請尊重智慧財產權,合法影印資料並使用正版教科書。 Please consult Intellectual Property Rights before making a photocopy. Please use the textbook of copyrighted edition.

## ③国o支束華大學 教學計劃表 Syllabus

課程名稱(中文) Course Name in Chinese	大數據系統				學年/學期 Academic Year/Semester			
課程名稱(英文) Course Name in English	Big Data Systems							
科目代碼 Course Code	CSIEM0410	系級 Department 碩士 & Year		開課單位 Course-Offering Department	資訊工程學系			
修別 Type	選修 Elective	學分數/時 Credit(s)/Hou		;	3.0/3.0			
授課教師 Instructor	/吴秀陽							
先修課程 Prerequisite								
課程描述 Course Description								

The term "big data" is now commonly used to mean that the growth of data in volume, velocity, variety and veracity are in such an unprecedented scale that traditional database management systems can no longer handle it properly. New technologies, artificial intelligence(AI), machine learning(ML) and internet of things(IoT) in particular, rely heavily on the processing of huge data sets. Online services (ChatGPT, YouTube, Meta, IG, …) need to handle hundreds of millions of users issuing billions of request at the same time. We therefore need new technologies (big data processing/analysis) and new tools (big data systems) to deal with extremely large data sets and service requests. This is an introductory course on big data concepts, processing, analytics and systems. You will learn the latest development in big data technologies and get hands on experience in using popular open source big data systems such as Hadoop, Spark, HBase, MongoDB, Neo4j, Kafka, Flink, etc. The objectives of this course can be summarized as follows.

- . Understand big data concepts, challenges and trends.
- . Learn the technological foundations of big data science and engineering.
- . Learn the principles and practices behind popular open source big data systems.
- . Get hands on experiences of using open source big data systems for solving big data problems.

This is a lecture-oriented course. The system part of the course will be executed through in-class example discussion, homework assignments and term project. Due to the time limit, the lectures will focus mostly on the technological innovation of each system rather than how to use them. With brief introduction to the basic operations of various big data systems, students are expected to learn to use them on their own.

Regular Topics

The topics to be covered in the lecture are listed as follows (\*\*: topics to be covered depending on the time and progress):

- . Introduction
- . General purpose big data platforms
- . Big data storage architecture and systems
- . Big data systems for structured/semi-structured data
- . Big graph processing
- . RDF processing systems\*\*
- . Big stream processing
- . Big data pipelining tools\*\*
- . Big data ETL(extract, transform, and load) tools
- . Big data analytics, other systems and trends\*\*
- . Open data\*\*
- . Big data system landscape\*\*

Special Topic(s)

Based on current practices and emerging trends, we will select one or two special topics to provide a brief overview (if time permits, of course). This semester, the special topic we plan to talk about is big data and AI (Artificial Intelligence):

- . Relationship between big data and AIML(Artificial Intelligence and Machine Learning)
- . How ChatGPT and other similar Large Language Models(LLMs) work?
- . LLaMA, the open source LLM from  $\ensuremath{\mathsf{Meta}}$
- . The future of big data and generative AI

課程目標 Course Objectives

大數據處理在當前資訊爆炸的時代,已成為必備技能。特別是席捲全世界的人工智慧與機器學習浪潮,都必須仰賴對 大數據的解析、分類、辨識和歸納。本課程探討大數據特性,讓同學了解大數據處理尖端技術發展脈絡,深入探索各 種大數據系統和工具的運作原理和應用實例,培育同學們具備未來將大數據處理技術應用在任何領域所需的理論知識 和實務技能。

In the age of information explosion, big data processing is already a must-have skill. Especially on the new waves of AI and machine learning, all systems rely heavily on big data analysis, classification, recognition and induction. The purposes of this course are to study big data characteristics, to understand the evolution of big data processing technologies, and to explore the underlying principles and real-world applications of big data systems/tools. Students will learn the theoretical knowledge and practical skills necessary for applying big data technologies on any future application domains.

