A faint, light blue world map is visible in the background of the slide, centered behind the text.

Arquitectura de sistemes Big Data pel record Linkage de xarxes socials

Treball de final de grau

Alumne: Alejandro García Carballo

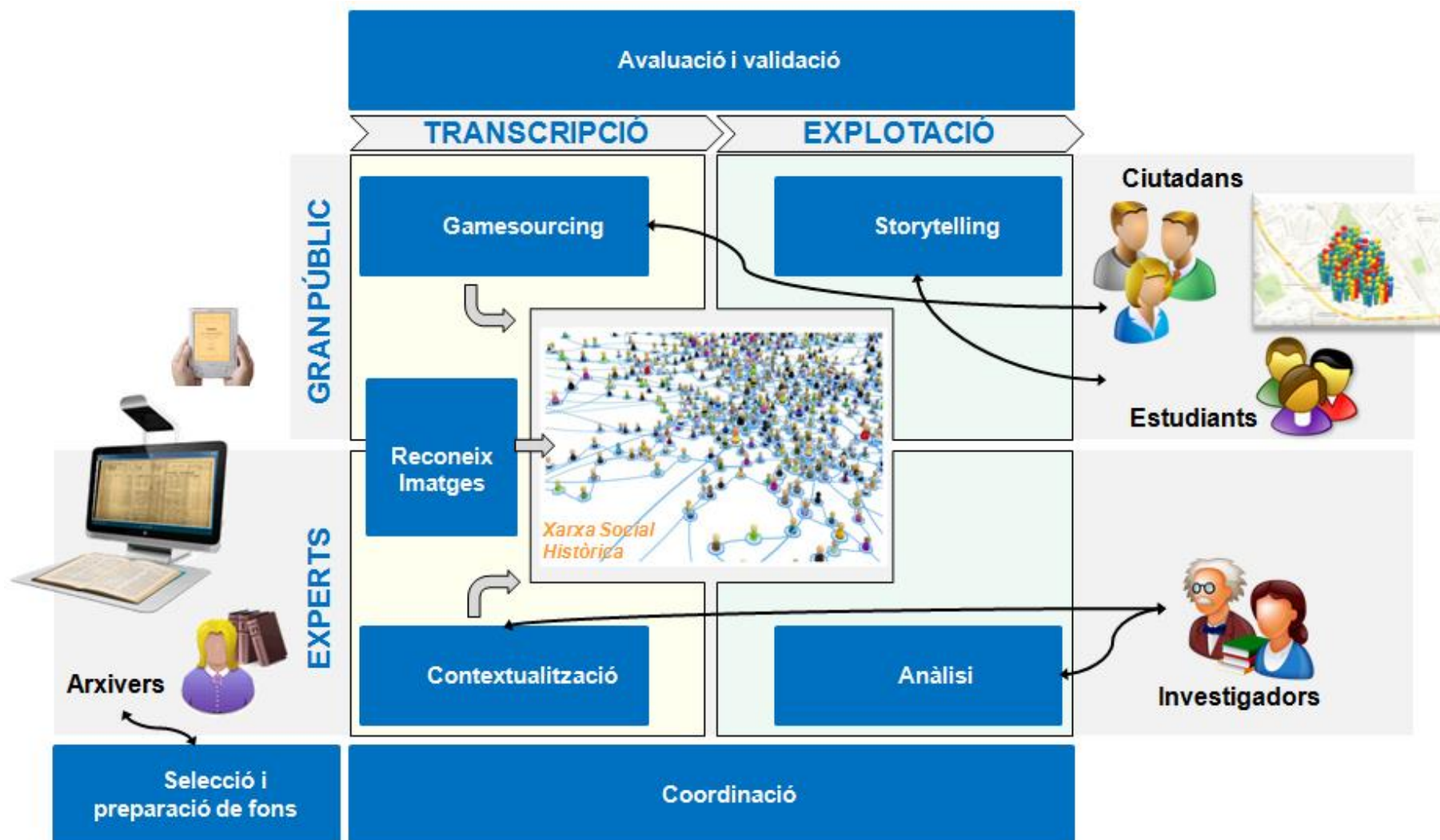
Tutor: Oriol Ramos

Contingut de la presentació

- 1** Context del TFG
- 2** Definició del projecte
- 3** Metodologia
- 4** Tecnologies existents

