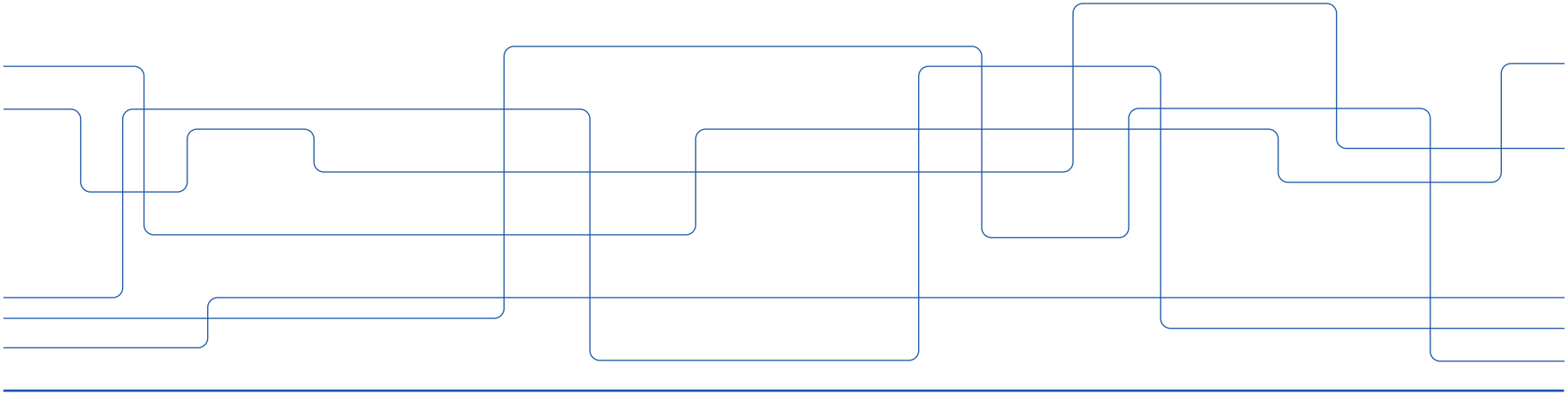
A Real-time Log Correlation System for Security Information and Event Management

Clémence Dubuc

KTH NSS Group

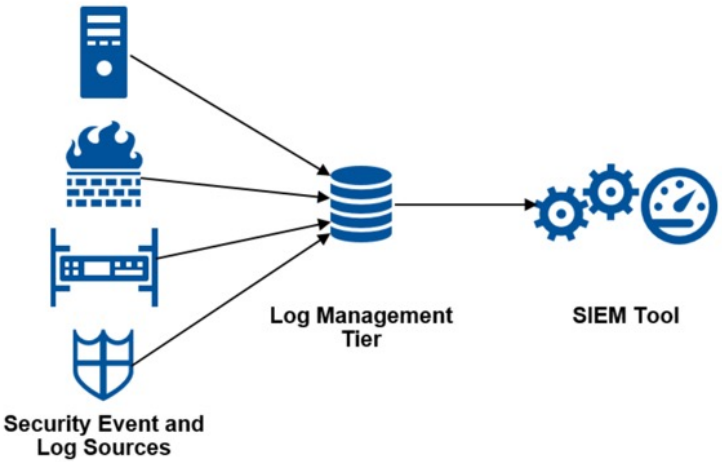


Plan of the presentation

1. Security Information and Event Management
2. Problem Statement
3. Solution implemented based on Apache Spark
4. Solution implemented based on Elastic Search
5. Results
6. Conclusion

Security Information and Event Management

- Collect security logs
- Monitor systems
- Analyse security logs
- Detect potential attacks based on detection rules

