ITAB
Academy

Présentation de

BIG DATA

ITAB ACADEMY

MapReduce & YARN

- MAPREDUCE V1
- YARN and MAPREDUCE V2
- LAB 1: Creating an application to find max temperature
- LAB2: Creating a WordCount Application

ITAB ACADEMY

Kinani

Date	Article N°	Amount
01012010	26	1000 DH
18012010	27	1500 DH
30012010	26	1000 DH
02022010	27	1500 DH
03022010	26	1000 DH
04022010	26	1000 DH
02012010	25	900 DH
19012010	26	1000 DH
28012010	25	2000 DH
10022010	26	1000 DH
11022010	25	500 DH
12022010	26	700 DH

- **Kinani**

Can we help Kinani ? The shoes store?

- Kinani wants to know his income (CA) per month and per product.
- Can we work all together and answer this question in the minimum time?
- Suppose we have 1000 lines, this can be done in 1 hour by 1 person.
- If we are 20 then we can do it in 3 min.

C’est quoi MapReduce?

Hadoop MapReduce est un framework de programmation permettant d'écrire facilement des applications qui traitent de grandes quantités de données en parallèle.

Les données sont traitées de manière fiable et tolérante aux pannes.

Les blocs de fichiers dans HDFS sont lus et traités par les tâches map s'exécutant sur les DataNodes où les blocs de données sont stockés.

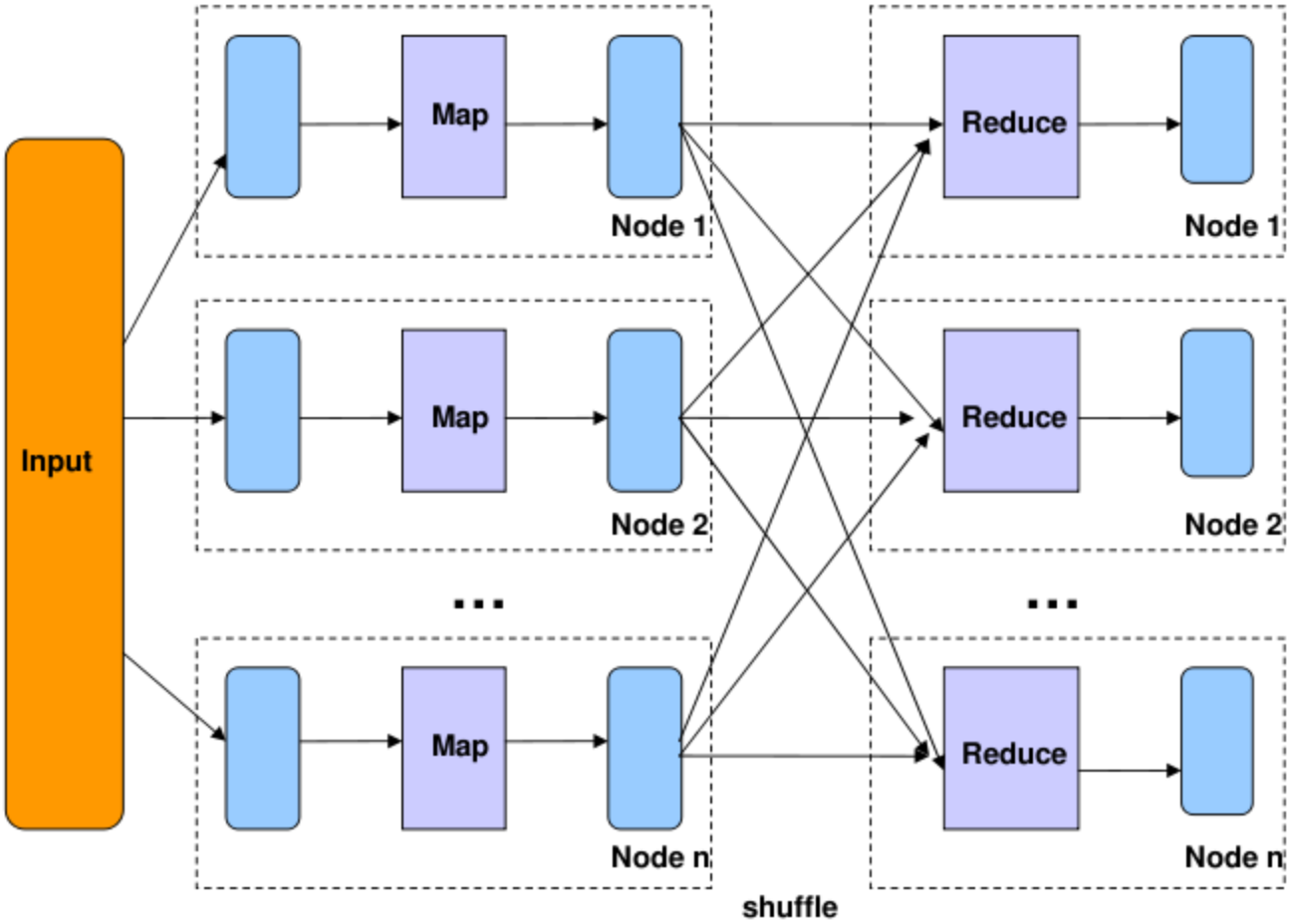
Les sorties des tâches map sont shuffled, puis envoyées aux tâches Reducer.

ITAB Academy

- **What is MapReduce?**

- Hadoop MapReduce is a software framework for easily writing applications which process big amounts of data in-parallel.
- The data is processed in a **reliable, fault-tolerant manner.**
- File blocks in HDFS are read and processed by **Map tasks** running on the **DataNodes** where the blocks of data are stored.
- The output of the Map tasks are **shuffled**, then sent to the Reducer tasks.

Data flow example with parallelism



Qu'est-ce qu'un mappeur?

Un mappeur est généralement un programme relativement petit avec une tâche fonctionnelle simple.

Il est chargé de lire une partie des données d'entrée - un bloc d'un fichier - d'interpréter, de filtrer ou de transformer les données si nécessaire.

Il produit un flux de paires **<clé, valeur>.**

- **What is a Mapper?**
- A **mapper** is typically a relatively small program with a simple functional task.
- It is responsible for reading a portion of the input data — one block of one file —
- interpreting, filtering or transforming the data as necessary.
- It produce a stream of **<key, value>** pairs.

Qu'est-ce qu'un réducteur?

Les réducteurs sont de petits programmes (bien, généralement de petite taille) chargés de réduire les valeurs associées à la clé (ou aux clés) affectée au nœud réducteur.

Toutes les données émises par les mappeurs sont d'abord regroupées localement par la **<clé>** choisie par votre programme.

Pour chaque clé unique, un nœud (réducteur) est choisi pour traiter toutes les valeurs de tous les mappeurs de cette clé.

- **What is a Reducer?**
- Reducers are small programs (well, typically small) that are responsible for reducing over the values associated with the key (or keys) assigned to the reducer node.
- All of the data that is being emitted from the mappers is first locally grouped by the **<key>** that your program chose.
- For each unique key, a node (reducer) is chosen to process all of the values from all of the mappers for that key.

Comment définir le nombre de map pour chaque job?