- 5** Disseny
- 6** Testing
- 7** Resultats
- 8** Conclusions

Context del TFG



Definició del projecte: Objectius del projecte

01

02

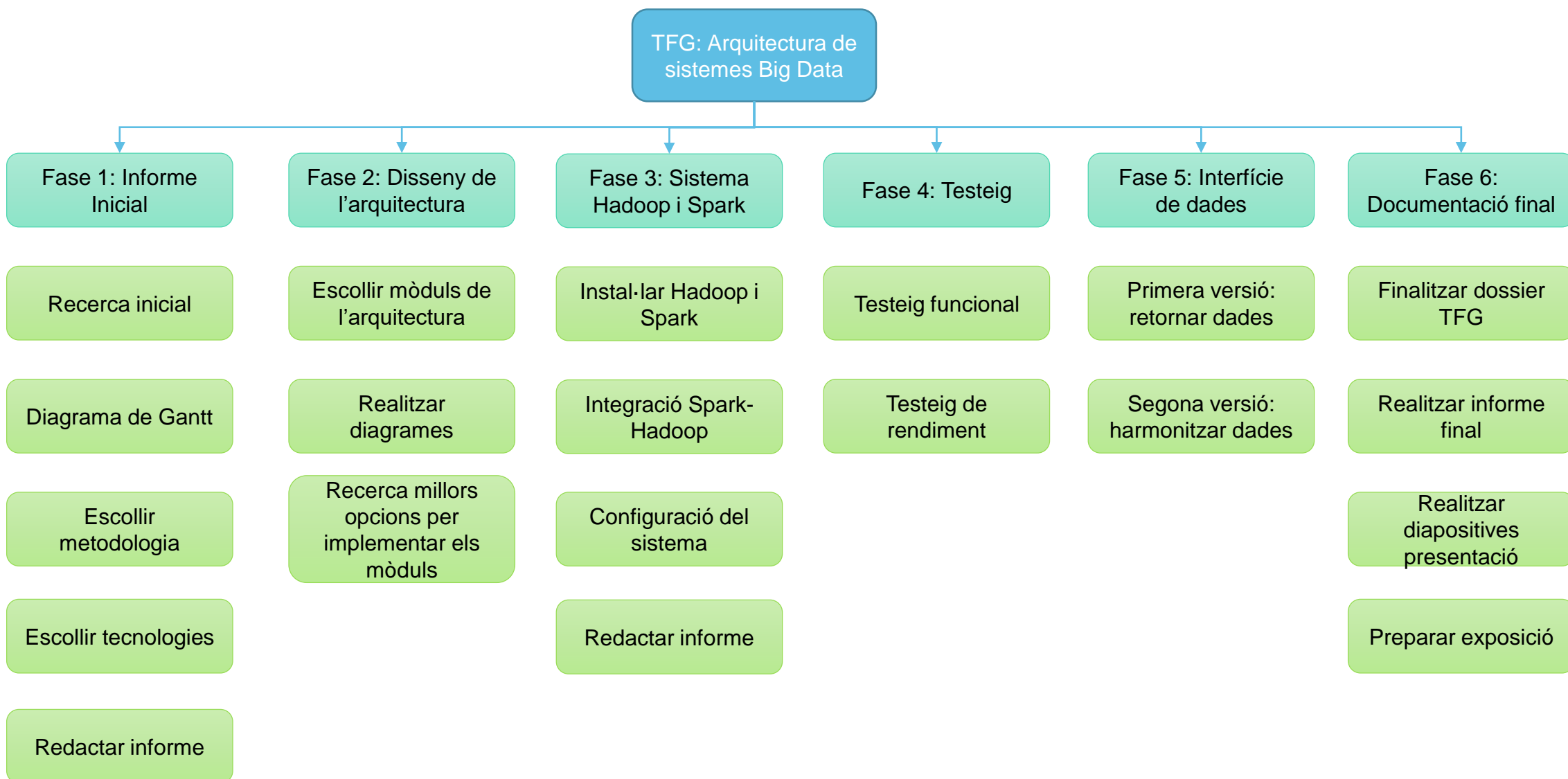
03

**Millorar la qualitat del projecte
global XARXES**

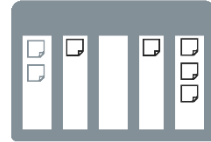
**Augmentar l'eficiència del projecte
XARXES**

