



elasticsearch

elasticsearch

# Elasticsearch

Pau, Miguel i Ferran

Sistema de gestió de BD, GEINF

# Minimum Viable Product (MVP)

- Aplicació Android de consulta, compra i venda d'articles d'usuaris.
- Obtenir beneficis a través d'un percentatge de les transaccions o anuncis que es poden afegir a l'aplicació.

## Les 6 W

- What?  
Producte relativament econòmic amb grans capacitats per a obtenir beneficis en el futur.
- Why?  
Globalització. Capitalisme. Senzillesa.
- Where?  
Per regions; continents, països.
- When?  
ARA.
- How?  
Aplicació i registre gratuït.
- Who?  
Qualsevol usuari.

## Què comporta?

- Aplicació ràpida i fàcil de fer servir.
- Disponibilitat i consistència de les dades.
- Una o vàries bases de dades grans.
- Consultes àgils.
- Transaccions eficients.
- Monitorització.

# Entorn de treball

## Dades

- Elasticsearch
- Kibana
- Logstash
- MySQL
- Ngrok

## Aplicació

- IntelliJ IDEA
- Android
- Retrofit 2



---

Data  
Aggregation  
& Processing

Indexing &  
storage

Analysis &  
visualization



elasticsearch

# ElasticSearch

## Que és?

**Motor de cerca i anàlisis** (Java - codi obert)

**Cerques molt ràpides**

**Servei API REST** (Index, Get, Delete, Update)



Ara és el més potent

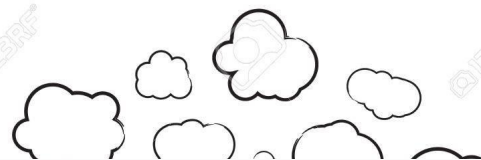
# ElasticSearch per dins

## Com funciona l'estructura?



NOTE

Each Elasticsearch shard is a Lucene index. There is a maximum number of documents you can have in a single Lucene index. As of [LUCENE-5843](#), the limit is 2,147,483,519 (= Integer.MAX\_VALUE - 128) documents. You can monitor shard sizes using the [\\_cat/shards](#) API.





# ElasticSearch

## Característiques

- **Cerques ràpides** en Big Data
- **Visualització de dades** immediata
- **Desnormalització:** proximitat entre els documents guardats = rapidesa
- **Distribuïda i altament escalable**

# AVANTATGES

## ElasticSearch



- **Construït sobre lucene**
- **Autocompletat i recerca instantània**
- **Fuzzy searching:** recerca difusa
- Descobriment de **nous productes i serveis**

# Organitzacions que l'utilitzen

## Mercado Libre

**4 milions** de venedors actius

**20 milions** de productes en temps real

**No cal reindexar** totes les dades cada vegada que s'afegeix un atribut nou



# Organitzacions que l'utilitzen

## Telefónica

Solució per al **creixement** dels serveis de **vídeo sota demanda**

**Identifica problemes** que podrien quedar ocults d'un altre mode



*Telefónica*

**Altres:** Facebook, Activision, IBM, ebay, justEat, tinder ...

# ELK

## Què és Kibana?

- Eina open-source.
- Permet visualitzar dades indexades a Elasticsearch.
- Múltiples packs de software i eines que faciliten tasques als usuaris.
- Monitorització de dades: fluctuacions, rendiment, dashboards, etc.

