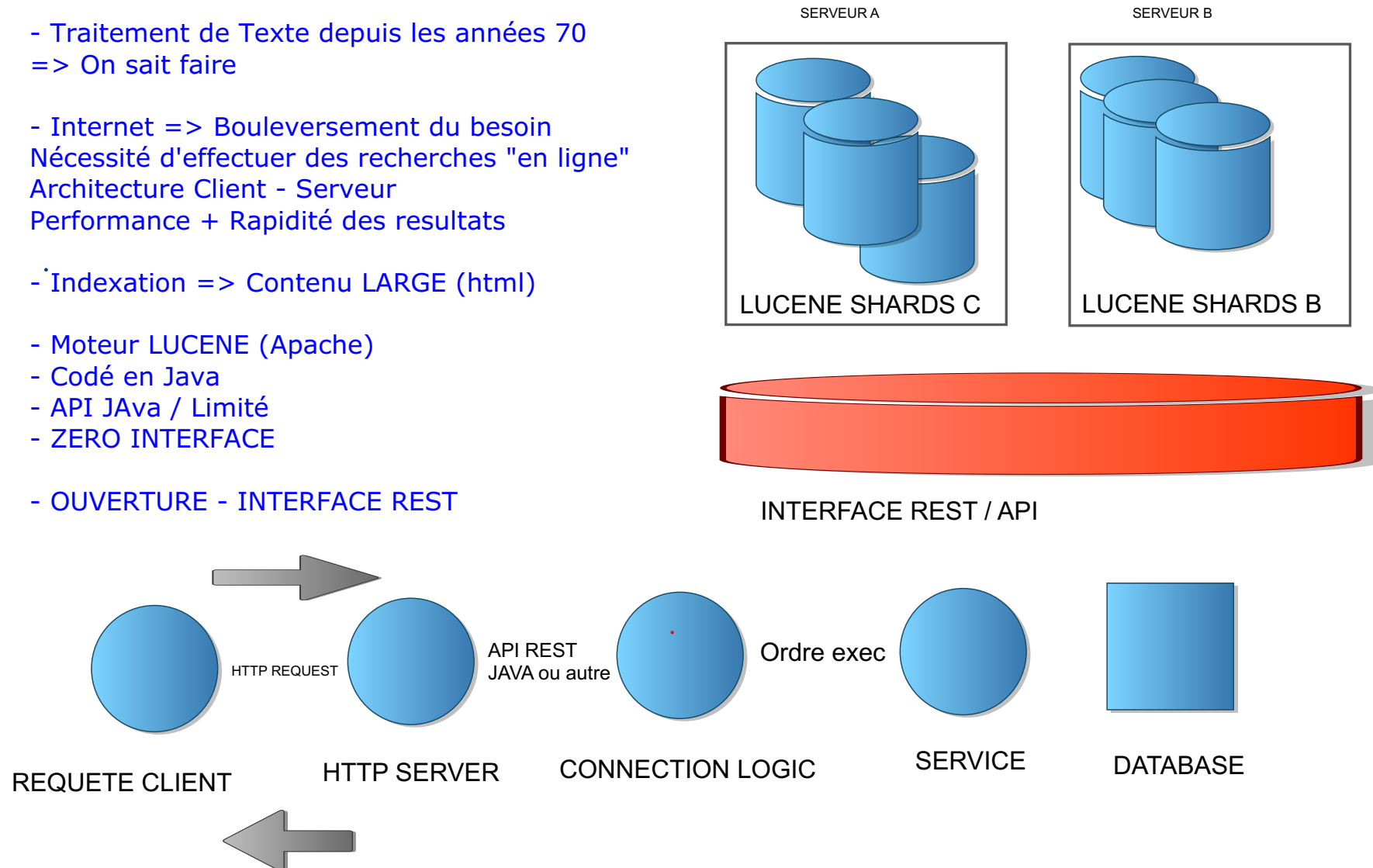
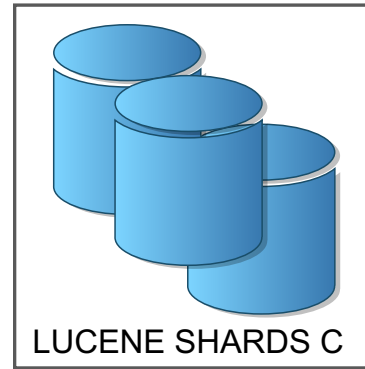


ELASTICSEARCH

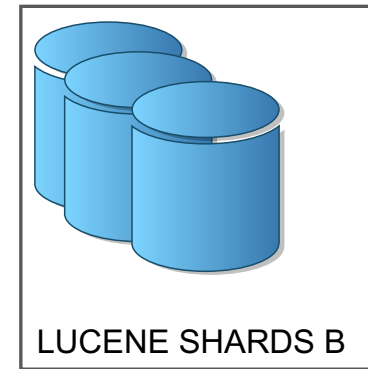
- Traitement de Texte depuis les années 70
=> On sait faire
- Internet => Bouleversement du besoin
Nécessité d'effectuer des recherches "en ligne"
Architecture Client - Serveur
Performance + Rapidité des resultats
- Indexation => Contenu LARGE (html)
- Moteur LUCENE (Apache)
 - Codé en Java
 - API JAVa / Limité
 - ZERO INTERFACE
- OUVERTURE - INTERFACE REST



