

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or a data network, extending from the top to the bottom.

PROJET 8

OPENCLASSROOMS: FORMATION
INGÉNIEUR DATA



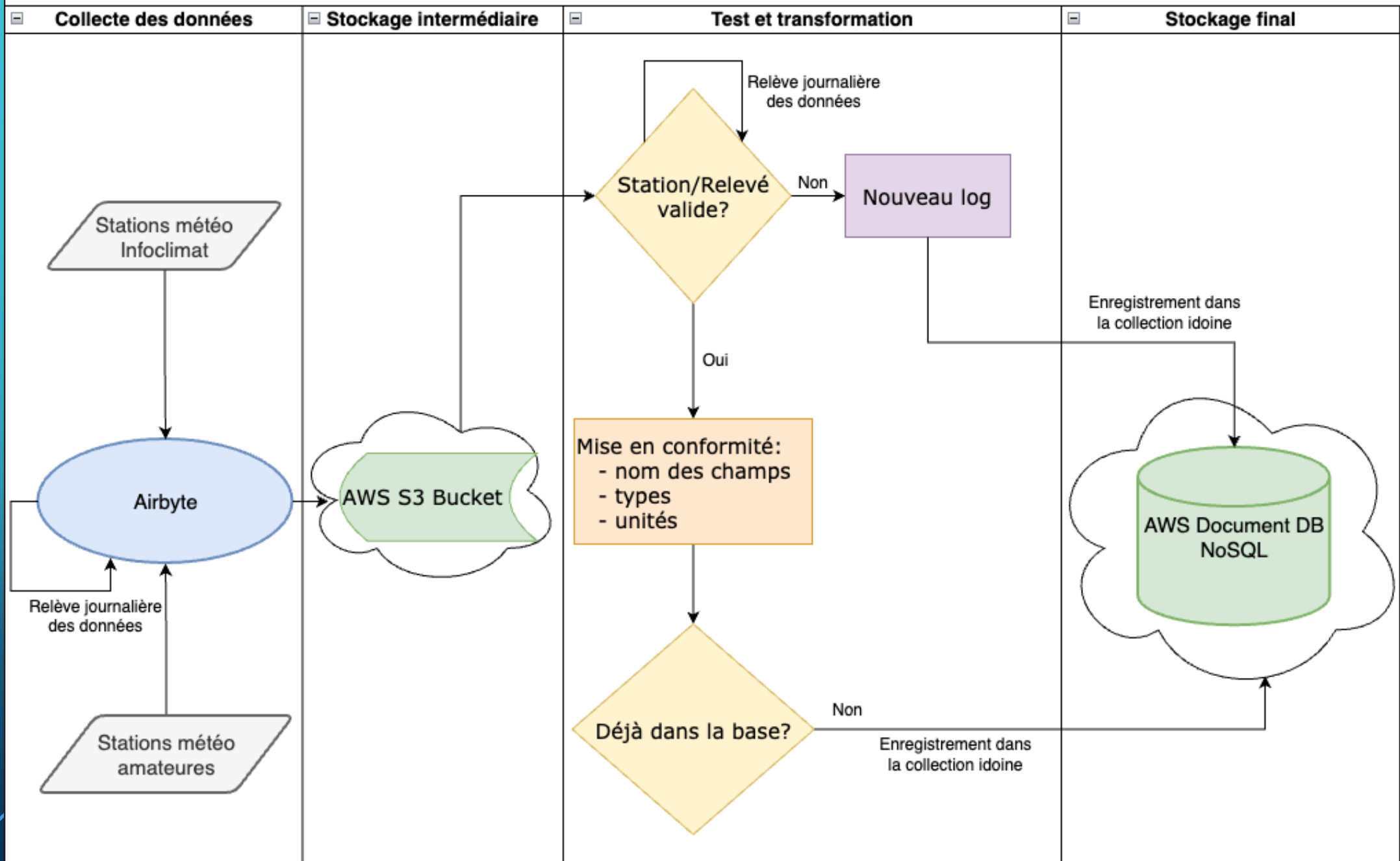
Contexte:

Data Engineer dans l'entreprise GreenAndCoop, un fournisseur coopératif français d'électricité d'origine renouvelable dans les Hauts-de-France.