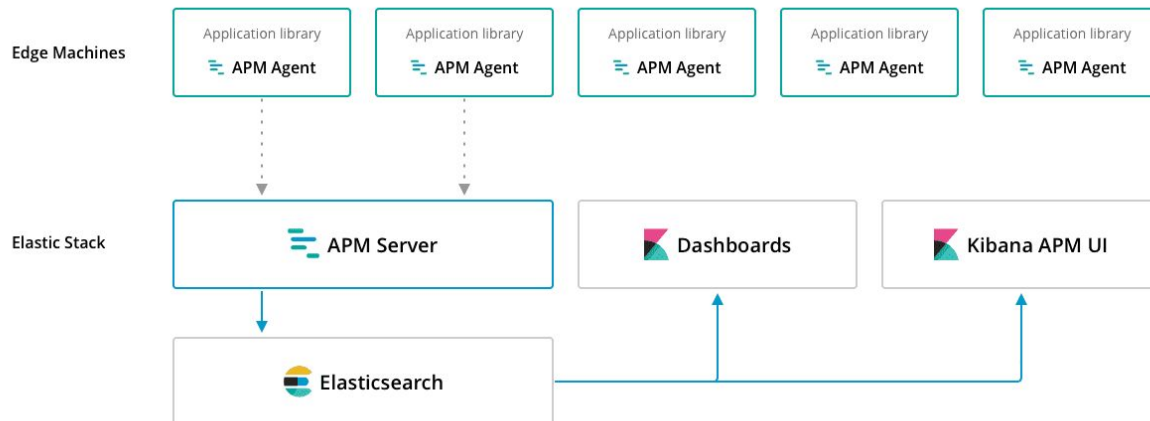


# kibana

# Kibana APM



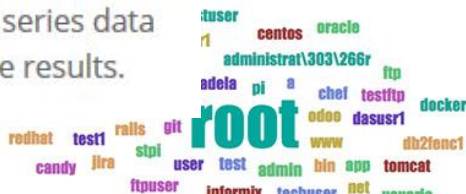
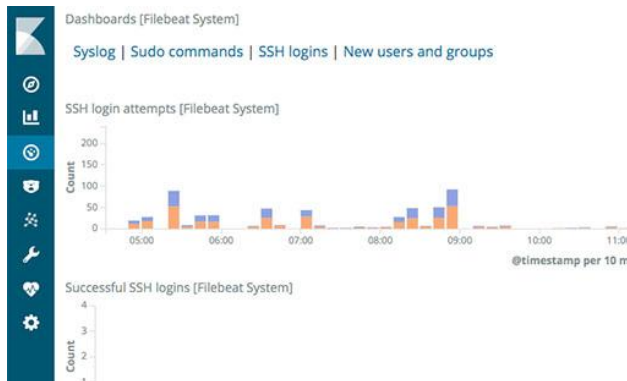
APM



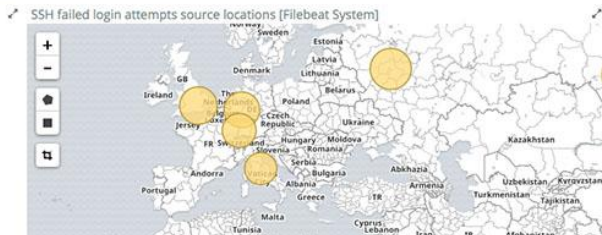
## Security Analytics



Use an expression language to analyze time series data and visualize the results.



## Metrics





```
1 # {dpcindex}/{doctype}/{id}
2
3 # Get les categories amb les subcategories
4 GET categories/categoria/_search
5
6 # Get les subcategories de la categoria Electronics
7 GET categories/categoria/Electronics
8
9 # Get igual que el anterior, pero fent ús de els
  filtres del querydsl
10 GET categories/categoria/_search
11 {
12   "query": {
13     "match": {
14       "_id": "Electronics"
15     }
16   }
17 }
18
19 # Get els productes guardats
20 GET productes/producte/_search
21
22 # Get del producte amb id (substituir id pel que
  calgui)
23 GET productes/producte/12245
24
25
26
27
28
29
```

```
1 {
2   "took": 9,
3   "timed_out": false,
4   "_shards": {
5     "total": 5,
6     "successful": 5,
7     "skipped": 0,
8     "failed": 0
9   },
10   "hits": {
11     "total": 5,
12     "max_score": 1,
13     "hits": [
14       {
15         "_index": "categories",
16         "_type": "categoria",
17         "_id": "Apple CarPlay",
18         "_score": 1,
19         "_source": {
20           "subcategories": "Auto & Tires",
21           "type": "categoria",
22           "nom": "Apple CarPlay",
23           "@version": "1",
24           "@timestamp": "2018-12-18T23:44:00.100Z"
25         }
26       },
27       {
28         "_index": "categories",
29         "_type": "categoria",
30         "_id": "Electronics",
31         "_score": 1,
32         "_source": {
33           "subcategories": "Audio & Video Accessories,Stereos,Computers,Portable Bluetooth Speakers,Surround Speakers
,Internal Solid State Drives,Bluetooth & Wireless Speakers,Electronics,Headphones,Audio Power Conditioners
,Parts & Accessories,Accessories,In-Wall & In-Ceiling Speakers,Office,Marine Audio,Computer Accessories &
Peripherals,Computers & Accessories,Straps & Hand Grips,Backpacks ffvzrebezbzuqvcddwzzxeuwva,LCD TVs,Frys
,Computers/Tablets & Networking,Consumer Electronics,Subwoofers,Car Electronics & GPS,LED & LCD TVs,Auto &
Tires,Video Games & Consoles,Satellite Radio,Desktop Memory,Samsung Tax Time Savings,Sports & Outdoors,Audio
,Floor Speakers,TV & Video,Home & Garden,Mobile,Outdoor Speakers,Towers,See more Samsung Ubd-m9500 4k Ultra
HD Blu-ray Player,TV,Photography,Digital Cameras,TVs & Electronics,Camera & Photo Accessories,Controllers,All
TVs,Used:Film Camera Lens Accessories,Home Theater Systems,Speaker Separates,Cameras & Photo,Samsung Smart
TVs,4K Ultra HD TVs,Tablets,Receivers Amplifiers,Touch-Screen All-in-One Computers,Home Audio,Android Auto
Receivers,Audio Visual Presentation,Wireless Multi-Room Speakers uytueusqxdvcfrxftfeefvcxudq,Sports &
Handheld GPS The Wall Chosen Event Micro SD (SD Bluetooth Headsets Prime Lenses TVs & Entertainment Solar &
```