	系專業能力 Basic Learning Outcomes	課程目標與系專業能 力相關性 Correlation between Course Objectives and Dept.'s Education Objectives
A scien	資工知識技術之能力Ability to integrate knowledge and technologies of computer ce and information engineering.	•
<sup>B</sup> valid	支術理論驗證實驗之能力Ability to design and conduct science experiments and to ate hypotheses.	•
C 資訊車 hardw	欢硬體設計開發之能力Ability to design and develop computer software and are.	•
D 團隊專	專案開發之能力Ability to design and develop team projects.	0
	生思考與創新研發之能力。Ability of analytical thinking, creative research ing, and innovative development.	•
-	llustration :● 高度相關 Highly correlated ○中度相關 Moderately	correlated
	授課進度表 Teaching Schedule & Content	
週次Week	內容 Subject/Topics	備註Remarks
1	Course Introduction: . Course description . Objectives . Syllabus . Textbook and references . Assignments . Independent study . Exam . Term project	
2	<pre>Introduction    Data -&gt; Knowledge -&gt; Intelligence    What is a Big Data?    Why Big Data?    Examples of Big Data    The opportunities and challenges for Big Data</pre>	
3	<pre>General purpose big data platforms I     Distributed and cluster computing     Apache Hadoop     Cloudera(CDH, Cloudera Distribution for Hadoop)     MapReduce and algorithm design</pre>	
4	<ul> <li>General purpose big data platforms II</li> <li>Apache Spark and in-memory computation</li> <li>High Performance Computing Cluster (HPCC), also referred to as DAS(Data Analytics Supercomputer)</li> </ul>	
5	Big data storage architecture . Distributed nodes . Scale-out NAS . All-solid-satae drive (SSD) arrays . Object-based storage . DNA storage	
6	Big data storage systems . Distributed file systems and big data storage . Google GFS and Apache HDFS . Cloud Storage . Data lake** . Big data storage security**	