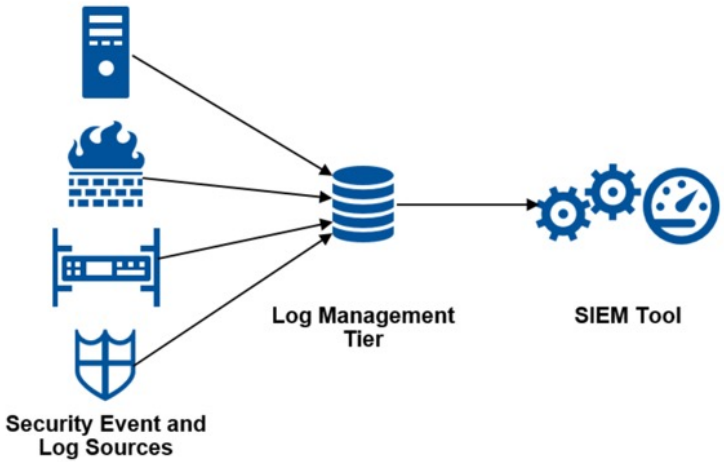


Source: Gartner (October 2016)

Security Information and Event Management

Detection rules

- Written by analysts
- Based on attributes
  - IP address
  - Username
  - DNS name



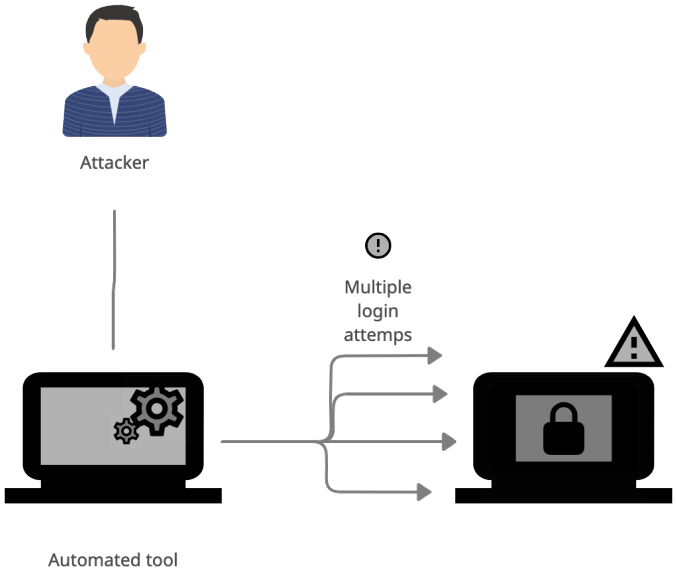
Source: Gartner (October 2016)

Example

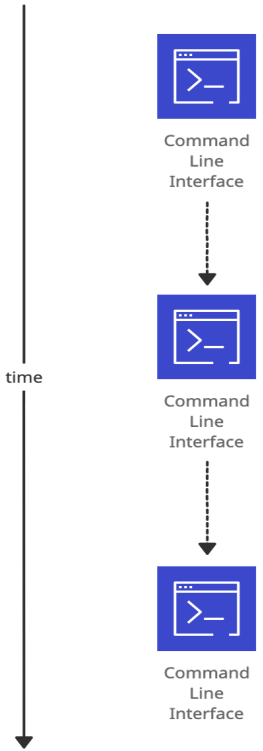
[ipv4-addr:value='127.0.0.0'] AND [process:name='cmd.exe']