# ELK

## Que és LogStash?

- Eina intermitja entre Kibana i ElasticSearch.
- Disposa d'una gran quantitat de plugins
- Multi servidor. Vàries pipelines.



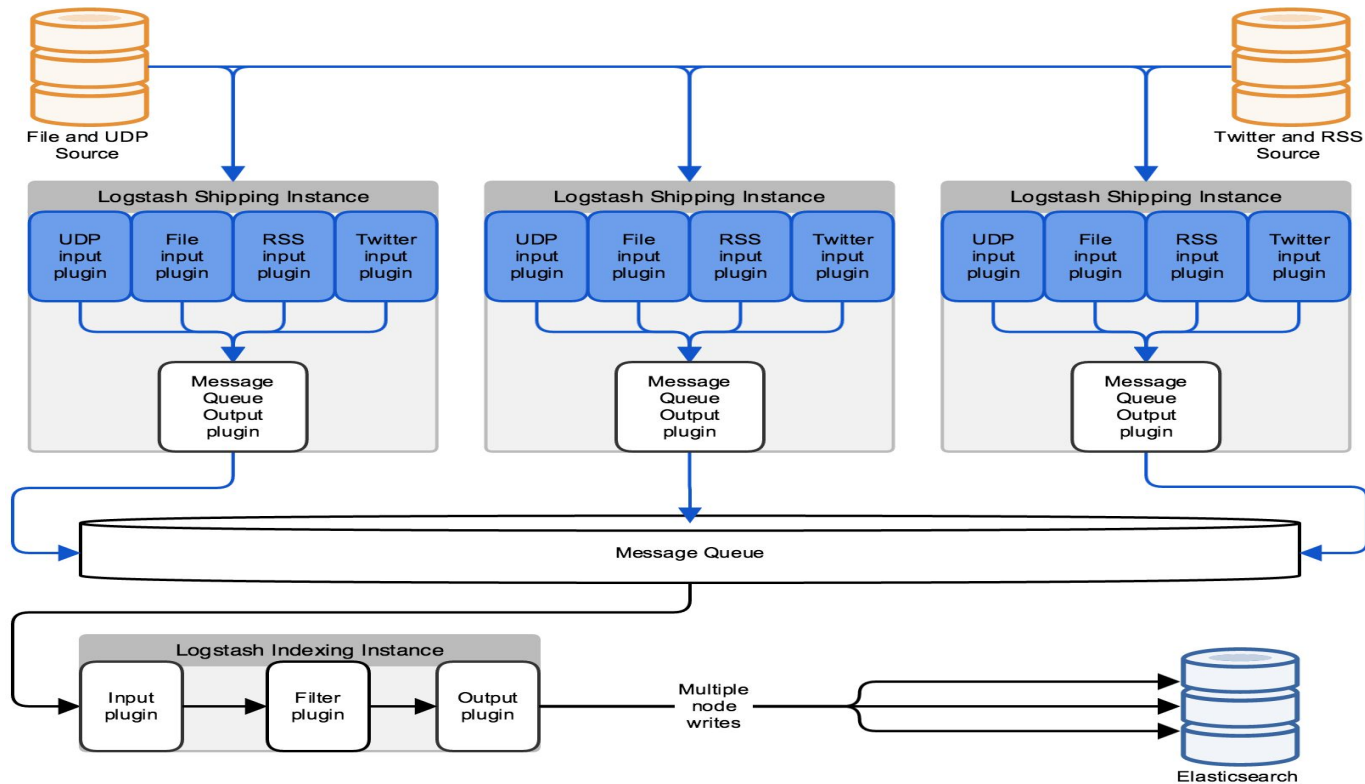
logstash



logstash

# LogStash

## Com funciona?



# Beats

## L'últim afegit a ELK

- ElasticStack, fa relativament poc, va ampliar-se amb un producte nou:
  - ElasticSearch
  - Kibana
  - LogStash
  - (+) Beats
- Beats és una família d'agents encarregats de recollir dades.