Le nombre de map dépend de la taille du fichier.

Par exemple, un fichier de 1 Go avec une taille de bloc HDFS de 128 Mo, vous aurez 8 map à executer.

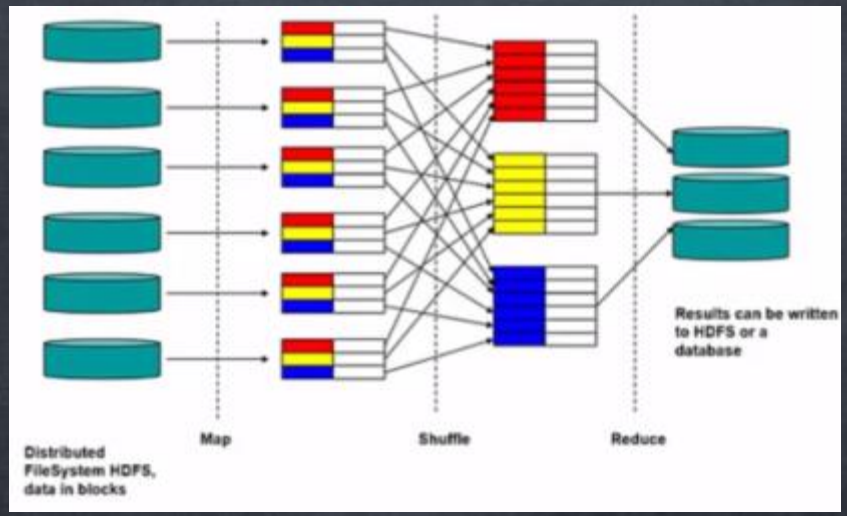
- **How we define the number of maps for each Job?**
- The number of maps depends on the size of the file.
- For example a file with 1 Go and an HDFS block size of 128 MB, you will have 8 maps.

Combien de fichiers chaque réducteur génère?

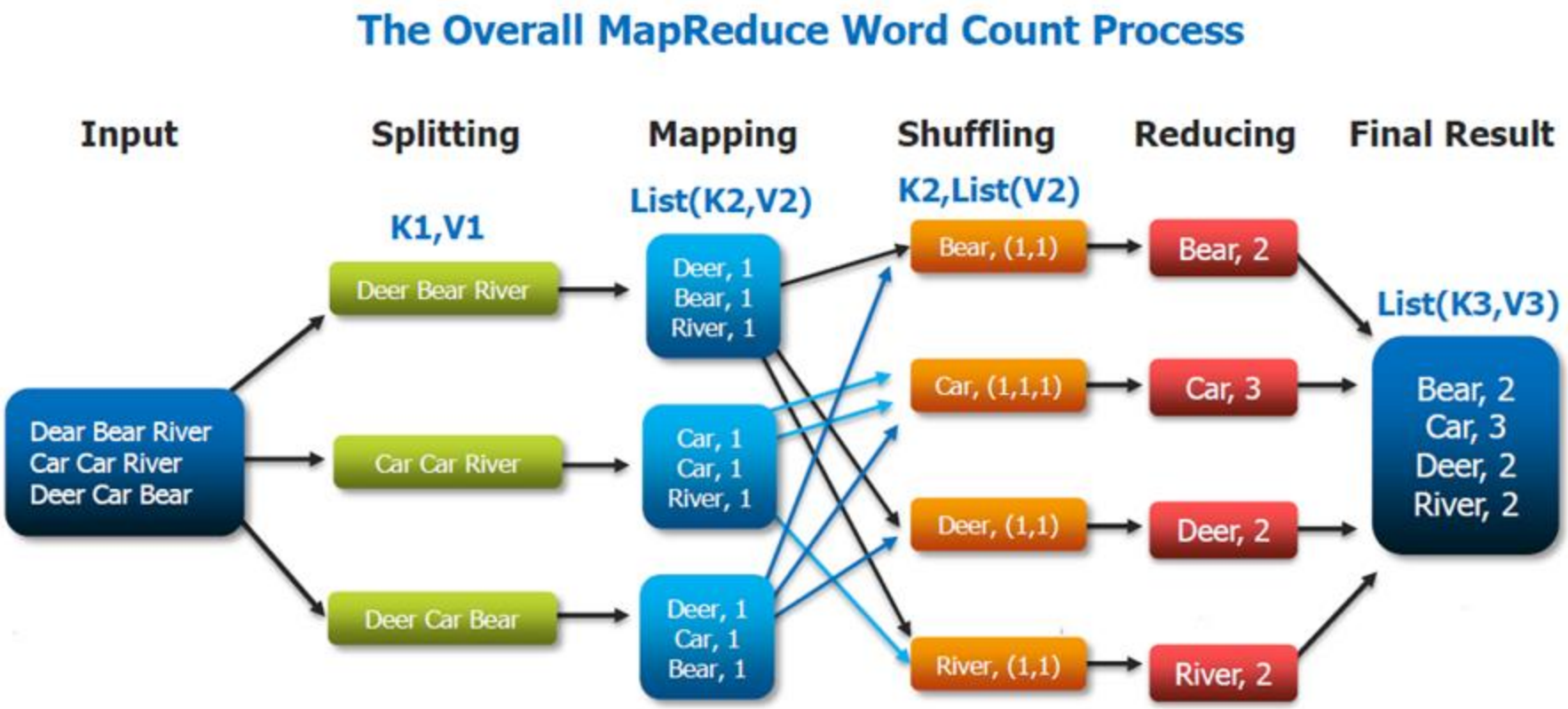
Les reducers produisent la sortie qui est stockée dans HDFS, avec un fichier pour chaque réducteur.

- **How many files each reducer generates ?**

- The Reducers produces the output and that output is stored in HDFS, with one file for each Reducer.