Problem Statement: Example

Force Brute



Multi-step attacks

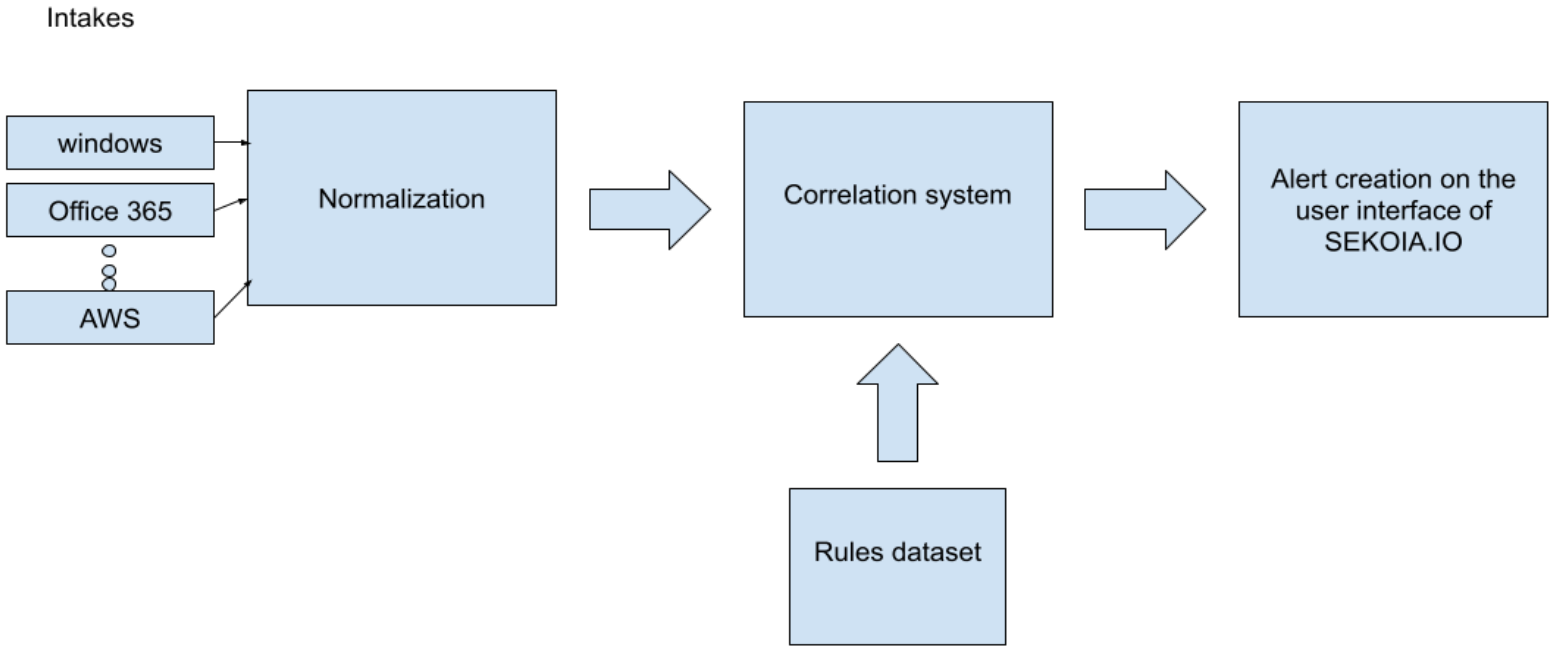


Problem Statement

Implementation of a new correlation system

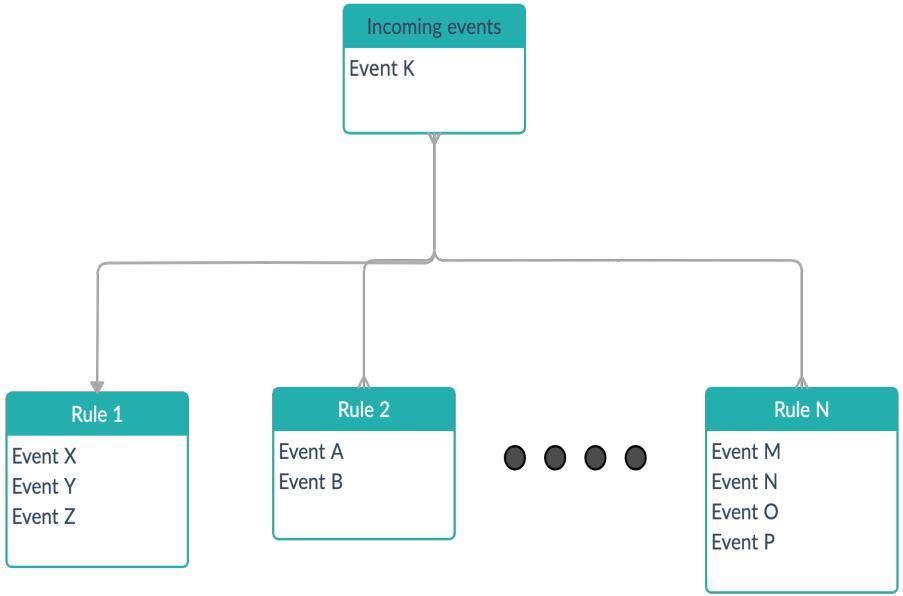
1. Correlating security logs
   - Help reduce false positive
   - Detection of attacks more precise
   - Group events by attributes
2. Scalable system
   - Increasing number of security logs
   - Increasing number of detection rules
3. Real-time system
   - Detection of potential threats in the allowed limit of time

