**姓名:**

**選擇/填空: (80%)**

1. **下列何者不是促進大數據技術演進的主要原因?**
2. 量產化硬體
3. **函式語言的興起**
4. 摩爾定律
5. 儲存設備成本降低
6. **請問下列何者不是大數據三個V中的其中一員?**
7. Velocity
8. **Victory**
9. Volume
10. Variety
11. **Hadoop 是以什麼程式語言撰寫? Java**
12. **關於Hadoop 的特色，下列何者為非?**
13. 高容錯性
14. 可以使用量產化硬體
15. **可以解決任意數據問題**
16. 採用主從式架構
17. **關於Hadoop 的敘述，下列何者為真?**
18. 可以用來取代資料庫
19. **批次處理大數據**
20. 可以即時產生出統計報表
21. 快速進行機器學習
22. **關於MapReduce 的敘述，下列何者為誤?**
23. **一定都會有Mapper及Reducer**
24. 可以使用Java 以外程式語言實做MapReduce
25. 可容錯
26. 單一Slave工作失敗時，Master 會重送工作
27. **關於HDFS 的敘述，下列何者為誤?**
28. **可以快速存取單筆資料**
29. 建立於原生檔案系統上
30. 一次寫入，多次讀取
31. 高容錯性
32. **如果我要讓公司所有人都可以透過瀏覽器操作Hadoop，我可以使用哪個模組建立操作介面?**
33. **HUE**
34. Hive
35. MapReduce
36. HDFS
37. **下列何者非HDFS的優點?**
38. **適合儲存小檔案**
39. 適合建立在量產化硬體上
40. 高容錯性
41. 適合寫入串流資料(Streaming)
42. **請問如果要修改HDFS儲存的複本數，請問hdfs-site.xml 中的 name 中應該填入什麼值?** **dfs.replication**

**<property>**

**<name>\_\_\_\_\_\_\_\_\_</name>**

**<value>3</value>**

**</property>**

1. **客戶端是向哪個代理服務(Daemon) 存取資料?**
2. Name Node
3. Secondary Name Node
4. Node Manager
5. **Data Node**
6. **如果有一個檔案大小為199MB，在Block Size 為128 MB及副本數為3 的設置下，該檔案會被分成多少個資料塊(**Block**)?6**
7. **如果要透過HDFS fs command下載檔案，請問指令 hadoop fs \_\_\_ test.txt 空白處該填入哪個指令?**
8. –du
9. –df
10. –put
11. **–get**
12. **如果我要檢視HDFS中 data目錄下的內容，空白處應該填?**

**hadoop fs \_\_\_  /data -ls**

1. **如果我要讓HDFS 可以即時備援Name Node，以防止單點毀損(Single Point Failure)，我該採取哪種架構?**
2. 在同一台機器上開啟兩個Name Node服務
3. **建立HDSF HA架構**
4. 使用Secondary Name Node 備援
5. 使用HDFS Federation 備援
6. **關於Secondary Name Node 的敘述，下列何者為非?**
7. 最好是將Secondary Name Node 配置於第二台機器
8. Secondary Name Node的所儲存最新的Meta 資料大概晚Name Node 約一個小時左右
9. Secondary Name Node 需要跟Name Node 一樣大的記憶體
10. **Secondary Name Node 可以即時備援**
11. **如果我要讓HDFS 可以使用多個Name Node 管理不同Data Node，以避免Client大量存取同一Name Node時遇到效能瓶頸 ，我該採取哪種架構?**
12. 使用Journal Node
13. 在同一台機器上開啟兩個Name Node服務
14. **使用HDFS Federation**
15. 使用Secondary Name Node
16. **如果要透過Java 呼叫API 存取HDFS，可以去哪個目錄夾找到相關JAR 檔(假設路徑已在hadoop 安裝目錄下)?**
17. bin
18. etc
19. **share**
20. conf
21. **如果我要將HDFS內的某檔案(test.txt)權限設為 rw-r--r--，則下列指令**