SERVEUR A

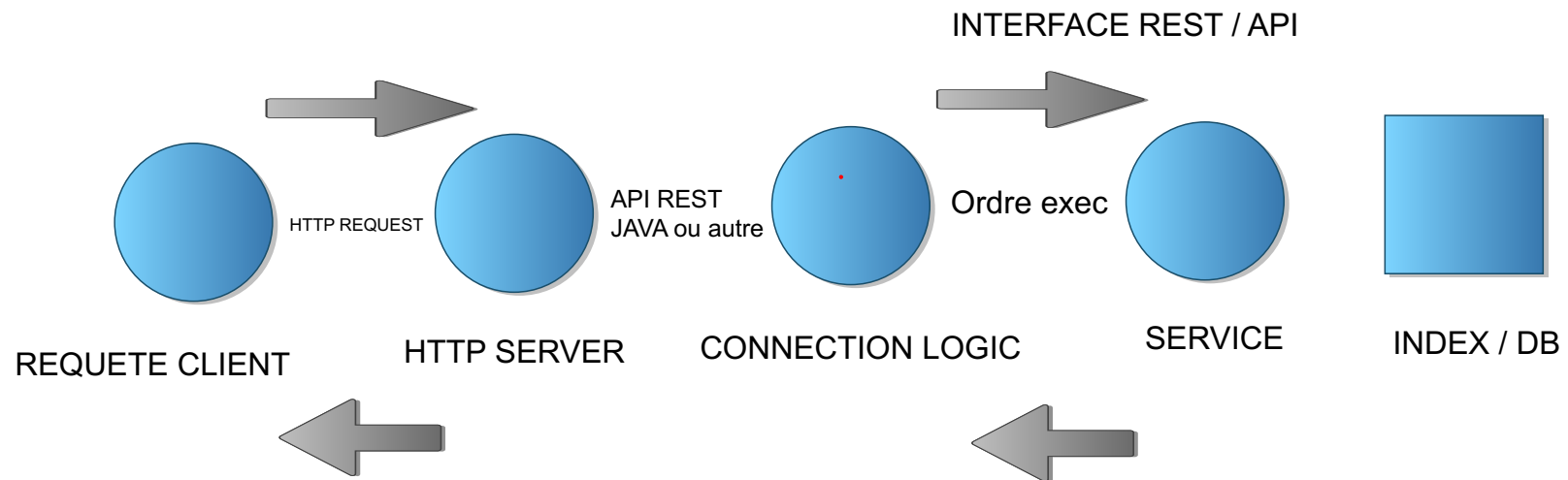


SERVEUR B

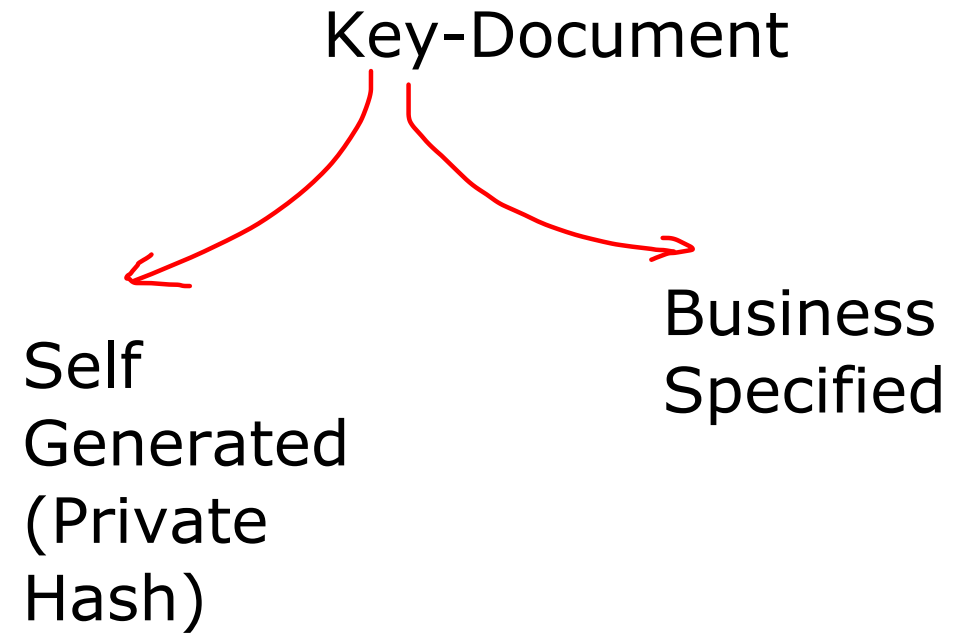


INTERFACE REST / API

REST API



Key-Value Store

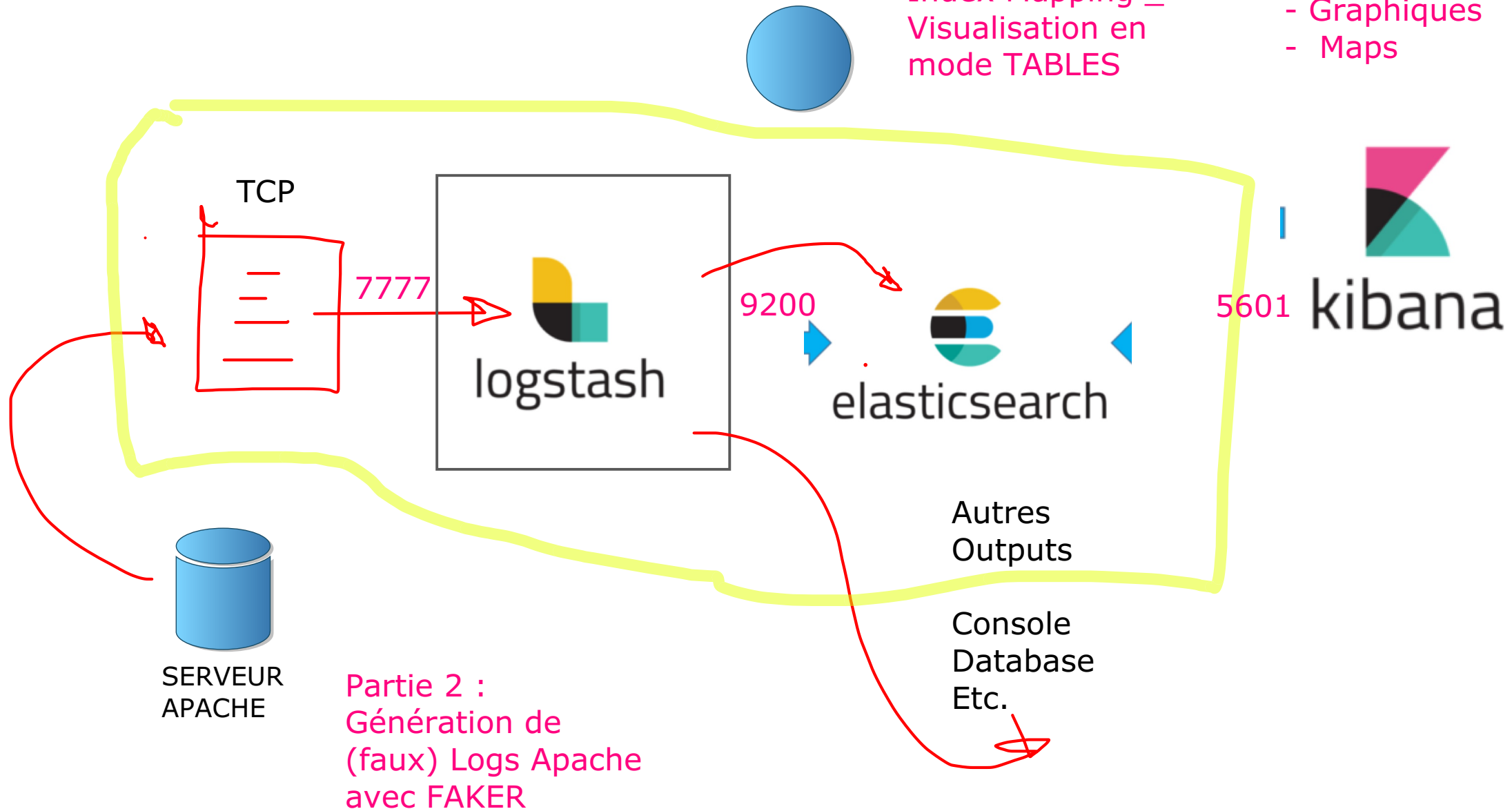


Partie 3 : KIBANA

- Gestion des Index Mapping – Visualisation en mode TABLES

Partie 4 : KIBANA

- Visualisation
- Charts
- Graphiques
- Maps



DISTRIBUTED
ARCHITECTURE

