Teaching Design of Distributed Systems Based on Process Objectives in the Context of New Engineering

Dongxin Wen[a]*, Tianyi Zang[b], Jiu Lou[c]

[a]wdongxin@hit.edu.cn*, [b]zty@hit.edu.cn, [c]lj@hit.edu.cn

School of Computer Science, Harbin Institute of Technology, Harbin, China

Abstract. The establishment of new engineering disciplines is a significant policy initiative in higher engineering education. This paper summarizes the pedagogical approach for the 'Distributed Systems' course within this new engineering context. The course integrates the core tenets of new engineering disciplines by establishing process objectives to achieve the course objectives. The teaching design encompasses process objectives aligned with the new engineering content, instructional content, experimental teaching, and assessment methods. This process-oriented approach transitions into interdisciplinary teaching content, with experimental teaching guiding students to proactively explore cutting-edge topics. This fosters innovative thinking and teamwork among students, providing new methods and perspectives for cultivating talent in new engineering disciplines.

Keywords: New Engineering; Process Objectives; Distributed Systems; Teaching Design

1 Introduction

The construction of new engineering disciplines is a significant strategic decision and deployment in higher engineering education, tailored to adapt to new economic and developmental trends. It aims to cultivate students' innovation and entrepreneurial concepts, building an education system that integrates innovation and entrepreneurship. This approach melds theory with practice, creating a practical teaching environment that embeds innovation and entrepreneurship education throughout the engineering training process for university students. The essence of new engineering disciplines encompasses moral education as the guiding principle, adapting to changes and shaping the future as the foundational concept, and leveraging inheritance and innovation, interdisciplinary integration, and coordination and sharing as the primary pathways to nurture diverse and innovative engineering talents for the future [1].

In line with the new engineering disciplines initiative, Harbin Institute of Technology has introduced the "HIT New Engineering π-Type Plan"[2]. The computer science curriculum is guided by demand and integrates teaching philosophies of new engineering disciplines with industry and research application [3]. It aims to foster a sense of national pride and service to society among students. Through interdisciplinary cooperation between schools and enterprises, the program cultivates students' abilities in knowledge application, lifelong learning, engineering practice, and innovation [4]. The course group has developed an innovative talent training system for new engineering disciplines, focused on learning

EIMT 2024, March 29-31, Wuhan, People's Republic of China

DOI 10.4108/eai.29-3-2024.2347678

outcomes and has established diverse, integrated evaluation methods. This system places equal emphasis on cultivating engineering skills and the ongoing learning process, with a focus on developing students' thinking, communication, management skills, and character, thus providing a more equitable and comprehensive assessment of student abilities [5].

2 Introduction to the ‘Distributed Systems’ Course

As chip integration continues to improve, the power consumption and complexity of chips are increasing. The method of enhancing performance merely by increasing the clock frequency and complexity of the architecture is becoming insufficient [6]. A distributed system is a loosely coupled system constituted by multiple processors interconnected by communication lines. From the perspective of any single processor in the system, all other processors and their resources are remote, while only its resources are local. This system facilitates long-distance communication between nodes, greatly simplifying information exchange among people, allowing users from different regions to collaborate on a project. Distributed systems are superior for many types of applications compared to other architectures such as multiprocessors with shared memory. Given their growing importance, distributed systems have been integrated into the undergraduate curriculum for computer science majors. The "Distributed Systems" course at Harbin Institute of Technology is offered as an elective for computer science students. The course aims to introduce the classification, characteristics, distributed file systems, distributed transaction processing, basic concepts and foundational theories of distributed consensus, typical distributed consensus algorithms, and typical distributed systems, enabling students to understand the complexity and uncertainty inherent in distributed systems. It cultivates the ability to select and apply optimized distributed algorithms to solve problems in various distributed scenarios, thereby enhancing students' proficiency in distributed systems.

3 Process-Oriented Teaching Design

The teachers play a key role with their competences in using new technologies constituting an essential prerequisite for the effective implementation of technology-related skills. [7] This article introduces a process-oriented teaching design, as shown in (Figure 1). It focuses on the collaborative and complementary aspects of process objectives, course objectives, and social objectives. Highlighting the concept of 'industry-education integration', it aims to develop engineering talents that closely match industry requirements. This approach satisfies the urgent need for professional technical talent in businesses, while utilizing the research strengths of academic faculty to encourage student involvement in innovative frontiers of technology. The objective is to nurture exceptional talents who will lead in new industries and future technological developments.