beats

# Beats

## La familia beats



### The Beats family



Packetbeat

Network data



Metricbeat

Metrics



Winlogbeat

Windows Event Logs



Auditbeat

Audit data



Filebeat

Log files



Heartbeat

Uptime monitoring

**+40**  
**community**  
**Beats**



**Preparació entorn**

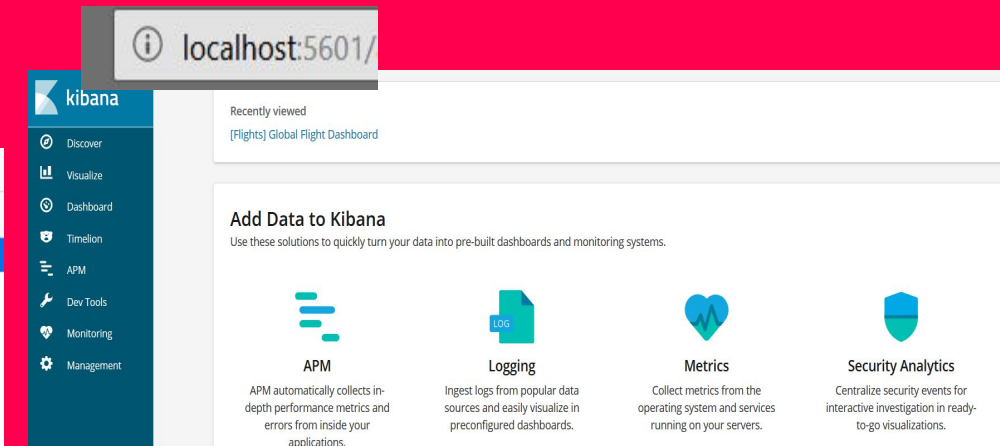
# Muntatge del sistema ELK

## ElasticSearch i Kibana

```
localhost:9200

Guardar Copiar Contraer todo Expandir todo

name: "45Zhssz"
cluster_name: "elasticsearch"
cluster_uuid: "kDfZxwjSRT-RAv28bHTNVg"
version:
  number: "6.5.2"
  build_flavor: "default"
  build_type: "zip"
  build_hash: "9434bed"
  build_date: "2018-11-29T23:58:20.891072Z"
  build_snapshot: false
  lucene_version: "7.5.0"
  minimum_wire_compatibility_version: "5.6.0"
  minimum_index_compatibility_version: "5.0.0"
tagline: "You Know, for Search"
```



# Elasticsearch

## (elasticsearch.yml)

```
##### Elasticsearch Configuration #####
#
# NOTE: Elasticsearch comes with reasonable defaults for most settings.
#       Before you set out to tweak and tune the configuration, make sure you
#       understand what are you trying to accomplish and the consequences.
#
# The primary way of configuring a node is via this file. This template lists
# the most important settings you may want to configure for a production cluster
#
# Please consult the documentation for further information on configuration of
# https://www.elastic.co/guide/en/elasticsearch/reference/index.html
#
# ----- Cluster -----
#
# Use a descriptive name for your cluster:
#
#cluster.name: my-application
#
# ----- Node -----
#
# Use a descriptive name for the node:
#
#node.name: node-1
#
# Add custom attributes to the node:
#
#node.attr.rack: r1
#
# ----- Paths -----
#
# Path to directory where to store the data (separate multiple locations by comma)
#
#path.data: /path/to/data
#
# Path to log files:
#
#path.logs: /path/to/logs
#
```

```
# ----- Memory -----
#
# Lock the memory on startup:
#
#bootstrap.memory_lock: true
#
# Make sure that the heap size is set to about half the memory available
# on the system and that the owner of the process is allowed to use this
# limit.
#
# Elasticsearch performs poorly when the system is swapping the memory.
#
# ----- Network -----
#
# Set the bind address to a specific IP (IPv4 or IPv6):
#
#network.host: 0.0.0.0
#
# Set a custom port for HTTP:
#
#http.port: 9200
#
# For more information, consult the network module documentation.
#
# ----- Discovery -----
#
# Pass an initial list of hosts to perform discovery when new node is started:
# The default list of hosts is ["127.0.0.1", "[::1]"]
#
#discovery.zen.ping.unicast.hosts: ["host1", "host2"]
#
# Prevent the "split brain" by configuring the majority of nodes (total number of master-eligible nodes / 2 + 1):
#
#discovery.zen.minimum_master_nodes:
#
# For more information, consult the zen discovery module documentation.
#
# ----- Gateway -----
#
# Block initial recovery after a full cluster restart until N nodes are started:
#
#gateway.recover_after_nodes: 3
#
# For more information, consult the gateway module documentation.
#
# ----- Various -----
#
# Require explicit names when deleting indices:
#
#action.destructive_requires_name: true
```

