

28 Janvier 2024



Année Universitaire 2023-2024

Compte rendu

R5.10 : Mini projet

Présenté par

Wassim LOMRI

Table des matières

1. Mise en place du stack en local.....	3
2. Initialisation de votre cluster Elasticsearch.....	5
2.1 Création de notre index de données.....	5
2.2 Définition du mapping de nos données.....	5
2.3 Importer les données dans notre index.....	7
3. Exploration des données à travers quelques requêtes.....	7
4. Construction de 2 Dashboards Kibana.....	15
4.1 Configuration de votre Space et de votre Data View dans Kibana.....	15
4.2 Construction de 2 dashboards.....	17
Dashboard « Synthèse des données ».....	17
Dashboard « Analyse des retards ».....	18

1. Mise en place du stack en local

Nous avons mis en place **une stack avec un cluster Elasticsearch avec 2 nodes et un serveur Kibana** qui peut être initialisée grâce une simple commande « docker-compose up » (Si code erreur Exit 78 exécuter la commande « sudo sysctl -w vm.max_map_count=262144 »).

Nous avons fais un premier service **‘es001’** qui est associé au conteneur **“conteneur_es001”** basé sur l’image officielle “elasticsearch:8.11.3” et correspondant à un node nommé **“node_es001”**, avec comme cluster « cluster_lomriw » sur le port 9200 en local.

Par la suite, nous avons fait un deuxième service **‘es002’** qui est associé au conteneur **“conteneur_es002”** basé sur l’image officielle “elasticsearch:8.11.3” et correspondant à un node nommé **“node_es002”** avec le même cluster que le service d’avant sur le port 9201 en local.

Pour Kibana, nous avons fais un troisième service **“kibana000”** qui est associé à un conteneur **“conteneur_kibana000”** basé sur l’image officielle “kibana:8.11.3” et **qui dépend des 2 précédents service “es001” et “es002”**, ce serveur à pour nom « **kibana_local** » au port 5600 en local.

Sur ces 3 conteneurs, nous avons désactiver la sécurité et la limité de la mémoire alloué à la heap de la JVM en définissant deux variables d’environnement :

- xpack.security.enabled=false
- "ES_JAVA_OPTS=-Xms512m -Xmx512m"

Fichier docker-compose.yml :

```
version: '3'

services:
# 1er service : Conteneur Elasticsearch "conteneur_es001"
es001:
image: elasticsearch:8.11.3
container_name: conteneur_es001
environment:
- node.name=node_es001
- cluster.name=cluster_lomriw
- cluster.initial_master_nodes=node_es001,node_es002
- discovery.seed_hosts=conteneur_es002
- xpack.security.enabled=false
- "ES_JAVA_OPTS=-Xms512m -Xmx512m"
ports:
- 9200:9200

# 2nd service : Conteneur Elasticsearch "conteneur_es002"
es002:
image: elasticsearch:8.11.3
container_name: conteneur_es002
environment:
- node.name=node_es002
- cluster.name=cluster_lomriw
- cluster.initial_master_nodes=node_es001,node_es002
```

```
- discovery.seed_hosts=conteneur_es001
- xpack.security.enabled=false
- "ES_JAVA_OPTS=-Xms512m -Xmx512m"
```

ports:

```
- 9201:9200
```

3ème service : Conteneur Kibana "conteneur_kibana000"

kibana000:

image: kibana:8.11.3

container_name: conteneur_kibana000

depends_on:

```
- es001
```

```
- es002
```

environment:

```
- SERVERNAME=kibana_local
```

```
- ELASTICSEARCH_HOSTS=http://conteneur_es001:9200
```

```
- "ES_JAVA_OPTS=-Xms512m -Xmx512m"
```

ports:

```
- 5601:5601
```

The screenshot displays the Elasticvue web interface for a cluster named 'cluster_lomriw'. The interface includes a navigation bar with links to HOME, NODES, SHARDS, INDICES, SEARCH, REST, and SNAPSHOTS. The main content area is divided into several sections:

- Cluster Overview:** Shows the cluster name 'cluster_lomriw' with ID 'tQVE8UtIS0OpZ-ZFg7G9Pw' and a green status. It also displays key metrics: 2 nodes (2 master, 2 data), 56 shards (28 primaries, 28 replicas), and 28 indices (153 docs, 5.94 MB on disk).
- Cluster Information:** A table listing various cluster details:

Node Name	node_es001
Cluster name	cluster_lomriw
Cluster uuid	ycsoYtEBQR6Ttf8f3XuiCA
Tagline	You Know, for Search
Version	
Number	8.11.3
Build flavor	default
Build type	docker
Build hash	64cf052f3b56b1fd4449f5454cb88aca7e739d9a
Build date	2023-12-08T11:33:53.634979452Z
Lucene version	9.8.0
Minimum wire compatibility version	7.17.0
Minimum index compatibility version	7.0.0
- Cluster Health:** A table showing the overall health and shard status:

Status	green
Timed out	false
Relocating shards	0
Initializing shards	0
Unassigned shards	0
Delayed unassigned shards	0
Active shards	56
Active shards percent	100.00%
Pending tasks	0
In flight fetch	0

2. Initialisation de votre cluster Elasticsearch

Dans cette partie, nous avons créé un index qui stocke les données, défini un mapping de l'index et importé les données dans notre cluster Elasticsearch.

2.1 Création de notre index de données

Nous avons créé un index de nom « `sample_flights_lomriw` » avec 4 primary shards et le nombre maximum de replicas possible soit 1.

2.2 Définition du mapping de nos données

Nous avons récupéré le fichier « [sample_flights_data.json](#) » pour pouvoir faire le mapping au format JSON à l'aide des propriétés suivantes :

- `boolean`
- `integer`
- `float`
- `keyword`
- `text`
- `date`
- `geo_point`

Fichier `sample_flights_mapping.json` :

```
{
  "properties": {
    "fields": {
      "properties": {
        "FlightNum": {
          "type": "keyword"
        },
        "DestCountry": {
          "type": "keyword"
        },
        "OriginWeather": {
          "type": "keyword"
        },
        "OriginCityName": {
          "type": "keyword"
        },
        "AvgTicketPrice": {
          "type": "float"
        },
        "FlightDelay": {
          "type": "boolean"
        },
        "DestWeather": {
          "type": "keyword"
        }
      }
    }
  }
}
```

```
"Dest": {
  "type": "keyword"
},
"FlightDelayType": {
  "type": "keyword"
},
"OriginCountry": {
  "type": "keyword"
},
"DistanceKilometers": {
  "type": "float"
},
"timestamp": {
  "type": "date",
  "format": "strict_date_optional_time"
},
"DestLocation": {
  "type": "geo_point"
},
"DestAirportID": {
  "type": "keyword"
},
"Carrier": {
  "type": "keyword"
},
"Cancelled": {
  "type": "boolean"
},
"FlightTimeMin": {
  "type": "float"
},
"Origin": {
  "type": "keyword"
},
"OriginLocation": {
  "type": "geo_point"
},
"DestRegion": {
  "type": "keyword"
},
"OriginAirportID": {
  "type": "keyword"
},
"OriginRegion": {
  "type": "keyword"
},
"DestCityName": {
  "type": "keyword"
},
"FlightDelayMin": {
```

```
"type": "integer"
}
}
}
}
}
```

Après avoir créé ce fichier, nous l'avons défini pour l'index « **sample_flights_lomriw** » à l'aide de cette commande :

```
curl localhost:9200/sample_flights_lomriw/_mapping \
-X PUT \
-H "Content-Type: application/json" \
--data @sample_flights_mapping.json
```

2.3 Importer les données dans notre index

Commande pour insérer les données dans `sample_flights_lomriw` :

```
curl localhost:9200/sample_flights_lomriw/_bulk \
-X POST \
-H "Content-Type: application/json" \
--data-binary @sample_flights_data.json
```

Après avoir importé, nous vérifions que nous avons bien 13014 documents avec cette commande :

```
curl "localhost:9200/sample_flights_lomriw/_count?pretty"
```

Page des indices de l'extension Elasticvue :

The screenshot shows the Elasticvue web interface. At the top, there's a navigation bar with links for HOME, NODES, SHARDS, INDICES (which is active), SEARCH, REST, and SNAPSHOTS. Below the navigation bar, the 'Indices' page is displayed. It features a table with the following columns: Name, Health, Status, UUID, Aliases, Shards, Segments, Lucene docs, and Storage. The table contains one entry for the index 'sample_flights_lomriw', which has a 'green' health status, is 'open', and contains 13014 Lucene documents. To the left of the table, there are buttons for 'NEW INDEX' and 'Index templates'. To the right, there's a search bar and a settings icon. At the bottom left of the table area, there's a 'BULK ACTION' button. The footer of the page includes the Elasticvue logo, version information (1.0.3-stable), a changelog link, and a GitHub homepage link.