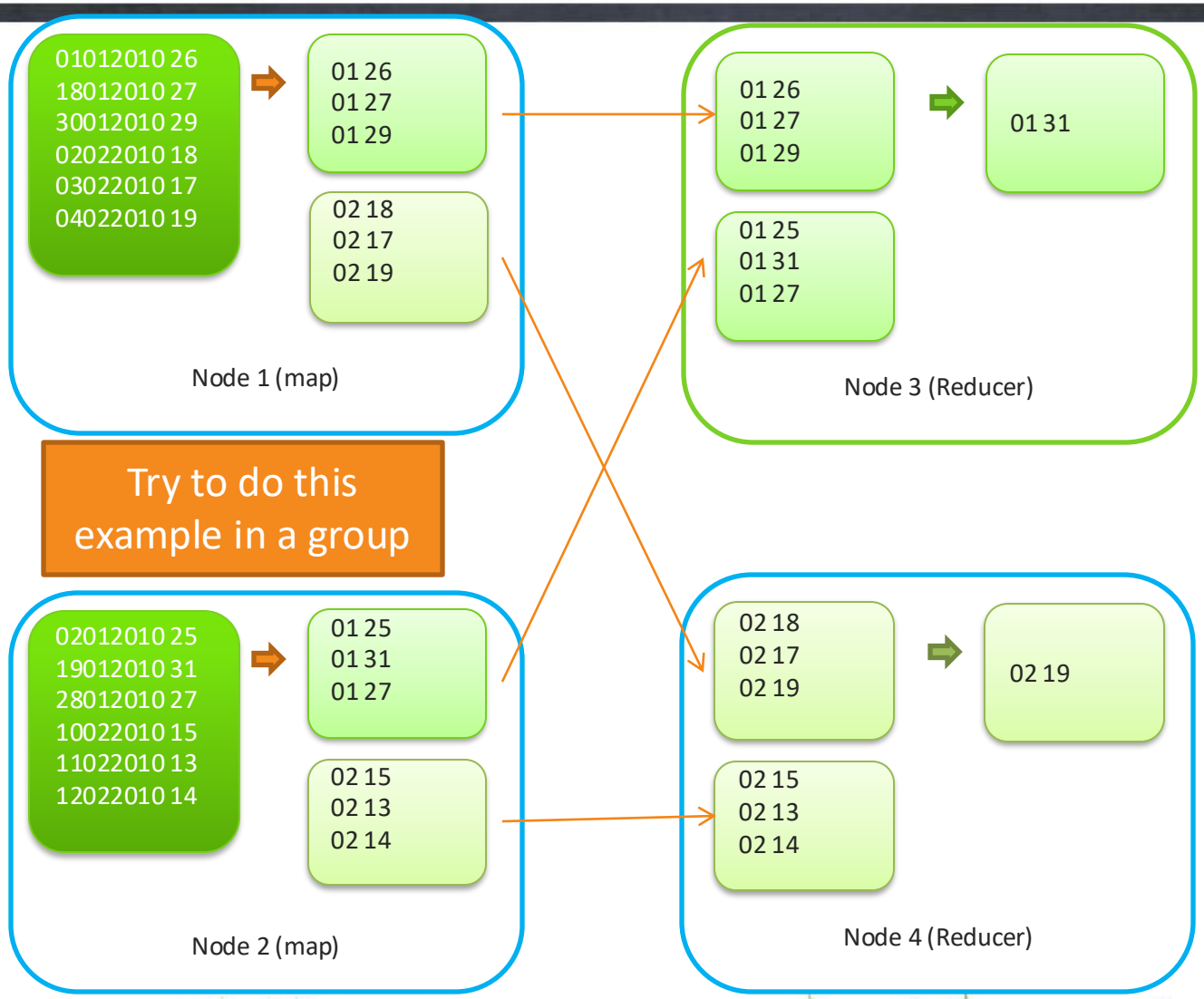
MapReduce & YARN



What is the minimum and maximum number of reducers that I can have here ?

Min – 1
Max - 4

- Example
  - Calculate the max temperature of each month given each days temperature.



Qu'est-ce qu'un combinateur?

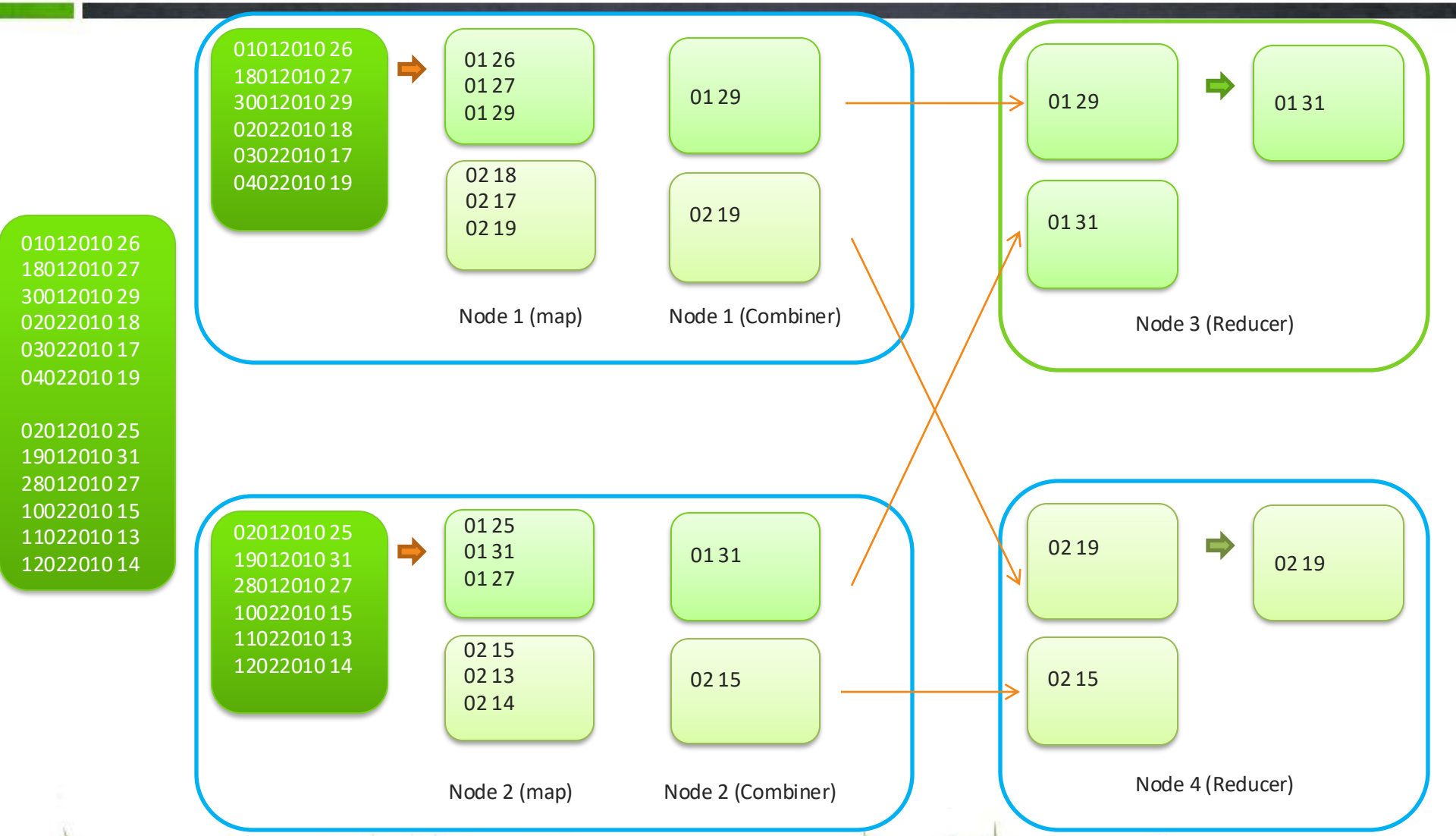
Ceci est connu sous le nom de «mini-réducteur». C'est optionnel.

Il peut traiter les données de sortie du mappeur avant de les transmettre à Reducer.

Le combinateur est exécuté sur le même nœud que le mappeur.

Il peut également être utilisé pour minimiser les données transférées entre les mappeurs et les réducteurs afin de réduire l'encombrement du trafic réseau.

- **What is a Combiner?**
- This is known as 'Mini-reducer'. It is optional.
- It can process the output data from the Mapper, before passing it to Reducer.
- The combiner gets executed in the same node that did the mapper.
- It can be used also to minimize the data transferred between mappers and reducers to minimize network traffic congestion.


