The all presented herein are not for any purpose except for students' study reference in this class only, and not to be reproduced without prior consent.

The copyright, names and brands may be claimed as the property of others, please fully respect them all in the published media.

NTUT – 1st Classday Check List Big Data Administration & Technology(BDAT)

- 1. Course Curriculum
- 2. Syllabus and Lecturing Disciplines (attendance, copyright, time table, etc.)
- 3. Grading Policy
- 4. Textbook and Reference
- 5. Individual Study & Learning Assessment 40%
- 6. Topic for Teamwork presentation 50% BD Hadoop Ecosystems
- 7. Study & Presentation Assignment Cover Page Introduction
- 8. Teaming (Max.3 students per team)
- 9. Class Communications LINE Group (next slide) & Google Drive
- 10. Tools for Remote Class TBA (Microsoft Teams and/or Google Meet)
- 11. Lecturer Introduction
- 12. Questions and Answers

Class Attention + Self Studies + Team works + Peer Learnings

Textbook and Readings:

- In-class handout and research papers downloaded from Google Drive
- Research Paper Readings (Google's issued papers)
 - 1. The Google File System; https://static.googleusercontent.com/media/research.google.com/zh-TW//archive/gfs-sosp2003.pdf
 - 2. MapReduce: Simplified Data Processing on Large Clusters; https://static.googleusercontent.com/media/research.google.com/zh-TW//archive/mapreduce-osdi04.pdf
 - 3. Big Data associated, To-Be-Advised in class
- Additional Reading Reference
 - 1. Hadoop: The Definitive Guide, Fourth Edition, April 2015, Published by O'Reilly Media, Inc., Editor: Tom White
 - 2. Hadoop Essentials, First Edition, April 2015, Published by O'Reilly Media, Inc., Editor: Shiva Achari.
 - 3. Hadoop Operation, First Edition, September 2012, Published by O'Reilly Media, Inc., Editor: Eric Sammer
 - 4. 大數據大時代-新一代儲存技術及實作,作者:查偉,發行所:佳魁資訊,出版日期:2017,3月
 - 5. Yusuf Aytas, Designing Big Data Platform, How to Use, Deploy and Maintain Big Data Systems, Wiley, February 2021



Why I need to study the course?

Is it just of mandated courses for pursuing my further study/degree?



- Enrich my IT Knewledge
 Widel my Viewscope and up-to-date
 Deepen my Career Vision
 Strengthen my Professional Capacity
 - Make me Feel Better (奇樣子) & More Successful

To prove -> I can meet the above all

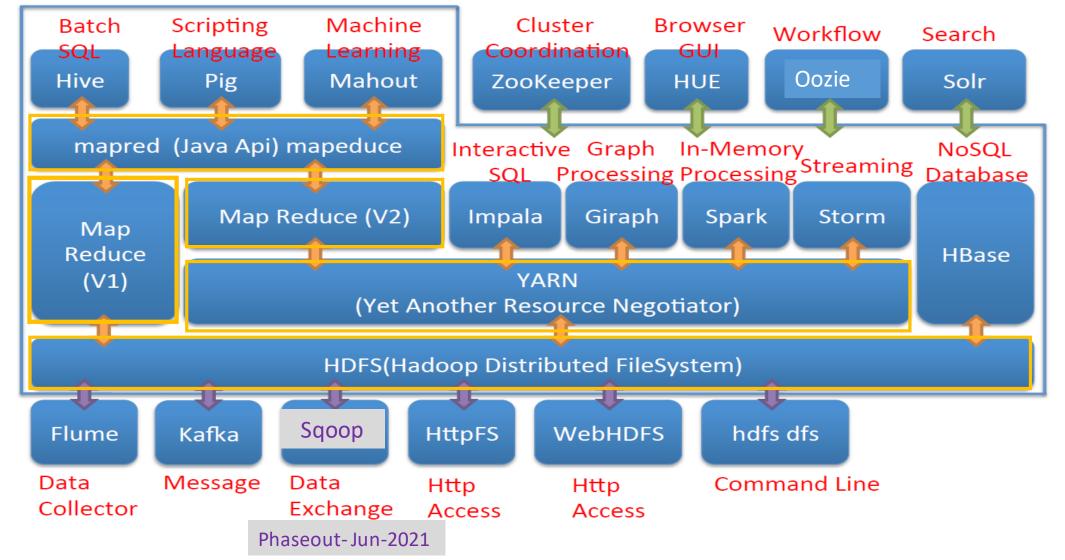
```
public class ConnectionSpout implements IRichSpout {
    private SpoutOutputCollector collector;
    // Define the name of each column in the training data
    private String[] field_names = new String[]{"duration", "protocol_type"
        , "service", ... };
    @Override
    public void open (Map conf, TopologyContext context,
        SpoutOutputCollector collector) {
        // Define the collector to be used for emitting tuples
                                           = null) {
                                           ");
                                           (fields[0], ..., fields[40]));
                                           FieldsDeclarer dec) {
                                           of the tuples
```

Storm ConnectionSpout.

[0], ..., field_names[40]));

Hadoop EcoSystem

All below w/ red-color marks



6

The Hadoop Ecosystem

- The part of related tools integrated with Hadoop
 - Data base: HBase
 - Data processing: Spark
 - Data analysis: Hive, Pig and Impala
 - Data discovery: Solr (Cloudera Search)
 - Data ingestion: Sqoop, Flume, Kafka
 - Data real-time management: Storm
 - Machine learning: Mllib, Mahout and others
 - Coordination: ZooKeeper
 - User experience: Hue
 - Workflow management: Oozie
 - Cluster management: Vendors' solution like Cloudera Manager
- The above are not considered as "core technology of Hadoop"
 - They are part of the Hadoop ecosystem
 - Many are open source Apache projects

NTUT 111-02 BD LINE Group



Lecturer's Email Address: bnwntut8@ntut.edu.tw

https://line.me/R/ti/g/rb4Lyijn2L

Call to Actions

- Class Registration reconsideration and/or confirmation
- LINE Group participation for class communication
- Teaming and Login (Spreadsheet)
- Access the class folder "Released to Students"
- Class attendance in time

Folder Access

https://drive.google.com/drive/folders/1jMcn6d1tW0K11Ul_L QcRHEDV_bq2EzxO?usp=sharing