Name	Health	Status	UUID	Aliases	Shards	Segments	Lucene docs	Storage
sample_flights_lomriw	green	open	aT8n8nx6RdZlRrcSVYXLQ	[]	4p 1r	8	13014	13 MB

3. Exploration des données à travers quelques requêtes

1. Quel est le nom de la compagnie aérienne qui s'est occupée du vol numéro "U95ZN76" ?

```
curl -s -X GET "localhost:9200/sample_flights_lomriw/_search?pretty" \
-H "Content-Type: application/json" \
-d '{
  "query": {
    "match": {
      "FlightNum": "U95ZN76"
    }
  }
}' | jq -r '.hits.hits[0]._source.Carrier'
```

Kibana Airlines

2. Combien de vols avaient un temps ensoleillé au départ et à l'arrivée ("Sunny"), mais ont quand même eu un retard lié à la météo ?

```
curl -X POST "http://localhost:9200/sample_flights_lomriw/_count?pretty" -H
'Content-Type: application/json' -d '{
  "query": {
    "bool": {
      "must": [
        {"term": {"OriginWeather.keyword": "Sunny"}},
        {"term": {"DestWeather.keyword": "Sunny"}},
        {"term": {"FlightDelay": true}}
      ]
    }
  }
}'
{
  "count" : 101,
  "_shards" : {
    "total" : 4,
    "successful" : 4,
    "skipped" : 0,
    "failed" : 0
  }
}
```

3. Quels sont les numéros des 3 vols qui ont fait un trajet de la France aux USA entre le 8 et le 10 janvier 2024 (inclus) ?

```
curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H 'Content-Type: application/json' -d '{t-Type: application/json' -d
'{
  "size": 3,
  "sort": [
    {"timestamp": {"order": "asc"}}
  ],
  "_source": ["FlightNum"],
  "query": {
    "bool": {
      "must": [
        {"term": {"OriginCountry.keyword": "FR"}},
        {"term": {"DestCountry.keyword": "US"}},
        {"range": {"timestamp": {"gte": "2024-01-08T00:00:00", "lte": "2024-01-
```



```
10T23:59:59"}}}
    ]
  }
}
}'
{
  "took" : 8,
  "timed_out" : false,
  "_shards" : {
    "total" : 4,
    "successful" : 4,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 3,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [
      {
        "_index" : "sample_flights_lomriw",
        "_id" : "PbPdT40BrTDmD_eq2z8v",
        "_score" : null,
        "_source" : {
          "FlightNum" : "TLTKOGK"
        },
        "sort" : [
          1704715675000
        ]
      },
      {
        "_index" : "sample_flights_lomriw",
        "_id" : "QbPdT40BrTDmD_eq20Ew",
        "_score" : null,
        "_source" : {
          "FlightNum" : "XN4R38U"
        },
        "sort" : [
          1704832735000
        ]
      },
      {
        "_index" : "sample_flights_lomriw",
        "_id" : "N7PdT40BrTDmD_eq20Ix",
        "_score" : null,
        "_source" : {
          "FlightNum" : "L4B4GPZ"
        },
        "sort" : [
          1704897784000
        ]
      }
    ]
  }
}
```

```
}  
]
```