But du projet:

Intégrer dans la base des données météorologiques de nouvelles sources issues de différents serveurs.

Logigramme



SYNCHRONISATION DES DONNÉES SUR UN BUCKET S3 AVEC AIRBYTE

localhost:8000/workspaces/cd19c9ac-b721-443f-a5b5-6f0c03892ce4/connections

Airbyte

- Connections
- Sources
- Destinations
- Builder
- Settings

Connections

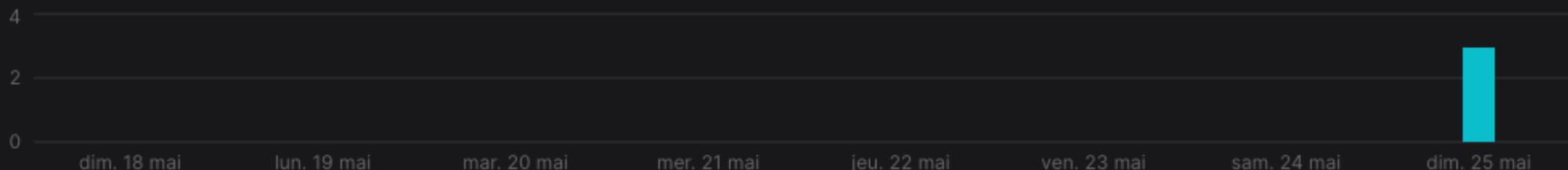
3 healthy

+ New connection

Search

All connections ▾ All statuses ▾ All sources ▾ All destinations ▾ Tags ▾

6h 24h 7d 30d



NAME	SOURCE NAME	DESTINATION NAME	FREQUENCY	TAGS	LAST SYNC	ENABLED
meteo_ichtegem → ...	meteo_ichtegem	S3-meteo-nau-csv	Manual		il y a 1 heure	<input checked="" type="checkbox"/>
meteo_infoclimat → ...	meteo_infoclimat	S3-meteo-nau-json	Manual		il y a 1 heure	<input checked="" type="checkbox"/>
meteo_madeleine → ...	meteo_madeleine	S3-meteo-nau-csv	Manual		il y a 1 heure	<input checked="" type="checkbox"/>

SERVICES AWS UTILISÉS

AWS

S3 Bucket

Stockage
intermédiaire des
données

AWS ECS

EC2 Instance

Traitement des données

AWS DocumentDB (MONGODB Compatible)

Replicaset

Primary

Secondary 1

Secondary 2

STOCKAGE INTERMÉDAIRE SUR UN BUCKET S3

The screenshot shows the Amazon S3 console interface. The top navigation bar includes the AWS logo, a search bar, and regional information (Europe (Paris)). The breadcrumb trail indicates the current location: Amazon S3 > Buckets > meteo-nau81 > meteo_sync/. The left sidebar contains a menu for 'Amazon S3' with options like 'General purpose buckets', 'Directory buckets', 'Table buckets', 'Access Grants', 'Access Points', 'Object Lambda Access Points', 'Multi-Region Access Points', 'Batch Operations', and 'IAM Access Analyzer for S3'. Below this is a section for 'Block Public Access settings for this account' and 'Storage Lens' with options for 'Dashboards', 'Storage Lens groups', and 'AWS Organizations settings'. The main content area displays the 'meteo_sync/' bucket. It has tabs for 'Objects' (selected) and 'Properties'. A 'Copy S3 URI' button is in the top right. Below the tabs, there are buttons for 'Refresh', 'Copy S3 URI', 'Copy URL', 'Download', 'Open', and 'Delete'. There are also buttons for 'Actions', 'Create folder', and 'Upload'. A text block explains that objects are fundamental entities in Amazon S3 and provides a link to 'Amazon S3 Inventory' for listing objects. Below this is a search bar 'Find objects by prefix' and a table of objects. The table has columns for 'Name', 'Type', and 'Last modified'. It lists three folders: 'ichtegem/', 'infoclimat/', and 'madeleine/'.

