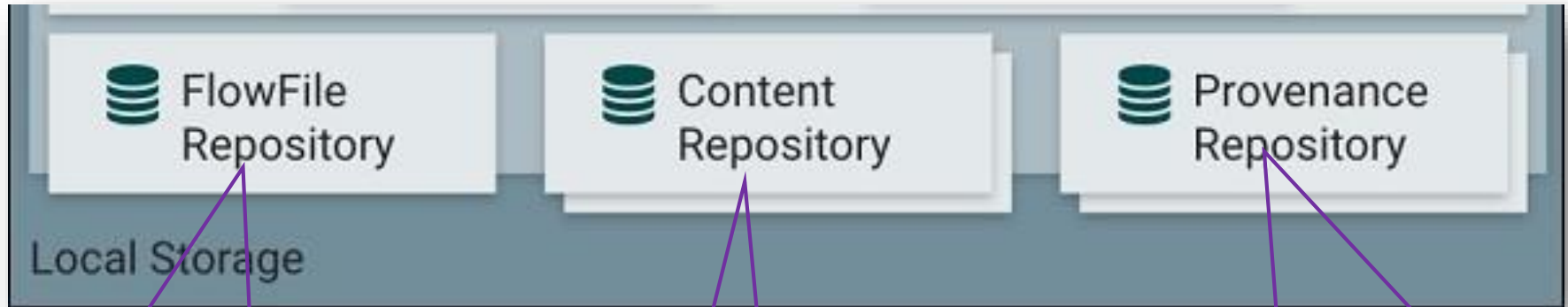
NiFi Architecture



Src: https://www.tutorialspoint.com/apache_nifi/apache_nifi_basic_concepts.htm

Apache Nifi Data Pipeline

NiFi Architecture - Repositories



stores the metadata of the FlowFiles during the active flow.

what happened to a particular data object (FlowFile) is kept in here. History of each FlowFile is stored here.

holds the actual content of the FlowFiles.

Src: https://www.tutorialspoint.com/apache_nifi/apache_nifi_basic_concepts.htm

Apache Nifi Key concepts

Key concepts

1. FlowFile

- represents each object moving through the system
- Include: data record (pointer to data payload)

2. Processor

- Processors actually perform the work
- E.g. processor to send email, upload data to S3 bucket, Reading data from FTP server, etc

3. Process Group

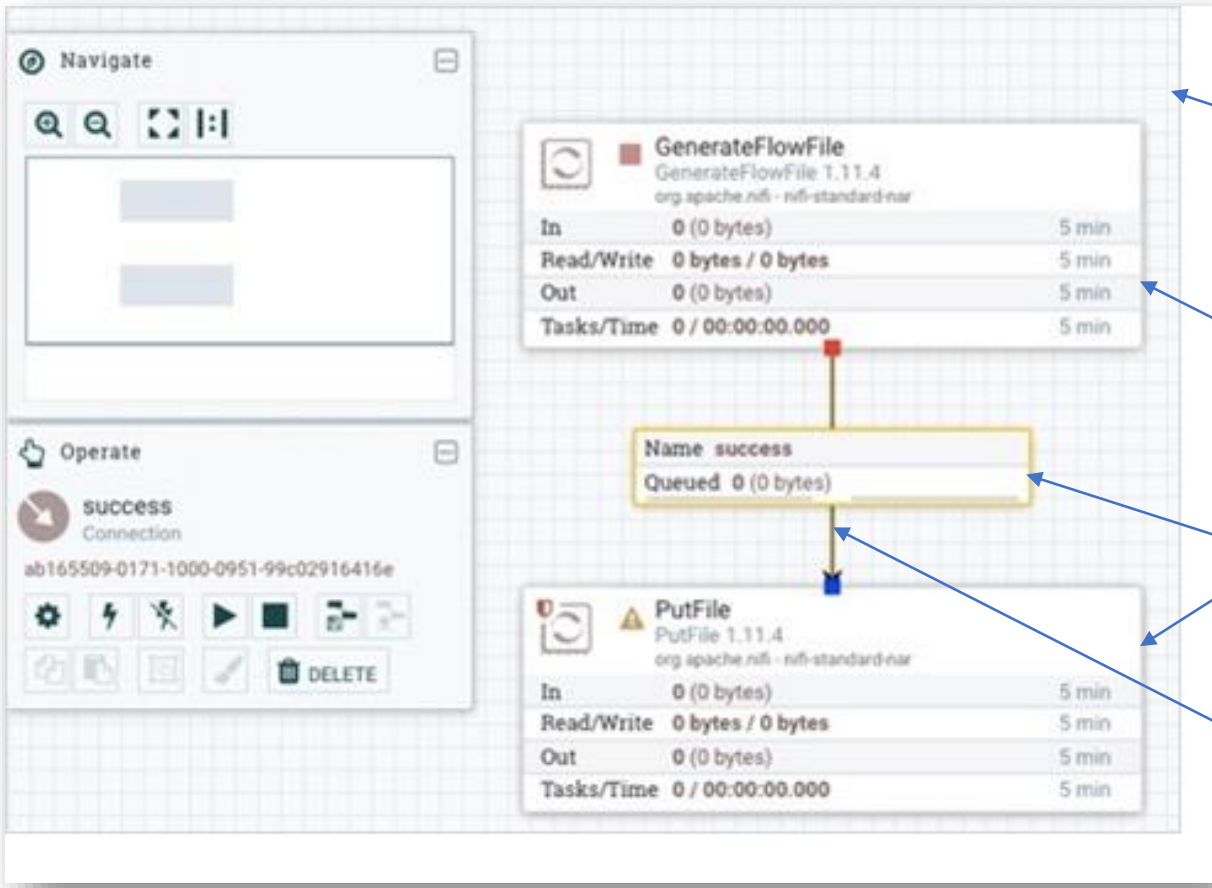
- Group of processors, connection, input/output, etc

