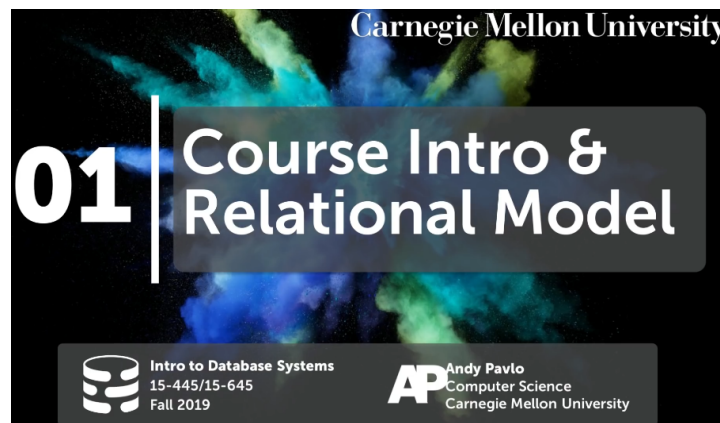


01-01 – Course Introduction & Relational...



2

00:00:31 – 00:00:37 winston

Welcome to CMU 15-445/645 introduction to database systems

欢迎来到 CMU 15-445/15-645 数据库系统简介

3

00:00:37 – 00:00:38

My name is Andy. I'm the instructor for this course

我叫Andy, 是这门课的讲师

0.39-0.44 虚生花

as I said on Piazza I can't be in Pittsburgh right now

正如我在Piazza(类似于wiki和论坛的混合)上所说, 我现在并不在匹兹堡

0.44-0.47

We are getting ready for Matt's cage fight on Wednesday

我们正在为周三的Matt's cage比赛做准备

6.

00:00:46 – 00:00:49

So we had to be out here in LA to prepare for that fight

所以我们必须到洛杉矶来为这场比赛做准备

7.

00:00:49 – 00:00:57

So rather than just not having classes this entire first week I wanted to record this

因此, 我不想在整个第一周都不上课, 所以抽空录了个视频

8.

00:00:57 – 00:01:07

While I'm on the road posted online and then that way when we start having in-class lectures next week, we can get started on the good stuff

我会发布到网上，这样，下周我们开始在课堂上讲课时，我们就可以开始学习好东西了
9.

00:01:07 – 00:01:14

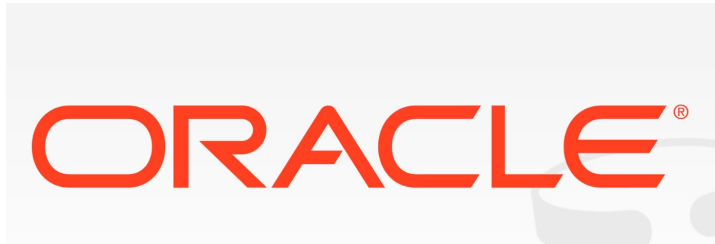
so before we get into today's lecture, I first want to talk about how Oracle is helping us
out this semester with course of course development

所以在开始今天的讲座之前，我首先要谈谈Oracle在本学期课程开发过程中对我们的帮助

1.14–1.21

So Oracle is actually won the first relational database management systems that came
out in the 1970s

诞生于1970年的Oracle是真正赢下第一关系型数据库宝座的存在



1.21–1.26 虚生花

And I'll explain what a relational database is in a few slides.

我会通过几张幻灯片来解释什么是关系型数据库

1.26–1.31 虚生花

There's one of the original ones, but it's still widely used and sold today right

Oracle是其中一种非常古老的关系型数据库，但它至今依然被广泛使用和出售

1.31–1.35 虚生花

It's probably the second most deployed database system in the world

它可能是目前世界上部署数量第二多的数据库系统

1.35–1.37

It's certainly the most deployed commercial system in the world

它确实是世界上部署数量最多的商业系统

1.37–1.46

And so even though it's from the 1970s And there's other systems of the 1970s that are
still sort of around, Oracle actually still under active development

即使它是在1970年代出现的，同时期还有其他数据库系统，但实际上Oracle依然在积极开发中

1.46–1.49 虚生花

And they're adding a lot of you know interesting new modern features to it

Oracle开发人员也增加了许多你所知道的有趣的与时俱进的新特性

1.49–2.02 !!!!