Amazon S3 > Buckets > meteo-nau81 > meteo_sync/

Amazon S3

- General purpose buckets
- Directory buckets
- Table buckets
- Access Grants
- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- IAM Access Analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

- Dashboards
- Storage Lens groups
- AWS Organizations settings

meteo_sync/

Copy S3 URI

Objects Properties

Objects (3)

Refresh Copy S3 URI Copy URL Download Open Delete

Actions Create folder Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified
<input type="checkbox"/>	ichtegem/	Folder	-
<input type="checkbox"/>	infoclimat/	Folder	-
<input type="checkbox"/>	madeleine/	Folder	-

ARCHITECTURE DE LA BASE DE DONNÉES

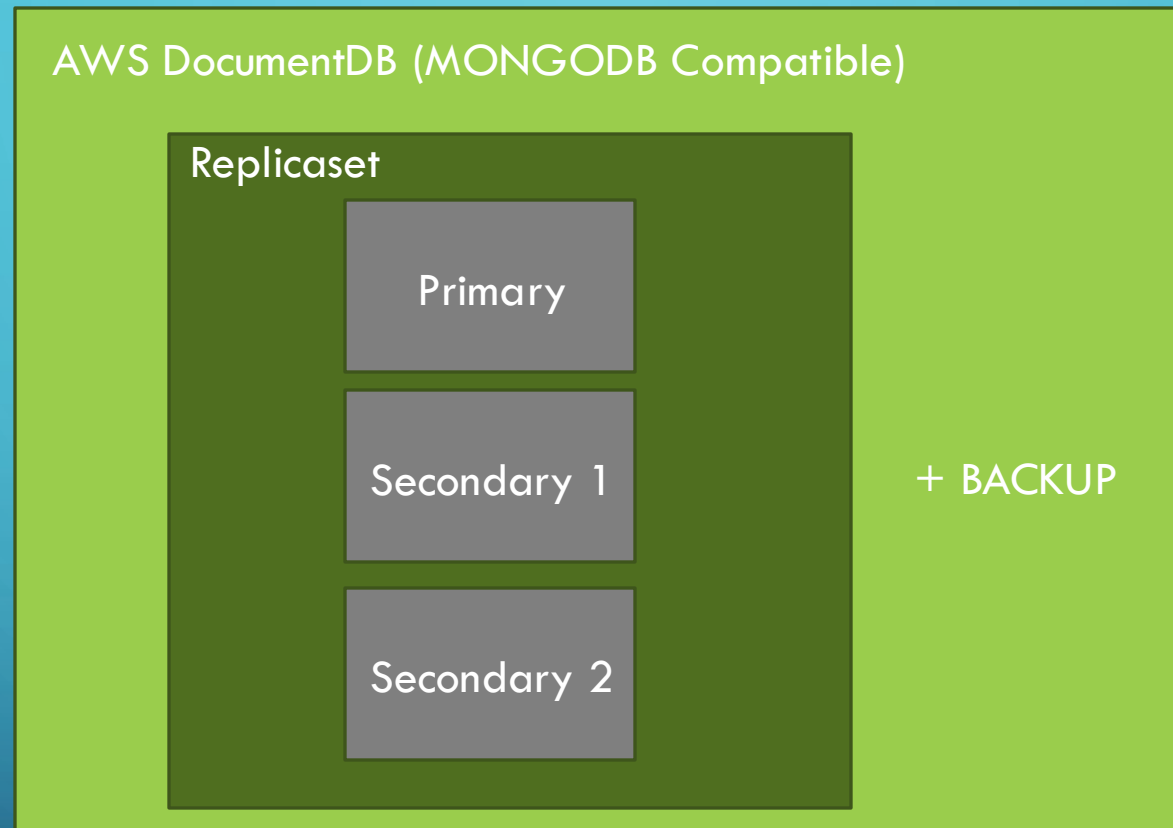


SCHÉMA DE LA BASE DE DONNÉES « METEO »