01012010 26
18012010 27
30012010 29
02022010 18
03022010 17
04022010 19
02012010 25
19012010 31
28012010 27
10022010 15
11022010 13
12022010 14
01 26
01 27
01 29
02 18
02 17
02 19
01 29
02 19
Node 1 (map)
Node 1 (Combiner)
01 25
01 31
01 27
02 15
02 13
02 14
01 31
02 15
Node 2 (map)
Node 2 (Combiner)
01 29
01 31
02 19
02 15
Node 3 (Reducer)
Node 4 (Reducer)

Qu'est-ce qu'un Job et une tâche?

Job

- ✓ Un Job est composé de nombreuses tâches et représente ce que le client doit exécuter.
- ✓ Pour chaque application, un seul Job est exécuté.

Tâche

- ✓ Un map ou Reduce représente une tâche.

- What is a Job and a Task?
- **Job**
- A job is composed of many tasks and represents what the client need to be performed.
- For each application a single Job is executed.
- **Task**
- A map or a reduce represents a task.

Qui gère les Jobs et les tâches dans MapReduce?

- **MapReduce comprend:**

✓ un seul maître JobTracker
✓ et un TaskTracker esclave par node

- Who manages the Jobs and Tasks in MapReduce?

- **The MapReduce framework consists of:**

  – a single master **JobTracker**
  – and one slave **TaskTracker** per cluster-node

ITAB Academy

ITAB Academy

MapReduce

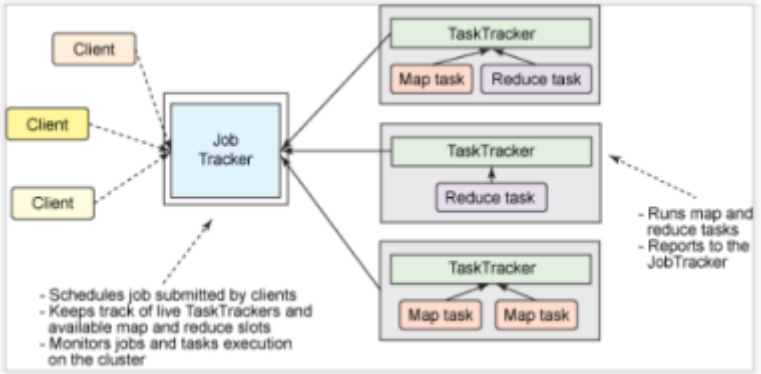
Le JobTracker est responsable de:

La gestion des ressources:

- ✓ Maintenir la liste des nœuds actifs.
- ✓ Maintenir la liste des map et reduce slot disponibles et occupées.
- ✓ Affectation des slots disponibles aux Job appropriés.

Le Traitement de données

- ✓ Demander aux TaskTrackers de démarrer les tâches map et reduce.
- ✓ Surveiller l'exécution des tâches.
- ✓ Redémarrage des tâches ayant échoué.



- The Single **JobTracker** is responsible of:

- **Resource management:**

- Maintaining the list of live nodes.
- Maintaining the list of available and occupied map and reduce slots.
- Allocating the available slots to appropriate jobs and tasks.

- **Data processing:**

- Instructing TaskTrackers to start map and reduce tasks.
- Monitoring the execution of the tasks.
- Restarting failed tasks.

ITAB Academy

MapReduce

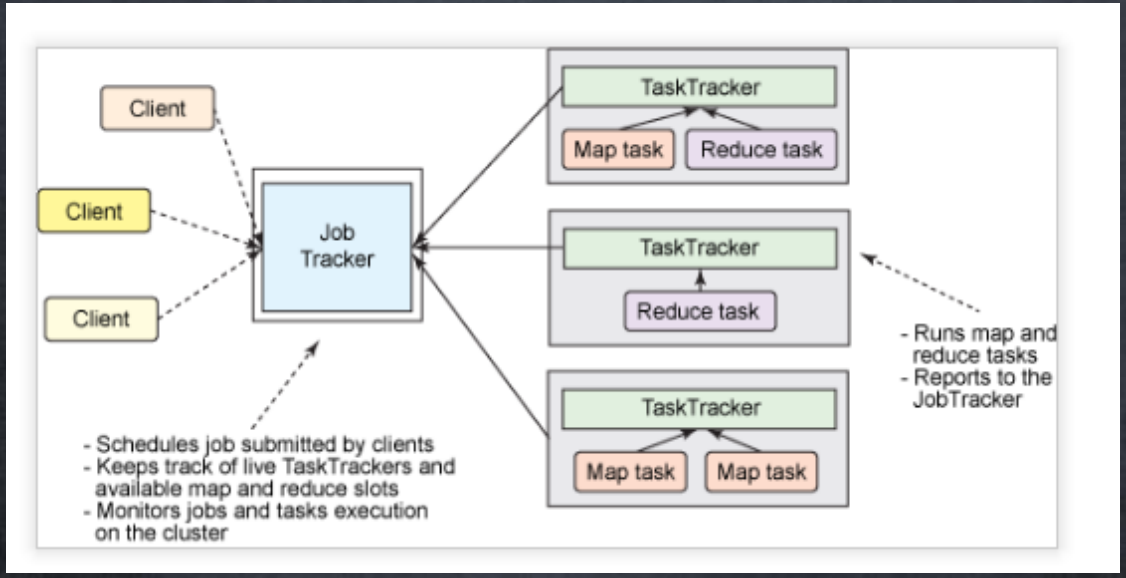
Le TaskTracker est responsable de:

- TaskTracker

✓ exécuter les tâches assignées
✓ et signaler périodiquement le progrès au JobTracker

- The **Task Tracker** is responsible of:

- run assigned tasks
- and periodically report the progress to the JobTracker



Limitations

- What are the limitations of this architecture?

MapReduce Limitation

Si le Single Job Tracker tombe en panne, tous les travaux seront perdus.

- **Limitations of MapReduce (1)**
- **Single Job tracker,** If job tracker fails, all jobs are lost SPOF.
- According to Yahoo, bottleneck is reached with:
  - **5,000 nodes** **(using single machine where job tracker runs)**
  - and **40,000 tasks** running concurrently.

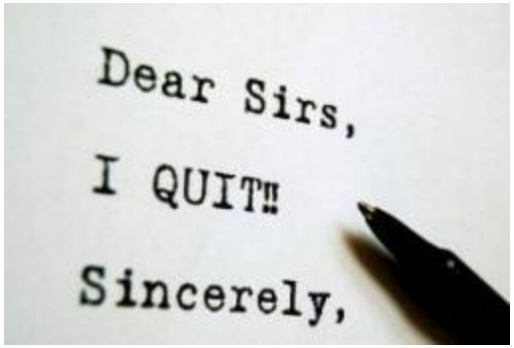
MapReduce Limitation

- ✓ Lorsque tous les slots map sont pris (et nous en voulons toujours plus), nous ne pouvons pas utiliser les slots reduce, même s'ils sont disponibles, ou vice versa.