Big data systems for structured/semi-structured data I         . Apache Blase, Cassandra, CochBB, Drill. Impala, Rive         7       Systems SQL, Distributed SQL         . Apache Blase, Cassandra, CochBB, Drill. Impala, Rive         Big data systems for structured/semi-structured data II         . Suerk SQL, DataFrames, Datasets         8       . MoncolB         . Konto Bigberry, Spanner, Fl         . Presto         . Term project proposal and selected tool demo instead.         . The challenges of big graph processing         . The challenges of big graph processing         . The challenges of big graph processing         . The challenges of big data streaming         . Soark Graph, GraphFrames         . Nody graph database         . Titan distributed graph database         . Nong processing systems         . Andoop-hased RDF systems         . Spark Graph Branes         . Notig strand database         . Titan distributed graph database         . Big stream processing 1         . Spark Streaming, Structured Streaming         . Spark Areaming, Structured Streaming         . Spark Areaming, Structured Streaming         . Big data streaming applications         . Big data streaming applications         . Big data streaming applications			
7       Spark SUL, DataFrames, Datasets         000008       Google BigQuery, Spanner, Fl         Presto       Big data systems for structured/semi-structured data 11         . Spark SQL, DataFrames, Datasets       . Google BigQuery, Spanner, Fl         . Fresto       Big dramas, Datasets         8       . Google BigQuery, Spanner, Fl         . Fresto       Big graph processing         9       No midterm.         7       The challenges of big graph processing         10       . Frequence of the systems (SR), Pregel, Giraph)         . GrapLab family of systems(SR), Pregel, Giraph)         . GrapLab family of systems(SR), Pregel, Giraph)         . The challenges of big garsp tracessing         11       . Spark Graph, Graph/Tames         . Neodj graph database         11       . Spark Graph, Structures         . Neody graph database         11       . Big dram frictioned strustures         . Big stream processing 11         . Spark-fused RDF systems         . Spark Graph/MOMA         . Big data fficterrect, transform, and load) tools		. SQL, NoSQL, NewSQL, Distributed SQL	
WongoDB         Google BigGuery, Spanner, F1           Presto         Big data systems for structured/semi-structured data 11           . Spark SQL, DataFrames, Datasets         MongoDB           . Google BigGuery, Spanner, F1         . Presto           Big data systems for structured/semi-structured data 11         .           . WongoDB         . Google BigGuery, Spanner, F1           . Presto         . Presto           Big graph processing 1         . The challenges of hig graph processing           . The challenges of hig graph processing         . Freqol family of systems(SP: Preget, Graph)           . GraphLab family of systems(SP: Preget, Graph)         . GraphChi)           . Spark Stock GraphK, GraphKPranees         . Neodig graph processing 11           . Spark Stock GP systems         . Biddoop-based RDF systems           . NoSQL-based RDF systems         . Spark Stock RDF systems           . Spark Stock RDF systems         . Spark Stock RDF           . Spark Stock RDF systems         . Spark Stock RDF           . Big data streaming applications         . Big data streaming applications           . Big data streaming applications         . Big data streaming applications           . Big data streaming applications         . Big data streaming applications           . Apache Stock Attra         . Apache Kina           . Apache	7		
Google BigQuery, Spanner, F1           Presto           Big data systems for structured/semi-structured data II           Spark SQL, DataFrames, Datasets           ManguDB           Google BigQuery, Spanner, F1           Presto           Presto           Big graph processing 1           There challenges of big graph processing           The challenges of big graph processing           Spark SQL, CraphTames           Spark Squery, Spanner, F1           Presto           Big graph processing 1           Spark Scrapk, CraphTames           Spark Scrapk, CraphTames           Need J graph database           Titan distributed graph database           Spark Scrapk RDF systems           Iddoop-based RDF systems           Big stream processing 11           Spark Scrapk RDF Systems           Spark Scrapk RDF Systems           Spark Scrapk RDF Systems           Spark Streaming, Structured Streaming           Spark Streaming, Structured Streaming           Spark Streaming, Structured Streaming           Spark Streaming, Structured Streaming           Big data TL(cektract, transform, and load) tools           Apache StMA           Big data TL(cektract, transform, and load) tools <t< td=""><td></td><td></td></t<>			
Big data systems for structured/semi-structured data II       . Spark SQL, DataFrames, Datasets       8       8       8       9       9       9       9       10       11       10       11       11       12       13       14       14       15       16       17       18       18       19       10       10       10       11       10       11       12       13       14       14       15       16       17       18       19       10       11       12       13       14       15       16       17       18       19       19       10       11       11       12       13       14       16       15       16       17       18       19       19 <t< td=""><td></td><td>5</td><td></td></t<>		5	