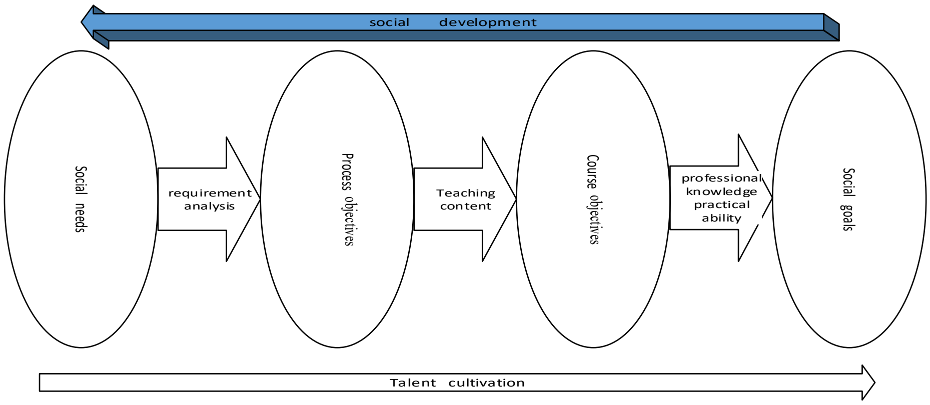

Figure 1: Process-Oriented Teaching Design

The design of the process objectives is as follows:

3.1 Process Objectives

The process objectives are structured to provide a comprehensive understanding of distributed systems.

Objective 1: Acquire knowledge of the classifications, characteristics, applications, and interconnectivity of distributed systems; comprehend the structure of naming services and domain name systems; master synchronization algorithms for both physical and logical clocks; attain proficiency in election algorithms and mutual exclusion algorithms; gain command over distributed file systems.

Objective 2: Grasp the inherent uncertainties of distributed systems and their models; understand the basic concepts of distributed consensus, communication protocols for consensus, and foundational theories.

Objective 3: Comprehend and apply modern distributed consensus algorithms based on voting and proofs, discern their characteristics and appropriate contexts, and develop the capability to select, analyze, optimize, and design effective consensus algorithms for real-world applications, enhancing problem-solving skills in consensus issues.

Objective 4: Understand the Apache Hadoop software library as a framework that facilitates distributed processing of large data sets across computer clusters using the distributed programming model MapReduce.

Objectives 1 and 2 focus on the foundational knowledge of distributed systems and consensus, whereas Objectives 3 and 4 are aimed at cultivating engineering qualities, fostering an engineering mindset and a comprehensive view of distributed systems to solve complex engineering problems in various distributed scenarios.

3.2 Instructional Content

In-depth analysis and rational categorization of the process objectives have been carried out to formulate the core strategy for the instructional design system. Based on this strategy, a teaching design system has been developed, which offers specific instructional content design solutions by meticulously analyzing the characteristics of different types of process objectives. This system provides new insights for effective teaching in emerging interdisciplinary fields [8]. The specific instructional content design is outlined as follows:

1. Relevant Discipline Courses: Review of key knowledge points in disciplines such as computer networks, data structures and algorithms, databases, and operating systems, with integration and comparison within the context of distributed systems.

2. Overview of Distributed Systems: Topics include name services and synchronization in distributed systems, aligning with Process objective 1, which focuses on understanding the application, classification, features, and interconnection of distributed systems, as well as mastering topics like clock synchronization, logical clocks, election algorithms, and mutual exclusion algorithms.

3. Distributed File Systems: Covers an overview of distributed file systems and topics related to network file systems (NFS), aligning with Process objective 1 and addressing concepts related to abstract models and design issues of distributed file systems, as well as understanding network file system structure models, remote procedure calls, file sharing, client caching, and server replication.

4. Distributed Transaction Processing: Provides an introduction to distributed databases, transaction types, concurrency control in distributed transactions, and distributed deadlock detection, in alignment with Process objective 2, which entails understanding the composition, classification, and views of distributed databases, mastering basic transaction models, classification, transaction serialization, transaction abort recovery, and various transaction concurrency control methods such as locking, timestamp-based, and optimistic concurrency control, as well as understanding deadlock detection messages and periodic detection methods, and comprehending deadlock prevention methods.