**hadoop fs –chmod \_\_\_ test.txt 中的空白處我該填入什麼? 644**

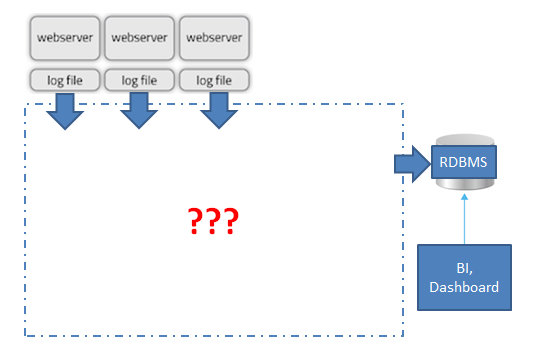
1. **預設連結HDFS所使用的連接port 為\_\_\_\_? 9000**
2. **下列何者不屬於YARN(MRV2)的元件?**
3. ResourceManager
4. NodeManager
5. **JobTracker**
6. ApplicationMaster
7. **在一MapReduce 程式中, Reducer 最少可以有 \_\_\_ 個？0**
8. **在Mapper 類別: class mapper<I1,I2,I3,I4>{…} 中, 請問參數I2 代表什麼意思？**
9. output key
10. **input value**
11. input key
12. output value
13. **以下關於Sqoop 的敘述，下列何者為非？**
14. Sqoop 支援匯入csv 檔
15. Sqoop 可以交換資料庫與HDFS 中的資料
16. **可以使用Sqoop 過濾匯入/匯出的資料**
17. Sqoop 支援增量匯入(Incremental Update
18. **請問Sqoop 是用什麼驅動程式存取資料庫? \_\_\_ (請小寫) jdbc**
19. **請問如果要將資料庫資料匯入至HDFS 中，該使用下列哪個Sqoop 指令?**
20. sqoop select
21. **sqoop import**
22. sqoop insert
23. sqoop export
24. **請問下列何者並非Flume 所包含的元件?**
25. **exec**
26. channel
27. source
28. sink
29. **以下關於Flume 的敘述，下列何者為非？**
30. 可以加值資料
31. 可以過濾傳輸中的資料
32. 可以清除資料中重複的部分
33. **可以從HDFS匯出資料至Flume 所在的Server**
34. **假設如果我要在Flume中設定從將資料串流寫進HDFS中，我該在Flume 中設定哪個元件?**
35. Source
36. Channel
37. **Sink**
38. **請問下列何者並非Hive 的優點?**
39. **使用者可以Ad-Hoc分析資料**
40. 可以依時間做資料分割(partition)
41. 執行速度快
42. 簡單易用
43. **請問下列關於Hive 跟資料庫比較的敘述，何者為是?**
44. **Hive 是以MapReduce 存取資料**
45. 使用Hive 可以即時完成資料處理與分析
46. 在Hive 中可以使用MySQL全部的語句
47. Hive 可以取代關聯式資料庫
48. **請問下列關於Hive 的敘述何者為非?**
49. Hive 可以使用類似MySQL的語句操作資料
50. Hive 是將資料存放在HDFS 上
51. **使用Hive 跟使用MapReduce 的執行效能相近**
52. Hive 是透過MapReduce 來操作資料
53. **請問如果要修改對Metastore資料庫連線的設定，請問要修改哪個檔案?** **hive-site.xml**
54. **請問如果要建立Hive 的外部表格, 下列空白處應填入(請全部以小寫表示)?**

**create \_\_\_\_\_\_ table External**

1. **以下對Spark描述何者為非?**
2. 是一個叢集式運算框架
3. 以記憶體(Memory)作為運算的儲存媒體
4. **是用來取代Hadoop的產品**
5. 運算速度較Hadoop快10~100倍
6. **以下對RDD的轉換類(Transformations)操作的描述何者為非 ?**
7. 用來將輸入的RDD轉換為另一個RDD
8. 是Lazy的運作
9. 具有容錯的特性
10. **主要以磁碟作為運算資料儲存媒體**
11. **以下操作何者不為RDD的轉換類(Transformations)操作 ?**
12. map
13. filter
14. **count**
15. sortBy
16. **Which one is Start point for SparkSQL?送分**
17. scContext
18. SQLContext
19. HiveContext
20. myContext
21. **Spark Streaming不是即時運算的類庫，而是近即時運算。**
22. **是**
23. 否
24. **Spark Streaming支援哪種資料來源？**
25. Kafka
26. HDFS
27. Flume
28. S3
29. **以上皆是**

**情境題1(15%) :**

目前有個系統建構需求，希望能接收前端WebServer的Log資料，由Log資料中截取感興趣的欄位進行分析及統計，並將分析結果存入關聯資料庫中，以便透過BI或Dashboard查看分析結果(如下圖):



1. 請問要完成上述功能，藍色虛線方框中需要使用那些Hadoop Ecosystem / Spark Library Stack 成員?
2. 承上題，Hadoop Ecosystem / Spark Library Stack 成員的使用順序為何 ? 每個成員負責什麼功能 ?
3. 試分析您規劃的成員組合的分析即時性為何 ? (即時性的可能選項:即時/近即時/批次作業)

**情境題2 (5%)(PS:本題有寫就有分，請勿放棄):**

1. 課程中您對那部份最有感覺/心得?
2. 課程中遇到最大的困難是那個部份 ?
3. 未來是否會考慮使用大數據平台?
4. 希望再對那個部份深入學習?