DISTRIBUTED
STORAGE

DISTRIBUTED
COMPUTING

MULTIDIMENSIONAL
SCALABILITY



NO SQL

Relachement de contraintes
=> Schema

INVERSION DE PARADIGME

STRUCTURE DEFINES CONTENT

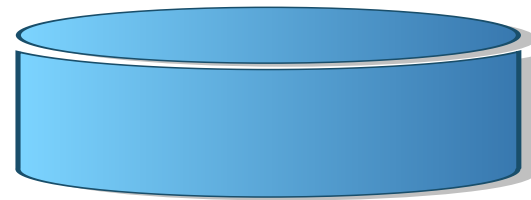
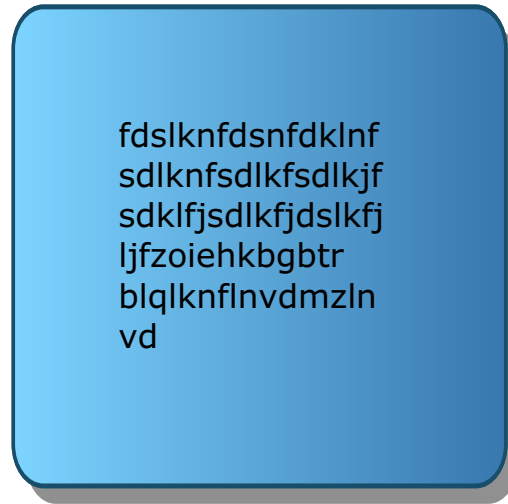


CONTENT DEFINES STRUCTURE

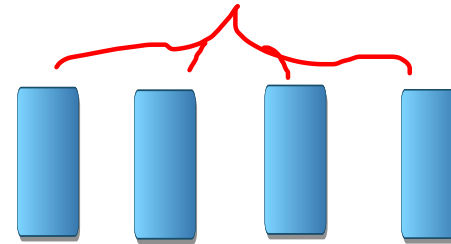
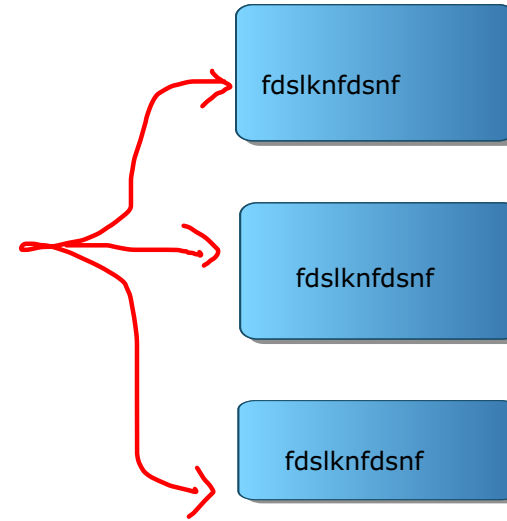