8       - Spark SQL, DataFrames, Datasets         8       . Mongo08         9       Soft 4, 5428 Midtern Exam         9       Morepton processing 1         10       Term project proposal and selected tool demo instead,         10       Big graph processing 1         10       The challenges of big graph processing         11       . The challenges of big graph processing 1         10       . Spark Graph2, GraphData family of systems(GraphLab, PowerGraph, GraphChi)         11       . Spark Graph2, GraphDatabase         11       . Spark Graph2, GraphDatabase         11       . NoSQL-based RDF systems         11       . Spark-based RDF systems         12       . The challenges of big data streaming         . Spark-based RDF systems         Big stream processing 1         13       . Apache Strom, Samza, Plink         . Apache Strom, Samza, Plink         . Apache Strom, Samza, Plink         . Apache Airflow         . Big data streaming applications         Big data pipelining tools**         . Pig data         . Tez         Big data analytics, other systems and trends         . Gogel C cloup Platform (GCP) vs Amazon Web Services (AWS)         . R Langmage and RStudio			
8		с .	
. Google BigQuery, Spanner, F1         . Presto         9       Ref # \$4.32 Widterm Exam         9       Term project proposal and selected tool demo instead.         10       The challenges of big graph processing         10       The challenges of big graph processing         10       The challenges of big graph processing         11       The challenges of big graph processing         12       Spark GraphLab family of systems(BSP, Pregel, Giraph).         . GraphLab family of systems(BSP, Pregel, Giraph)       Spark GraphChi)         . Big graph processing I       Spark GraphChi)         . Neodj graph database       . Neodj graph database         . Titan distributed graph database       . Titan distributed graph database         . Bidoop-based RDF systems       . Badoop-based RDF systems         . Badoop-based RDF systems       . Spark-based RDF systems         . Spark-based RDF systems       . Spark Streaming, Structured Streaming         . Spark Streaming applications       Big data streaming applications         Big data streaming applications       Big data pipelining tools**         . Pig latin       . Toz         . Apache Kirka       .         Big data pipelining tools**       .         . Pig latin       .         . RapidWiner	8		
9         No midterm.           9         No midterm.           10         Term project proposal and selected tool demo instead.           10         The challenges of big graph processing           10         The challenges of big graph processing           11         The challenges of big graph processing           12         GraphLab family of systems(GraphLab, PowerGraph, GraphChi)           11         Spark Graph database           11         . Nexdj graph database           11         . Nexdj graph database           11         . NoSQL-based RDF systems           11         Big stream processing I           12         . Hechallenges of big data streaming           13         . Apache Storm, Smura, Flink           14         Apache Sturm, Smura, Flink           13         . Apache KIMA           14         Big data streaming applications           15         . Apache KIMA           16         . Apache KIMA           17         . Apache KIMA           18         . Apache KIMA           19         . Apache KIMA           10         . Apache KIMA           11         . Apache KIMA           12         . The challenges of big data streaming		-	
9       No midterm. Term project proposal and selected tool demo instead.         10       The challenges of big graph processing Pregel family of systems(BP, Pregel, Giraph) . GraphLab family of systems(GraphLab, PowerGraph, GraphChi)         10       Spark GraphX, GraphPrames . Neodj graph database . Neodj graph database . NoSQL-based RDF systems         11       RDF processing systems** . NoSQL-based RDF systems . Back Stream processing I . The challenges of big data streaming . Spark-based RDF systems . Back Streaming, Structured Streaming . Spark-based RDF systems . Big stream processing I . The challenges of big data streaming . Spark-based RDF systems         12       Big stream processing I . The challenges of big data streaming . Spark-based RDF systems         13       Apache SAMOA . Big data streaming applications         14       Big stream processing I . Apache SAMOA . Big data streaming applications         14       Big data effl(extract, transform, and load) tools . Apache Kafka         14       Big data analytics, other systems and trends . Google Cloud Platform (GCP) vs Amazon Web Services (AWS) . RapidWiner . Tableau . Tableau . R Language and RStudio . Open data         15       . KIMME . Tableau . R Language and RStudio . Open data         16       . Sperial topic: Big data and Al . Now ChatGPT and other similar Large Language Models(LLMs) work? . LLaMA, the open source LLM from Meta . The future of big data and Al . Now ChatGPT and other similar Large Language Models(LLMs) work? . LLaMA, the open source LLM from Meta . The future of big data and generative AI			