So this course is about me what I'll call it to me is traditional database management
systems or a classical design, and the interesting to see when they come and in the
semester to give a guest lecture talk about some of the stuff that they're adding to
Oracle today

所以这门课程之于我，我称之为传统数据库管理系统或经典设计，并且你会想知道本学期他们
(这些开发人员)什么时候来这里做客座演讲，谈论他们往Oracle中所添加的新内容

2.02–2.10

you know how they sort of go beyond you know what we'll talk about in this course and
sort of look any more more advanced topics

你知道他们所谈论主题和我们本课程所涉及的相比，会更多更高级

TODAY'S AGENDA

Wait List
Overview
Course Logistics
Relational Model
Relational Algebra

2.10–2.20

So for this lecture here, I want to first talk about these or the overall outline of what's expected in the course, but you as the student taking it.

So 对于这节课来说, 我想首先谈下本课程的预期要讲内容的总体概述, 即你作为学生要学的范围

17

00:02:20 – 00:02:26 虚生花

and then we'll finish up giving them a half lecture on the relational model and the relational algebra

然后, 我们会花半节课时间来讲关系模型和关系代数

2.26–2.34 虚生花

and this sort of be the background you need to have for the various topics we'll discuss throughout the rest of the semester

这也是我们在剩余学期内要讨论的不同主题时你所需要了解的背景

这是我们在本学期剩余时间所要讨论各种主题的大背景

WAIT LIST

There are currently 150 people on the waiting list.
Max capacity is 100.

We will enroll people based on your S3 position.

18

00:02:34 – 00:02:42 winston

so if you're a student Carnegie Mellon then the probably the foremost thing that's on your mind is the wait list

所以, 如果你是一位CMU学生, 那么你最重要的事情可能就是等候名单了

19

00:02:42 – 00:02:51 winston

so unfortunately this year they gave me a smaller classroom in Margaret Morrison than I wanted. so that means we can't take a large number people

不幸的是今年他们给了我一个 Margaret Morrison 里比较小的教室, 我想这样就意味着我们不能有很多人

20

00:02:51 – 2.55 虚生花

I think the max it says max capacity is 100,

我想最多也就能容纳100人

2.55-2.57 虚生花

I think it's like 96 is in the room

我觉得可能就96个人吧

2.57-3.00 虚生花

we can bring flood that number a little bit ,

我们可以让这个人数变多一点

3.00-3.03 虚生花

but I mean the main take away is here we can't accept everyone

但我的意思是我们不能接收等待人员列表上的每一个人

3.03-3.08 虚生花

,so the wait list now I think is a 115 when I checked earlier today

在今天早些时候我查看等待人员列表时，上面有115人

3.08-3.15 虚生花

So unfortunate that means that if you're not enrolled in the course at this point, the likelihood that you're gonna get in is is very low

因此不幸的是，如果你此时还没有注册这门课，那么你上这门课的几率就非常低

3.15-3.18 虚生花

the people will drop the course over the next week or so,

有些人会在下周drop这门课

3.18-3.24 虚生花

and then we'll take people off the wait list ,but unfortunately cant take everyone

然后我们才会从等待人员列表上选人，但不幸的是我们并不会选所有人

3.24-3.32 虚生花

so you're more than welcome to order the course just let me know that you're

ordering, but we can't officially hold that have that many people

因此，非常欢迎你注册这门课，我也知道你注册了，但正式的讲，我们没办法容纳这么多人

3.32-3.42 虚生花

so the way we will enroll students as current students drop out will be just on your waitlist position on s3 on a registers website

因此，只有当现有的学生放弃这门课的时候，我们才会从后补人员名单上选人然后让他们注册这门课

3.42-3.46 虚生花

so again I'm sorry, but it's just everything's available online

对于这点我很抱歉，但是所有的东西网上都有

3.46-3.48 虚生花

so you're more than welcome to follow along if you want

因此，如果你想的话，我们欢迎你去关注

COURSE OVERVIEW

This course is on the design and implementation of disk-oriented database management systems.

This is **not** a course on how to use a database to build applications or how to administer a database.
→ See [CMU 95-703](#) (Heinz College)

Database Applications ([15-415/615](#)) is **not** offered this semester.