HETEROGENEOUS
SOURCES / FORMATS

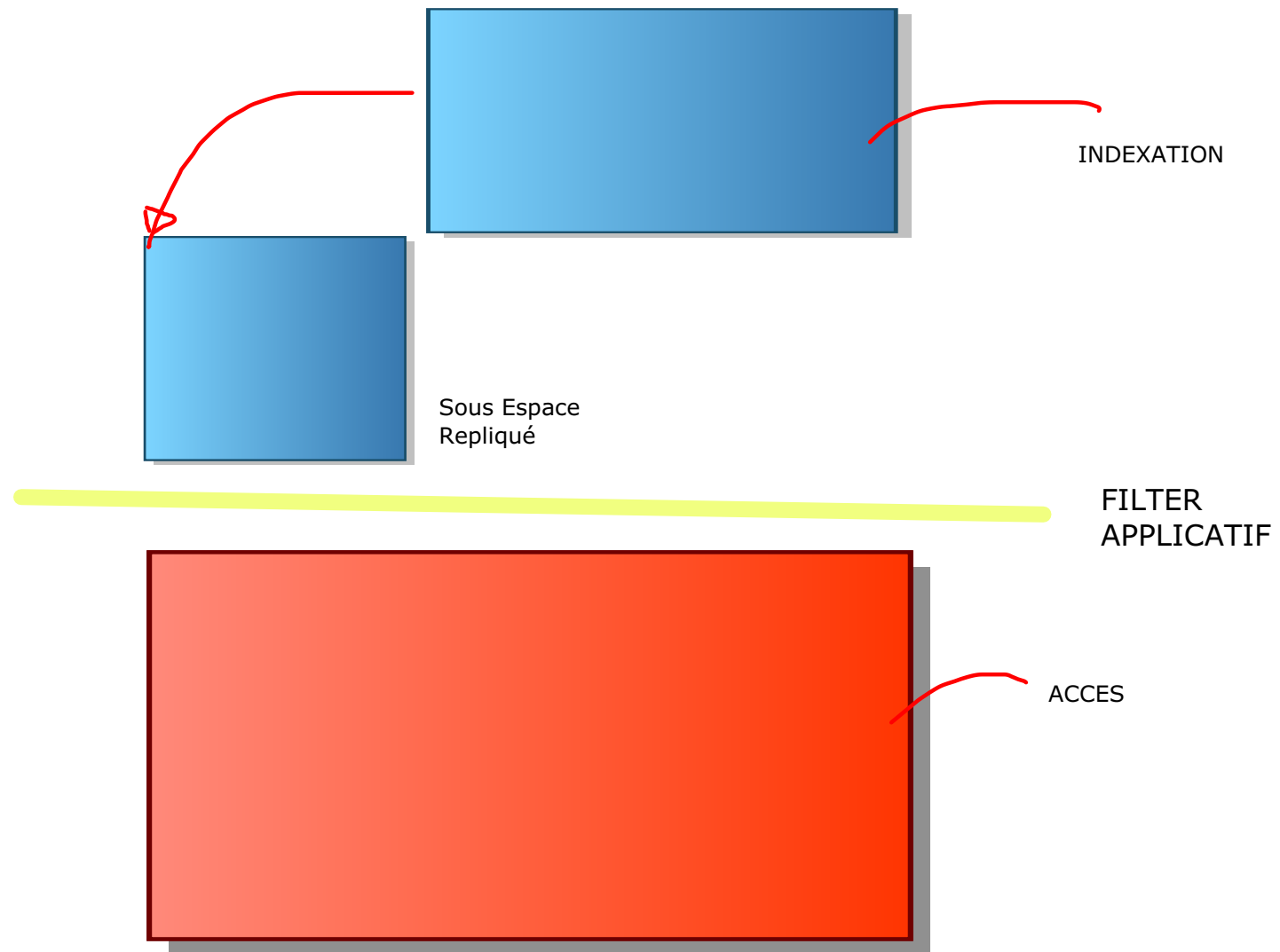


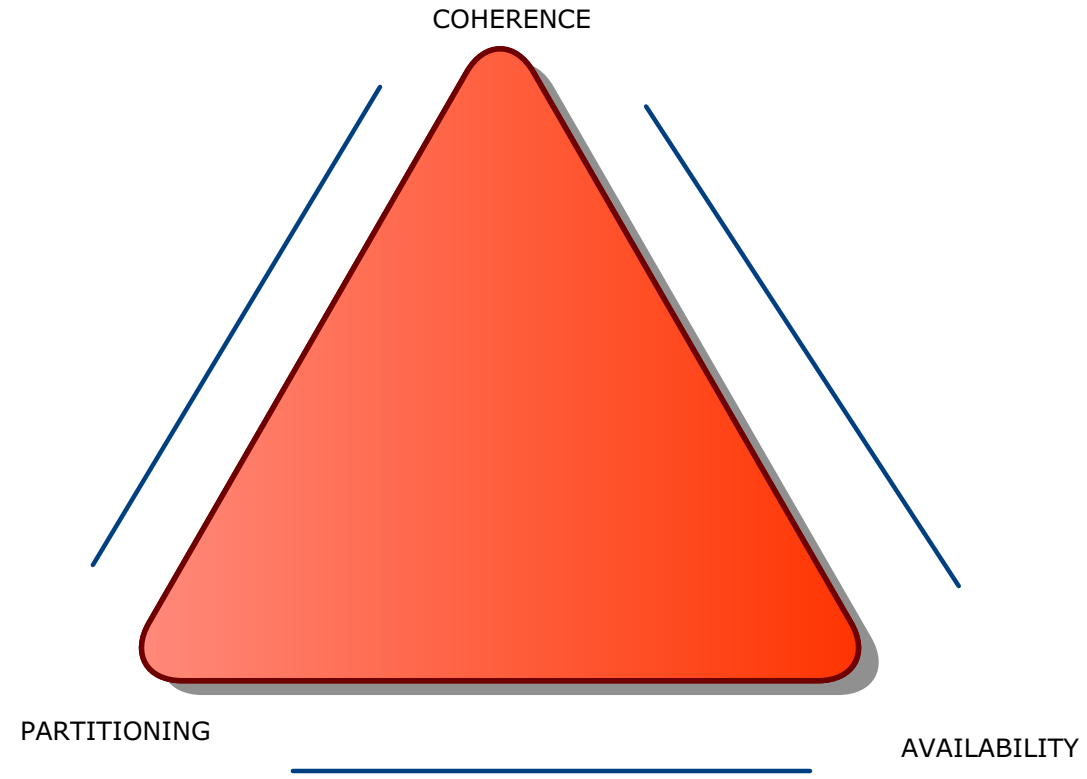


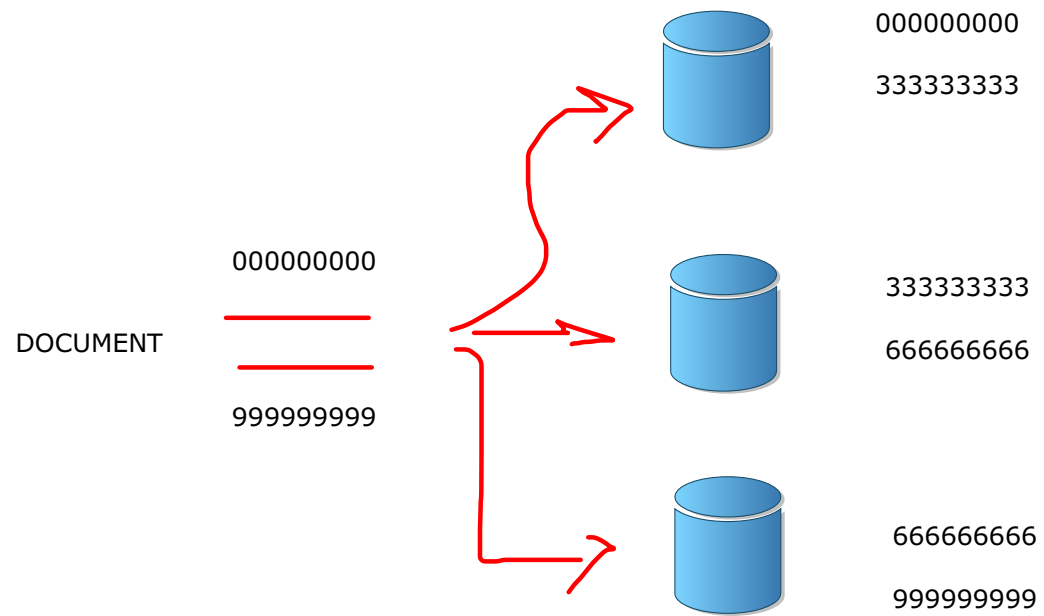
ENTREPOT MASSIF



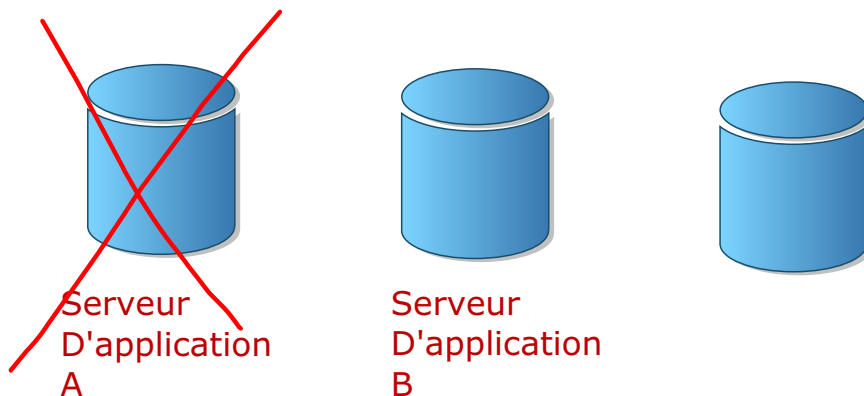
ENTREPOT IKEA







REPARTITION
HOMOGENE
DES CLES



BY DESIGN

Cas d'usages

- Répartition de charge - Load Balancing
- Tolérance aux Pannes - Fail Over



Integrated +
Automated

Gestion Manuelle - Third Party (RAID etc.)

Shit Happens.... (EVENTUALLY)

Shit WILL
Happen
(FOR SURE)