4. Event

5. Data provenance



Apache Nifi – An example



IDE

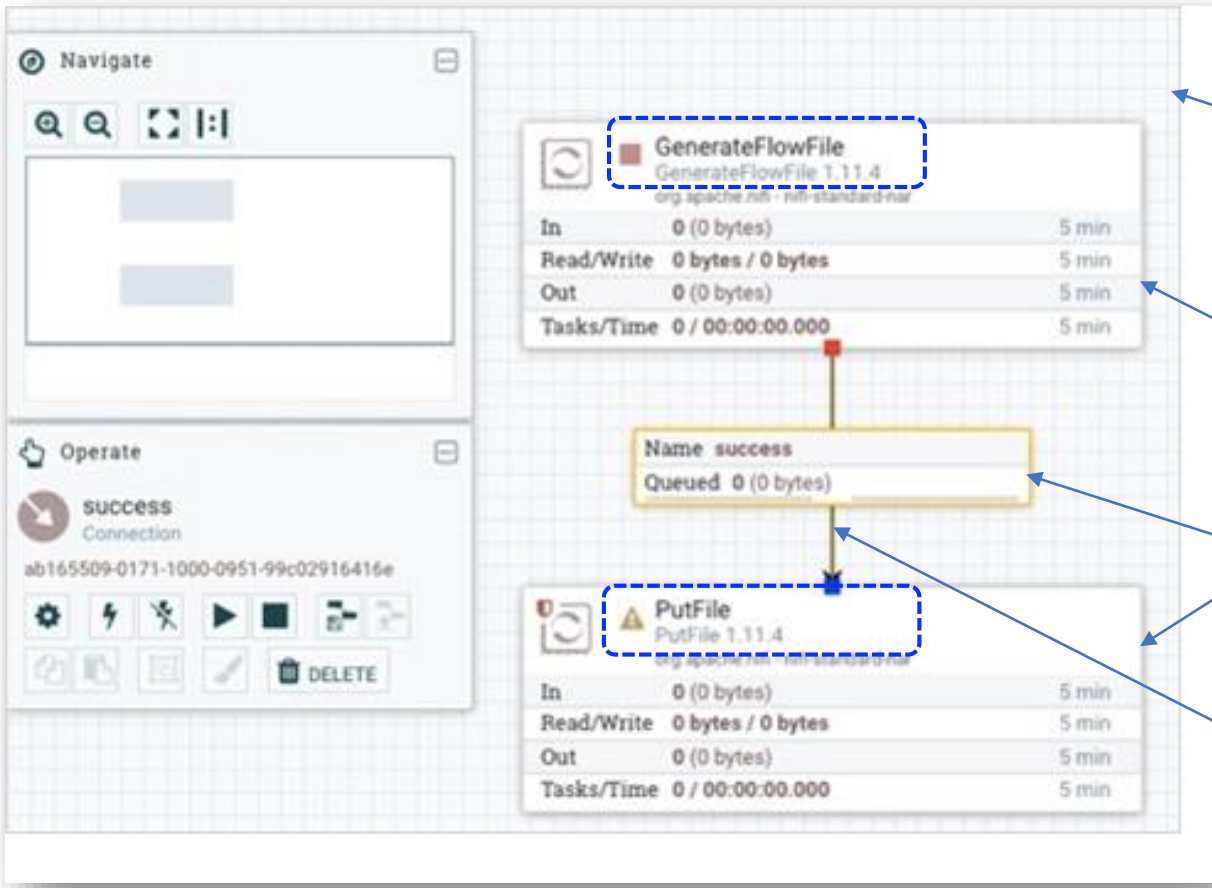
Processors

Buffer

Connection



Apache Nifi – An example



IDE

Processors

Buffer

Connection



Apache Nifi Data Pipeline

1. Major components

- I. Processors (execute the task)
- II. Queue (between processors)

2. Additional components

- I. Input Port
- II. Output Port
- III. Process Group (Groupism of multiple components such as processors)
- IV. Remote Process Group
- V. Template



Apache Nifi - Processors

1. Major components

I. Processors

283 processors

The screenshot shows the 'Add Processor' interface in Apache NiFi. It features a search bar on the left with various processor categories like 'amazon', 'attributes', 'avro', 'aws', 'azure', 'consume', 'csv', 'database', 'delete', 'fetch', 'get', 'hadoop', 'ingest', 'insert', 'json', 'listen', 'logs', 'message', 'pubsub', 'put', 'record', 'restricted', 'source', 'text', and 'update'. The main area displays a table of processors with columns for 'Type', 'Version', and 'Tags'. The 'AttributeRollingWindow' processor is highlighted. Below the table, there is a description for the selected processor.