# Muntatge ELK

## Dades emmagatzemades a MySQL DB

- Base de dades extretes de: <https://data.world/> amb petites modificacions
- Apache POI per xlsx i Java
- Taules producte, categoria i condicions.

### Table: **producte**

#### Columns:

<u><b>codi</b></u>	int(11) AI PK
dataInsercio	datetime
disponibilitat	int(11)
condicions	varchar(20)
nom	varchar(200)
marca	varchar(40)
fabricant	varchar(100)
preu	float(8,2)
preuEnviament	float(8,2)
categoriaPrincipal	varchar(30)
altreCategoria	varchar(200)
imageURL	varchar(400)

### Table: **categoria**

#### Columns:

<u><b>nom</b></u>	varchar(50) PK
subcategories	varchar(4000)

### Table: **condicions**

#### Columns:

<u><b>nom</b></u>	varchar(50) PK
-------------------	----------------



# Logstash: INPUT i OUTPUT

```
input {
  jdbc {
    jdbc_connection_string => "jdbc:mysql://localhost:3306/productes"
    jdbc_user => "root"
    jdbc_password => "root"
    schedule => "*" * * * *
    jdbc_validate_connection => true
    jdbc_driver_library => "C:\\Program Files (x86)\\MySQL\\Connector J 8.0\\mysql-connector-java-8.0.13.jar"
    jdbc_driver_class => "com.mysql.cj.jdbc.Driver"
    statement => "SELECT * from productes"
    type => "productes"
  }
  jdbc {
    jdbc_connection_string => "jdbc:mysql://localhost:3306/productes"
    jdbc_user => "root"
    jdbc_password => "root"
    schedule => "*" * * * *
    jdbc_validate_connection => true
    jdbc_driver_library => "C:\\Program Files (x86)\\MySQL\\Connector J 8.0\\mysql-connector-java-8.0.13.jar"
    jdbc_driver_class => "com.mysql.cj.jdbc.Driver"
    statement => "SELECT * from categoria"
    type => "categoria"
  }
  jdbc {
    jdbc_connection_string => "jdbc:mysql://localhost:3306/productes"
    jdbc_user => "root"
    jdbc_password => "root"
    schedule => "*" * * * *
    jdbc_validate_connection => true
    jdbc_driver_library => "C:\\Program Files (x86)\\MySQL\\Connector J 8.0\\mysql-connector-java-8.0.13.jar"
    jdbc_driver_class => "com.mysql.cj.jdbc.Driver"
    statement => "SELECT * from condicions"
    type => "condicions"
  }
}
```

```
output {
  if [type] == "productes" {
    elasticsearch {
      hosts => "localhost:9200"
      index => "productes"
      document_type => "%{type}"
      document_id => "%{codi}"
    }
  }
  if [type] == "categoria" {
    elasticsearch {
      hosts => "localhost:9200"
      index => "categories"
      document_type => "%{type}"
      document_id => "%{nom}"
    }
  }
  if [type] == "condicio" {
    elasticsearch {
      hosts => "localhost:9200"
      index => "condicions"
      document_type => "%{type}"
      document_id => "%{nom}"
    }
  }
}
```

Execució: logstash.bat -f logstash.conf

Rufus Scheduler:

minut hora dia mesos(int) UTC

MySQL DB:

SET @@global.time\_zone='+00:00';

SET @@session.time\_zone='+00:00';

## Filtres logstash(extres)

```
filter {  
  mutate { remove_field => [ "field1", "field2", "field3", ... "fieldN" ] }  
}
```

La gran majoria de camps es poden filtrar excepte els que són de sistema, com els `_index`, `_type`, `_id`, `_source`. Altres que directament es generen a l'hora de fer la cerca com és el `_score`. O altres com per exemple el `@timestamp`, que són necessaris.