25

00:03:48 – 00:03:58 winston

all right , so (15-415/615) is at it`s core the course is about the design and implementation of database management systems

好的, (15-415/615) 核心课程是有关数据库管理系统的设计与实现

26

00:03:58 – 00:04:09 虚生花

so that means that this is not a course on how to use a database to build like an application ,like a website or something, like that or have actually had administer a database how to deploy one set one up

这也就是说这并不是一门教你如何使用数据库去构建应用程序、网站或者其他东西的课, 也不是一门教你如何去管理和部署数据库的课

27

00:04:09 – 00:04:12 虚生花

we're not doing any of that

我们并不会教你那些

4.12-4.20 虚生花

we're really focusing on how do you actually build and design the software. that is a database management system

我们所注重的是教你如何去构建和设计软件, 这才是数据库管理系统

4.20-4.25 虚生花

so if that's not what you're looking for, if you're not looking to know how to have a better database management system

因此, 如果这不是你想要的课, 或者说你不想要知道如何拥有一个更好的数据库管理系统的话

4.25-4.31 虚生花

Then there's two other courses that you should consider at senior that are outside of SES
那么除了SES外, 你应该考虑其他两门课程

4.31-4.35 虚生花

and the one in particular you may want to look at is (95-703) at the Heinz College

你可能想要的一门是Heinz College的CMU 95-703

4.35-4.37 虚生花

I think is Information Systems

我觉得它是和信息系统相关的

4.37-4.42 虚生花

and again that's about how to like set up and administer a database management system

也就是说那是一门关于如何设置并对数据库管理系统进行管理的一门课

4.42-4.47 虚生花

but you don't really understand they don't really go into detail about how you actually built the software

但他们并不会真正详细介绍如何构建软件

但你真的不会理解而那门课也不可能真正深入构建数据库软件的内在细节

4.47-4.59 虚生花

the other thing to that people often ask is about, there's a usually another course at CMU (15-415/615) something that I and professor Christopher Lewis has taught in the past

还有另一件人们常问的事情，那就是我和Christopher Lewis教授以前所教过的CMU (15-415/615)

4.59-5.01 虚生花

that is not being offered the semester

但这学期并不提供这门课

5.01-5.04 虚生花

I don't know whether it's gonna be offered in the spring probably unlikely

我也不知道春季课程中有没有它

5.04-5.11 虚生花

so right, now I think the only course is available in SES would be this course 445

因此，我觉得现在在SES上唯一有的一门课就是445了

5.11-5.17

and then a 826 in machine learning machine learning or cross-listed in the CS department

以及CS系机器学习这块的15-826

5.17-5.24

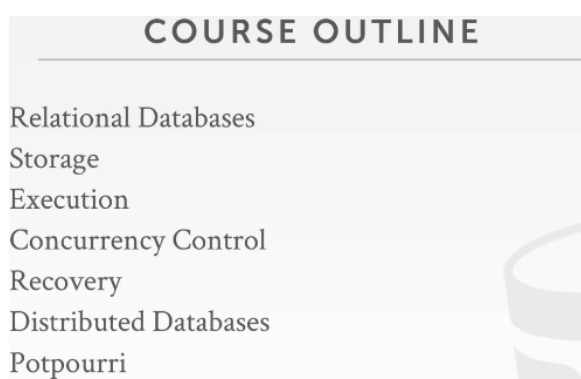
so again meets there's always there's only Christopher myself, we can't teach everything

因为现在只有我和Christopher，精力有限没办法去教所有东西

5.24-5.27

so unfortunately there's there's no other database class other than this one

因此，不幸的是除了这门课以外，没有别的数据库课程了



5.27-5.38

all right, so the course outline is that we're going to be going through and discussing ,then you know how to build a disk oriented database for an assessment

好了，我们接着要讨论下我们的课程大纲，然后你就会有有一个关于如何去构建一个用于评测的面向磁盘的数据库

5.38-5.40

and I'll explain what a disk oriented system (此处oriented部分存疑) is in a few lectures

我会花几节课时间来解释什么是disk oriented system

5.40-5.45

which basically we're just writing data out the disk, we assume the database is on disk
基本上，我们假设数据库是存放在硬盘上，然后我们在硬盘中写入数据
基本上，就是指我们假设数据库在磁盘上，然后将数据写到磁盘上，
5.45-5.52

so it's sort of broken up the outline for the the topics we'll be discussing is broken up
and sort of into layers of the system