- **Limitations of MapReduce (2)**
- In Hadoop MapReduce, we have a **fixed number of map and reduce slots** on each slave node.
- These are set by a cluster administrator.
- When all map slots are taken (and we still want more), **we cannot use any reduce slots,** even if they are available, or vice versa.
- A node **cannot run more map tasks** than map slots at any given moment, even if **no reduce tasks are running**.
- This harms the **cluster utilization**

MapReduce Limitation

- ✓ MapReduce supporte que les applications MapReduce, d'autre applications ne sont pas supportés.



- **Limitations of MapReduce (3)**
- Hadoop was designed to **run** MapReduce jobs only.
- It can not run other applications like:
  - **Giraph**
    - Apache Giraph is used to perform graph processing.
    - It is currently used at Facebook to analyze the social graph formed by users and their connections.
  - **Spark**
    - Apache Spark is a analytics engine for big data processing.
  - **Storm and Flink**
    - Streaming Data

In 2010, engineers at Yahoo! began working on a completely new architecture of Hadoop that addresses all the limitations.

ITAB Academy

YAHOO!

is our

Solution

Au lieu d'avoir un seul JobTracker:

1. Un gestionnaire de cluster (ou un gestionnaire de ressources) ayant pour seule responsabilité de suivre les nœuds actifs et les ressources disponibles dans le cluster et de les affecter aux tâches.

2. Pour chaque travail soumis à un cluster, un JobTracker (AM) dédié et de courte durée est démarré pour contrôler l'exécution des tâches de ce travail uniquement.

3. Les JobTrackers (AM) de courte durée sont démarrés par les TaskTrackers (NM ou Node Manager) s'exécutant sur des nœuds esclaves.

- **Instead of having a single JobTracker:**
- A **cluster manager** (or resource manager) with the sole responsibility of tracking live nodes and available resources in the cluster and assigning them to the tasks.
- For each job submitted to a cluster, a dedicated and **short-living JobTracker (AM)** is started to control the execution of tasks within that job only.
- The **short-living JobTrackers (AM)** are started by the **TaskTrackers (NM or Node Manager)** running on slave nodes.

Solution

La coordination du cycle de vie d'un travail (job) est répartie sur toutes les machines disponibles dans le cluster.

Grâce à ce comportement, davantage de tâches peuvent être exécutées en parallèle et l'évolutivité est considérablement accrue.

- **Solution**

- Thus, the coordination of a job's life cycle is **spread across** all of the available machines in the cluster.

- Thanks to this behavior, more jobs can run in parallel and **scalability is dramatically increased.**
  - Because JobTracker was working on a single machine.
  - Now short living jobtracker is running on datanodes. That are many.

C’est quoi YARN

Dans YARN, MapReduce est simplement dégradé en un rôle d'application distribuée et s'appelle maintenant MRv2.

MRv2 est simplement la ré-implémentation du moteur classique MapReduce qui s'exécute sur YARN.

- **What is YARN?**
- YARN is the ressource manager.
- In YARN, MapReduce is simply degraded to a role of a distributed application and is now called MRv2.
- MRv2 is simply the re-implementation of the classical MapReduce engine that runs on top of YARN.

Once upon a time, in 2000 there was a Start-up company. This company had only one project manager that manages all the projects within this company. The projet manager had to manage also the HR. i.e employees at holiday, and the staffing, employees working at which project. The projects are done by the developers under the lead of the single project manager.

In 2018, the start-up company became big and started to gain many projects. They decided to have many project manager. One project manager per project. They added also a role of resource manager. Any project manager who needs developers can get them from the resource manager. He has a visibility on the global staffing sheet of the company. There was also a team lead or manager responsible of a small number of developers.

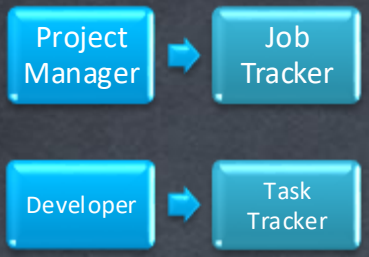
ITAB Academy

Il était une fois une société start-up. Cette société avait un seul chef de projet qui gère tous les projets au sein de la société. Le chef de projet devait également gérer les ressources humaines et le staffing. c'est-à-dire les employés en congés, et sur quel projet il travaillent. Les projets sont réalisés par les développeurs sous la direction du chef de projet.

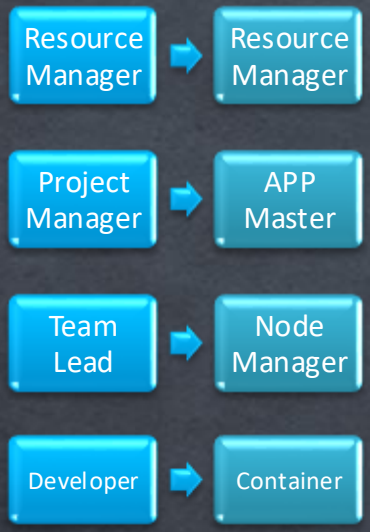
La start-up commence à gagner de nombreux projets. Ils ont décidé d'avoir plusieurs chefs de projet. Un chef de projet pour chaque nouveau projet. Ils ont également ajouté un rôle de gestionnaire de ressources. Tout chef de projet ayant besoin de développeurs peut en obtenir auprès du responsable des ressources. Il a une visibilité sur la fiche d'effectifs globale de la société. Il y avait également un chef d'équipe ou responsable d'un petit nombre de développeurs.