stations (collection)

```
{
  _id          <string>
  name         <string>
  city         <string>
  latitude     <double>
  longitude    <double>
  elevation    <integer>
  type         <string>
  hardware     <string>
  software     <string>
  license_info <string>
  reports      {...}
}
```

logs (collection)

```
{
  _id          <string>
  timestamp    <date>
  type         <string>
  msg          <string>
}
```

```
{
  utc_time          <date>
  temperature       <double>
  dew_point         <double>
  humidity          <integer>
  wind_direction    <string>
  mean_wind_speed   <double>
  wind_gust         <double>
  pressure          <double>
  precipitation_3h   <double>
  precipitation_1h   <double>
  precipitation_acc  <double>
  precipitation_rate <double>
  snow_depth        <integer>
  ncloud_cover      <integer>
  temps_omm         <double>
  uv               <integer>
  solar            <double>
}
```



EC2 INSTANCE

Tous les jours à minuit (via systemd timer)



- Récupération des données depuis S3 Bucket
- Traitement des données, enregistrement des nouveaux relevés
- Enregistrement des erreurs éventuelles de traitement





EC2 INSTANCE

  [Option+S]

 [EC2](#) > Instances

Instances (1/1) [Info](#)

All states ▼

<input checked="" type="checkbox"/>	Name  ▼	Instance ID	Instance state ▼	Instance type ▼	Status check	Alarm status	Availability Zone ▼
<input checked="" type="checkbox"/>	meteo-nau-1	i-098c96ee81861bb0c	 Running  	t2.micro	 2/2 checks passec	View alarms +	eu-west-3a






TEMPS D'ACCESSIBILITÉ


Chercher tous les relevés du 01/10/2024 de la station d'Ichtegem



288 documents in 0.000027 seconds

AMAZON DOCUMENTDB

  [Option+S]   





 DocumentDB > Clusters


Amazon DocumentDB <

[Dashboard](#)
[Clusters](#)
[Performance Insights](#)
[Snapshots](#)

[Subnet groups](#)
[Parameter groups](#)

Clusters (1)

<input type="checkbox"/>	Cluster identifier	Status	Role	Engine version	Region & AZ	Instance health	CPU
<input type="checkbox"/>	 nau-meteo-cluster	✓ Available	Regional cluster	5.0.0	eu-west-3	-	-
<input type="checkbox"/>	 nau-meteo-primary	✓ Available	Primary Instance	5.0.0	eu-west-3a	✓ Healthy	<div><div></div></div> 16.01%
<input type="checkbox"/>	 nau-meteo-replica-1	✓ Available	Replica Instance	5.0.0	eu-west-3b	✓ Healthy	<div><div></div></div> 15.71%
<input type="checkbox"/>	 nau-meteo-replica-2	✓ Available	Replica Instance	5.0.0	eu-west-3b	✓ Healthy	<div><div></div></div> 15.44%

 ☒ Group Resources

DOCUMENTDB BACKUP

Maintenance

Maintenance window
mon:04:00-mon:05:00 UTC (GMT)

Pending maintenance
Available

Pending Maintenance (1)



Apply now

Apply at next main

🔍 Filter pending maintenance

	Description ▲	Type	Status ▼	Apply date ▼
<input type="radio"/>	Bug Fixes	system-update	-	Mon Jun 09 2025

Backup

Automated backups
Enabled (1 day)

Latest restore time
6/7/2025, 4:05:04 PM UTC+2

Copy tags to backups
Disabled

Backup window
09:55-10:25 UTC (GMT)

Snapshots (1)

🔍 Filter snapshots

	Snapshot I... ▲	Snapshot creation time ▼	Status ▼	Progress ▼	VPC ▼	Type ▼	Engine vers... ▼	Storage
<input type="radio"/>	rds:nau-meteo...	6/7/2025, 11:59:16 AM UTC+2	✅ Available	Completed	vpc-098255f5...	automated	docdb 5.0.0	0