Facilitar l'escalabilitat de les dades

Definició del projecte: Fases i tasques



Metodologia



Kanban

Kanban

- No hi ha rols prescrits
- Entrega continua
- Els canvis es poden produir en qualsevol moment

1

2

Scrum

- Hi ha rols definits (Scrum master, product owner ...)
- Sprint de temps parcial
- Els canvis no es poden produir a meitat d'un sprint

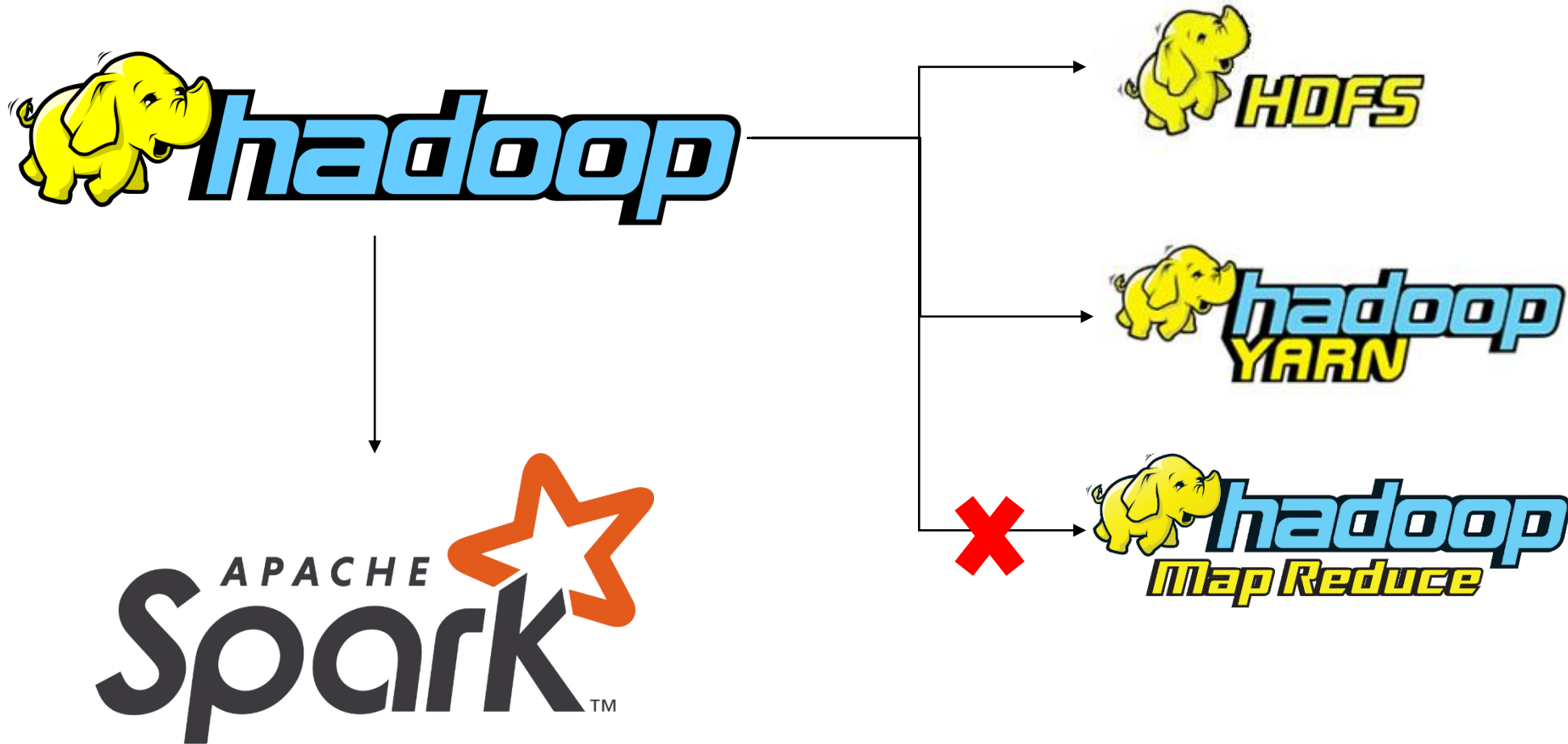


Scrum

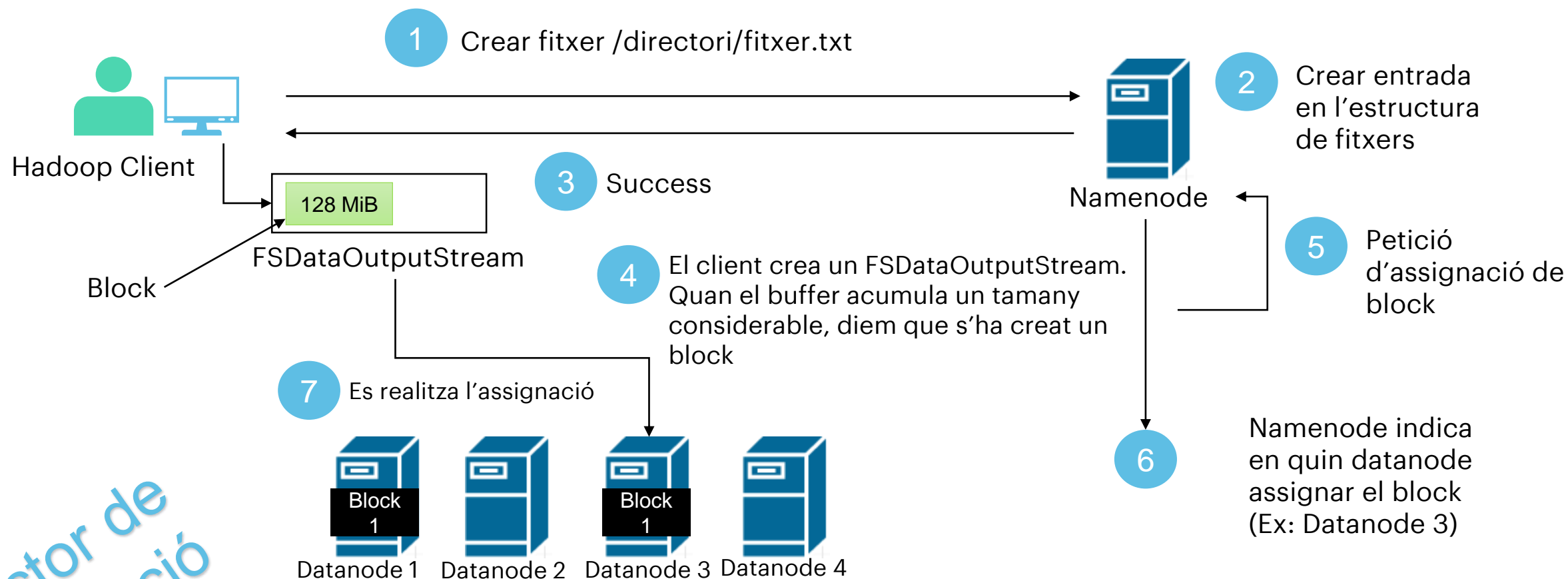


KANBAN

Tecnologies existents

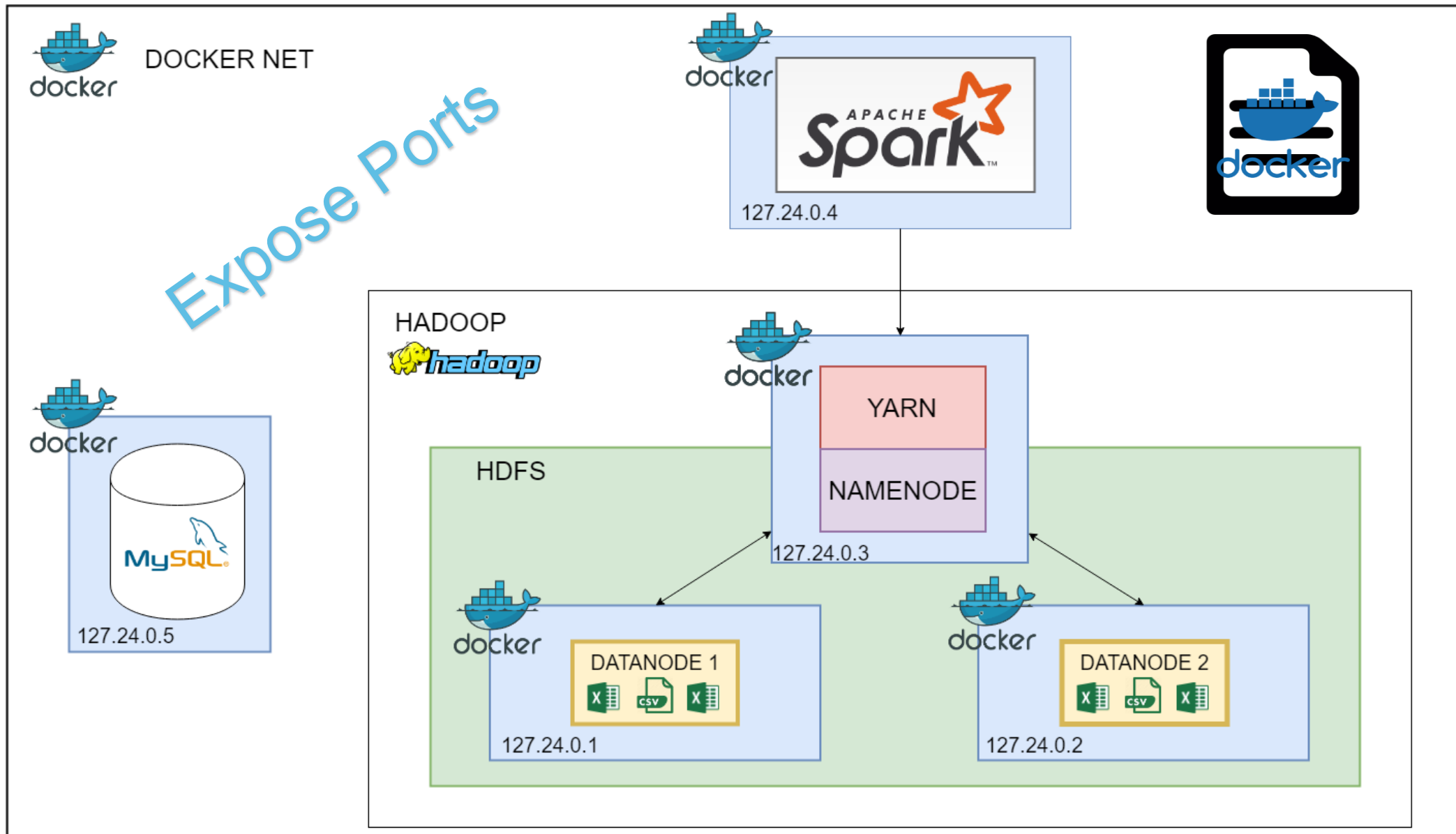


Hadoop Distributed File System (HDFS)