4. Quels sont les numéros et les villes de destination des 5 vols des compagnies autres que "Logstash Airways" qui ont fait France - Canada sur toute la période des données ?

```
curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H 'Content-Type: application/json' -d '{  
  "size": 5,  
  "_source": ["FlightNum", "DestCityName"],  
  "query": {  
    "bool": {  
      "must": [  
        {"term": {"OriginCountry.keyword": "FR"}},  
        {"term": {"DestCountry.keyword": "CA"}},  
        {"range": {"timestamp": {"gte": "2000-01-01T00:00:00", "lte": "2100-01-01T23:59:59"}}},  
        {"bool": {"must_not": [{"term": {"Carrier.keyword": "Logstash Airways"}}]}}  
      ]  
    }  
  }  
}'  
{  
  "took" : 5,  
  "timed_out" : false,  
  "_shards" : {  
    "total" : 4,  
    "successful" : 4,  
    "skipped" : 0,  
    "failed" : 0  
  },  
  "hits" : {  
    "total" : {  
      "value" : 5,  
      "relation" : "eq"  
    },  
    "max_score" : 8.156712,  
    "hits" : [  
      {  
        "_index" : "sample_flights_lomriw",  
        "_id" : "QLPdT40BrTDmD_eq2yUf",  
        "_score" : 8.156712,  
        "_source" : {  
          "FlightNum" : "PYK4TVQ",  
          "DestCityName" : "Winnipeg"  
        }  
      },  
      {  
        "_index" : "sample_flights_lomriw",
```

```

    "_id" : "pLPdT40BrTDmD_eq2zMp",
    "_score" : 8.032415,
    "_source" : {
      "FlightNum" : "MUFJ77J",
      "DestCityName" : "Edmonton"
    }
  },
  {
    "_index" : "sample_flights_lomriw",
    "_id" : "3rPdT40BrTDmD_eq2z8w",
    "_score" : 8.032415,
    "_source" : {
      "FlightNum" : "0G503MR",
      "DestCityName" : "Edmonton"
    }
  },
  {
    "_index" : "sample_flights_lomriw",
    "_id" : "bLPdT40BrTDmD_eq2yYg",
    "_score" : 7.805272,
    "_source" : {
      "FlightNum" : "XH9H5H3",
      "DestCityName" : "Edmonton"
    }
  },
  {
    "_index" : "sample_flights_lomriw",
    "_id" : "VbPdT40BrTDmD_eq2yki",
    "_score" : 7.805272,
    "_source" : {
      "FlightNum" : "V1TCR2A",
      "DestCityName" : "Edmonton"
    }
  }
]
}
}

```

5. Quels sont les numéros de vol et le nom de l'aéroport de départ des 2 vols qui sont partis à moins de 50 km du London Eye (Latitude = 51.5034 / Longitude = -0.1195) le 09/12/2023 entre 9H et 15H ?

```

curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H
'Content-Type: application/json' -d '{
  "size": 2,
  "_source": ["FlightNum", "Origin"],
  "query": {
    "bool": {
      "must": [
        {"range": {"timestamp": {"gte": "2023-09-12T09:00:00", "lte": "2023-09-12T15:00:00"}}},
        {"geo_distance": {
          "distance": "50km",

```

```

        "OriginLocation": {"lat": 51.5034, "lon": -0.1195}
    }}
]
}
}
}'

```

6. Quels sont les noms des 4 différentes compagnies aériennes présentes en base, avec le nombre de vols par compagnie ?

```

curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H
'Content-Type: application/json' -d '{

```

```

    "size": 0,
    "aggs": {
        "compagnies": {
            "terms": {
                "field": "Carrier.keyword",
                "size": 4
            }
        }
    }
}'

```

```

{
  "took" : 5,
  "timed_out" : false,
  "_shards" : {
    "total" : 4,
    "successful" : 4,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,
      "relation" : "gte"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "compagnies" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 0,
      "buckets" : [
        {
          "key" : "Logstash Airways",
          "doc_count" : 3323
        },
        {
          "key" : "JetBeats",
          "doc_count" : 3261
        }
      ]
    }
  }
}

```

```

    {
      "key" : "Kibana Airlines",
      "doc_count" : 3219
    },
    {
      "key" : "ES-Air",
      "doc_count" : 3211
    }
  ]
}
}
}

```

7. Quels sont les noms des 4 différentes villes de départ des vols qui décollent depuis l'Allemagne ?

```

curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H
'Content-Type: application/json' -d '{

```

```

  "size": 0,
  "query": {
    "match": {
      "OriginCountry.keyword": "DE"
    }
  },
  "aggs": {
    "villes_de_depart": {
      "terms": {
        "field": "OriginCityName.keyword",
        "size": 4
      }
    }
  }
}'