Term project proposal and selected tool demo instead.         Big graph processing I         The challenges of big graph processing         GraphLab family of systems(BxP, Pregel, Giraph)         GraphLab family of systems(GraphLab, PowerGraph, GraphChi)         Big graph processing II         Spark GraphA, GraphFrames         Neodj graph database         Titan distributed graph database         III         Big stream processing I         Isadop-based RDF systems         Big stream processing I         The challenges of big data streaming         Spark Streaming, Structured Streaming         Big data efficient processing II         Apache Storm, Samza, Flink         Apache Storm, Samza, Flink         Apache Storm, Samza, Flink         Apache Kafka         Big data streaming applications         Big data streaming applications         Big data analytics, other systems and trends         Goagle Cloud Platform (GCP) vs Amazon Web Services (AWS)         Apache Kafka         Big data analytics, other systems and trends         Coggle Cloud Platform (GCP) vs Amazon Web Services (AWS)         Apache Kafka         Big data analytics, other systems and trends         Coggle Cloud Platform (GCP) vs Amazon Web Services (AWS)	0		
Big graph processing 1         10       . Tree challenges of big graph processing         . Pregel family of systems(BSP, Pregel, Giraph)         . GraphLab family of systems(GraphLab, PowerGraph, GraphChi)         Big graph processing 11         . Spark GraphX, GraphFrames         . Needj graph database         11         RDF processing 11         . Needj graph database         11         RDF processing 11         . Needj graph database         11         RDF processing 11         . Needel RDF systems         . Biddoop-based RDF systems         . Spark-Lased RDF systems         . Spark-Lased RDF systems         . Spark-Lased RDF systems         . Spark Streaming. Structured Streaming         . Spark Streaming applications         Big stream processing 11         . Apache SAMOA         . Apache Afrika         Big data streaming applications         Big data extreaming applications         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . Tableau	9		
10       . Pregel family of systems(GraphLab, PowerGraph, GraphChi)         Big graph processing II       . Spark GraphX, GraphFrames         . Neodj graph database       . Titan distributed graph database         11       RDF processing systems**         . Noodj graph database       . Spark Active database         11       RDF processing systems**         . Noodj graph database       . Spark-based RDF systems         . Big stream processing I       . The challenges of big data streaming         . Spark Streaming. Structured Streaming       . Spark Streaming. Structured Streaming         13       . Apache Storm. Samza. Flink         . Apache Storm. Samza. Flink       . Apache Storm. Samza. Flink         . Apache Kafka       . Pig latin         14       Big data streaming applications         Big data pipelining tools**       . Pig latin         . Tecz       . Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)       . RapidWiner         . NNME       . RapidWiner         . The relationship between big data and Al       . The relationship between big data and Al         . Now Obig data and Al work together?       . Dowerful synergy of big data and Al         . Now ChatGPT and other similar Large Language Models(LLMs) work?       . LadAl, the open source LLM from			
. Pregel Tamily of systems(USA), Pregel, Giraph)         . GraphLa family of systems(GraphLab, PowerGraph, GraphChi)         Big graph processing II         . Spark GraphX, GraphTrames         . Neodj graph database         . Titan distributed graph database         . Spark-based RDF systems         . Badoop-based RDF systems         . Spark-based RDF systems         . Spark-based RDF systems         . Spark Stream processing I         . The challenges of big data streaming         . Spark Stream processing II         . Apache Storm, Samza, Flink         . Apache Storm, Samza, Flink         . Apache Storm, Samza, Flink         . Apache Airflow         . Apache Kafka         Big data nipelining tools**         . Pig Latin         . Tecz         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . Ro vchatGPt regry of big data and AI         .	10		
Big graph processing II         . Neoti graph database         . Titan distributed graph database         . Titan distributed graph database         . Titan distributed graph database         . RDF processing systems**         . NoSQL-based RDF systems         . Badoop-based RDF systems         . Badoop-based RDF systems         . Spark-based RDF systems         . Spark-based RDF systems         . Spark-based RDF systems         . Spark Streaming. Structured Streaming         . Spark Streaming. Structured Streaming         . Apache Storm, Samza, Flink         . Apache Kafka         14         Big data streaming applications         Big data inpelining tools**         . Pig Latin         . Tec         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidWiner         . RapidWiner         . Robic data and Alton         . Recial topic: Big data and Al         . Recial topic: Big data and Al         . Recial topic: Big data and Al         . Bew do big data and Al work together?			