Type	Version	Tags
AttributeRollingWindow	1.13.2	rolling, data science, Attribute ...
AttributesToCSV	1.13.2	flowfile, csv, attributes
AttributesToJSON	1.13.2	flowfile, json, attributes
Base64EncodeContent	1.13.2	encode, base64
CalculateRecordStats	1.13.2	stats, record, metrics
CaptureChangeMySQL	1.13.2	cdc, jdbc, mysql, sql
CompareFuzzyHash	1.13.2	fuzzy-hashing, hashing, cyber...
CompressContent	1.13.2	lzma, snappy-hadoop, deflate, ...
ConnectWebSocket	1.13.2	subscribe, consume, listen, We...
ConsumeAMQP	1.13.2	receive, amqp, rabbit, get, cons...
ConsumeAzureEventHub	1.13.2	cloud, streaming, streams, eve...
ConsumeFWS	1.13.2	FWS Exchange, Email Consu...

AttributeRollingWindow 1.13.2 org.apache.nifi - nifi-stateful-analysis-nar

Track a Rolling Window based on evaluating an Expression Language expression on each FlowFile and add that value to the processor's state. Each FlowFile will be emitted with the count of FlowFiles and total aggregate value of values processed in the current time window.



Apache Nifi - Processors

1. Major components

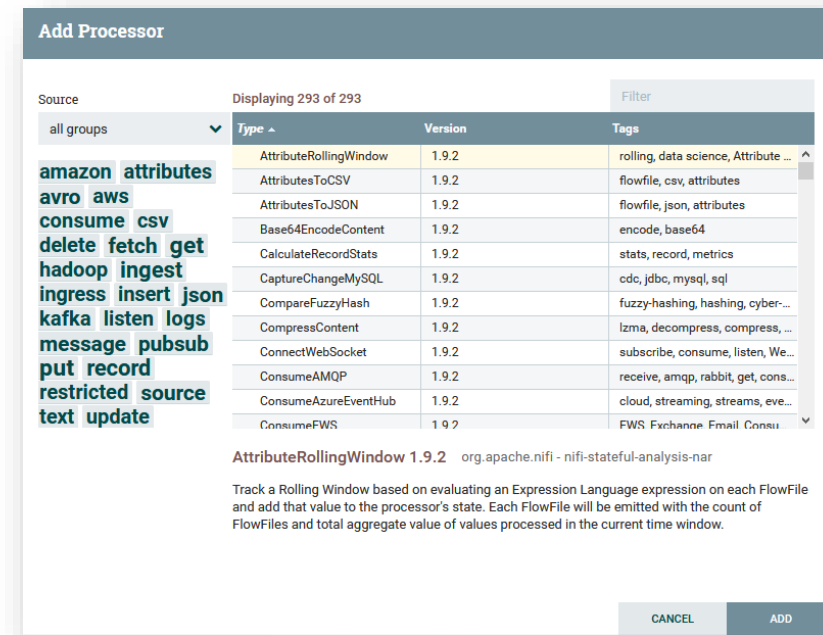
I. Processors

Different States of a Processor:

Start, Stop, Enable, & Disable

Disable processor can not be started.

When a group of Processors is started, this (disabled) Processor should be excluded



Add Processor

Source: all groups | Displaying 293 of 293 | Filter

Type	Version	Tags
AttributeRollingWindow	1.9.2	rolling, data science, Attribute ...
AttributesToCSV	1.9.2	flowfile, csv, attributes
AttributesToJSON	1.9.2	flowfile, json, attributes
Base64EncodeContent	1.9.2	encode, base64
CalculateRecordStats	1.9.2	stats, record, metrics
CaptureChangeMySQL	1.9.2	cdc, jdbc, mysql, sql
CompareFuzzyHash	1.9.2	fuzzy-hashing, hashing, cyber...
CompressContent	1.9.2	lzma, decompress, compress, ...
ConnectWebSocket	1.9.2	subscribe, consume, listen, We...
ConsumeAMQP	1.9.2	receive, amqp, rabbit, get, cons...
ConsumeAzureEventHub	1.9.2	cloud, streaming, streams, eve...
ConsumeFWS	1.9.2	FWS, Exchange, Email, Consu...

AttributeRollingWindow 1.9.2 org.apache.nifi - nifi-stateful-analysis-nar

Track a Rolling Window based on evaluating an Expression Language expression on each FlowFile and add that value to the processor's state. Each FlowFile will be emitted with the count of FlowFiles and total aggregate value of values processed in the current time window.

CANCEL ADD

Apache Nifi – Processors Setting

- 1. Major components
 - I. Processors

Configuring a Processor


SETTING:

Penalty duration: Time to wait, when the the data can not be processed for some reason.

Yield Duration: Time to wait, when the process can not progress.

