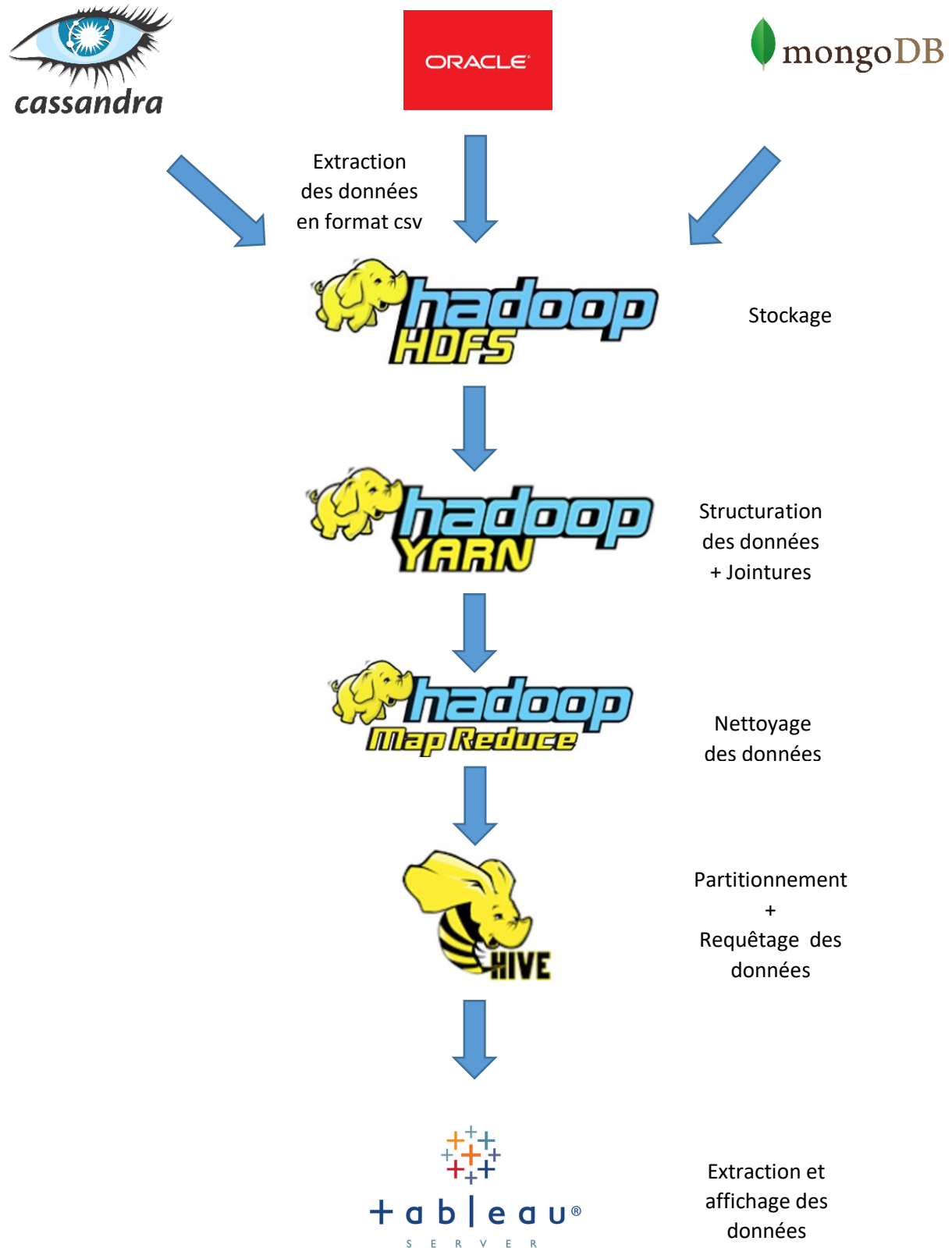


Documentation technique

Diagramme d'architecture



Description des flux

Les données une fois extraites des différentes bases de données sont stockées sur HDFS. Les données seront ensuite traitées par Yarn et Map Reduce mais elles resteront sur HDFS. Ensuite, les données seront organisées avec Hive (mais resteront stockées sur HDFS grâce à des tables externes). Nous créerons une table Hive où nous stockerons les différents indicateurs du SI pilotage. Depuis cette table, nous réaliserons des visualisations sur Tableau avec l'aide de Tableau Server.

Organisation des données

Les données dans les tables externes Hive seront organisées comme ceci :

```
CREATE EXTERNAL TABLE installation_k (  
  pdk STRING,  
  id_concentrateur STRING,  
  date_debut DATE,  
  date_fin DATE,  
  id_constructeur STRING,  
  id_commune STRING  
)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ';'   
STORED AS TEXTFILE  
LOCATION '/tmp/project/installation_k';
```

```
CREATE EXTERNAL TABLE installation_c (  
  pdc STRING,  
  id_compteur STRING,  
  date_debut DATE,  
  date_fin DATE,  
  id_constructeur STRING,  
  pdk STRING,  
  id_commune STRING  
)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ';'   
STORED AS TEXTFILE  
LOCATION '/tmp/project/installation_c';
```

```
CREATE EXTERNAL TABLE commune (  
  id_commune STRING,  
  code_insee INT,  
  libelle_commune STRING,  
  id_departement STRING,  
  id_region STRING,  
  libelle_departement STRING,  
  libelle_region STRING  
)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ';'   
STORED AS TEXTFILE  
LOCATION '/tmp/project/communes';
```

```
CREATE EXTERNAL TABLE evenement (  
  id_evt STRING,  
  date_system_evt TIMESTAMP,  
  type_evt STRING,  
  type_equipement STRING,  
  id_equipement STRING,  
  infos_div STRING  
)  
PARTITIONED BY (date_occur_evt TIMESTAMP)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ';'   
STORED AS TEXTFILE  
LOCATION '/tmp/project/evenements';
```