Spark GraphX, GraphPrames         Neo4j graph database         Titan distributed graph database         RDF processing systems**         NoSQL-based RDF systems         Big stream processing 1         The challenges of big data streaming         Spark Streaming, Structured Streaming         Big stream processing 1         Apache Storm, Samza, Flink         Apache Storm, Samza, Flink         Apache Storm, Samza, Flink         Apache Kafka         Big data streaming applications         Big data pipelining tools**         Pig Latin         Tez         Big data analytics, other systems and trends         Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         RapidMiner         Special topic: Big data and Al         Robig data and Al work together?         Powerful synergy of big data and Al         How ChatGPT and other similar Large Language Models(LLMs) work?         LLaMA, the open source LLM from Meta         The future of big data and generative Al			
11       . Titan distributed graph database         11       RDF processing systems**         NoSQL-based RDF systems       . Bidadop-based RDF systems         . Spark-based RDF systems       . Spark-based RDF systems         12       The challenges of big data streaming         . Spark-based RDF systems       . Spark Streaming, Structured Streaming         13       . Spark Streaming, Structured Streaming         13       . Apache Storm, Samza, Flink         . Apache SMOA       . Big data streaming applications         Big data streaming applications       . Apache Kafka         14       Big data pipelining tools**         . Pig Latin       . Tez         15       . KNIME         . Tez       . RapidMiner         15       . KNIME         . Tez       . RapidMiner         . RapidMiner       . Tableau         . R Language and RStudio       . Open data         Special topic: Big data and AI       . The relationship between big data and AI         . How Obig data and AI work together?       . Powerful synergy of big data and AI         . How ChaftPT and other similar Large Language Models(LLMs) work?       . LLaMA, the open source LLM from Meta         . The future of big data and generative AI       . The future of big data and generative AI		. Spark GraphX, GraphFrames	
11       RDF processing systems**         NoSQL-based RDF systems			
RDF processing systems**         NoSQL-based RDF systems         Hadcop-based RDF systems         Spark-based RDF systems         Big stream processing I         12         The challenges of big data streaming         Spark Streaming, Structured Streaming         Big stream processing II         . Apache Storm, Samza, Flink         . Apache SMOA         . Big data streaming applications         Big data ETL(extract, transform, and load) tools         . Apache Kafka         14         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . K Language and RStudio         . Open data         Special topic: Big data and AI         . How Ob big data and AI         . How do big data and AI         . How do big data and AI         . How chaftPT and other sini lar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17       Student presentation	11	. IItan distributed graph database	
. Hadoop-based RDF systems         . Spark-based RDF systems         Big stream processing I         . The challenges of big data streaming         . Spark Streaming, Structured Streaming         Big stream processing II         . Apache Storm, Samza, Flink         . Apache Katka         Big data streaming applications         Big data pipelining tools**         . Apache Katka         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . K Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI         . How ChatOPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI			
. Spark-based RDF systems         Big stream processing I         . The challenges of big data streaming         . Spark Streaming, Structured Streaming         Big stream processing II         . Apache Storm, Samza, Flink         . Apache SM0A         Big data streaming applications         Big data ETL(extract, transform, and load) tools         . Apache Kafka         I4         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI			
Big stream processing I         12       The challenges of big data streaming         13       Spark Streaming, Structured Streaming         13       Big stream processing II         13       Apache Storm, Samza, Flink         13       Apache Storm, Samza, Flink         14       Big data streaming applications         Big data ETL(extract, transform, and load) tools         . Apache Kafka         14         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         15         . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         16         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaWA, the open source LLM from Meta         . The future of big data and generative AI         17       Student presentation			
. Spark Streaming, Structured Streaming         Big stream processing II         . Apache Storm, Samza, Flink         . Apache SMOA         . Big data streaming applications         Big data ETL(extract, transform, and load) tools         . Apache Kafka         14         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How Chat6PT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI			-
Big stream processing II         13         13         13         13         13         13         13         14         14         14         14         14         14         14         14         14         14         14         14         14         14         15         16         17         18         19         19         11         12         13         14         14         14         15         16         17         18         19         10         11         11         12         13         13         14         15         15         16         17         18         19         10         11	12		