5. Fundamentals of Distributed Consensus: Covers the characteristics of distributed systems, Byzantine generals problem, distributed system models, CAP theorem, remote procedure calls, broadcast protocols, replication, and idempotence, and FLP theorem, in alignment with Process objective 3, which involves understanding and mastering basic concepts in distributed consensus, including distributed system characteristics, the Byzantine generals problem, distributed system models, and distributed consensus communication protocols. The curriculum also addresses practical application scenarios by establishing distributed system models for problem-solving.

6. Voting-Based Consensus Algorithms: Focuses on typical voting-based consensus algorithms such as Paxos, PBTF, and Raft, in alignment with Process objective 3, which emphasizes mastering algorithms like Paxos, PBTF, and Raft, understanding their characteristics and applicable scenarios, and grasping the key implementation points and challenges. The curriculum enables students to choose and optimize relevant consensus algorithms for practical applications.

7. Proof-Based Consensus Algorithms: Covers typical proof-based consensus algorithms including PoW, PoS, Algorand, and extends to four notable distributed consensus algorithms: Zab, HotStuff, Thunderella, and HoneyBadger, aligning with Process objective 3), which aims to understand and master typical proof-based consensus algorithms like PoW, PoS, and

Algorand, along with their characteristics and applicable scenarios. The curriculum also delves into the core ideas of Algorand, focusing on key implementation points and challenges. Additionally, it provides insights into four prominent distributed consensus algorithms: Zab, HotStuff, Thunderella, and HoneyBadger, by covering fundamental concepts, distributed consensus processes, understanding key application scenarios, and addressing critical problem-solving aspects through comparative analysis. This equips students with the ability to optimize consensus algorithms based on real-world application scenarios.

3.3 Experimental Teaching

In accordance with the requirements of cultivating new engineering talents, reforms have been introduced in the instructional design of experimental teaching in distributed systems. A series of progressively comprehensive experiments have been devised, starting from basic concepts and gradually advancing to more sophisticated functionalities. These experiments are driven by the background of big data processing and are aimed at providing students with hands-on experience throughout the entire process of designing distributed systems. This approach allows students to gain a deep appreciation for the complexity and significance of distributed systems, while also enhancing their practical skills and system design capabilities [9].

1) Experimental Objectives

(1) Emphasizing the Main Thread: Combining real-world application scenarios with abstract problem modeling, cultivating and enhancing students' ability to abstract complex problems in distributed settings, highlighting problem-solving approaches in distributed systems, and fostering a deep understanding and mastery of methods for solving problems in distributed systems.

(2) Problem-Driven Approach and Problem-Solving Skills Development: Addressing specific instances of distributed problems, gradually and systematically teaching how to analyze, optimize, and design effective algorithms, and through concrete examples and practical problems, explaining the process of solving problems in distributed systems. This approach aims to cultivate and enhance students' problem-solving capabilities.

2) Experimental Content

(1) Hadoop Installation and Configuration: Setting up a pseudo-distributed Hadoop cluster environment on a computer, checking the cluster's startup status using commands and web access after environment setup is complete.

(2) MapReduce Distributed Computing and Its Applications.

(3) Distributed Data Warehouse Hive and Its Applications.

Corresponds to Process Objective 4).

3) Experimental Methods

(1) Practice-Driven Approach: Cultivating the ability to understand the complexity and uncertainty of distributed systems and to apply learned knowledge to solve typical problems in distributed systems. On one hand, practical experience deepens the understanding and mastery of fundamental concepts, principles, techniques, and methods covered in the course. On the other hand, through experiments, students apply their knowledge to solve real-world problems,

achieving an organic integration of theory and practice.

(2) Group-Based Experiments: Determining group sizes based on the specific difficulty levels of experimental tasks. Fostering students' innovative thinking and teamwork awareness.

(3) Project Discussions and Presentations: Assessment of student experiment completion by teachers, with excellent experiments being presented, discussed, and shared within the class to inspire fellow students.

(4) Report Writing: In addition to submitting source code, experimental data, and results, students are required to write and submit experimental reports.

3.4 Course Assessment Methods

Students' knowledge and skills were assessed before and after integrated learning activities. [10] The new engineering education paradigm requires the establishment of a comprehensive and diverse evaluation system that assesses students' performance from various aspects, serving as a value judgment [11]. The assessment elements used in teaching include:

Regular Assignments: Regular assignments focus on key course content or extended topics, emphasizing the mastery of core content, fostering problem identification, problem-solving, self-learning, and problem-solving abilities. The evaluation criteria consider whether the logical thinking process in problem-solving is clear and efficient, and they actively encourage innovation. Corresponds to Process Objectives 1), 2), and 3).