Bulletin level: Level of bulletin, Nifi will display in the user interface. (e.g. Warn, error, info, debug)

Failure & Success

	CountText CountText 1.9.2 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

SETTINGS | SCHEDULING | PROPERTIES | COMMENTS

Name: CountText Enabled

Id: f95bdf29-016e-1000-429f-a4e643a35d9b

Type: CountText 1.9.2

Bundle: org.apache.nifi - nifi-standard-nar

Penalty Duration: 30 sec | Yield Duration: 1 sec

Bulletin Level: WARN

Automatically Terminate Relationships:
 failure
If the flowfile text cannot be counted for some reason, the original file will be routed to this destination and nothing will be routed elsewhere
 success
The flowfile contains the original content with one or more attributes added containing the respective counts

CANCEL APPLY



Apache Nifi – Processors Scheduling

1. Major components

I. Processors


Configuring a Processor

Scheduling :

Time vs Event vs CRON Driven

Concurrent Tasks: Number of FlowFiles

should be processed by this Processor at the same time.

	CountText CountText 1.9.2 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

SETTINGS | **SCHEDULING** | PROPERTIES | COMMENTS

Scheduling Strategy ⓘ
Timer driven ▼

Concurrent Tasks ⓘ
1

Execution ⓘ
All nodes ▼

Run Duration ⓘ
0ms 25ms 50ms 100ms 250ms 500ms 1s 2s
Lower latency Higher throughput

Run Schedule ⓘ
0 sec

CANCEL APPLY



Apache Nifi – Processors Properties

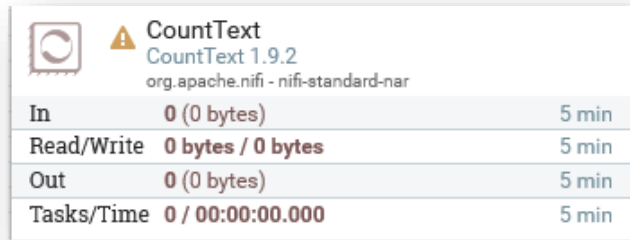
1. Major components

I. Processors

Configuring a Processor

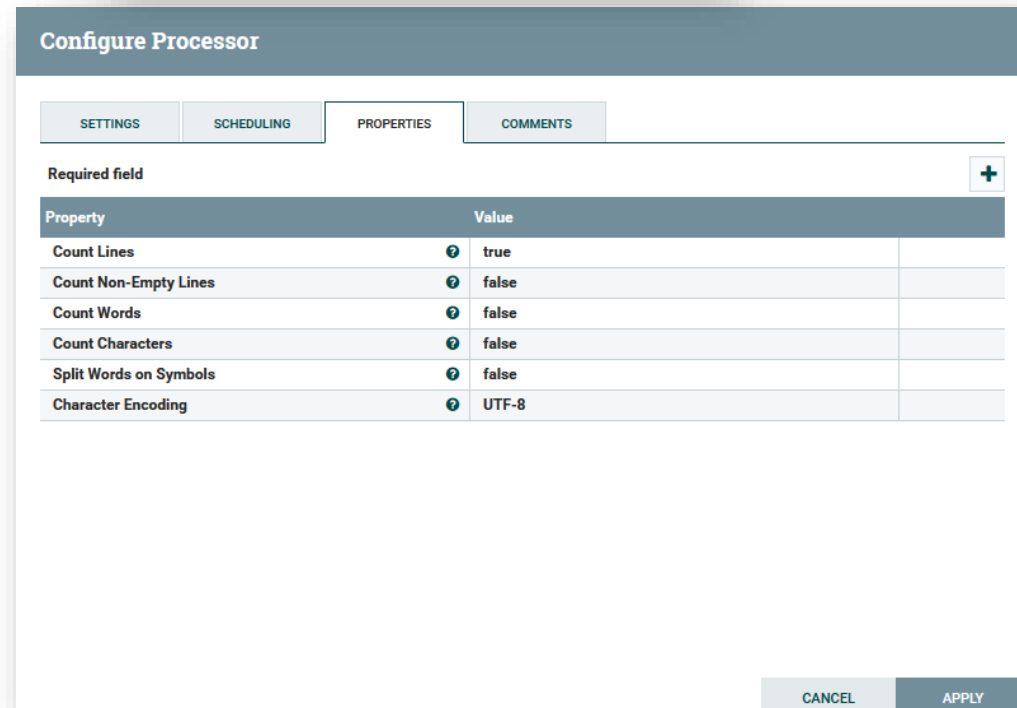
Properties :

- Provides a mechanism to configure Processor-specific behavior.
- There are no default properties.



CountText
CountText 1.9.2
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

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

Property	Value
Count Lines	true
Count Non-Empty Lines	false
Count Words	false
Count Characters	false
Split Words on Symbols	false
Character Encoding	UTF-8

CANCEL APPLY

Apache Nifi – Processor categories

Different categories of processors

- **Data Ingestion Processors:** GetFile, GetHTTP, GetFTP, etc
- **Routing and Mediation Processors:** RouteOnAttribute, RouteOnContent, ControlRate, RouteText, etc.
- **Database Access Processors:** ExecuteSQL, PutSQL, PutDatabaseRecord, ListDatabaseTables, etc.
- **Attribute Extraction Processors:** UpdateAttribute, EvaluateJSONPath, ExtractText, AttributesToJSON, etc
- **System Interaction Processors:** ExecuteScript, ExecuteProcess, ExecuteGroovyScript, ExecuteStreamCommand, etc