```

```

{
  "took" : 6,
  "timed_out" : false,
  "_shards" : {
    "total" : 4,
    "successful" : 4,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 546,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "villes_de_depart" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 0,
      "buckets" : [
        {

```

```

        "key" : "Frankfurt am Main",
        "doc_count" : 145
    },
    {
        "key" : "Berlin",
        "doc_count" : 135
    },
    {
        "key" : "Cologne",
        "doc_count" : 133
    },
    {
        "key" : "Munich",
        "doc_count" : 133
    }
]
}
}
}

```

8. Combien de vols correspondent aux 3 plages suivantes ?

- **“Aucun retard”** (9744 vols sont dans ce cas)
- **“Retard de [1, 60[min”** (412 vols sont dans ce cas)
- **“Retard de 60 min ou plus”** (2858 vols sont dans ce cas)

```

curl -X POST "http://localhost:9200/sample_flights_lomriw/_search?pretty" -H
'Content-Type: application/json' -d '{
  "size": 0,
  "aggs": {
    "retard_ranges": {
      "range": {
        "field": "FlightDelayMin",
        "ranges": [
          {"from": 0, "to": 1, "key": "Aucun retard"},
          {"from": 1, "to": 60, "key": "Retard de [1, 60[ min"},
          {"from": 60, "key": "Retard de 60 min ou plus"}
        ]
      }
    }
  }
}'
{
  "took" : 10,
  "timed_out" : false,
  "_shards" : {
    "total" : 4,
    "successful" : 4,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,

```

```

        "relation" : "gte"
    },
    "max_score" : null,
    "hits" : [ ]
},
"aggregations" : {
    "retard_ranges" : {
        "buckets" : [
            {
                "key" : "Aucun retard",
                "from" : 0.0,
                "to" : 1.0,
                "doc_count" : 9744
            },
            {
                "key" : "Retard de [1, 60[ min",
                "from" : 1.0,
                "to" : 60.0,
                "doc_count" : 412
            },
            {
                "key" : "Retard de 60 min ou plus",
                "from" : 60.0,
                "doc_count" : 2858
            }
        ]
    }
}
}
}

```

4. Construction de 2 Dashboards Kibana

Dans cette partie dédié à Kibana, nous avons configuré notre espace où nous avons défini notre **Data View** adaptée et deux dashboard pour mettre en forme nos données sur Elasticsearch.

4.1 Configuration de votre Space et de votre Data View dans Kibana

J'ai en premier lieu créé mon Espace Wassim Lomri, qui ont donnée les initiales WL, en laissant les deux fonctionnalité : **Analytics** et **Management**.

Je suis passé sur mon nouvelle espace et j'ai défini mon DataView qui a pour nom **"FLIGHTS_LOMRIW"** basé sur mon index **"sample_flights_lomriw"** en définissant le Timestamp field comme étant **la propriété "timestamp" de mon index** et configurer avec ces critères :

- Les valeurs de **timestamp** doivent s'afficher au format de date à la française.
Exemple pour le 1er février 2003 à 12H45 et 30sec => "01/02/2003 12:45:30".

