Team's Study and Presentation – 50%

- Title: Ecosystem Study of Big Data Hadoop
  - Topic selection
  - Teamwork: 3 (Max. 4) students per team
  - Coverage
  - Limited and ≤ 20 slides including reference per topic
  - Presentation time max. 20 minutes per team
  - Template(next slides)
  - In English only
- Report and Oral Presentation on stage
- File Submission with MS PPT format
  - Due Date: 12:00 noon time, by week#16
  - Send team's PPT to
    - https://drive.google.com/drive/folders/1sTzACfbGFdDd6AdCfhFaPhqOy4qMU6lV?usp=share_link

Grading Policy for Assignment – 50% of the total score

- Report and Oral Presentation: 30%
- Over-Spot Q&A : 10%
- Comments Making: 10%

Hadoop Big Data Ecosystem and Trend
Hadoop應用生態與趨勢 (max. 20 minutes)

Report-Topic: ???

Team ID#:

University: National Taipei University of Technology(NTUT)

Course Name: Big Data Administration and Technology

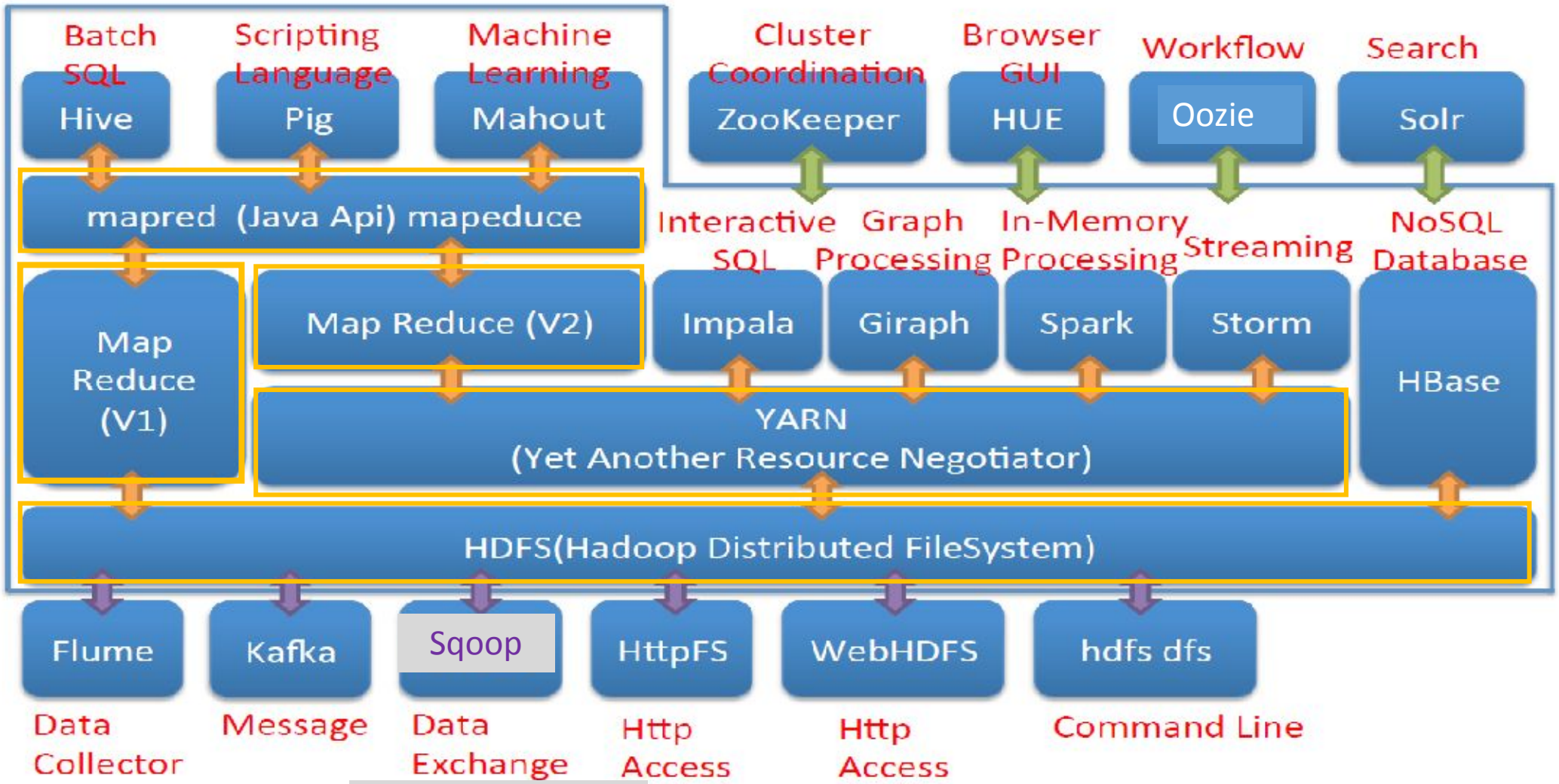
Student Names and ID# (per team):