ITAB
Academy

Start-up Company is MapReduce



Big Company is YARN

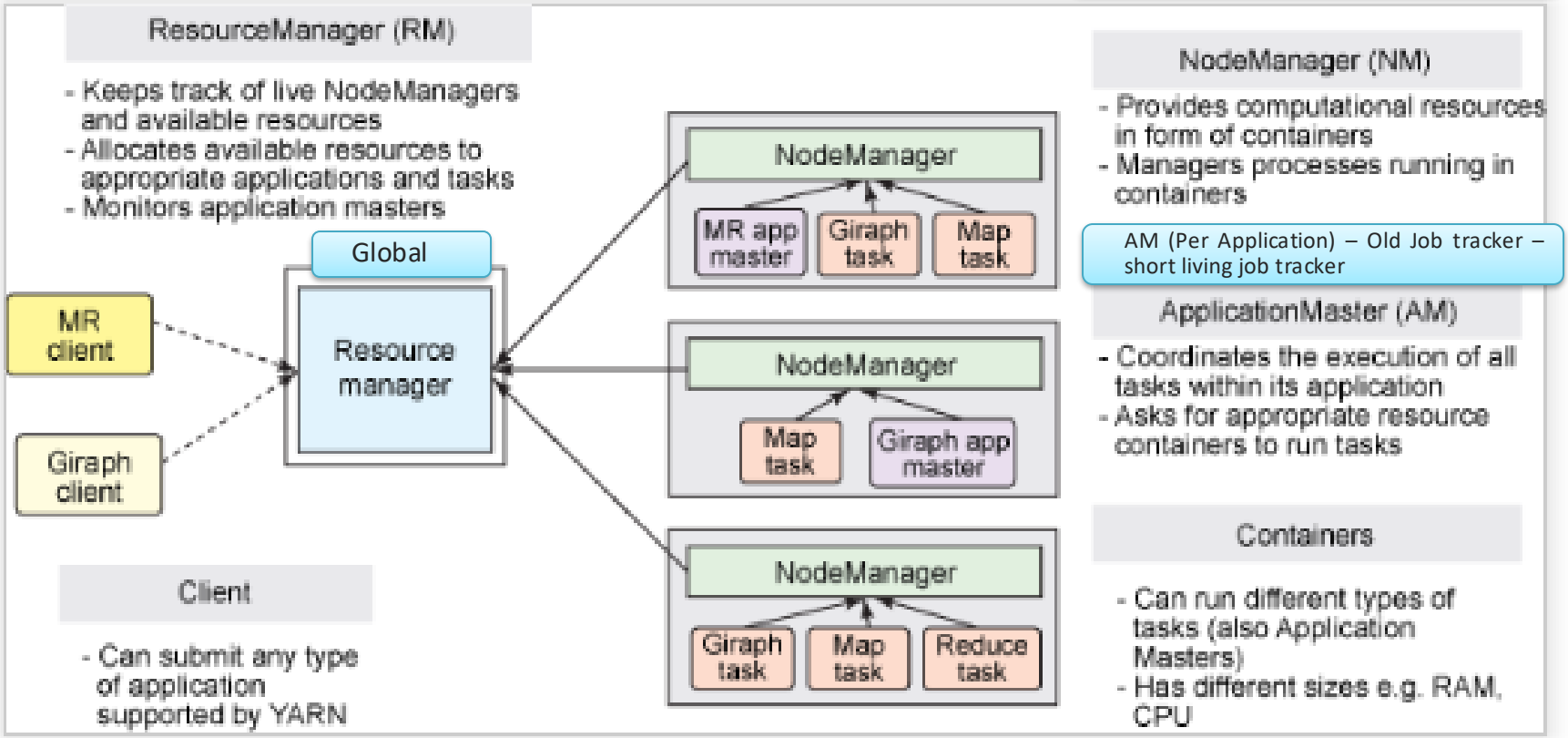


ITAB Academy

MapReduce & YARN

Architecture of YARN

NM (Per Node) – Old task tracker – Dynamic slots



ITAB Academy

Resource Manager (RM)

S'exécute en tant que master daemon sur une machine dédiée.

Il connaît les nœuds actifs et les ressources disponibles sur le cluster.

Affecte les ressources aux applications.

Resource Manager

- Visibilité sur les ressources (congès, Libres, ou affectées)
- Affecte les ressources sur des projets

- **Resource Manager (RM) - global**
- Runs as a **master daemon** on a dedicated machine.
- It knows the live nodes and resources that **are available** on the cluster.
- **Affects** the resources to the applications.

Who was responsible of this in MR v1 ?

Jobtracker

ITAB Academy

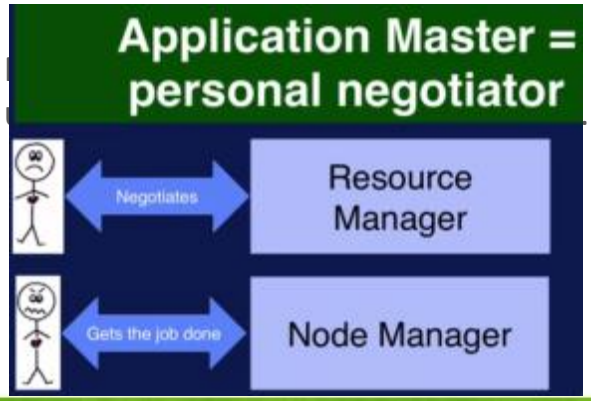
hadoop YARN

Application Master (AM)

Commence lorsque l'utilisateur soumet une application.

Il coordonne l'exécution de toutes les tâches de l'application.

Surveille les tâches et redémarre les tâches ayant échoué.



Chef de projet

- La durée du projet est la durée de l'affectation du chef de projet.
- Il est reponsable de l'execution de toute les taches du projet.

- **Application Master (AM) – per app**
- Started when user **submits** an application.
- It coordinates the execution of **all tasks** within the application.
- **Monitors** tasks and **restarts** failed tasks.
- Can run any type of task inside a **container** (MapReduce or Giraph).

Who was responsible of this in MR v1 ?

Jobtracker

ITAB Academy

Node Manager (NM)

Fournit des ressources de calcul sous forme de conteneurs.

Gère les processus à l'intérieur des conteneurs.

Au lieu d'avoir un nombre fixe de mappers et reducers, le NodeManager dispose d'un certain nombre de conteneurs de ressources créés dynamiquement.

Le gestionnaire de nœud lance le maître d'applications.

Team Lead

- **Node Manager (NM) – Per Node**
- Started when user **submits** an application.
- Provides computational resources in forms of **containers**.
- Manages **processes** inside the containers.
- Instead of having a fixed number of map and reduce slots, the **NodeManager** has a number of **dynamically** created resource containers.
- The Node Manager **launches** the Application Master.

Who was responsible of this in MR v1 ?

Tasktracker

ITAB Academy

hadoop YARN

Container

- ✓ Peut exécuter différents types de tâches.
- ✓ Ils sont créés dynamiquement.
- ✓ La taille d'un conteneur dépend de la quantité de ressources qu'il contient, telles que la mémoire, la CPU, le disque et les E / S réseau.

Developer

- **Container**
- Can run different type of tasks.
- They are dynamically created.
- The size of a container depends upon the amount of resources it contains, such as **memory, CPU, disk, and network IO.**

ITAB Academy

hadoop YARN

Hadoop V1 Vs V2

Yahoo ont utilisé Hadoop 2.0 et il confirment que:

- ✓ Ca a multiplié par deux le nombre de tâches exécutées par jour.
- ✓ Ca a multiplié par deux l'utilisation du CPU.
- ✓ Ca a multiplié par 2,5 le nombre de tâches de tous les JOB.

- **Hadoop 1.0 Vs 2.0**

- According to Yahoo:
  - It multiplied the number of executed Jobs and the CPU utilization by two.
  - It multiplied the number of tasks from all jobs by 2,5.

ITAB Academy

hadoop YARN

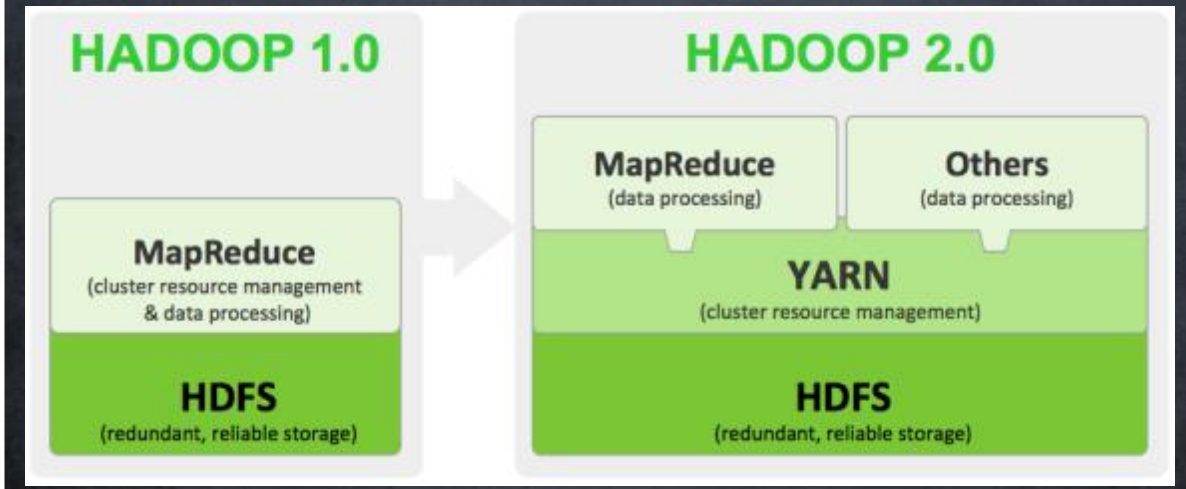
Container

Hadoop 1.0 contre Hadoop 2.0

✓ Dans HADOOP 2.0, MapReduce n'est qu'une application qui s'exécute au dessus du YARN.