Choix de
design de
topographie
de Cluster

NB SERVEURS



RISQUE DE
DEFAILLANCE

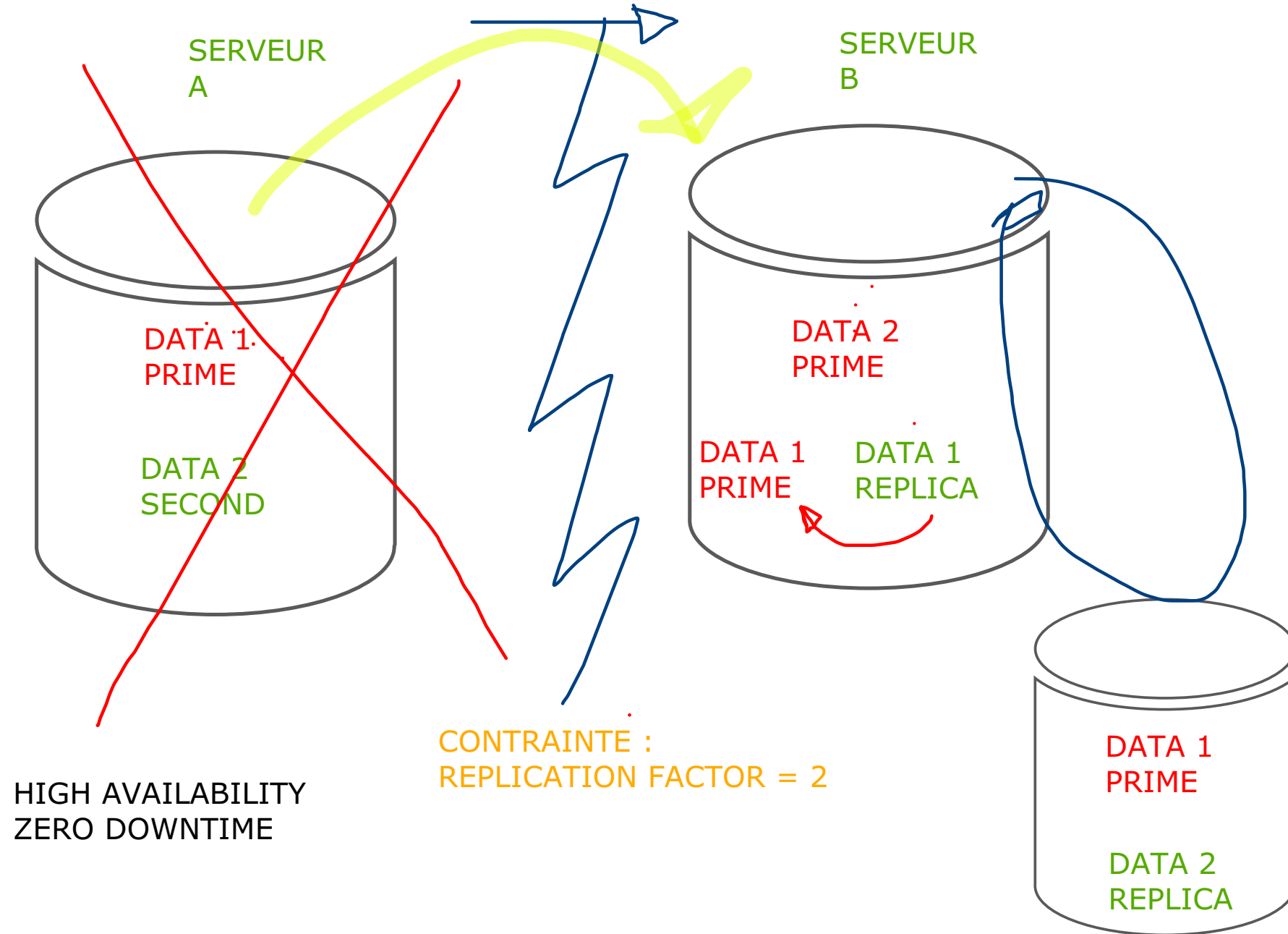
NOYER /
MITIGATION RISQUE

NECESSITE DE
REDONDANCE /
REPLICATION