DOCUMENTDB MONITORING

Amazon DocumentDB <

Dashboard

Clusters

Performance Insights

Snapshots

Subnet groups

Parameter groups

Event Subscriptions

Events

Recommendations

No-code machine learning

What's New

Tutorials

nau-meteo-cluster

▼ Summary

Engine version

docdb 5.0.0

Cluster status

Available

Instance status

3 / 3 Instances are available

NVMe-backed Instances

0 Instances

Memory optimized Instances

3 Instances

Pending maintenance

Available

Connectivity & security

Instances

Configuration

Monitoring

Events & tags

Maintenance & backups

CloudWatch

Resource Utilization

1h3h12h1d3d1wCustomUTC time zone

VolumeBytesUsed (Bytes)

Bytes

71.2M

35.6M

0

06/0406/0506/06

VolumeBytesUsed

TotalBackupStorageBilled (Bytes)

Bytes

1

0.5

0

06/0406/0506/06

TotalBackupStorageBilled

SnapshotStorageUsed (Bytes)

No unit

No data available.

Try adjusting the dashboard time range.

1

0.5

0

06/0406/0506/06

SnapshotStorageUsed

CPUUtilization (Percent)

Percent

32.9

24.1

15.3

06/0406/0506/06

CPUUtilization

FreeableMemory (Bytes)

Bytes

2.86G

2.60G

2.34G

06/0406/0506/06

FreeableMemory

FreeLocalStorage (Bytes)