这里我把我们要讨论的主题列成大纲并会将整个系统按层次逐步讲下去

so 我们将要讨论的主题的大纲分解，并且系统地分了几个层面

5.52-5.56

right so we'll talk at a high level what relational databases are

我们会以上层的角度来讨论关系型数据库是什么

5.56-6.04 ssss

and then we'll talk about how to store them , how to execute query on them ,how to run
transactions on them, how to recover them if there's a crash or we need to restart the
system

接着，我们会去讨论如何存储它们，如何在它们之上进行查询，如何在它们之上执行事务管理，
以及如果遇上冲突或者我们需要重启系统时该如何去恢复它们

6.04-6.11

so up to that point up to recovery that's the core knowledge you need to have to
understand how a database management system works

直到恢复这块内容为止，这些都是你所必须理解的关于数据库管理系统是如何工作的核心知识

6.11-6.15

and then from there we can then build on that and start talking about more,

接着，我们可以在此之上来讨论更多内容

6.15-6.25

you know more advanced topics like distributed databases ,or various other types of
databases that are out there ,or extensions of relational databases

就好比你所知道的那些高级主题，例如分布式数据库，或者一些其他类型的数据库以及关系型数据库的一些扩展

6.25-6.30

so again like the way to think about this is we'll go through every single layer of how to
actually build the system

我们可以通过每一层面的知识来告诉你如何构建这个系统

6.30-6.37

and we'll finish up for recovery and so ~~at that point~~ that's the basic you need to
understand of how a database system works,

直到恢复数据库这部分内容，这些是你所需要了解数据库系统是如何工作的基础知识

6.37-6.45

and then we'll talk about how to you know how to extend them to scale them up or
scale now ~~of running and~~ running them ,and in the cloud environment

然后，我们会去讨论如何对它们的规模进行扩展以及如何在云环境运行它们

COURSE LOGISTICS

Course Policies + Schedule:

→ Refer to [course web page](#).

Academic Honesty:

→ Refer to [CMU policy page](#).

→ If you're not sure, ask the professors.

→ Don't be stupid.

All discussion + announcements will be on [Piazza](#).

6.45–6.53 虚生花

so the right now the course website is online along with the syllabus and the schedule

现在，课程网站已经上线，上面有教学大纲以及课程表

6.53–6.56 虚生花

so the basic outline there's a lecture twice a week

一般来说，我们一周两节课

6.56–7.00 虚生花

and then with each lecture there's a there's readings that go along with it ,

然后每节课都会有相关阅读材料

7.00–7.07 虚生花

but better supplemental that extend the kind of things that that I'll be talking about as

well also provide some mention on a second

但一些更好的课后补充材料之后我也会提供

7.07–7.16 虚生花

so at all times please refer to the course webpage that should be up-to-date, you know

it have been up most up-to-date information on what's going on

因此，请访问课程主页，上面所放的内容始终都是关于课程的最新动态

7.16–7.20 虚生花

so unfortunately we always have to talk about academic honestly

~~可惜，我们始终必须诚实地来谈论学术问题~~

不幸的是，我们总是不得不去谈论学术诚信的问题

7.30–7.23 虚生花

I'll go a little bit more detail what I mean about this as we go along

关于这个问题，我会对它进行详细说明

7.23–7.29 虚生花

but when you ever have a question and you don't understand

但当你有不理解的地方

7.29–7.30 虚生花

please contact me

请联系我

7.30–7.34 虚生花

so that we can discuss whether what you're doing could be considered

so，我们来谈下关于学术诚信这方面的内容

7.34–7.40 虚生花

you know plagiarism or stealing somebody else's work right

你知道抄袭和剽窃他人作业是一种很糟糕的行为

7.40–7.42 虚生花

So in general again ,this is an advanced course

总体来说，这是一门高级课程

7.42-7.47 虚生花

so everyone should be aware that you don't copy coding fine don't randomly on the internet you don't copy from each other

每个人都应该意识到你们不应该从网上随机抄袭或者抄其他人的

7.47-7.50 虚生花

but just you know be very careful because we will check for these things

但你应该认真对待它们，因为我们会对这些进行检查

7.50-8.00 虚生花

and as I said , all the discussion and announcements for projects, lectures ,homeworks will be you on Piazza will do grading with great scope

正如我所说，所有的讨论和关于项目，课程，课后作业的公告都会放在Piazza上，然后我们也会在上面进行打分

8.00-8.02 虚生花

your final grade will be posted on Canvas

你的最终成绩会发布在Canvas上

8.02-8.05

because ~~that's what see me once~~ but the the detonated discussion will be on Piazza

但一些热门讨论也会放在Piazza上面

8.05-8.10

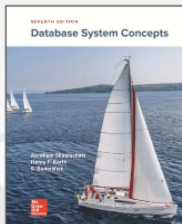
and there's a link on the course web page now a little that'll take you to our page

课程网站上有一个链接会把你带到我们的页面上

TEXTBOOK

Database System Concepts
7th Edition
Silberschatz, Korth, & Sudarshan

We will also provide lecture notes that covers topics not found in textbook.