System Model

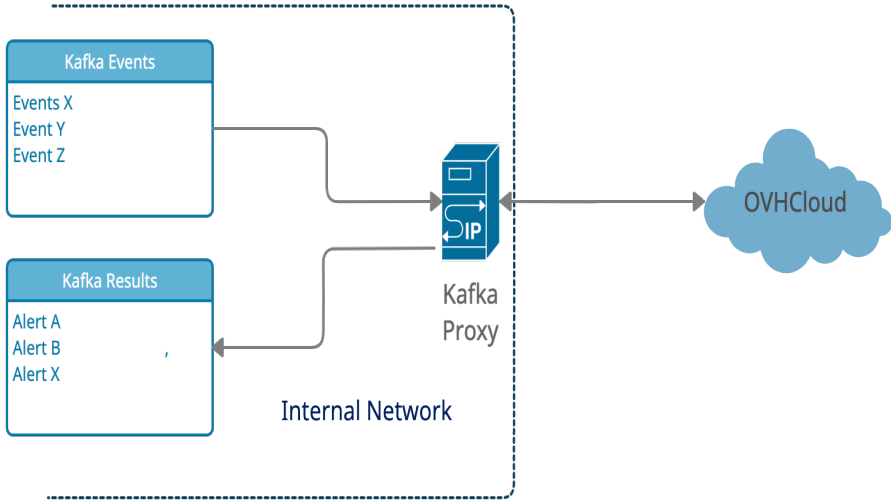


Actual Correlation System

- No correlation on a long period of time
- No correlation on a large number of events
- No aggregation per attributes