Course Experiments: Assess students' understanding of knowledge and their ability to apply it comprehensively, as well as their problem-solving skills. Course experiments should be completed outside of class and undergo in-class verification checks. Evaluation criteria include the quality of code, the completeness of functionality, the comprehensiveness of reports, and the clarity of charts, among others. Corresponds to Process Objective 4).

Final Examination: Covers fundamental and foundational theories of distributed systems, mastery of typical systems and core algorithms, and the application of basic algorithms. Corresponds to Process Objectives 1) to 3).

4 Course Development and Implementation

In the design of the distributed systems course, the guiding principles of moral education and character cultivation, as well as the construction concepts of adaptability to change and shaping the future, have been fully incorporated. The teaching philosophy focuses on inheritance and innovation, cross-disciplinarity, and harmony and sharing. Remarkable results have been achieved through teaching practice in this semester, leading to the rapid transformation of diverse and innovative engineering talents for the future [12].

1) Compilation of Teaching Resources: Teaching resources include electronic textbooks, course introductions, electronic courseware, and teaching designs. For each lecture, the problems to be addressed and related content are listed, and a certain number of valuable reference materials are provided to achieve the teaching objectives that combine basic standards with optimal standards.

2) Establishment of QQ Group for Teaching Classes: A QQ group is established to share teaching content, such as PPT slides, explanatory audio, and electronic teaching materials. It facilitates online tutoring, interactive communication, and Q&A sessions.

3) Case-Based Teaching Method: Case-based teaching is the primary teaching method of the course. It involves real or virtual scenarios based on typical cases, allowing students to immerse themselves in case scenarios and engage in interactive teaching processes such as teacher-student Q&A and discussions [13]. The content includes the introduction of the background of actual systems and algorithms, and learning about the pursuit of truth and the rigorous scientific spirit from Turing Award winners and other scientists.

4) Flexible Teaching Modes: In addition to traditional offline teaching, two additional teaching methods are introduced: live broadcasting and recording. QQ group classrooms and Tencent Meetings are used as teaching software for live broadcasting. QQ group classrooms are optimized for remote teaching and can also use Tencent Meetings for remote audio and video conferences, online document collaboration, screen sharing, and more, to complete various aspects of the class. To avoid network failures and congestion, SuperStar Learning software can be used to record classroom teaching activities.

5) Hands-On Practices: The following practical exercises are designed:

Content 1:

Understanding the MapReduce execution process through word frequency statistics:

1) Install the same version of JDK on Windows as in the virtual machine, use Eclipse or IntelliJ IDEA as the MapReduce programming tool, and configure the MapReduce integrated environment. Add all the jar files under the /share/hadoop directory in the Hadoop installation package. It is recommended to use the highly compatible Hadoop 3.3.1 version.

2) Run the WordCount program example source code provided by Hadoop officially separately on the local machine and the Hadoop platform to verify the experimental steps.

Content 2: Movie Website User Review Analysis:

Using the MapReduce distributed computing architecture to solve application problems. Analyze user reviews on a movie website: The movie website provides two sets of data, namely user ratings data (ratings.dat) and movie information data (movies.dat). Using MapReduce, perform statistical analysis on the two sets of movie review data based on the number of evaluations for each movie to analyze user viewing preferences.

Content 3: Hive Data Warehouse:

Hive is a distributed data warehouse based on HDFS and MapReduce, with massive data storage and processing capabilities. It is a common tool for offline batch data processing in the big data field. It parses SQL statements into MapReduce jobs and relies entirely on HDFS and MapReduce. Using Hive, calculate the dropout rate for each base station, sort it in descending order, and find the top 10 base stations with the highest dropout rates.

6) Assessment and Evaluation: Course grades = Regular Assignments (10%) + Course Experiments (30%) + Final Examination (60%)