8.10-8.15 虚生花

there is a textbook assigned for this for this class Database Systems Concepts

这门课中我们所使用的教材是Database Systems Concepts

8.15-8.19 虚生花

so this is actually a new edition that came out this year

这实际上是今年出的最新版

8.19-8.22 虚生花

I've looked at pretty much every single database systems textbook fit that's out there

我看了不少符合我想法的数据库系统相关的教科书

8.22-8.26 虚生花

in my opinion this one's actually the the best one

实际上，这一本是我认为最好的一本

8.26-8.30 虚生花

it's it's it's the most up-to-date

它的内容也是最新的

We will also provide lecture notes
that covers topics not found in textbook.

8.30–8.36 虚生花

and as I said, well we've above I've lecture notes for topics that aren't covered in the textbook

正如我所说，我们也会提供一些这本书所没有涉及内容的课程笔记

8.36–8.44 虚生花

~~I have to MIT me~~ I haven't looked in a detail of the seventh edition to understand how much it differs in the sixth edition

我还没有仔细去看第七版和第六版之间有什么区别

8.44–8.46 虚生花

so if you want to get the sixth edition ,I'm fine with that

如果你用的是第六版，我觉得也没什么问题

8.46–8.48 虚生花

I don't think there should be any major difference

我不觉得这里面有什么太大差别

8.48–8.58 虚生花(待修改)

~~I just may not know exactly how to you could look at last semesters chap last semesters.~~ oh sorry ,last year's course and see for the different topics with the chapter numbers

你可以去浏览下去年的课程，并和这些章节进行对比来查看其中不同的主题

8.58–8.59 虚生花

because they have changed

因为它们的内容不少已经变了

8.59–9.02 虚生花

and I'm actually not sure whether you can buy this book anymore

实际上，我也不确定你还能不能买到这本书

9.02–9.04 虚生花

like you can't buy it as a bound book

你买不到这本书的精装版

9.04–9.10 虚生花

they sent me you know a bunch of page loose loose pages or ,you know three three-ring hole punches in them

你们也知道，他们送了我一本，但里面缺了不少页，甚至还有几个洞在书上

9.10–9.13 虚生花

so I don't know what the bookstore has ,I don't know whether to available online it,

我不知道书店有没有，我也不知道网上能不能找到它

9.13–9.22 虚生花

in the sixth edition but it's probably good enough, right we won't do any homeworks or any problems out of the book directly

对于我们来说，第六版就够用了。我们不会去做任何超出这本书范围的作业或者习题

9.22–9.24 虚生花

,right well we'll provide everything for you

我们会为你们提供一切

COURSE RUBRIC

Homeworks (15%)

Projects (45%)

Midterm Exam (20%)

Final Exam (20%)

Extra Credit (+10%)

9.24–9.29 虚生花

all right ,so the breakdown for your grade in the class will be the following

好了，这门课的成绩是由以下几部分组成的

9.29–9.35 虚生花

so homeworks will be 15%, then it would be course projects which I'll discuss in a second that'll be 45%

作业占了分数的15%，之后我所要谈的课程项目占了45%

9.35–9.44 虚生花

so for those of you that are CS undergrads ,because the project grade your final grade is comprised of 45 percent projects

对于你们那些CS本科生来说，因为项目的分数已经占了你最终成绩的45%

9.44–9.52 虚生花

that's why this course counts for the system software elective for the CS undergrad curriculum

这就是为什么这门课作为CS本科生阶段系统软件选修课程的原因了

9.52–9.55 虚生花

and then there'll be a midterm and final exam both at 20% ,

接着，期中考试和期末考试的成绩都各占了20%

9.55–9.58 虚生花

and then there'll be an extra credit which I'll announce in a few weeks

接着还有一个额外分数加成，关于这点我会在之后几周宣布

9.58–10.03 虚生花

when you can get an additional 10% bonus points and that's that's entirely optional

你可以得到额外的10%的加分，但这取决于你要不要

HOMEWORKS

Five homework assignments throughout the semester.