Apache Nifi – Processor categories

Different categories of processors

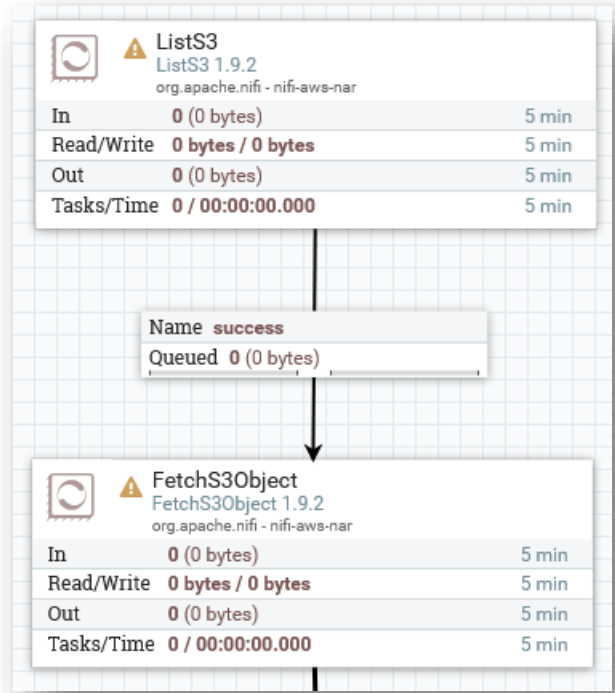
- **Data Transformation Processors:** ReplaceText, JoltTransformJSON, etc
- **Sending Data Processors:** PutEmail, PutSFTP, PutFile, PutFTP, etc.
- **Splitting and Aggregation Processors:** SplitText, SplitJson, SplitXml, MergeContent, SplitContent, etc.
- **HTTP Processors:** InvokeHTTP , ListenHTTP, etc
- **AWS Processors:** GetSQS, PutSNS, PutS3Object, FetchS3Object, etc.

Apache Nifi - Queue

1. Major components

II. Queue

- To handle the large amount of data inflow.
- Possible to see the content, ID, Filename, FileSize etc of a flowfile



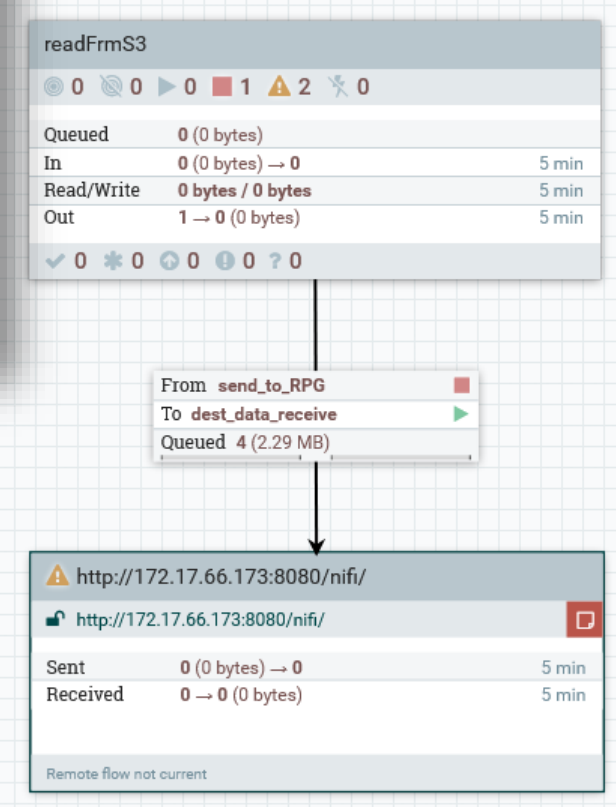
ListS3
ListS3 1.9.2
org.apache.nifi - nifi-aws-nar

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

Name **success**
Queued 0 (0 bytes)

FetchS3Object
FetchS3Object 1.9.2
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



readFromS3

Queued 0 (0 bytes)

In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	1 → 0 (0 bytes)	5 min

From **send_to_RPG**
To **dest_data_receive**
Queued 4 (2.29 MB)