Submitted Date: DD/MM/YYYY

Weight of the total score: 50%

Hadoop EcoSystem

All below w/ red-color marks



Phaseout- Jun-2021

Source: Cloudera Hadoop https://www.cloudera.com

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: ~~Sqoon~~, 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 but prominent of the Hadoop ecosystem
  - Many are open source Apache projects

Report and Presentation - 30%
Development Backgrounds 6 Ws (10%)

- Who/Owner(s)
- When
- Why
- Where
- What/Abstract
- How

Key Features & Benefits (20%)

- Provide the key and prominent description

How Does It Work (20%)

- Architecture
- Algorithm
- Flow Chart or Work Flow
- Its Ecosystem (if any)

Application Case Study (25%)

- Work Example(s) and Briefing

Study Comments (20%)

- Your abstract/evaluation?
- Any impact to you from your study and understanding?
- Any comparison and alternative better technology?
- Any questions? How (methodology) to find the answers by yourself?
- Q&A, follow ups

Reference(s) (5%)

- Title, Author(s), Publisher, and Published Date

Team's Action Request (AR)

- Finalize the topic associated w/ Hadoop Ecosystems, by *week#5*
- Book a timeslot for presentation, by week#5
- *1st in-class Team Discussion for the draft*, *in week#9*
- Present it with your team's submission(PPT), by *week#16, (max. 20 minutes per)*
- Score multipliers for presentation vs. timeline
  - 110% in week#14, max. 2 timeslots
  - 105% in week#15, max. 4 timeslots
  - 100% in week#16, no limits
  - 0% after , no limits

Assignment Q&A and Comments – 20%

Actions per Team – Max. 20 scores per team

Q. Issue questions per presented topic/team – 10 scores

- Non-duplicated question – neither nor in/externally
- 2 score per hit on 1 topic
- Max. 1 question per Topic
- Max. 5 topics by your team's selection; Max. 5 questions in total
- Max. 10 scores in team's scope
- Action: Issue questions while in Q&A session per topic/team presentation

C. Make ***constructive*** comments on a presented topic in respects of its sections – 10 scores

1. **How does it work? 5 scores**
2. **What about Its applications? 5 scores**

- Non-duplicated comments per
- 1 score by bullet comment per section per topic/team
- Max. 5 topics (2 sections per topic); Max. 10 bullet comments
- Max. 10 scores in team's scope
- Actions:
  - Edit them in English by file with PPT format in 1 slide per topic
  - File title: Team(ID#)-Project Comments
  - Upload the file by week#17 by team to
    - https://drive.google.com/drive/folders/1iHVXW-XyGKGlg9EqgbBoW_7ycDGQZzn6?usp=sharing

Big Data Ecosystem Study
Comment Summary -Template

Team ID#:

University: National Taipei University of Technology(NTUT)

Course Name: Big Data Administration and Technology

Student Names and IDs (of all students per team):

Submitted Date: DD/MM/YYYY

Weight of the total score: 10% (of Class Report and Activity)

Presented Team ID# - Topic (1 slide per topic)

1. How does it work?

- *

3. What about its Applications?

- *