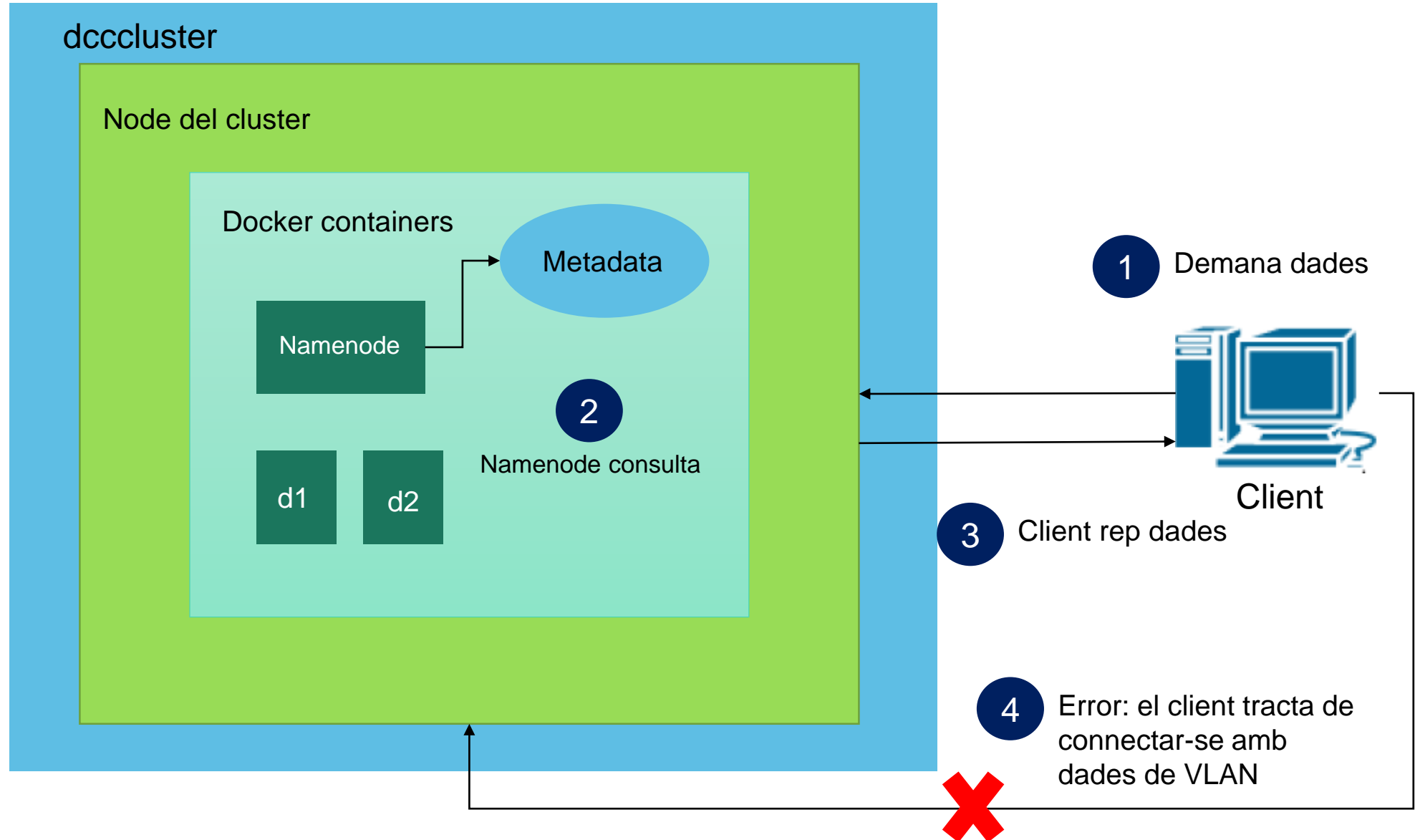


Factor de
replicació

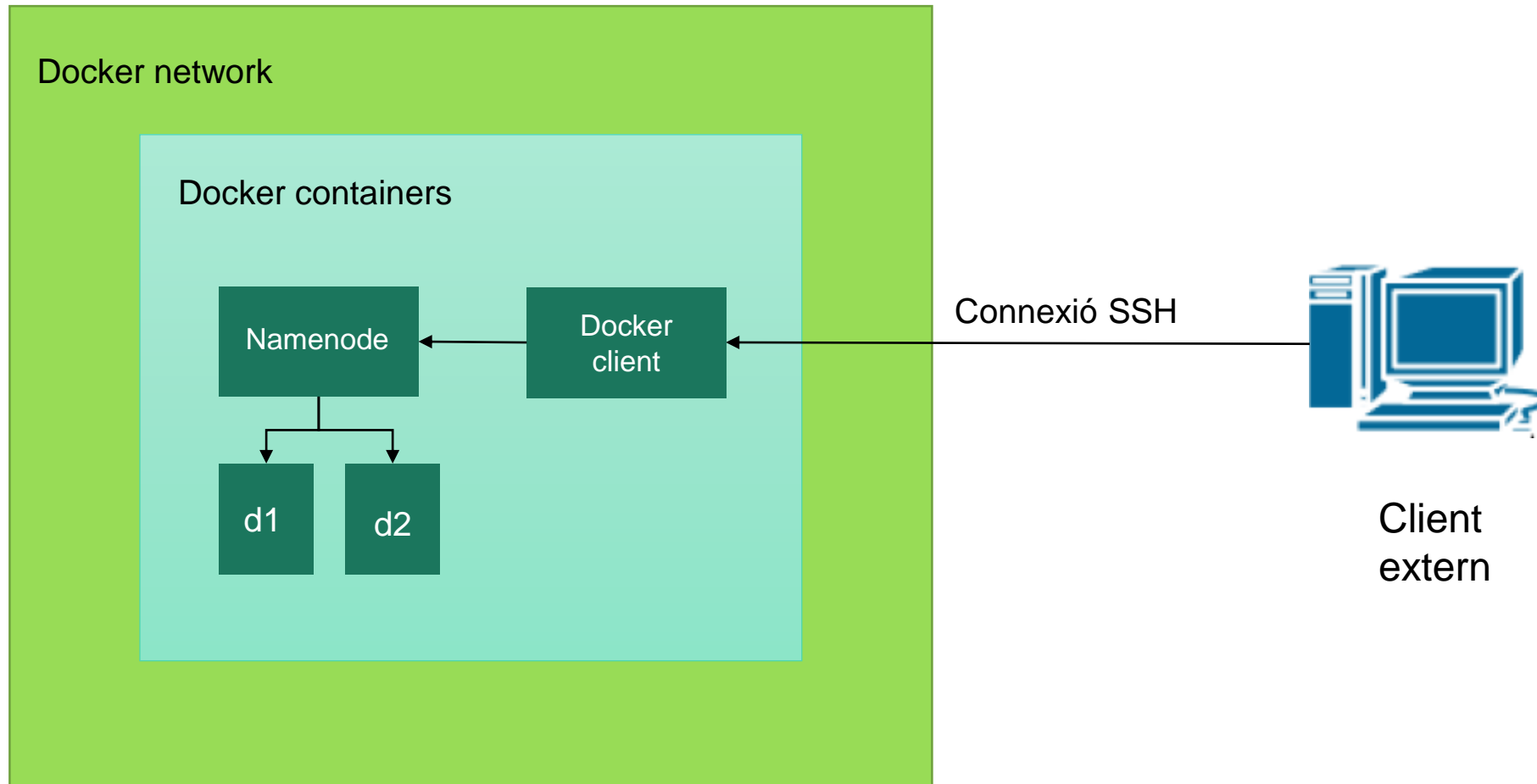
Disseny: Arquitectura del sistema implementat



Xarxa Docker



Xarxa Docker



Interfície de dades