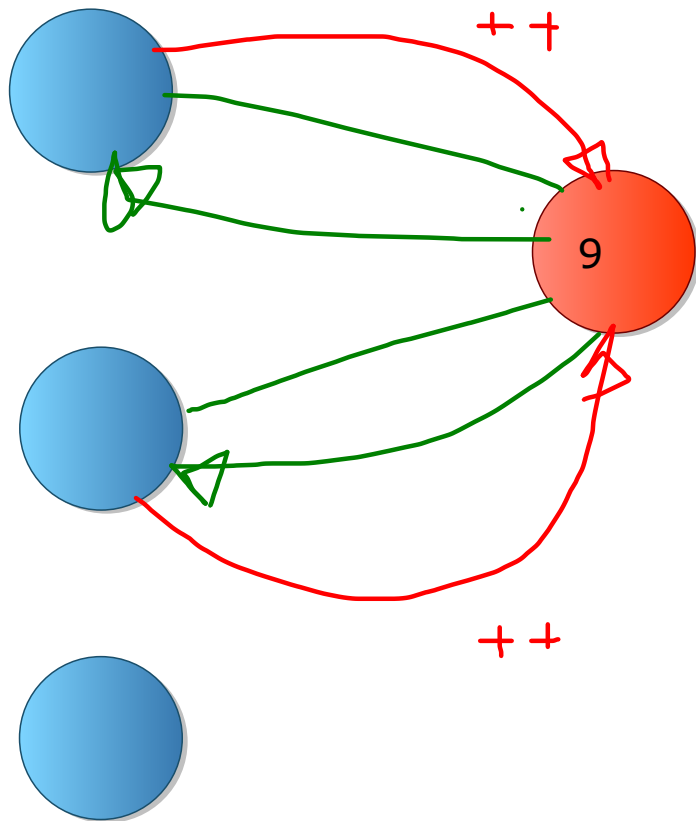
DISTRIBUTION

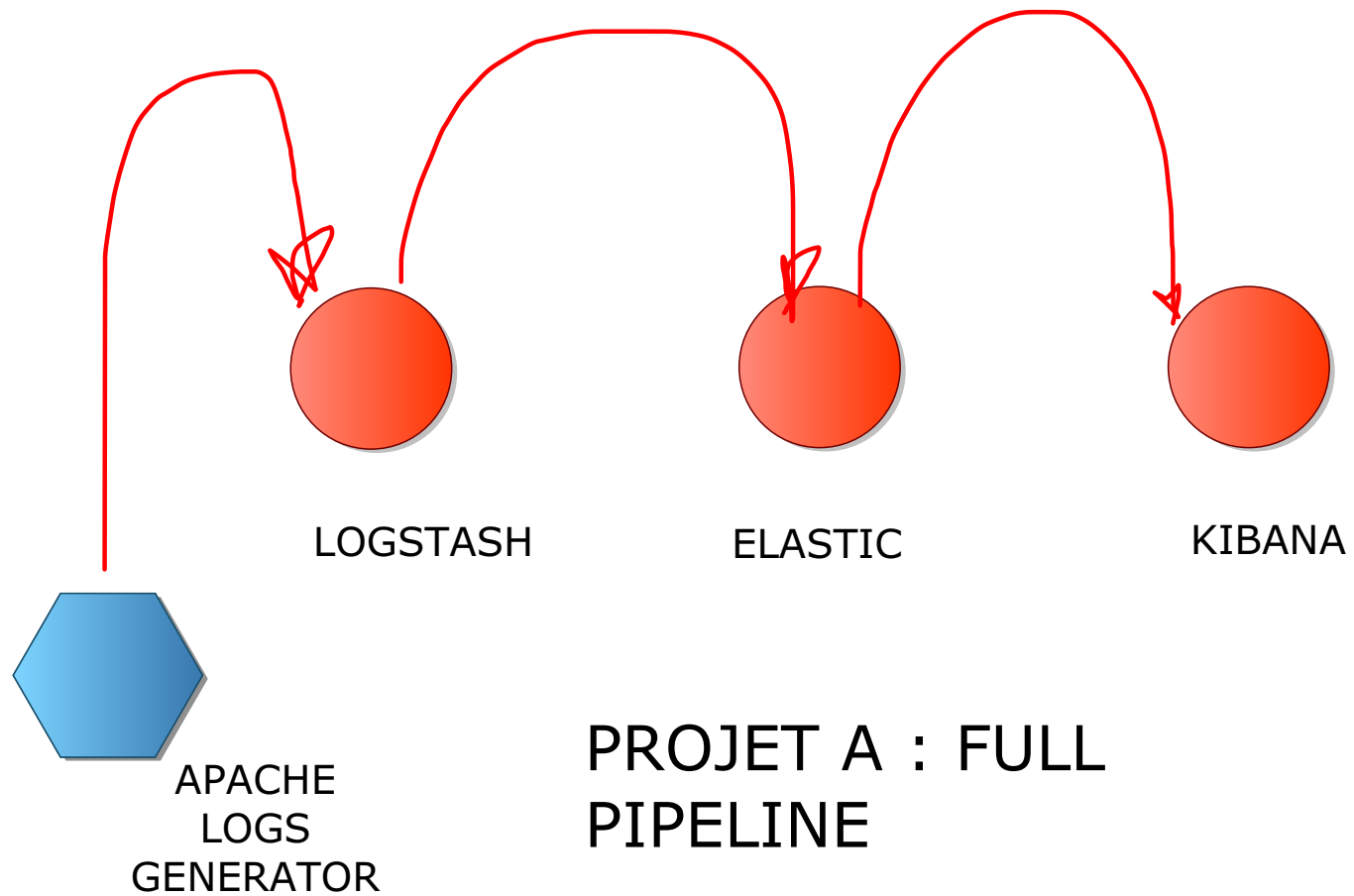
2D

DUPLIQUER -
REPLICATION
FACTOR



— R
— W



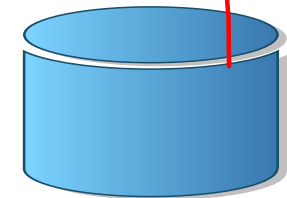
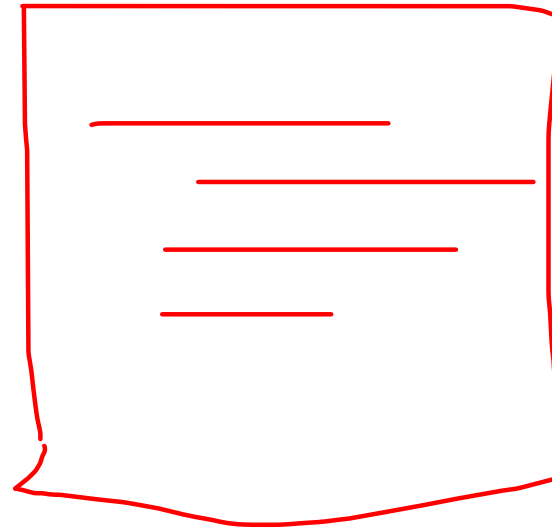


