

GIỚI THIỆU

Dữ liệu lớn

ThS. Nguyễn Hồ Duy Trí
trinhhd@uit.edu.vn



TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THỐNG THÔNG TIN

2020



THÔNG TIN GIẢNG VIÊN

Thạc sĩ

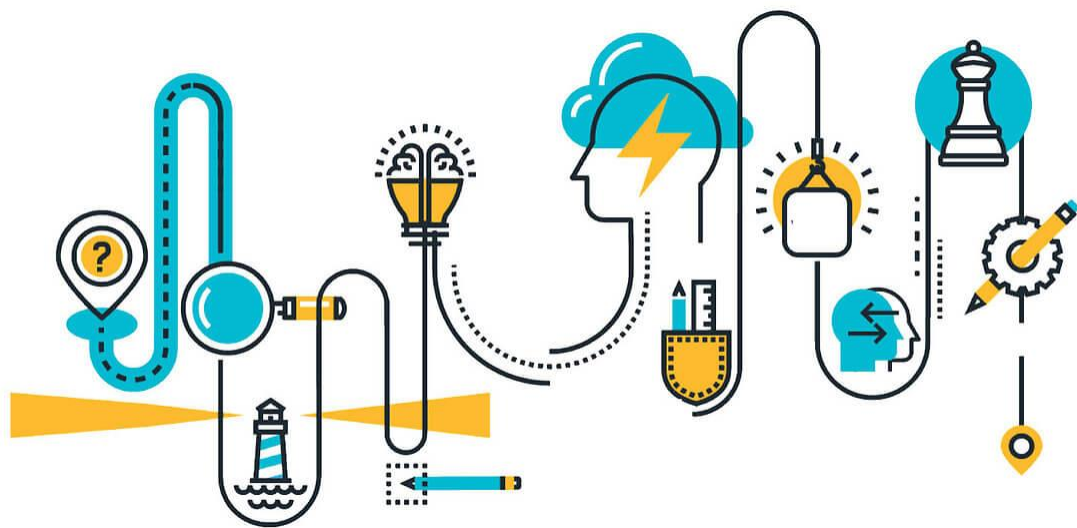
NGUYỄN HỒ DUY TRÍ

Email: trinhhd@uit.edu.vn

NỘI DUNG

- Mô tả môn học
- Mục tiêu
- Nội dung môn học
- Phân bố thời lượng
- Một số lưu ý khi học lý thuyết
- Đánh giá
- Nhóm
- Đồ án
- Tài liệu tham khảo