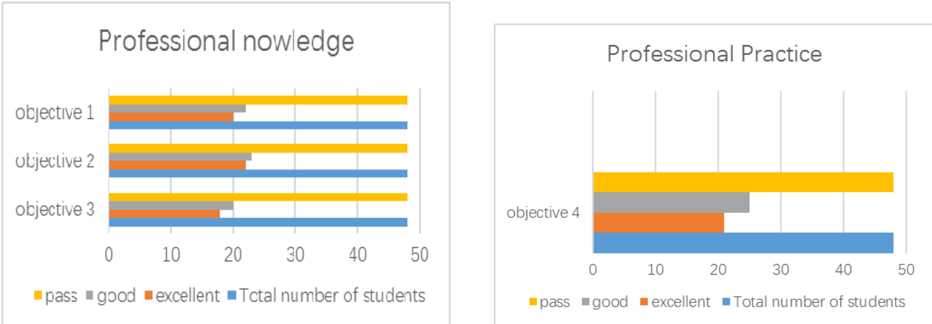

Figure 2: Completion Status of Process Objectives

Figure 2 shows the grade analysis of the distributed systems course in the spring semester of 2023. The total number of students is 48, with 18 outstanding students in objective 1 in the course assessment, and an excellent rate of 37.5%; 20 students are in good condition, with a good rate of 41.7%; The pass rate is 100%. There are 22 outstanding students in objective 2, with an excellent rate of 45.8%; 23 students are in good condition, with a good rate of 47.9%; The pass rate is 100%. There are 20 outstanding students in objective 3, with an excellent rate of 41.7%; 22 students are in good condition, with a good rate of 45.8%; The pass rate is 100%. In the course practice assessment, there are 21 outstanding students in objective 4, with an excellent rate of 43.8%; 25 people are in good condition, with a good rate of 52.1%; The pass rate is 100%.

5 Conclusion

This paper has presented the teaching design of the distributed systems course in the context of the new engineering education. Firstly, it established process objectives corresponding to the connotations of the new engineering education. Secondly, it developed interdisciplinary teaching content guided by process objectives. Next, it employed experimental teaching to guide students to actively explore cutting-edge topics, fostering innovative thinking and teamwork awareness. This research has explored and contributed to the cultivation of talents in the new engineering education. Outstanding results have been achieved in student teaching evaluations for the semester, with a satisfaction rate of 100%.

References

[1] Zhong Denghua, Connotation and Action of New Engineering Construction, Higher Engineering Education Research, 2017, Issue 3

[2] Xu Xiaofei et al. New Perspectives of New Engineering: Agile Teaching System for Sustainable Competitiveness, Chinese University Teaching, 2018 (10)

[3] Li Xuewei, Wang Wenjun, Xi Haixia, et al. Research on the Multiple Integration Evaluation Method of New Engineering Project based Teaching - Taking the Course of "Smart Society and Big Data Intelligence" as an Example [J]. Higher Engineering Education Research, 2023 (06): 27-33

[4] Xia Na, Wan Jiashan, Du Huazheng, et al. Exploration of a Progressive New Engineering Innovation Talent Training System [J]. Journal of Higher Education, 2023,9 (35): 35-38
[5] Liu Yue, Research and Practice on Cultivating the Quality and Abilities of Engineering College Students for Multivariate Evaluation, Higher Education Journal 2024, 2:151-155
[6] Liu Xiping, Current Situation and Reflection on the Construction of Parallel and Distributed Computing Courses, Computer Education, 2020, 2:109-113
[7] Kerstin Drossel, Birgit Eickelmann. Teachers' participation in professional development concerning the implementation of new technologies in class: a latent class analysis of teachers and the relationship with the use of computers, ICT self-efficacy and emphasis on teaching ICT skills [J] . Large-scale Assessments in Education. 2017,01
[8] Zhang Yiqun, Luo Huiting, Gu Fangqing Exploring the Path of Knowledge Modular Teaching Design System in Emerging Interdisciplinary Areas under the Background of New Engineering [J] China Modern Education Equipment, 2023, (19): 84-86
[9] Liu Chen, Deng Qingxu, Chen Guocheng, et al. Reform and Practice of Embedded System Experimental Teaching in IoT Majors under the Background of New Engineering [J]. Laboratory Research and Exploration, 2023,42 (10): 196-199
[10] BozzolanM., SimoniG., BalboniM., Undergraduate physiotherapy students' competencies, attitudes and perceptions after integrated educational pathways in evidence-based practice: A mixed methods study [J] . Physiotherapy theory and practice. 2014,08
[11] Liu Yue, Yang Xinzhe, Zhai Yiming, et al. Research and practice on cultivating the quality and ability of engineering college students based on multiple evaluations [J]. Journal of Higher Education, 2024,10 (02): 151-155
[12] Zheng Lina Emerging Engineering: New Opportunities for the Development of Basic Disciplines under the Background of the New Industrial Revolution Higher Science Education, 2020, (01): 16-17
[13] Zhou Xiaojie, Design of Case Teaching in Higher Mathematics under the Background of New Engineering [J]. Science and Education Guide, 2023 (32): 110-112