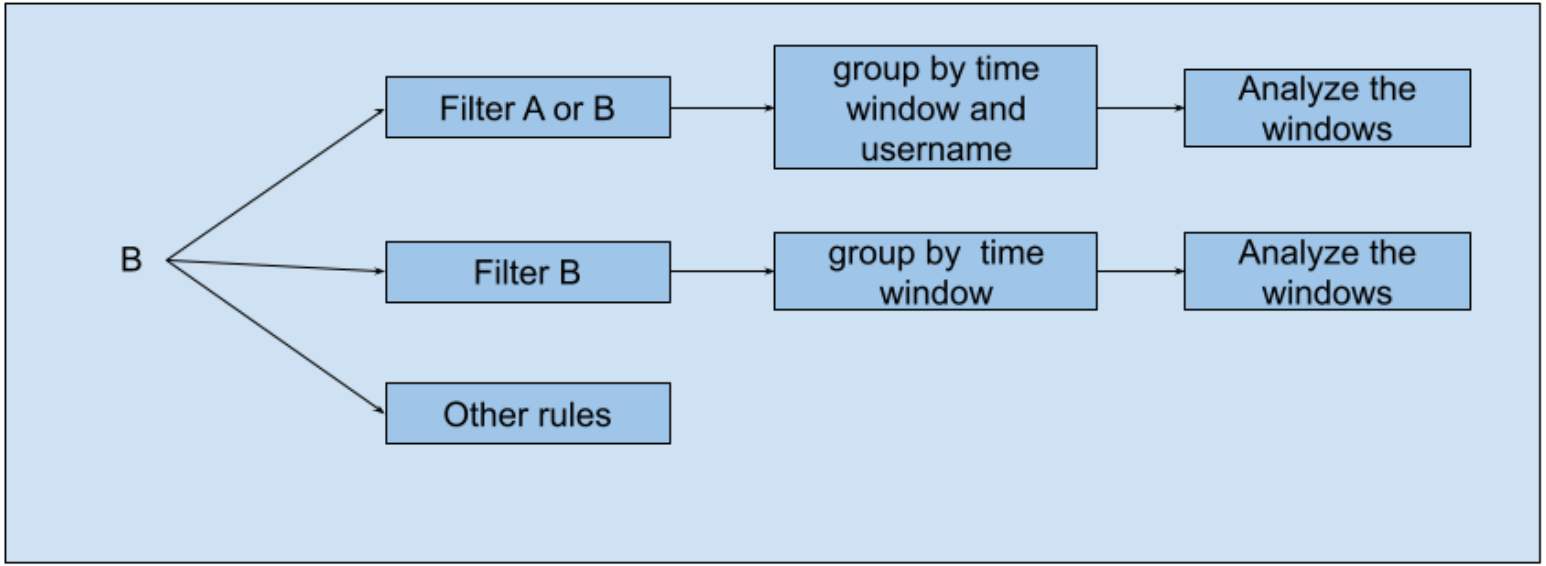


Solution implemented based on Apache Spark



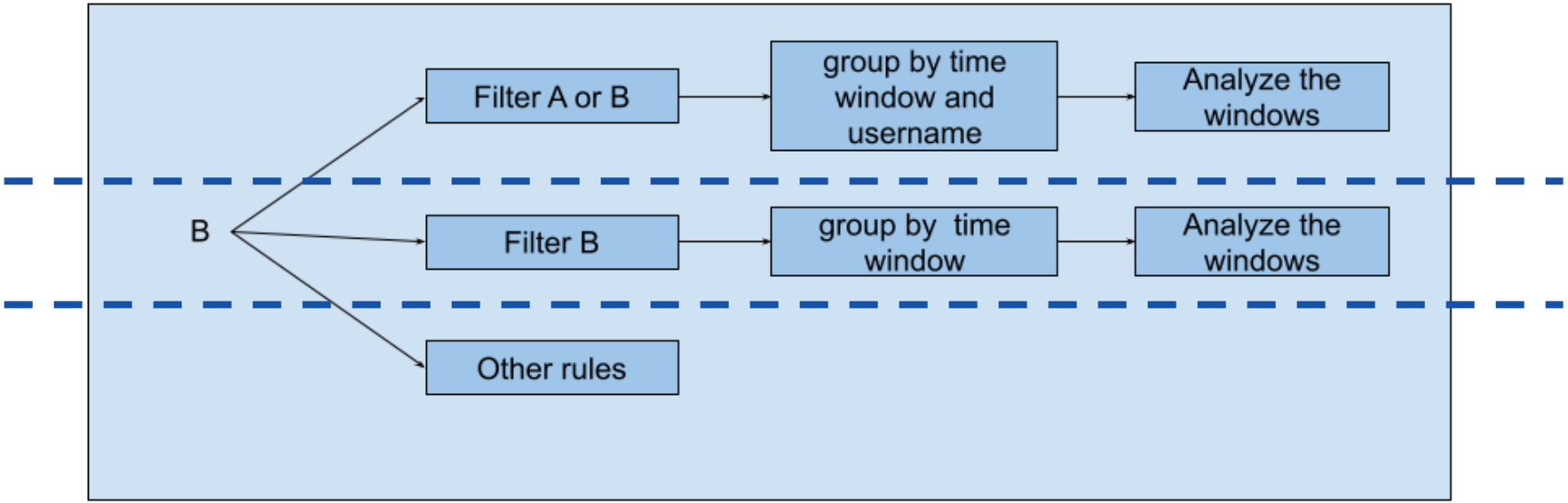
- Apache Spark Structure Streaming: micro batch treatment
- Distributed In-Memory Computation
- Implemented on OVHCloud
- Rules implemented in application run all the time

Solution implemented based on Apache Spark Implementation



A FOLLOWEDBY B GROUPBY username WITHIN 60 SECONDS

Solution implemented based on Apache Spark Implementation



Solution implemented based on Apache Spark

Demonstration

Solution implemented based on Apache Spark

1 rule in 1 application

Activités Éditeur de texte 16 févr. 19:12

Documents Ouvrir *Document 1 sans titre Enregi

test_video.txt

Document 1 sa...