MÔ TẢ MÔN HỌC



Cung cấp các kiến
thức cơ bản



Trang bị các mô hình,
kỹ thuật, công cụ hỗ trợ

MỤC TIÊU

1

Hiểu và trình bày được các mô hình,
khái niệm cơ bản

2

Sử dụng được các nền tảng lưu trữ

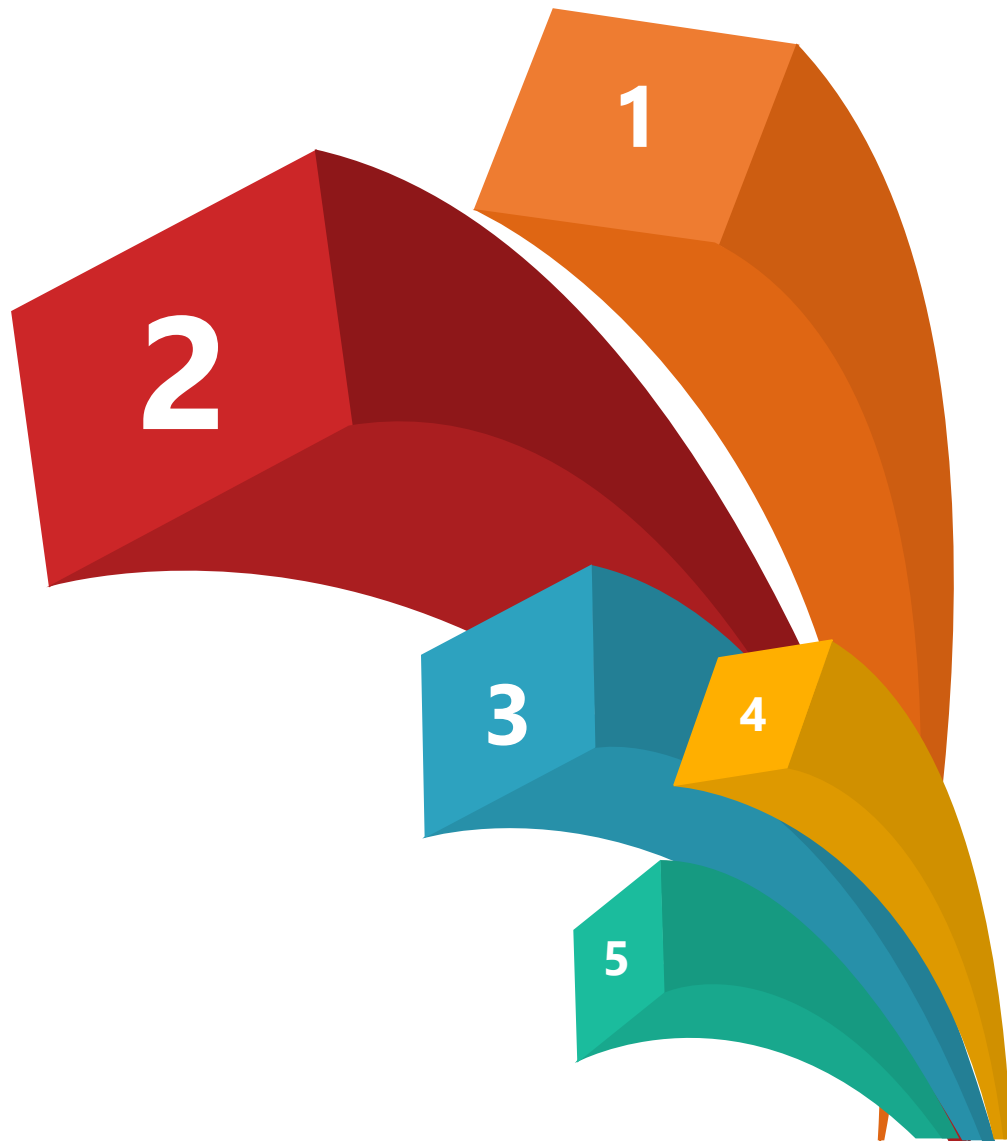
3

Áp dụng giải thuật khai phá dữ liệu
trên dữ liệu lớn

4

Tìm hiểu những hệ thống và
hệ sinh thái xử lý dữ liệu lớn

NỘI DUNG MÔN HỌC



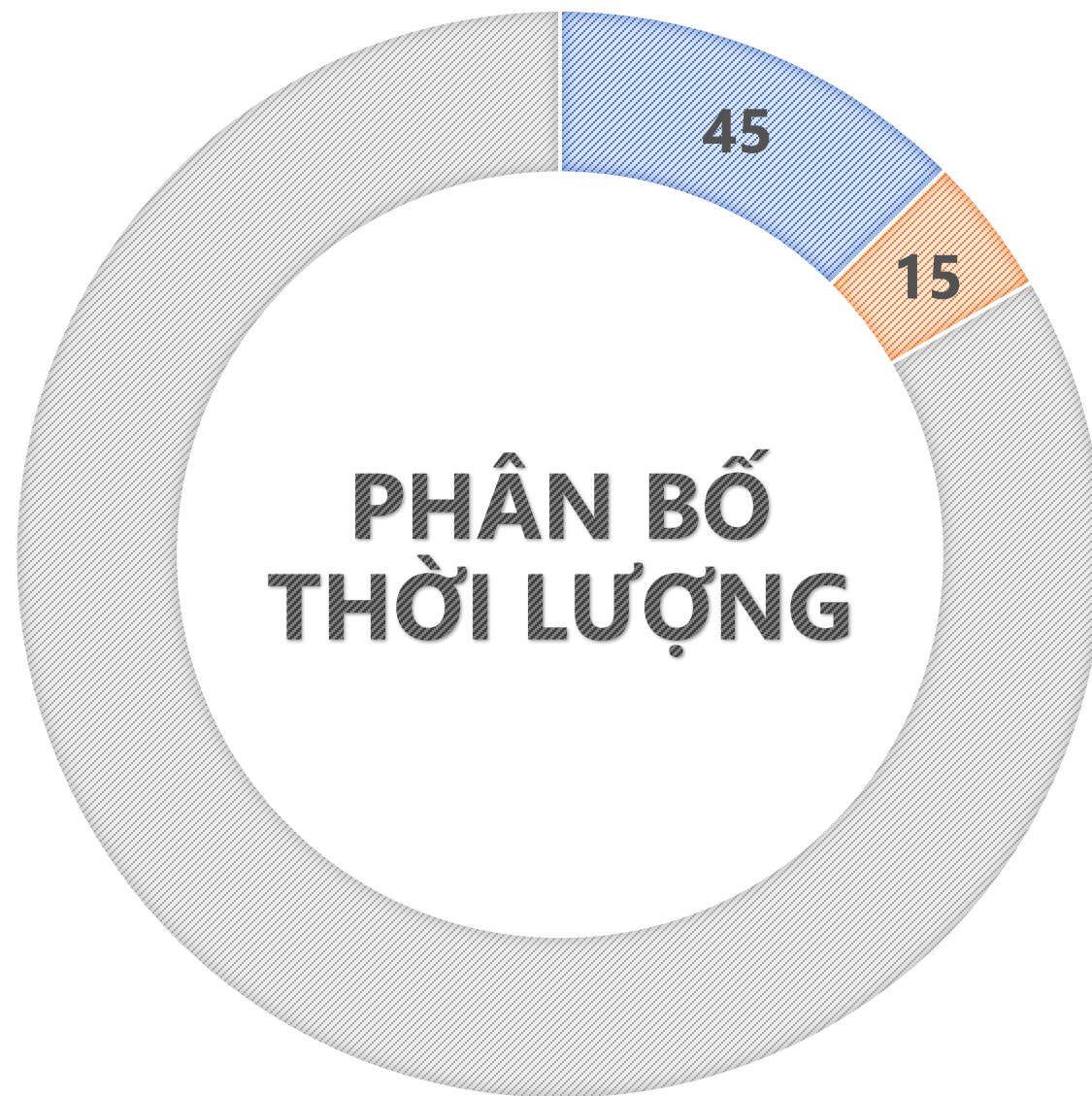
TỔNG QUAN

**MÔ HÌNH MAPREDUCE VÀ
FRAMEWORK APACHE HADOOP**

HỆ SINH THÁI HADOOP

FRAMEWORK APACHE SPARK

HỆ THỐNG XỬ LÝ DỮ LIỆU LỚN



■ Lý thuyết

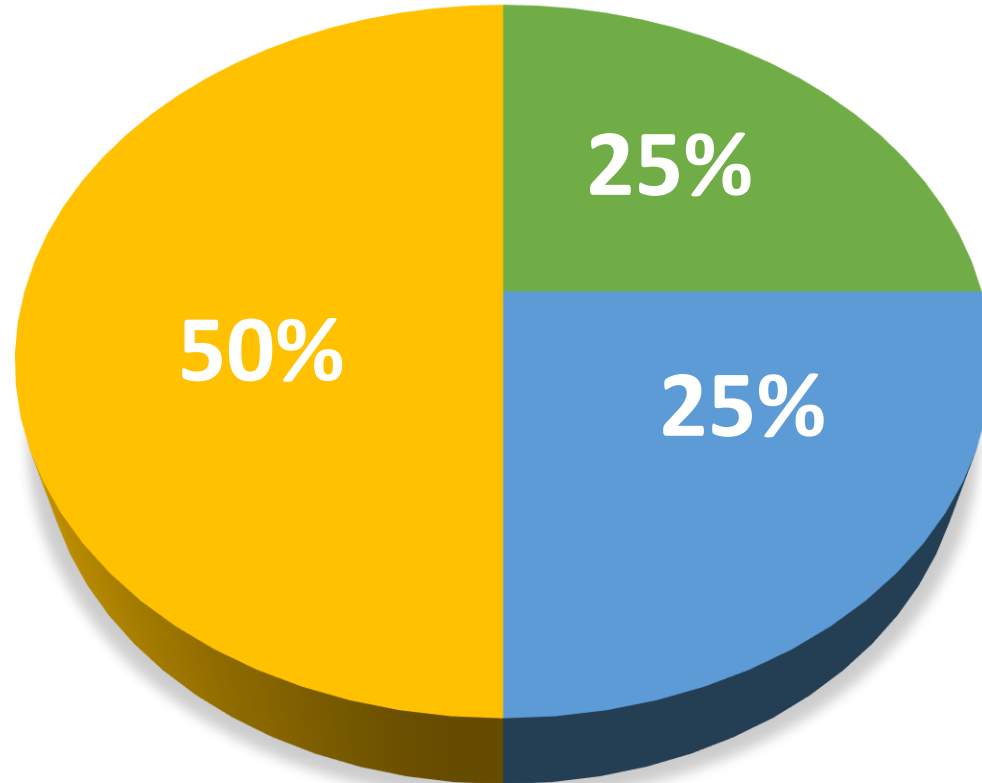
■ Thực hành

■ Tự học

MỘT SỐ LƯU Ý

- Tham dự lớp thường xuyên
- Tích cực tham gia trao đổi, thảo luận, làm việc nhóm
- Làm bài tập đầy đủ, đúng deadline
- Chủ động tìm kiếm, đọc thêm tài liệu
- Theo dõi, cập nhật tình hình lớp học

ĐÁNH GIÁ



■ Đồ án ■ Seminar + Chuyên cần ■ Thi cuối kỳ



NHÓM



NHÓM

ĐỒ ÁN



Viết ứng dụng minh họa xử lý và khai thác dữ liệu lớn.



Cài đặt hệ thống xử lý một bài toán dữ liệu lớn theo thời gian thực.

TÀI LIỆU THAM KHẢO

1. Tom White. 2015. **Hadoop: The Definitive Guide (4th ed.)**. O'Reilly Media, Inc.
2. Donald Miner and Adam Shook. 2012. **MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems (1st ed.)**. O'Reilly Media, Inc.
3. Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia. 2015. **Learning Spark: Lightning-Fast Big Data Analytics (1st ed.)**. O'Reilly Media, Inc.
4. Sandy Ryza, Uri Laserson, Sean Owen and Josh Wills. 2017. **Advanced Analytics with Spark: Patterns for Learning from Data at Scale (2nd ed.)**. O'Reilly Media, Inc.