# Ngrok

ngrok.exe http port

ngrok by @inconshreveable

```
Session Status      online
Account             Miguel Ángel Bueno Rivera (Plan: Free)
Version             2.2.8
Region              United States (us)
Web Interface       http://127.0.0.1:4040
Forwarding           http://c3c61862.ngrok.io -> localhost:9200
Forwarding           https://c3c61862.ngrok.io -> localhost:9200
```

Connections	ttl	opn	rt1	rt5	p50	p90
	0	3	0.00	0.00	0.00	0.00

HTTP Requests

-----

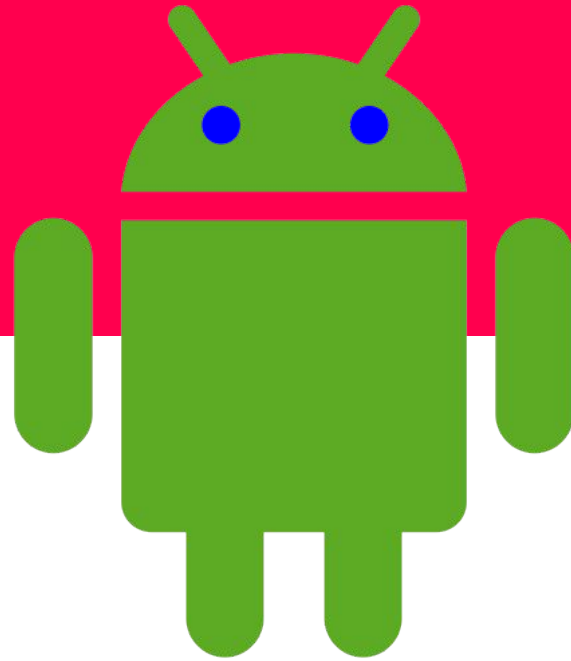
```
GET /favicon.ico          200 OK
GET /productes/producte/_search 200 OK
GET /productes/producte/_search 200 OK
GET /condicions/condicio/_search 200 OK
GET /productes/producte/_search 200 OK
GET /categories/categoria/_search 200 OK
GET /productes/producte/_search 200 OK
```

# Muntatge ELK

## Crides REST

Algunes funcionalitats de la aplicació, requereixen agafar dades dels documents indexats a l'elasticsearch fent ús de crides REST, com per exemple:

- productes (per o id o tots)  
`@GET("/productes/producte/{id}")`  
`@GET("/productes/producte/_search")`
- categories (subcategories, categories)  
`@GET("/categories/categoria/_search")`  
`@GET("/categories/categoria/{id}")`

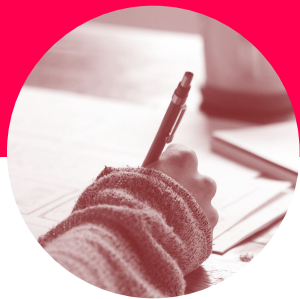


## Futurs canvis en la aplicació - ELK

- Utilitzar l'eina APM per rebre informació d'execució de l'aplicació i possibles errors que es produeixin en elles
- Implementar un servei de login i utilitzar el FileBeat per captar informació que ens interressi sobre aquests. I/o utilitzar Logging per utilitzar ja les dashboards pre-dissenyades.
- Millorar la distribució dels shards, repliques i ampliar informació.
- Ampliar la bases de dades, establir un schedule, exportar les dades d'elasticsearch a MySQL.
- Logs, millorar app, transaccions, seguretat, custom url keys, etc.

# Exemple query

```
GET _search
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "nom": "tv"
        }
      },
      "filter": {
        "range": { "disponibilitat": { "gte": 30 }}
      }
    }
  }
}
```



# Gràcies!!

Pau Recacha Borrell  
Ferran Rodríguez  
Miguel Ángel Bueno Rivera



# Webgrafia

<https://github.com/jmettraux/rufus-scheduler>

<https://www.elastic.co/guide/en/logstash/5.0/plugins-inputs-jdbc.html>

[https://www.elastic.co/\\*](https://www.elastic.co/*)

[https://stackoverflow.com/\\*](https://stackoverflow.com/*)

[https://lucene.apache.org/core/2\\_9\\_4/queryparsersyntax.html](https://lucene.apache.org/core/2_9_4/queryparsersyntax.html)

<https://developer.android.com/>

<https://square.github.io/retrofit/>

<https://www.elastic.co/guide/en/elasticsearch/reference/current/docs-get.html>