First homework is a SQL assignment. The rest will be pencil-and-paper assignments.

All homework should be done individually.

10.03–10.07 虚生花

so there'll be five homeworks throughout the semester

整个学期有5次家庭作业

10.07–10.14 虚生花

the first one will be a sequel assignment , you will give you a SQL Lite database you have to write some queries for us

第一次（作业）是数据集分配，会给你像数据库一样的数据集，你要为我们写一些查询语句

第一个作业是一个SQL相关的作业，我会给你一个SQL Lite数据库，你需要写一些数据库查询语句

10.14-10.18 虚生花

but then everything after that will be pencil and paper ,

但之后的作业都是你用纸和笔去完成的

10.18-10.24 虚生花

because it's a way to work through those sort of more theoretical side of some of the things that we're talking about

因为这是一种去理解我们所讨论的理论知识的方法

10.24-10.33 虚生花

but it'll be like filling out multiple-choice ,and then you just take a picture of it and upload it to could grade scope and will provide grade that way

但就像填写选择题那样，你只需把你的作业拍个照然后上传，接着我们就会对它打分

就比如做选择题，你只要把你的答案拍个照然后上传，然后我们就会对它打分

10.33-10.35 虚生花

so again, the first assignment is SQL

So再说一遍，第一次作业是SQL相关

10.35-10.40 虚生花

just because of you know we actually won't be writing SQL for the rest of the semester, 就像你知道的，实际上在这个学期剩下的时间里，我们不会一直去写SQL

10.40-10.42 虚生花

because the course projects don't require it

因为课程项目中并不需要它

10.42-10.48

and I think it's good for you to guys to touch it at least once at least my big band

stuff we'll talk about it next semester

我觉得对于你们来说至少玩一下也是好的，下学期我们会讨论它

10.48-10.52 虚生花

so again because is about saying all these homeworks should be done individually as well as the projects

再说一遍，所有这些作业以及项目都必须独立完成

10.52-10.57 虚生花

but it's one emphasized this that like you're not allowed to work in groups and try to figure things out

但有一点要强调，我不允许你们组队完成作业

10.57-10.58 虚生花

it's not like a Theory class yet

它并不是一门理论课

10.58-11.02 虚生花

all these are assignments can be done and should be done individually

所有的作业都应该独立完成

PROJECTS

You will build your own storage manager from scratch of the course of the semester.

Each project builds on the previous one.

We will not teach you how to write/debug C++17

11.02–11.07 虚生花

All right ,for the projects, this is the one I'm pretty excited about

好了，我最兴奋的事情就是项目了

11.07–11.13 虚生花

so throughout the course of the semester, you will build your own database storage manager from scratch

在整个学期中，你需要从头开始构建你自己的数据库存储管理器

11.13–11.20 虚生花

So you'll start adding ~~you know~~ pieces one by one and start building out aids or a full-featured a database storage manager

So你可以一点一点的来构建出一个功能完整的数据库管理器

11.20–11.24 虚生花

so the keyword there storage manager and not a database system

注意，这里的关键字是存储管理器而不是数据库系统

11.24–11.27 虚生花

because you're not gonna be able to run SQL or have a query parser

因为在这里面，你无法运行SQL语句，也没有查询解析器

11.27–11.32 虚生花

but you'll be able to run queries that are hand coded that will provide you

但你能够通过提供给你的手写代码来执行查询

11.32–11.36 虚生花

So it's it's it's more complex than the sort of that simple key value store

因此，这要远比简单的键值存储来的复杂得多

11.36–11.38 虚生花

but it's not like a full-fledged system

但它并不是一个功能完善的系统

11.38–11.46 虚生花

So the key thing about this is that it's very important for you to keep up to date with the projects

对你来说，实时更新你的项目非常重要，这也是最关键的事情

11.46–11.49 虚生花

because every project is gonna build after one after another

因为每个项目都是基于另一个项目构建出来的

11.49–11.55 虚生花

so you sort of have to have the first project working correctly in order for the second project to work correctly and the third and so forth

因此，你必须保证你第一个项目正确工作，这样你第二个项目才能正确工作，第三个项目也是如此，以此类推

11.55–12.01 虚生花

so this particular system that we're using this semester is written entirely in C++ 17