- **Hadoop 1.0 Vs 2.0**

- Hadoop 1.0 Vs Hadoop 2.0
  - In HADOOP 2.0 MapReduce is just an application that runs on top of YARN.



ITAB Academy

hadoop YARN

Container

Hadoop 1.0 contre Hadoop 2.0

- ✓ Dans Hadoop V1, nous ne pouvions pas exécuter d'application qui sont pas compatible avec MapReduce tel que:
  - ✓ Giraph (graph processing),
  - ✓ Storm et Flink (streaming),
  - ✓ Spark (computing framework).

- **Hadoop 1.0 Vs 2.0**

hadoop YARN

Container

Hadoop 3.0

- **Hadoop V3**
- Hadoop released hadoop v3.0 almost last year.
- It contains a set of features.
- Minimum Required Java Version in Hadoop 3 is Increased from 7 to 8

Container

Hadoop 3.0

- **Hadoop V3**
- Support for Erasure Encoding in HDFS.
- Erasure Coding is mostly used in ***Redundant Array of Inexpensive Disks (RAID)***.

ITAB Academy

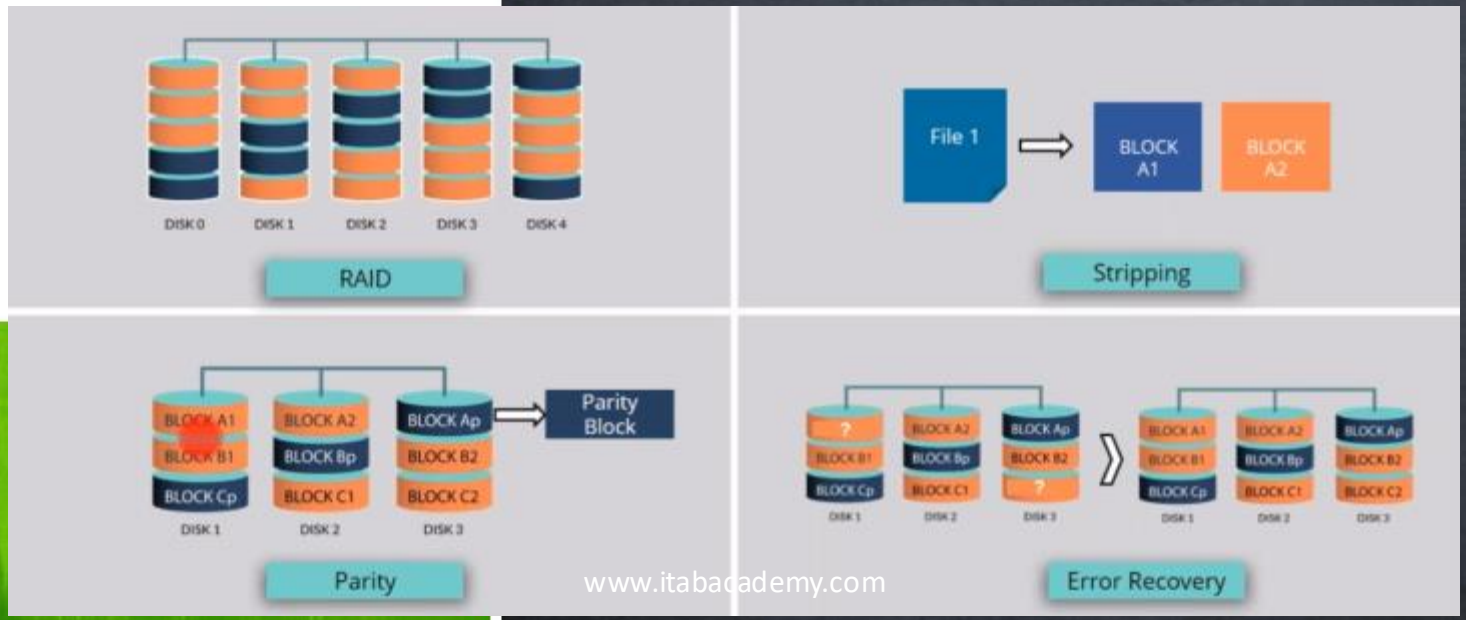
hadoop YARN

Container

Hadoop 3.0

- **Hadoop V3**

- RAID implements EC through ***striping***:
  - A file is divided into smaller units and stores consecutive units on different disks.



ITAB Academy

hadoop YARN

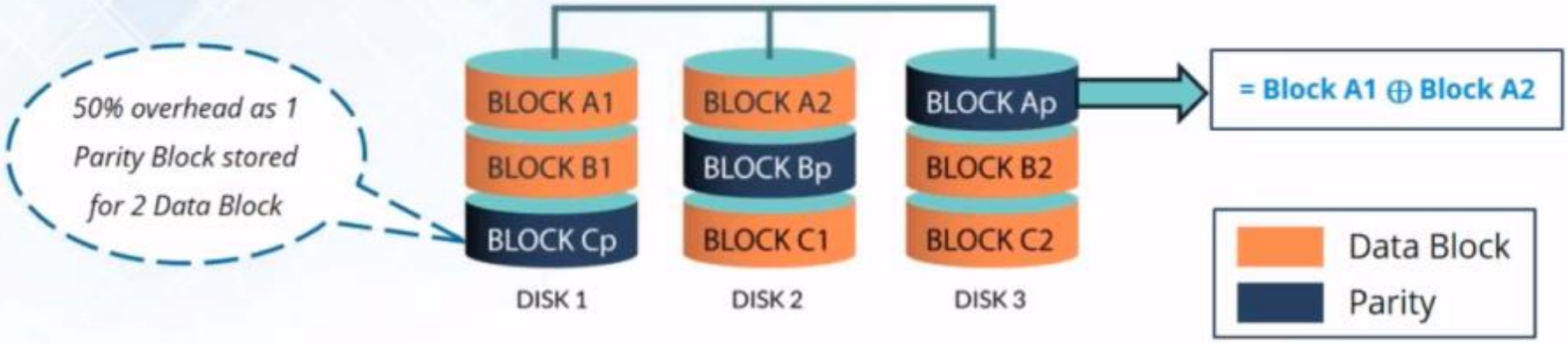
Container

Hadoop 3.0

- **Hadoop V3**

- Integrating EC with HDFS can maintain the same fault-tolerance with improved storage efficiency.

