2023/3/21

* 沒有期中考，僅有期末考。

〔期末考〕

* 全部選擇題
* 有複選題
* 總題數不超過15題
* A→Artificial

B→Big Data  
C→Cloud Computation

* HPC→High Performance Computing
* Bottleneck for Data Processing  
  Worse as more compute nodes are added.
* System Requirement: Failure Handling

1. Automatic: without manual intervention
2. Transparent: failed component are picked up by alternatives.
3. Graceful: only in a proportional loss
4. Recoverable: can be reclaimed
5. Consistent: Failure does not produce corruption of invalid results.

2023/3/28

* Fault Tolerance → 排除錯誤  
  →Hadoop realize that failure is inevitable.  
  →Trying to minimize the effect of failure.
* Scalability→規模格局之延展性
* Data永遠不會過水NameNode
* SPOF → Single Point of Failure  
  若NameNode掛了，則會發生SPOF

2023/4/11

* 重點：Replication increases reliability & **performance**.
* Performance：More opportunities for data locality.

→ 不用跨網路，Band-Width

* **Record-Reader是用Key-Value**Each Record has a **Key & Value.**
* **Map結果是放在本地磁碟**的內部檔案  
  Intermediate Data is written by the Mapper to local disk （Mapper Local Disk）.
* The **developer** specifies the number of Reducers.
* 例外：Combiner Functions
  + Like the reduce function, finds the maximum for each map output.
  + **提前在Map Node做**
* Job = Many Tasks

〔Presentation簡報〕

* Maximum 20 minutes  
  每組最多20分鐘
* 期末報告 50%

1. Report & Presentation → 30%
2. Question → 10%
3. Comment → 10%

Week #14 → 110%

Week #15 → 105%

Week #16 → 100%

2023/4/18

一張含有 文字 的圖片

自動產生的描述

* Node Manager → 倉庫管理員  
  Resource Manager → 負責調度安排  
  Node Manager聽從Resource Manager在資源調度上的吩咐。

一張含有 圖表 的圖片

自動產生的描述

* Log儲存在HDFS

一張含有 文字 的圖片

自動產生的描述

* Application Master → 領隊、帶頭的。

一張含有 文字 的圖片

自動產生的描述

* Manage Application Masters  
  → Create a Container#0 for Application Masters.   
  為Application Master創立一個編號為0的Container

一張含有 文字 的圖片

自動產生的描述

* 資源放在Node Manager（倉庫）

一張含有 文字 的圖片

自動產生的描述

* MR缺點：Each Container is deleted once a task completes.  
  任務一結束，就刪除Container。
* Spark優點：Each Container stays alive for the lifespan of the application. 保留Container。

一張含有 圖表 的圖片

自動產生的描述

IM File資料放在**Mapper**的**Local Disk**

若僅能備份一份，HA一定要備份在Name Node（Meta Data），因為Name Node掌管資料（Data）

2023/5/2

* Resource Manager只管兩個資源：CPU & Memory

一張含有 文字 的圖片

自動產生的描述

2023/5/9

先開NameNode、再開DataNode

先開Master、再開Worker