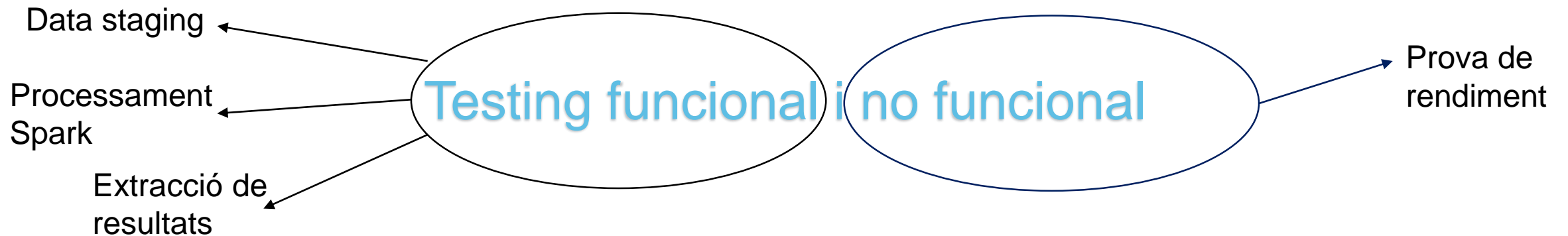


PYSPARK



Retornar dades
HDFS

Testing



Resultats

Show entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	root	supergroup	12.45 KB	May 14 14:45	2	128 MB	100 Sales Records.csv	