Bytes

8.07G

7.87G

7.68G

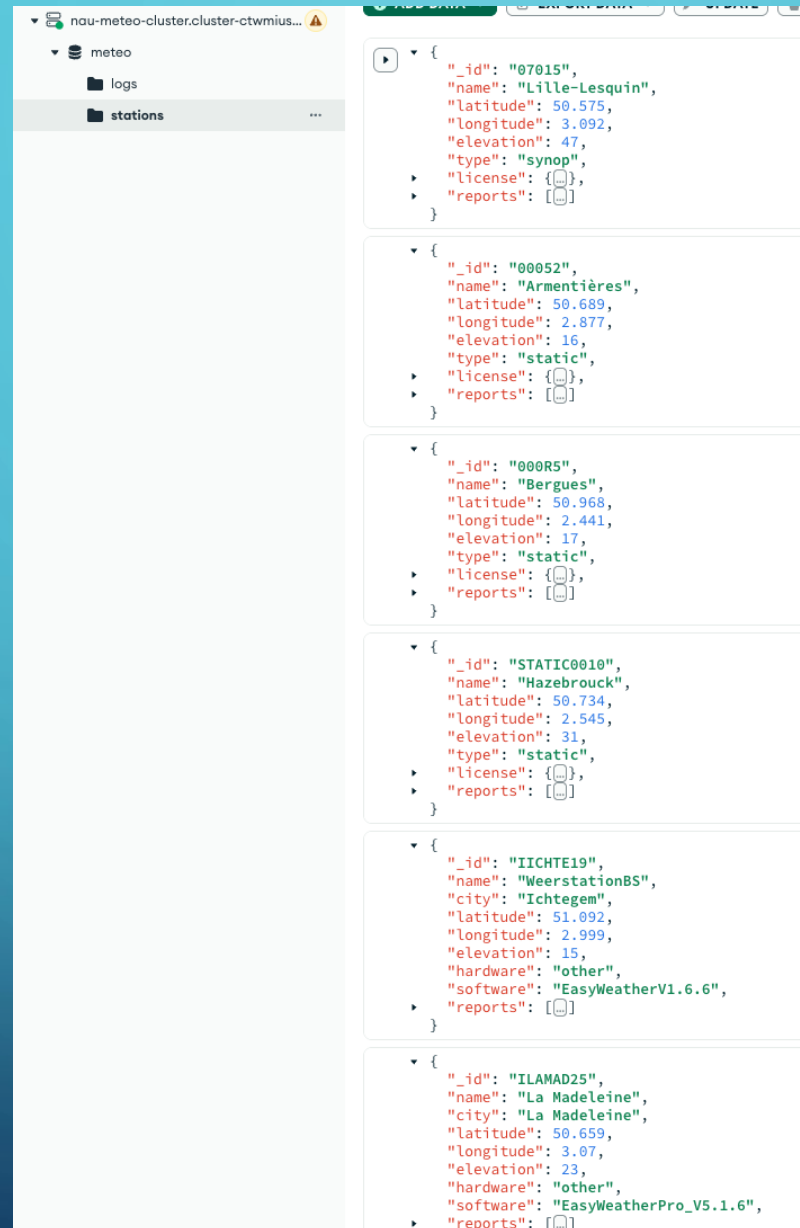
06/0406/0506/06

FreeLocalStorage

DOCUMENTDB STATUS

```
[(base) nicolas@Mac-Nicolas AWS % ./test-replica-status.sh
Statut réplication:
{
  set: 'rs0',
  date: ISODate('2025-06-08T14:13:25.000Z'),
  myState: 1,
  members: [
    {
      _id: 0,
      name: 'nau-meteo-primary.ctwmius4qph4.eu-west-3.docdb.amazonaws.com:27017',
      health: 1,
      state: 1,
      stateStr: 'PRIMARY',
      self: true,
      uptime: 345962
    },
    {
      _id: 1,
      name: 'nau-meteo-replica-1.ctwmius4qph4.eu-west-3.docdb.amazonaws.com:27017',
      health: 1,
      state: 2,
      stateStr: 'SECONDARY',
      uptime: 345962
    },
    {
      _id: 2,
      name: 'nau-meteo-replica-2.ctwmius4qph4.eu-west-3.docdb.amazonaws.com:27017',
      health: 1,
      state: 2,
      stateStr: 'SECONDARY',
      uptime: 345962
    }
  ],
  ok: 1,
  operationTime: Timestamp({ t: 1749392005, i: 1 })
}
```

MONGODB COMPASS



The screenshot shows the MongoDB Compass interface. On the left, the database structure is displayed: 'meteo' > 'stations'. The main panel shows a list of documents from the 'stations' collection. Each document represents a weather station with fields for _id, name, latitude, longitude, elevation, type, license, and reports. The documents are expanded to show their full JSON structure.

```
{
  "_id": "07015",
  "name": "Lille-Lesquin",
  "latitude": 50.575,
  "longitude": 3.092,
  "elevation": 47,
  "type": "synop",
  "license": {},
  "reports": []
}
```

```
{
  "_id": "00052",
  "name": "Armentières",
  "latitude": 50.689,
  "longitude": 2.877,
  "elevation": 16,
  "type": "static",
  "license": {},
  "reports": []
}
```

```
{
  "_id": "000R5",
  "name": "Bergues",
  "latitude": 50.968,
  "longitude": 2.441,
  "elevation": 17,
  "type": "static",
  "license": {},
  "reports": []
}
```

```
{
  "_id": "STATIC0010",
  "name": "Hazebrouck",
  "latitude": 50.734,
  "longitude": 2.545,
  "elevation": 31,
  "type": "static",
  "license": {},
  "reports": []
}
```

```
{
  "_id": "IICHTE19",
  "name": "WeerstationBS",
  "city": "Ichtegem",
  "latitude": 51.092,
  "longitude": 2.999,
  "elevation": 15,
  "hardware": "other",
  "software": "EasyWeatherV1.6.6",
  "reports": []
}
```

```
{
  "_id": "ILAMAD25",
  "name": "La Madeleine",
  "city": "La Madeleine",
  "latitude": 50.659,
  "longitude": 3.07,
  "elevation": 23,
  "hardware": "other",
  "software": "EasyWeatherPro_V5.1.6",
  "reports": []
}
```