http://172.17.66.173:8080/nifi/

Sent	0 (0 bytes) → 0	5 min
Received	0 → 0 (0 bytes)	5 min

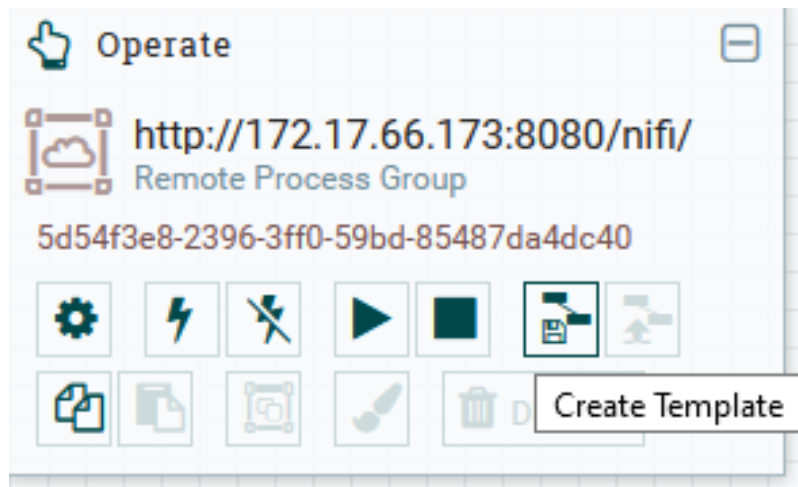
Remote flow not current



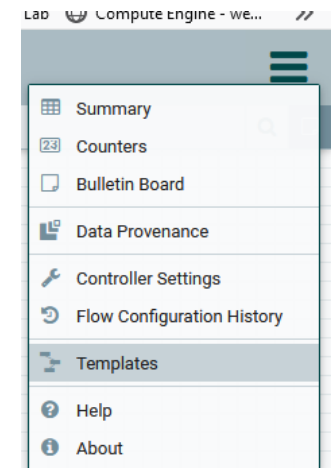
Apache Nifi - Flow Template

Templates:

- Can be thought of as a reusable sub-flow.
- Any properties that are identified as being Sensitive Properties (such as a password that is configured in a Processor) will not be added to the template.



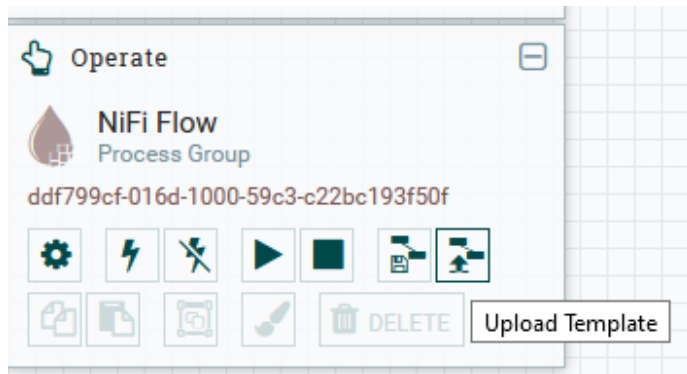
Create Template



Download Template

Apache Nifi - Flow Template

Templates:



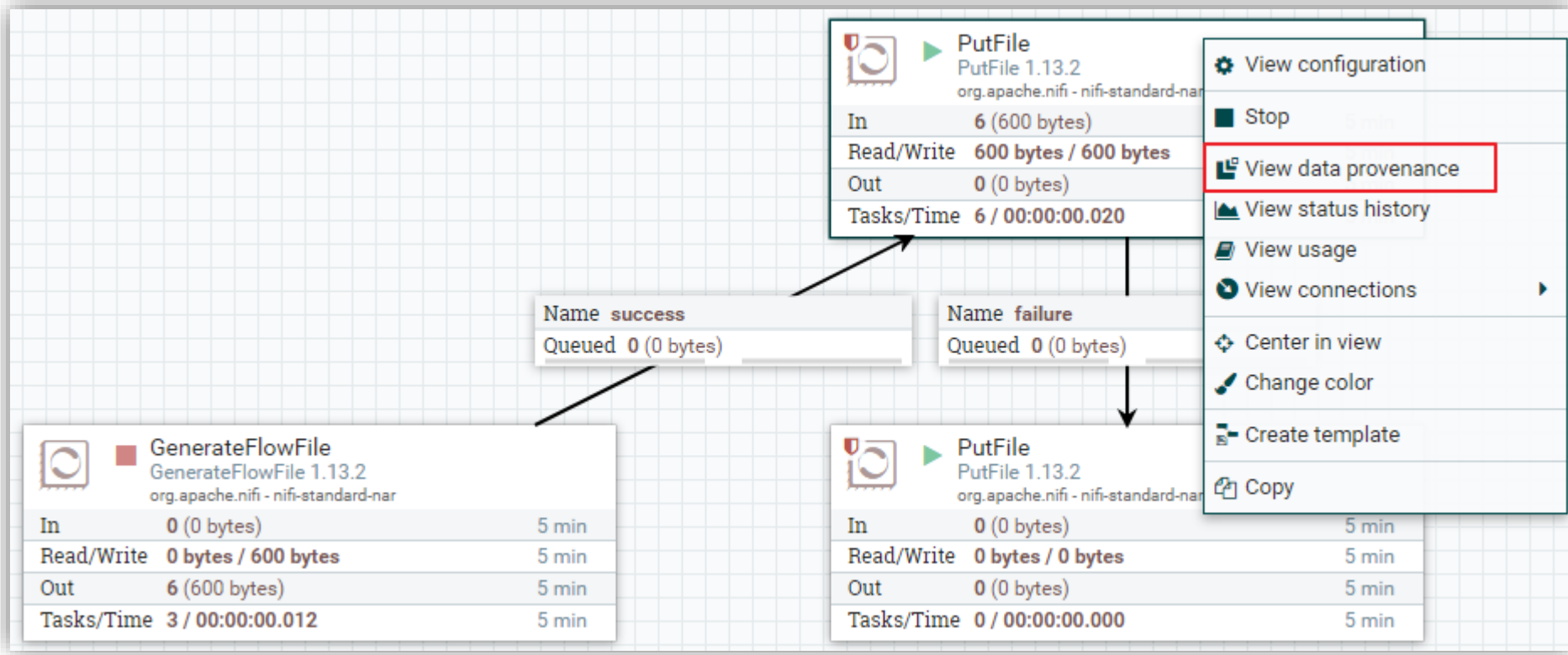
[Upload Template](#)



[Add Template](#)

Apache Nifi – Data Provenance

- Snapshots of each FlowFile.
- Event type, FlowFile Lineage Graph, Provenance event Details
- In-depth discovery of the chain of events.