13       . Apache Storm, Samza, Flink         . Apache SAMOA       . Big data streaming applications         Big data streaming applications       Big data streaming applications         14       Big data pipelining tools**         14       Big data pipelining tools**         14       Big data analytics, other systems and trends         15       Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How Ob big data and AI work together?         16         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI			
13       . Apache SAMOA         . Big data streaming applications         Big data full(extract, transform, and load) tools         . Apache Airflow         . Apache Airflow         . Apache Kafka         14         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How ChatCPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI	10		
Big data ETL(extract, transform, and load) tools         Apache Airflow         Apache Kafka         14         Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         15       . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         16       . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17	13	. Apache SAMOA	
<ul> <li>Apache Airflow <ul> <li>Apache Kafka</li> </ul> </li> <li>14</li> <li>Big data pipelining tools** <ul> <li>Pig Latin</li> <li>Tez</li> </ul> </li> <li>Big data analytics, other systems and trends <ul> <li>Google Cloud Platform (GCP) vs Amazon Web Services (AWS)</li> <li>RapidMiner</li> <li>KNIME <ul> <li>Tableau</li> <li>R Language and RStudio</li> <li>Open data</li> </ul> </li> <li>Special topic: Big data and AI <ul> <li>How do big data and AI work together?</li> <li>Powerful synergy of big data and AI</li> <li>How ChatGPT and other similar Large Language Models(LLMs) work?</li> <li>LLaMA, the open source LLM from Meta</li> <li>The future of big data and generative AI</li> </ul> </li> </ul></li></ul>			_
14       Apache Kafka         14       Big data pipelining tools**         . Pig Latin       . Tez         Big data analytics, other systems and trends       . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner       .         15       . KNIME         . Tableau       . R Language and RStudio         . Open data       .         16       Special topic: Big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17       Student presentation			
Big data pipelining tools**         . Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI		•	
. Pig Latin         . Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         15       . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         16       . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI	14		
. Tez         Big data analytics, other systems and trends         . Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         15       . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI			
. Google Cloud Platform (GCP) vs Amazon Web Services (AWS)         . RapidMiner         15       . KNIME         . Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17			
15       . RapidMiner         15       . KNIME         . Tableau       . R Language and RStudio         . Open data       .         Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         16       . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17			
15       . KNIME         . Tableau       . R Language and RStudio         . Open data       . Open data         Special topic: Big data and AI       . The relationship between big data and AI         . How do big data and AI work together?       . How chatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta       . The future of big data and generative AI         17       Student presentation		-	
. Tableau         . R Language and RStudio         . Open data         Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         16       . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17	15	1	
. Open data         Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17			
Special topic: Big data and AI         . The relationship between big data and AI         . How do big data and AI work together?         16       . Powerful synergy of big data and AI         . How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17			
<ul> <li>The relationship between big data and AI</li> <li>How do big data and AI work together?</li> <li>Powerful synergy of big data and AI</li> <li>How ChatGPT and other similar Large Language Models(LLMs) work?</li> <li>LLaMA, the open source LLM from Meta</li> <li>The future of big data and generative AI</li> </ul>			
<ul> <li>How do big data and AI work together?</li> <li>Powerful synergy of big data and AI</li> <li>How ChatGPT and other similar Large Language Models(LLMs) work?</li> <li>LLaMA, the open source LLM from Meta</li> <li>The future of big data and generative AI</li> </ul>	16		
. How ChatGPT and other similar Large Language Models(LLMs) work?         . LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17         Student presentation		. How do big data and AI work together?	
. LLaMA, the open source LLM from Meta         . The future of big data and generative AI         17         Student presentation			
. The future of big data and generative AI       17       Student presentation			
18 期末考試週 Final Exam	17	Student presentation	
	18	期末考試週 Final Exam	