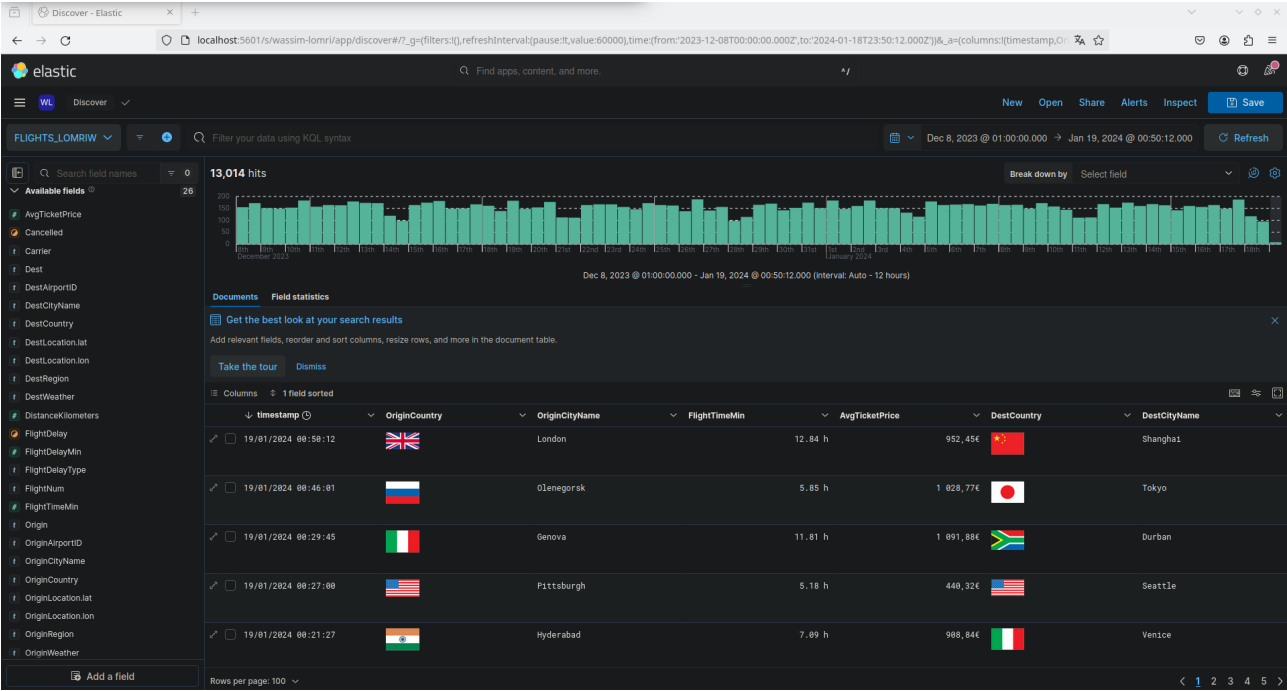
- Les valeurs de **DistanceKilometers** doivent s’afficher en nombre de kilomètres arrondi à l’entier.
Exemple pour 123.45678 kilomètres => “123”.
- Les valeurs de **FlightTimeMin** doivent s’afficher comme des **durées précises avec 2 chiffres après la virgule et unité courte**.
Exemple pour 63 minutes => “1.05 h”.
- Les valeurs de **FlightDelayMin** doivent s’afficher comme des **durées approximatives avec unité**.
Exemples : Pour un vol sans retard => “0 seconds”, et pour un vol avec 2.2 heures de retard => “2 hours”
- Les valeurs de **AvgTicketPrice** doivent s’afficher comme des prix en € au format français.
Exemple pour 1234,56 => “1 234.56 €”.
- Les valeurs de **OriginCountry** et **DestCountry** (qui sont des codes de pays sur 2 lettres) doivent être affichées sous la forme d’**images de drapeaux**.
Les images des drapeaux de tous les pays sont hébergées à l’URL : <http://www-etu-info.iut2.univ-grenoble-alpes.fr/~vinccarl/flags-mini/> avec des fichiers sous la forme “CODE_PAYS.gif”.
La taille des différentes images variant un peu, la taille d’affichage doit être harmonisée à **une largeur de 50 pixels**.

Page configuration Data View avec comme nom « FLIGHTS_LOMRIW » :

The screenshot shows the Elastic Kibana interface for configuring a Data View named 'FLIGHTS_LOMRIW'. The left sidebar contains navigation links for Management, Data, Alerts and Insights, and Kibana. The main panel displays the configuration for the Data View, including the index pattern 'sample_flights_lomriw' and the time field 'timestamp'. Below this, there is a table of fields with columns for Name, Type, Format, Searchable, Aggregatable, and Excluded. The table lists 14 fields, including 'AvgTicketPrice', 'Cancelled', 'Carrier', 'Carrier.keyword', 'Dest', 'Dest.keyword', 'DestAirportID', 'DestAirportID.keyword', 'DestCityName', and 'DestCityName.keyword'. The 'Rows per page' is set to 10, and the page number is 1.