As an example, with EC (6 data, 3 parity) deployment, it will only consume 9 blocks of disk space



ITAB Academy

hadoop YARN

Container

Hadoop 3.0

- **Hadoop V3**
- As an example, a 3x replicated file with 6 blocks will consume 6*3 = 18 blocks of disk space.
- But with EC (6 data, 3 parity) deployment, it will only consume 9 blocks (6 data blocks + 3 parity blocks) of disk space.
- This only requires the storage overhead up to 50%.

Container

Hadoop 3.0

- **HDFS Erasure Coding**

- **Erasure coding policies:**
- Each policy is defined by the following pieces of information:

- *The EC schema:* This includes the numbers of data and parity blocks in an EC group (e.g., 6+3), as well as the codec algorithm (e.g., Reed-Solomon, XOR).

- *The size of a striping cell:* This determines the granularity of striped reads and writes, including buffer sizes and encoding work.

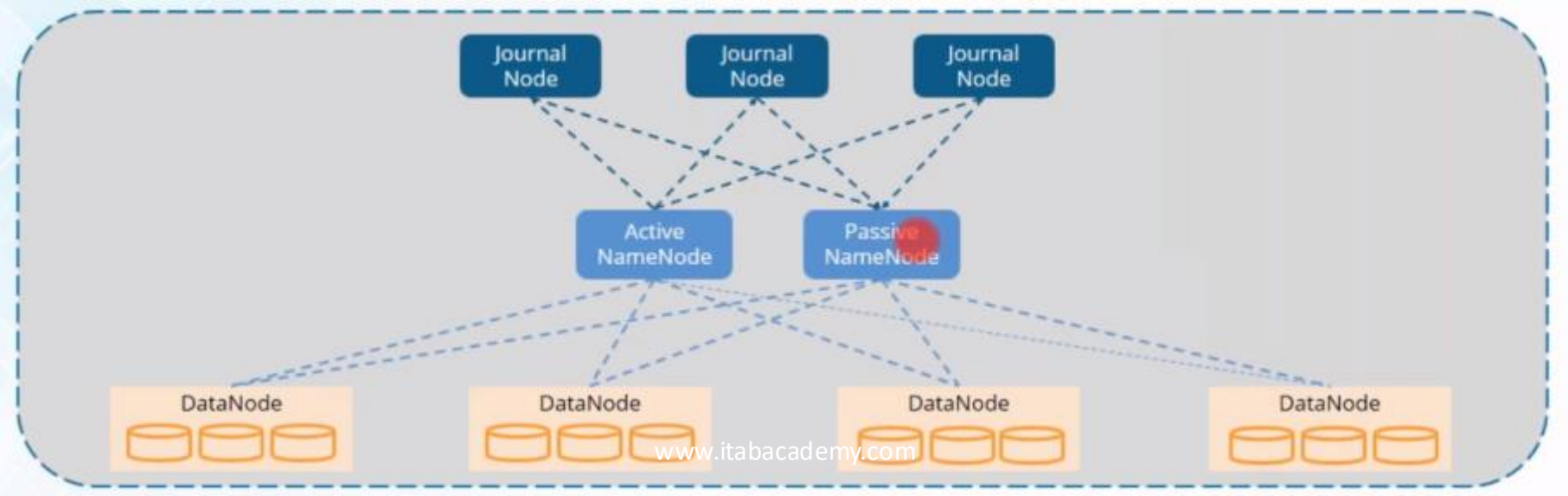
ITAB Academy

hadoop YARN

Container

Hadoop 3.0

- **Hadoop V3**
- In Hadoop V2, we had two name nodes.
- Journal nodes are used to synchronize name nodes.



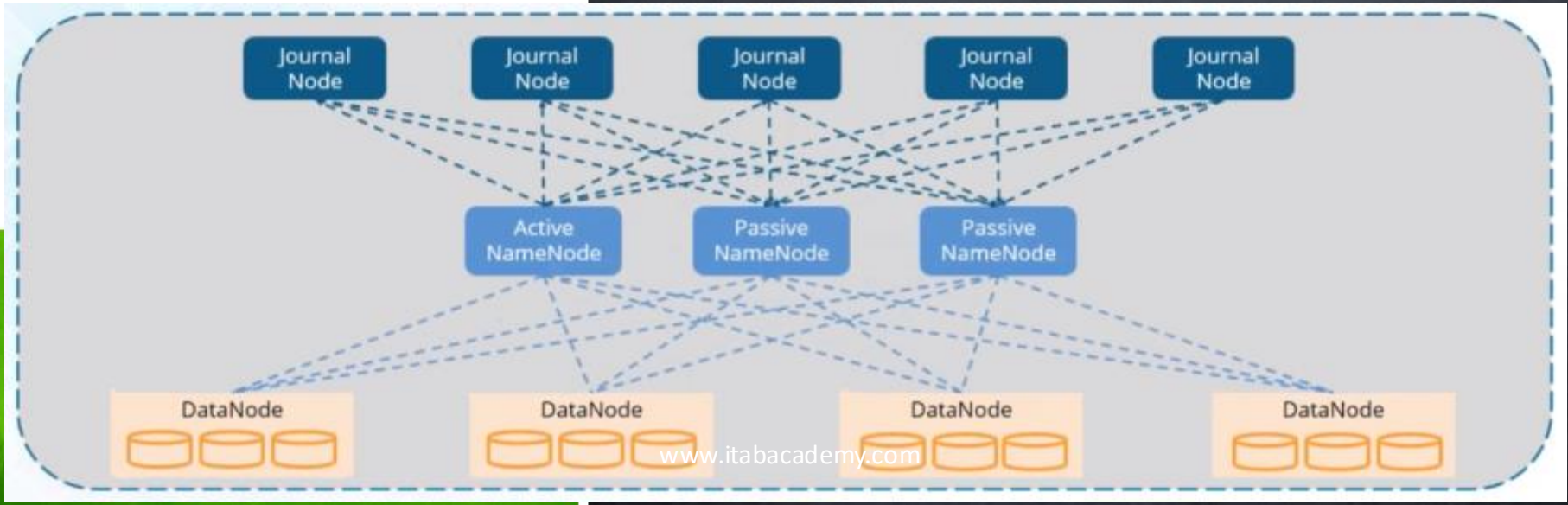
ITAB Academy

hadoop YARN

Container

Hadoop 3.0

- **Hadoop V3**
- In Hadoop V3, we can have more than two name nodes.
- This architecture provides more availability.



Container

Hadoop 3.0

- **Hadoop V3**

- Many enhancements were added to YARN for a better scalability and usability.

ITAB Academy

Container

Hadoop 3.0

- Hadoop V 3.0

https://youtube.videoken.com/embed/N4WzQ1H5h5I
https://hadoop.apache.org/docs/r3.0.0/
https://www.edureka.co/blog/hadoop-3/
https://activewizards.com/blog/hadoop-3-comparison-with-hadoop-2-and-spark/
https://data-flair.training/blogs/hadoop-2-vs-hadoop-3/