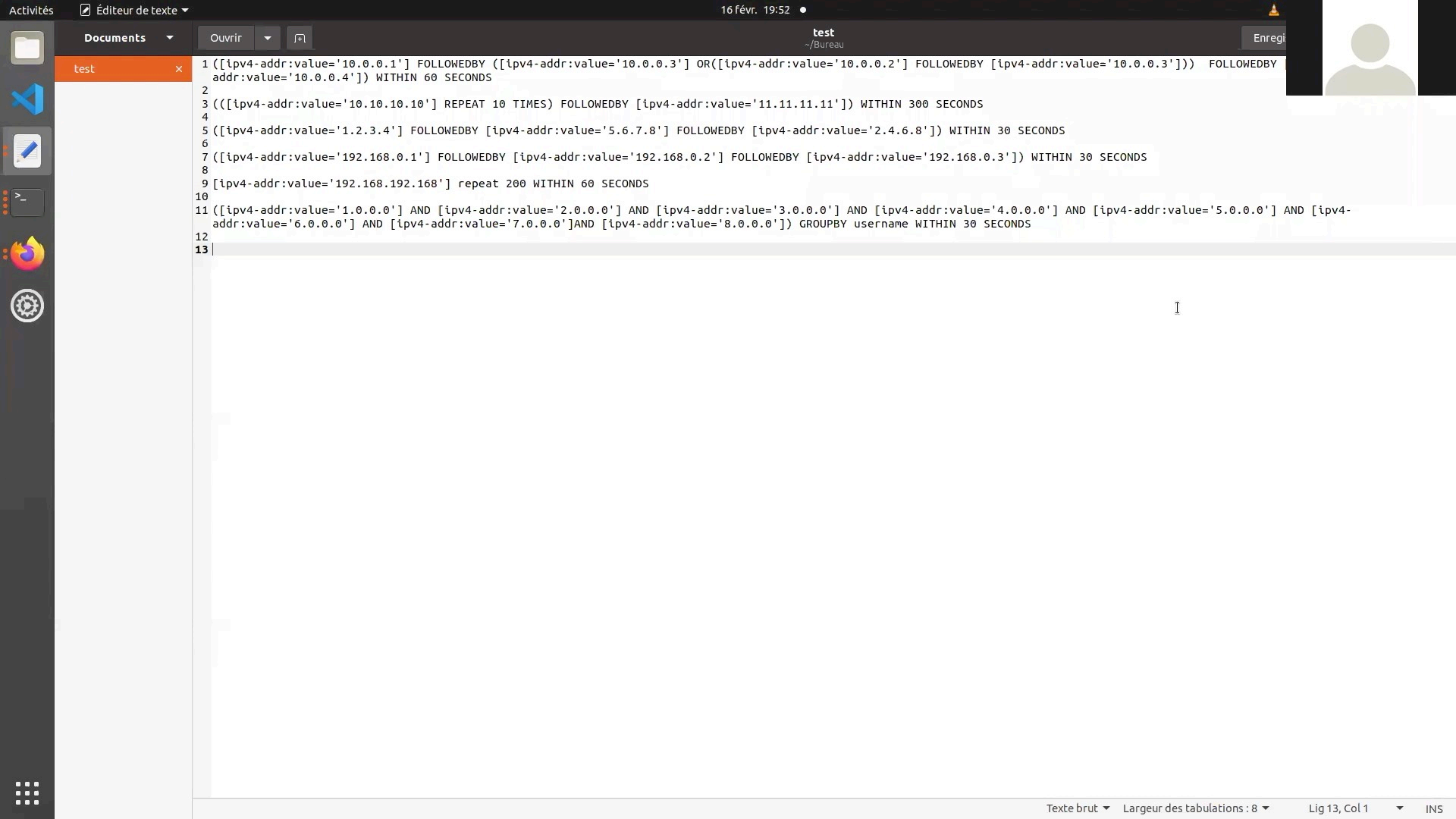
*test_video.txt *Document 1 sans titre

```
1 ([ipv4-addr:value='10.0.0.1'] FOLLOWEDBY ([ipv4-addr:value='10.0.0.3'] OR([ipv4-addr:value='10.0.0.2'] FOLLOWEDBY [ipv4-addr:value='10.0.0.3']))  FOLLOWEDBY [
  addr:value='10.0.0.4']) WITHIN 60 SECONDS
```