Name	Type	Format	Searchable	Aggregatable	Excluded
AvgTicketPrice	float	Number	•	•	
Cancelled	boolean		•	•	
Carrier	text		•		
Carrier.keyword	keyword		•	•	
Dest	text		•		
Dest.keyword	keyword		•	•	
DestAirportID	text		•		
DestAirportID.keyword	keyword		•	•	
DestCityName	text		•		
DestCityName.keyword	keyword		•	•	

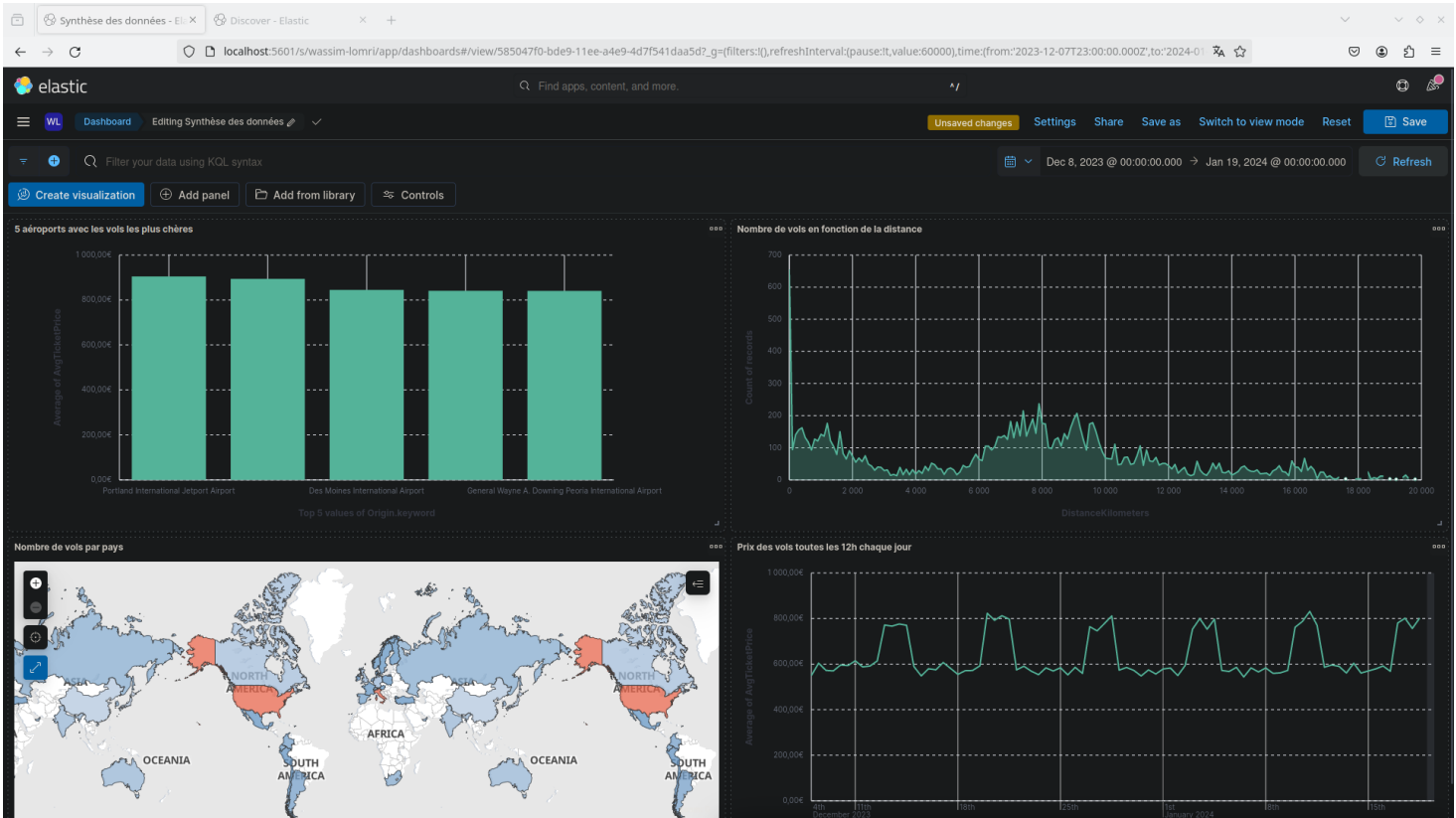
Page discover :



4.2 Construction de 2 dashboards

Dashboard « Synthèse des données »

Ce premier dashboard permet à l'utilisateur d’avoir rapidement et facilement données les plus importantes.



Dashboard « Analyse des retards »

Ce second dashboard permet à l'utilisateur de pouvoir analyser les différents retards enregistrés sur les vols.