Apache Nifi – Data Provenance

List of Events

NiFi Data Provenance

Displaying 56 of 56
Oldest event available: 04/12/2021 13:42:13 UTC

Showing the events that match the specified query. [Clear search](#)

Filter: by component name

	Date/Time	Type	FlowFile UUID	Size	Component Name	Component Type	
	04/12/2021 13:46...	DROP	a6377ad4-7305-4...	100 bytes	PutFile	PutFile	
	04/12/2021 13:46...	SEND	a6377ad4-7305-4...	100 bytes	PutFile	PutFile	
	04/12/2021 13:46...	DROP	b1529b0c-3ae4-4...	100 bytes	PutFile	PutFile	
	04/12/2021 13:46...	SEND	b1529b0c-3ae4-4...	100 bytes	PutFile	PutFile	
	04/12/2021 13:46...	DROP	fd7312aa-c1ce-47...	100 bytes	PutFile	PutFile	
	04/12/2021 13:46...	SEND	fd7312aa-c1ce-47...	100 bytes	PutFile	PutFile	

Provenance event details

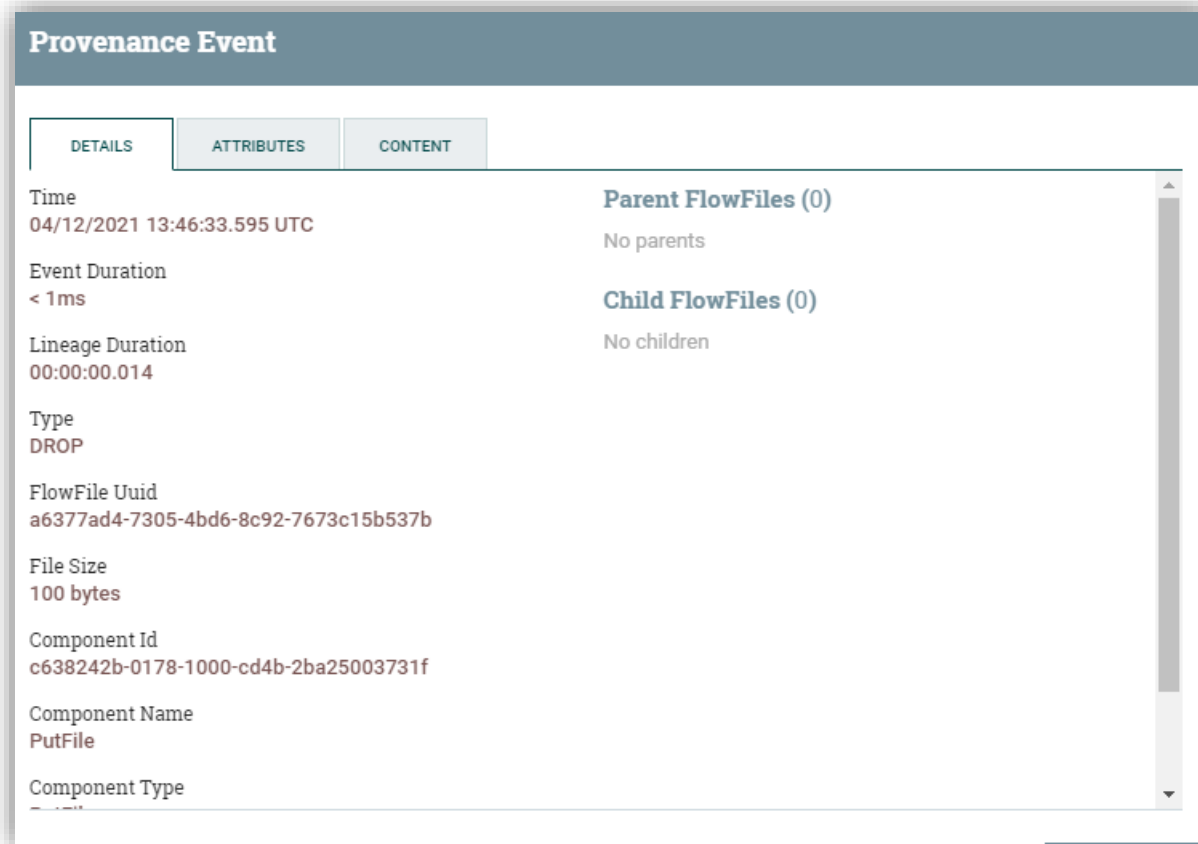
Event type: RECEIVE, SEND, DROP, JOIN, CONTENT_MODIFIED, ATTRIBUTES_MODIFIED, FORK, CLONE, ROUTE, etc.