Texte brut Largeur des tabulations : 8 Lig 1, Col 205 INS

Solution implemented based on Apache Spark

8 rules in 1 application


Activités
Éditeur de texte
16 févr. 19:52
Documents
Ouvrir
test
~/Bureau
Enregi
test
([ipv4-addr:value='10.0.0.1'] FOLLOWEDBY ([ipv4-addr:value='10.0.0.3'] OR([ipv4-addr:value='10.0.0.2'] FOLLOWEDBY [ipv4-addr:value='10.0.0.3'])) FOLLOWEDBY [ipv4-addr:value='10.0.0.4']) WITHIN 60 SECONDS
(([ipv4-addr:value='10.10.10.10'] REPEAT 10 TIMES) FOLLOWEDBY [ipv4-addr:value='11.11.11.11']) WITHIN 300 SECONDS
([ipv4-addr:value='1.2.3.4'] FOLLOWEDBY [ipv4-addr:value='5.6.7.8'] FOLLOWEDBY [ipv4-addr:value='2.4.6.8']) WITHIN 30 SECONDS
([ipv4-addr:value='192.168.0.1'] FOLLOWEDBY [ipv4-addr:value='192.168.0.2'] FOLLOWEDBY [ipv4-addr:value='192.168.0.3']) WITHIN 30 SECONDS
[ipv4-addr:value='192.168.192.168'] repeat 200 WITHIN 60 SECONDS
([ipv4-addr:value='1.0.0.0'] AND [ipv4-addr:value='2.0.0.0'] AND [ipv4-addr:value='3.0.0.0'] AND [ipv4-addr:value='4.0.0.0'] AND [ipv4-addr:value='5.0.0.0'] AND [ipv4-addr:value='6.0.0.0'] AND [ipv4-addr:value='7.0.0.0']AND [ipv4-addr:value='8.0.0.0']) GROUPBY username WITHIN 30 SECONDS
Texte brut
Largeur des tabulations : 8
Lig 13, Col 1
INS

Solution implemented based on Apache Spark

8 rules in 8 applications

Activités Éditeur de texte 16 févr. 19:24

Documents Ouvrir *Document 1 sans titre Enregi

test_video.txt

Document 1 sa...

*test_video.txt *Document 1 sans titre

```
1 ([ipv4-addr:value='10.0.0.1'] FOLLOWEDBY ([ipv4-addr:value='10.0.0.3'] OR([ipv4-addr:value='10.0.0.2'] FOLLOWEDBY [ipv4-addr:value='10.0.0.3'])) FOLLOWEDBY [
  addr:value='10.0.0.4']) WITHIN 60 SECONDS
2
3 (([ipv4-addr:value='10.10.10.10'] REPEAT 10 TIMES) FOLLOWEDBY [ipv4-addr:value='11.11.11.11']) WITHIN 300 SECONDS
4
5 ([ipv4-addr:value='1.2.3.4'] FOLLOWEDBY [ipv4-addr:value='5.6.7.8'] FOLLOWEDBY [ipv4-addr:value='2.4.6.8']) WITHIN 30 SECONDS
6
7 ([ipv4-addr:value='192.168.0.1'] FOLLOWEDBY [ipv4-addr:value='192.168.0.2'] FOLLOWEDBY [ipv4-addr:value='192.168.0.3']) WITHIN 30 SECONDS
8
9 [ipv4-addr:value='192.168.192.168'] repeat 200 WITHIN 60 SECONDS
10
11 ([ipv4-addr:value='1.0.0.0'] AND [ipv4-addr:value='2.0.0.0'] AND [ipv4-addr:value='3.0.0.0'] AND [ipv4-addr:value='4.0.0.0'] AND [ipv4-addr:value='5.0.0.0'] AND [ipv4-
  addr:value='6.0.0.0'] AND [ipv4-addr:value='7.0.0.0']AND [ipv4-addr:value='8.0.0.0']) GROUPBY username WITHIN 30 SECONDS
12
13
```

Texte brut Largeur des tabulations : 8 Lig 13, Col 1 INS

Solution implemented based on Apache Spark

- Correlate events in real time
- Too many used resources used
- Too expensive for production