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information --

Block ID: 1073741825

Block Pool ID: BP-1731354512-192.168.128.4-1557839554919

Generation Stamp: 1001

Size: 12744

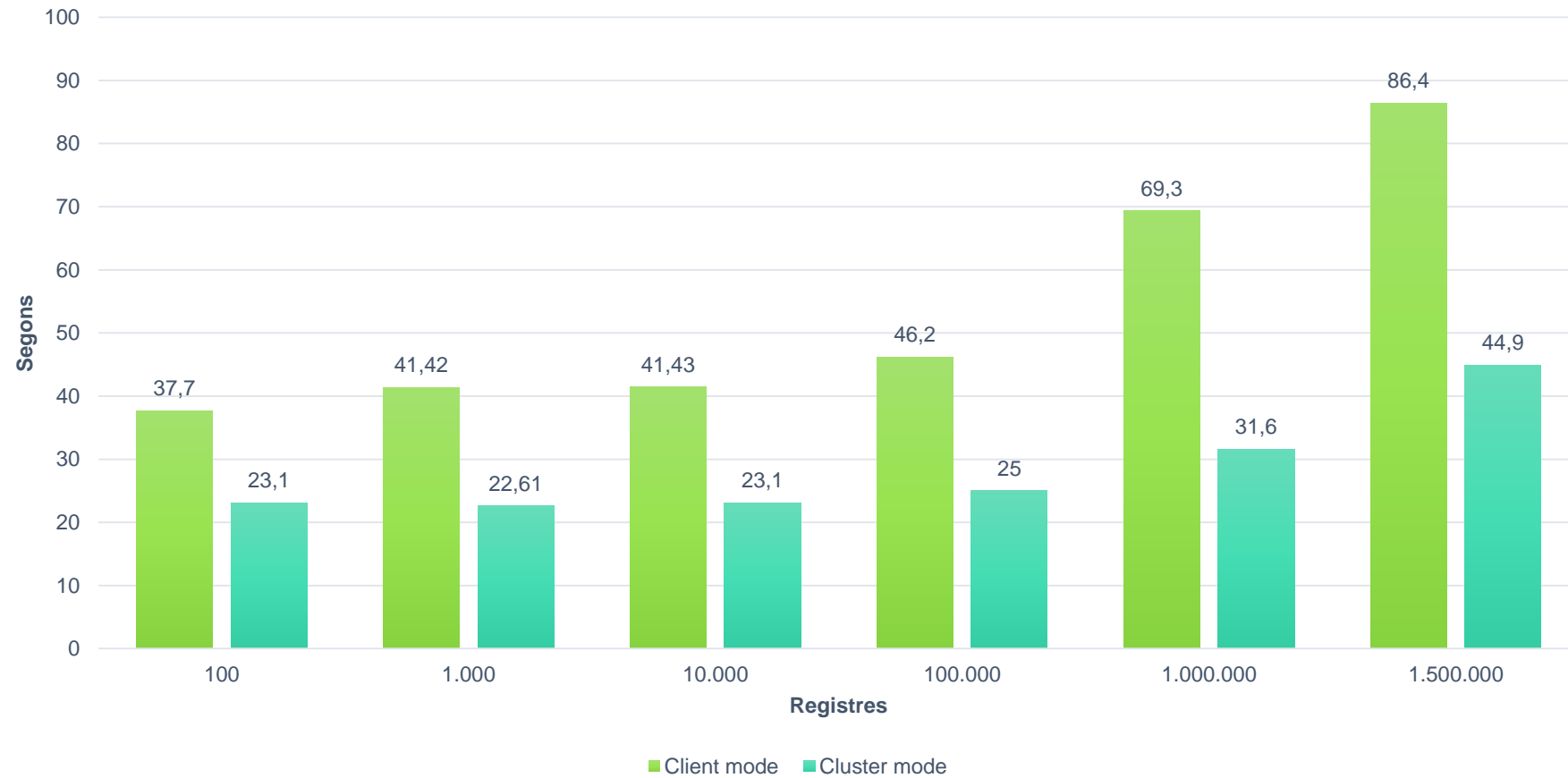
Availability:

- node2
- node1

```
2019-05-12 18:02:31 INFO Client:54 -
    client token: N/A
    diagnostics: N/A
    ApplicationMaster host: node2
    ApplicationMaster RPC port: 46135
    queue: default
    start time: 1557684118412
    final status: SUCCEEDED
    tracking URL: http://namenode:8088/proxy/application_1557595668802_0001
/
    user: root
```

Resultats

Comparativa temps execució



Conclusions

- Gran part dels objectius principals s'han assolit.
- Possibles millores:
 - Modificar tamany de bloc.
 - Connexió directa sense shell.
- Aconseguir objectius secundaris.



Treball final de grau

Gràcies per la vostra atenció!