PROJET A : FULL
PIPELINE

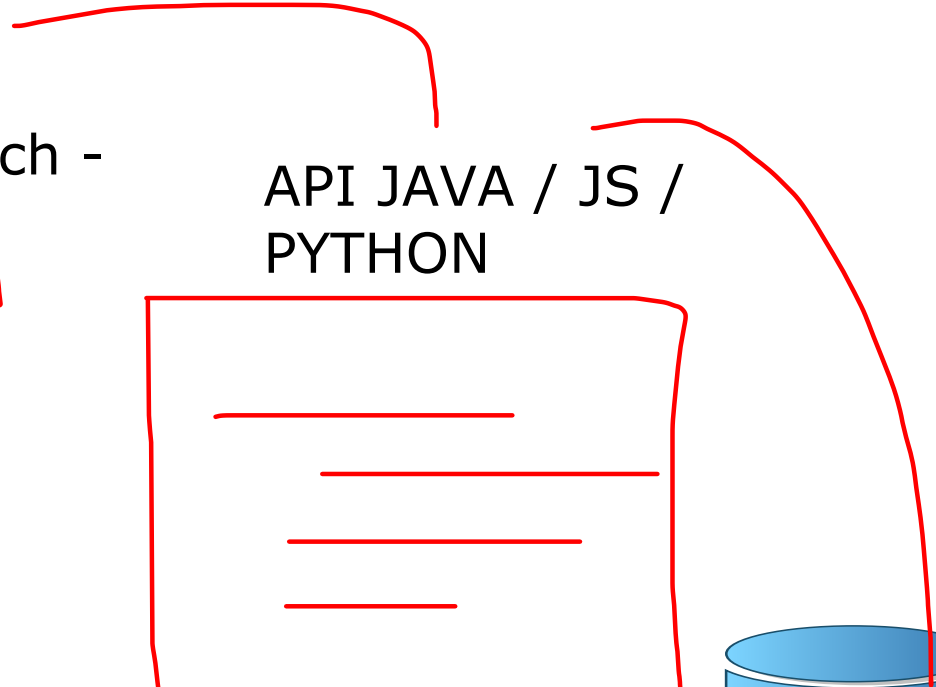
WEB



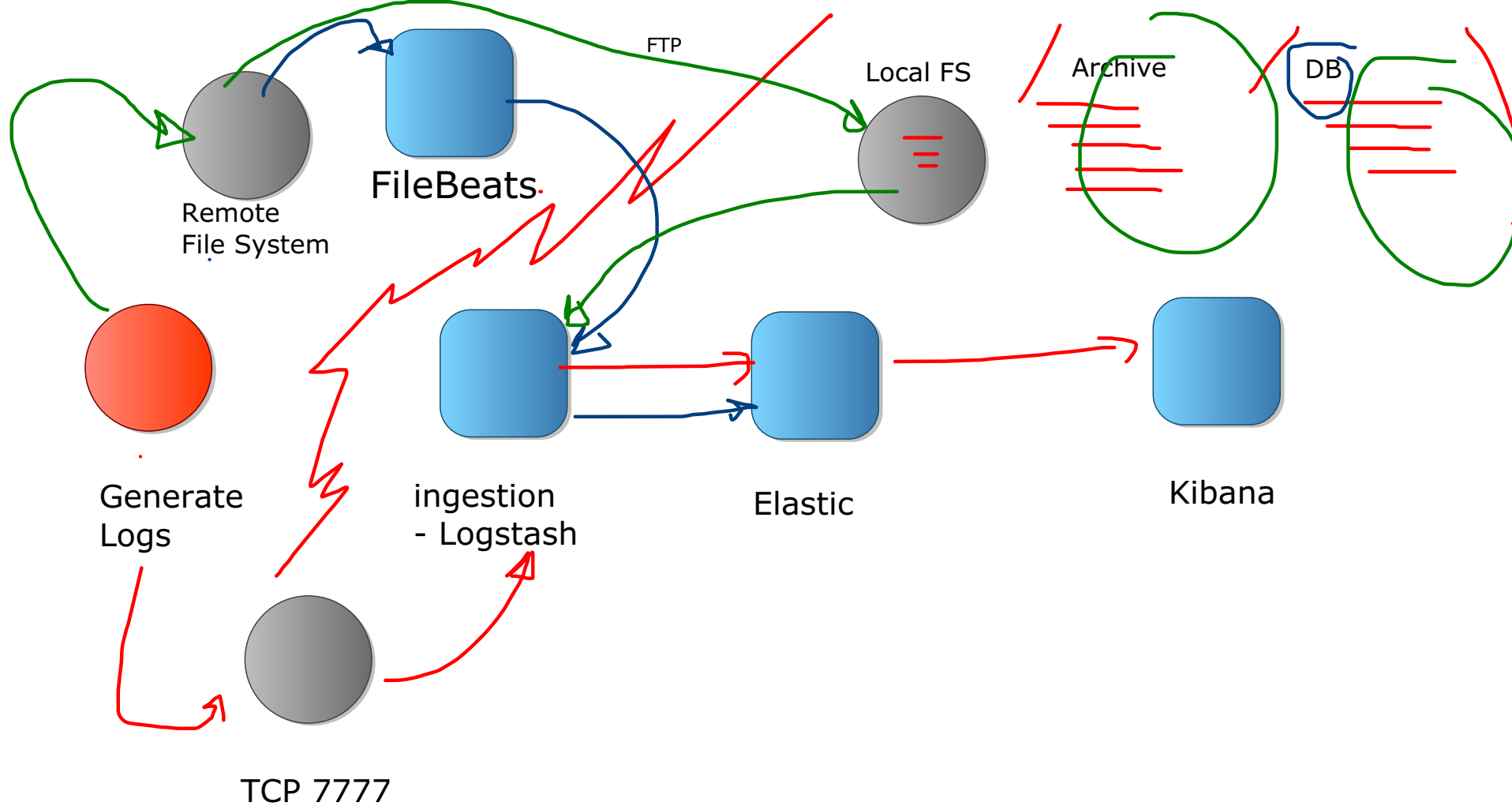
API JAVA / JS /
PYTHON



ELASTIC



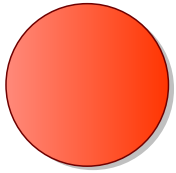
- PERFORMANCE
- BENCHMARK
- MONTEE EN CHARGE
- CLUSTERING
(Connecter en mode Cluster)
- Autres problematiques
 - Concurrency
 - Accès par niveau d'autorisation



GEO POINT PROCESSING

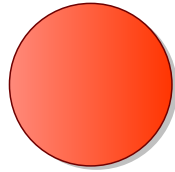
- COMPLEX ASPECT
- BLACK BOX
- NOT WELL DOCUMENTED
- MANY WAYS TO IMPLEMENT
 - FILTER
 - PROCESSOR
- VERY VERSION SENSITIVE
- VERY SYNTAX SENSITIVE

=> WHICH LEAD TO :
NECESSITY OF TRIAL AND ERROR



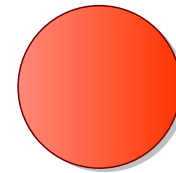
Input

- Sources
- Fichiers
- Directories
- Ports
- JDBC
- Databases
- Streaming
- REST



Filter

- grok
- processors
- geoip
- etc.



Output

- Console
- Log
- Elasticsearch
- Others