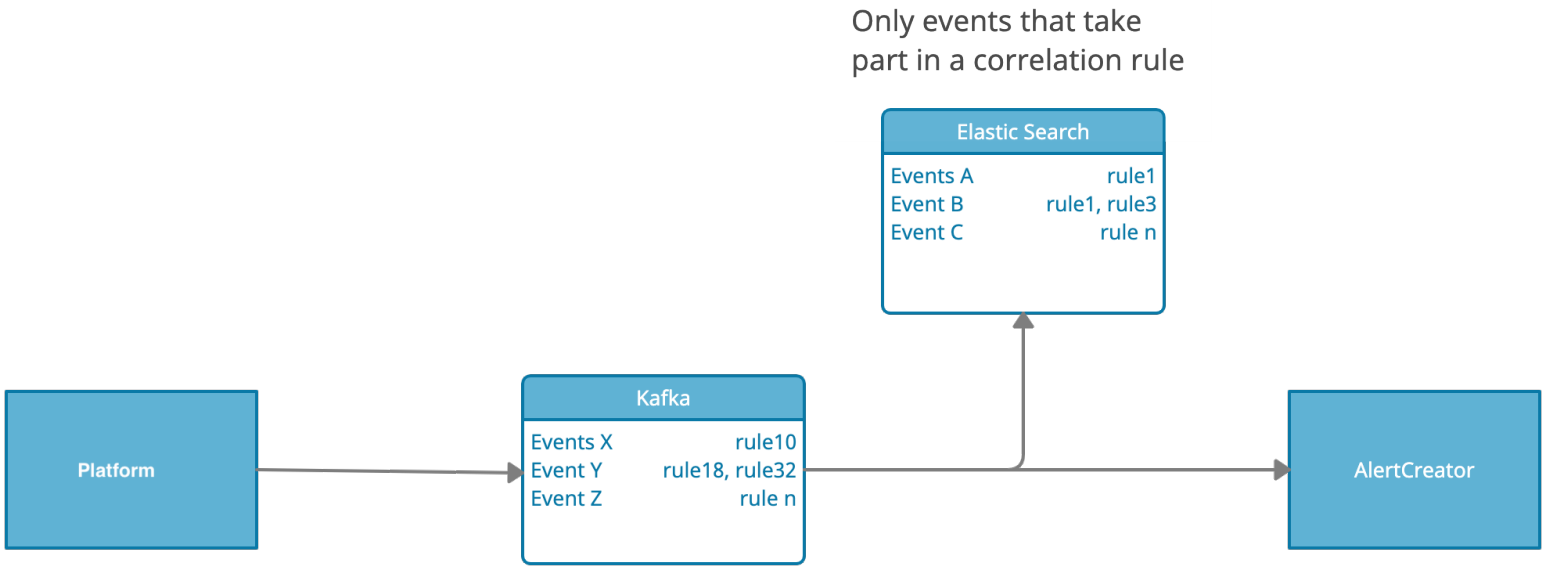
Implementation of another solution

System based on Elastic Search

- Distributed Search & Analysis Engine
- Event store in Elastic Search
- Queries launch when an event associated to them are received

```
{"filter": {
    "range" : {
      "@timestamp" : {
        "gte" : "now-5m",
        "lte" : "now"
      }
    }
  },
  "query": """
  sequence by user.name
      [ any where action.id == "524" ]
      [ any where action.id == "506" ]
      [ any where action.id == "703" ]
  """
}
```

System based on Elastic Search



Results

	Actual System	Apache Spark System	Elastic Search System
GroupBy Operator	No	Yes	Yes
Advantages	Develop in intern	In memory computation and **streaming** computation	Stable storage solution and **optimization** of the run time
Drawbacks	Unpredictable behavior of the rules	**Too many resources** needed to launch hundreds of correlation rules	Maintain Elastic Search cluster and no real-time treatment of the events

Conclusion

1. Correlating security logs
   - The two systems correlate the events
   - The two systems group the events by attributes
2. Scalable system
   - The system based on Apache Spark is scalable in term of number of events but not in term of number of rules
   - The system based on Elastic Search is scalable in both case
3. Real-time system
   - Apache Spark System analyses stream of events
   - Elastic Search system gives an impression of real-time

KTH ROYAL INSTITUTE
OF TECHNOLOGY

Thank you for your attention