FlowFile Lineage Graph



Apache Nifi – Data Provenance

Provenance event Details

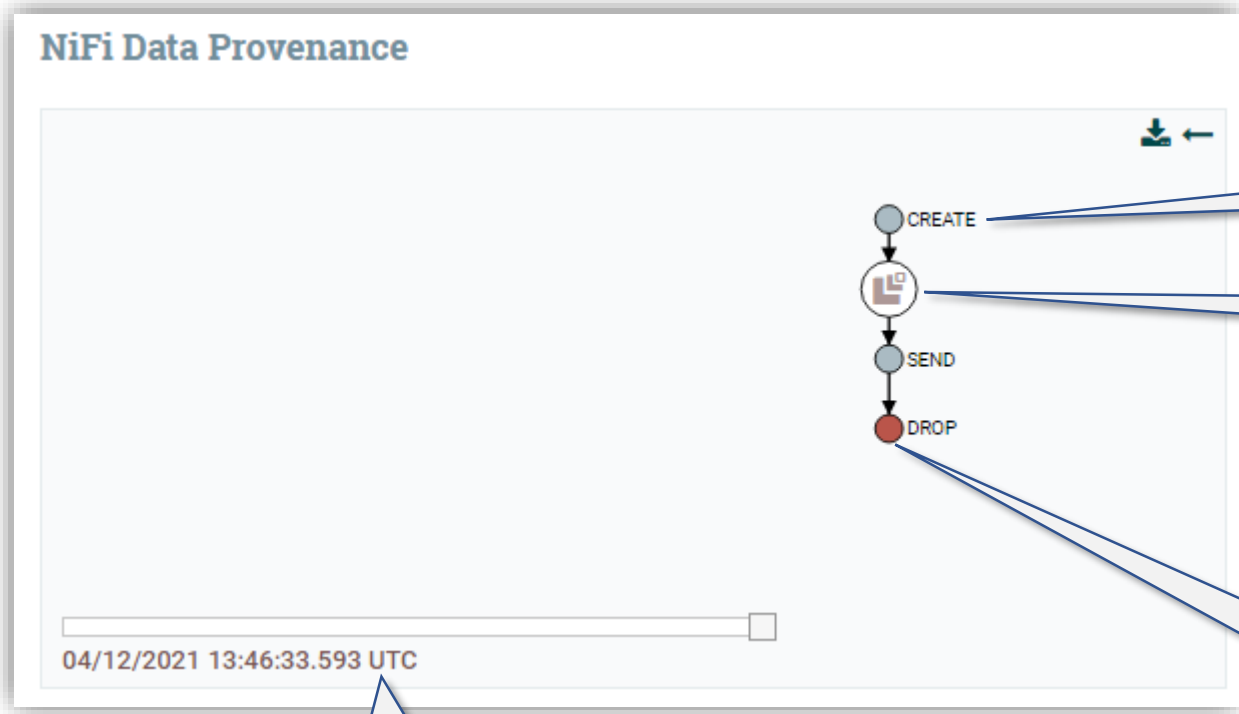


The screenshot shows the 'Provenance Event' details page in Apache NiFi. It features three tabs: 'DETAILS' (selected), 'ATTRIBUTES', and 'CONTENT'. The 'DETAILS' tab displays the following information:

Time	04/12/2021 13:46:33.595 UTC	Parent FlowFiles (0) No parents
Event Duration	< 1ms	Child FlowFiles (0) No children
Lineage Duration	00:00:00.014	
Type	DROP	
FlowFile Uuid	a6377ad4-7305-4bd6-8c92-7673c15b537b	
File Size	100 bytes	
Component Id	c638242b-0178-1000-cd4b-2ba25003731f	
Component Name	PutFile	
Component Type		

Apache Nifi – Data Provenance

FlowFile Lineage Graph



Event

FlowFile

Timestamp of the event

Event whose graph was selected (red color)



What next ???

Let's move to lab session...
(Introduction data pipelines using Apache NiFi)

References

1. Lyko, Klaus, Marcus Nitzschke, and Axel-Cyrille Ngonga Ngomo. "Big data acquisition." *New Horizons for a Data-Driven Economy*. Springer, Cham, 2016. 39-61.
2. Casale, G., Artač, M., van den Heuvel, W. *et al.* RADON: rational decomposition and orchestration for serverless computing. *SICS Softw.-Inensiv. Cyber-Phys. Syst.* (2019). <https://doi.org/10.1007/s00450-019-00413-w>
3. <http://radon-h2020.eu/>
4. <https://nifi.apache.org/docs/nifi-docs/html/overview.html>
5. <https://nifi.apache.org/docs.html>
6. <https://nifi.apache.org/powered-by-nifi.html>
7. <https://courses.cs.ut.ee/2021/cloud/spring/Main/Practice10>
8. <https://aws.amazon.com/datapipeline/>
9. <https://aws.amazon.com/streaming-data/>
10. Lyko, K., Nitzschke, M., Ngonga Ngomo, AC. (2016). Big Data Acquisition. In: Cavanillas, J., Curry, E., Wahlster, W. (eds) *New Horizons for a Data-Driven Economy*. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_4
11. <https://www.teradata.com/Trends/Data-Management/4-Common-Data-Migration-Mistakes-and-How-to-Avoid-Them>
12. <https://www.experian.co.uk/blogs/latest-thinking/data-quality/8-hurdles-of-a-data-migration/>
13. <https://www.ibm.com/blogs/systems/storage-data-migration-101-online-versus-offline-migration/>

Thank you