教 學 策 略 Teaching Strategies						
✓ 課堂講授 Lecture 分組討論Group Discussion 参觀實習 Field Trip						
✓ 其他Miscellaneous: <u>Students will get hands-on experience with popular open source big data</u>						
教 學 創 新 自 評 Teaching Self-Evaluation						
創新教學(Innovative Teaching)						
□問題導向學習(PBL)      團體合作學習(TBL)						
─ 翻轉教室 Flipped Classroom ── 磨課師 Moocs						
社會責任(Social Responsibility)						
□ 在地實踐Community Practice □ 產學合作 Industy-Academia Cooperation						
跨域合作(Transdisciplinary Projects)						
□跨界教學Transdisciplinary Teaching □跨院系教學Inter-collegiate Teaching						
業師合授 Courses Co-taught with Industry Practitioners						
其它 other: Play with open-source big data tools on VM.						

學期成績計算及多元評量方式 Grading & Assessments									
配分項目 配分比例 多元評量方式 Assessments									
Items	Percentage	測驗 會考	實作 觀察	口頭 發表	專題 研究	創作 展演	卷宗 評量	證照 檢定	其他
平時成績 General Performance	15%			~	~				Independent study and presentation
期中考成績 Midterm Exam	0%								
期末考成績 Final Exam									
作業成績 Homework and/or Assignments	35%		~						
其他 Miscellaneous (Term Project)	25%		~		~	~			
	Creation 0 4			補充說明		muctio			
	Grading & A	ssessme	ents Sup	plement	al inst	ruction	S		
<ul> <li>Independent Study and Presentation <ul> <li>All students are to conduct independent study on self-selected topics (discussed with me first).</li> <li>Pick an open source big data system not discussed in the class as the study target.</li> <li>Prepare a presentation and a demonstration of the system in class.</li> <li>Every student must present and demo.</li> </ul> </li> </ul>									
<ul> <li>Term Project <ul> <li>There will be a modest scale term project for you to show your creativity.</li> <li>May use any big data tools for your project.</li> <li>Prepare a project proposal and a tool(s) demo on VMs by the end of the midterm week.</li> <li>Prepare a project demo at the end of the semester to explain your project to me.</li> <li>Turn in the project and report one week after the final exam.</li> </ul> </li> </ul>									
教科書與參考書目(書名、作者、書局、代理商、說明) Textbook & Other References(Title, Author, Publisher, Agents, Remarks, etc.)									
No required textbook.									
References: . Jawwad Ahmad Shamsi and Muhammad Khojaye. Big Data Systems: A 360-degree Approach. Chapman & Hall/CRC, 2021. . Balamurugan Balusamy, Nandhini Abirami R, Seifedine Kadry, Amir H. Gandomi. Big Data: Concepts,									
<ul> <li>Technology, and Architecture. Wiley, 2021.</li> <li>Sherif Sakr. Big Data 2.0 Processing Systems: A Systems Overview, 2nd Edition, Springer, 2020.</li> <li>S. Sasikala and D. Renuka Devi (Authors), Raghvendra Kumar (Editor). Research Practitioner's Handbook on Big Data Analytics. Apple Academic Press, 2023.</li> </ul>									
. Guido Dartmann, Houbing Song, et al. Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things. Elsevier Science Publishing Co Inc, 2019. . Kai Hwang and Min Chen. Big Data Analytics for Cloud, IoT and Cognitive Computing. John Wiley &									
Sons Ltd., 2017. . Tom White. Hadoop: The Definitive Guide, 4th Edition, O'reilly, 2015. . Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Cambridge University Press, 2010~2014.									
. Mohammed J. Zaki and Wagner Meira JR. Data Mining and Machine Learning - Fundamental Concepts and Algorithms, 2nd Edition. Cambridge University Press, 2020.									
		s Websi nal webs	te(Incl site car	uding o n be lis	nline t sted her	eaching e.)	inform		
Instructor' s homepage: ht									
All lecture notes will be available online.									

其他補充說明(Supplemental instructions)