我们这学期所使用的系统是完全用C++17所编写的

12.01–12.05 虚生花

it doesn't mean we're using all the advanced features of C++ 17

这并不意味着我们会去使用C++17的所有高级特性

12.05-12.14 虚生花

it just means that it's you know it's not like you're you know C99, so traditional single plus you may have learned in other classes

它和你所知的C99不同, 你可能会在其他课上学到传统的C++

12.14-12.19 虚生花

so because this is Carnegie Mellon and I assume this is advanced class

So因为这里是卡内基梅隆大学, 我觉得这也是一门高级课程

12.19-12.26

I'm not ignore the TA are going to be teaching you or teaching how to write or debug C++

我不确定助教们是不是要去教你们如何编写C++代码或者对代码进行debug

12.26-12.28

I sent up a self-exam on Piazza

我在Piazza上设置了一次自我测验

12.28-12.34

if you feel like you're uncomfortable with ,you know those are the gnarly aspects of C++

如果你对上面关于C++相关方面的知识不熟悉的话

12.34-12.38

and then you should try to figure out how we try to start learning this stuff now

那么你就应该试着去弄懂我们开始要学的东西

12.38-12.44

right ,you can't come to us as the TA'S office hours and say, hey you know what is the stack trace mean

你也不要来助教办公室问一些诸如stacktrace这种简单问题

12.44-12.53

you know the this time is really expect it to be discussing the serve them the more high-level important database concepts that you're trying to implement in your code


你知道我们希望这些时间可以用来和你们讨论那些你尝试用代码所想实现的数据库概念

BUSTUB

All projects will use the new **BusTub** academic DBMS.
→ Source code will be released on Github.

Architecture:

- Disk-Oriented Storage
- Volcano-style Query Processing
- Pluggable APIs
- Currently does not support SQL.



12.53-13.04

so the ,all the projects this year will be implemented on this new academic system that we've been working on called BusTub

今年所有的项目都会在我们所开发的新学术系统BusTub上进行实现

13.04-1308 虚生花

so all the source code will be released on Github

所有的源码都会在Github上进行开源

13.08-13.12 虚生花

of course obviously it won't have the implementation of the piece that you chose to implement

当然，很明显，它上面并没有你选择去实现的那部分功能

13.12-13.14 虚生花

but you sort of fill that in

但你需要去实现它们

13.14-13.21 虚生花

so at a high level it's a disk based or disk oriented a data management system that'll support volcano-style query processing

从上层的角度来讲，这是一个面向硬盘的数据管理系统，它支持Volcano式风格的查询处理

13.21-13.26 虚生花

Different parts of the system have a sort of a Pluggable API

该系统中的部分内容还支持插拔式API

13.26-13.34

so that we can Pluggable in different you know replacement algorithms are given

,different index data structures ,or different logging schemes or control schemes

这样我们就可以替换系统中已经给出的算法，使用不同的索引数据结构或者不同的日志格式和控制方案

13.34-13.40

so it's designed to be that way so that every year will switch up the project's entirely

该系统就是使用这种思路来实现的，因此每年都会将整个项目完全换掉

13.40-13.46

and have it be different form one year to the next ,So the slowly build out the system further and further with new features and functionality

这样它就会每年都不一样，这样一步步通过添加新特性和功能去构建整个系统

13.46-13.51

so that you know after a couple years we'll have a full-fledged database management system

这样你就知道，数年之后，我们就会有一个功能完善的数据库管理系统

13.51-13.55 （飞哥既视感）

So you guys are sort of the first ones starting off with these first set of projects

因此，你们这批人就是第一批开始做这期项目的人

13.55-13.59

and then next year what will modify them in it'll be different

然后，在下一年我们会对它们进行修改，这样又让它们变得不一样了

13.59-14.02

All right, so this for this reason, we can make it open source

基于这个原因，我们将它开源

14.02-14.10

because I'm not worried about people you know people next year finding your crappy projects implementations and copying their code

因为我并不担心下一年的学生找到你那垃圾的项目实现并抄袭你的代码

14.10-14.12

because it's all the project of being entirely different

因为整个项目都会变得完全不一样

14.12-14.16

so this is what I was saying before the last slide that you're good in basically storage manager

So就是我上张幻灯片上所说的一样，它基本就是一个存储管理器

14.16–14.21 虚生花

,the database system doesn't support SQL at this point and nor will and this the semester
这个数据库系统这学期并不支持SQL

14.21–14.29 虚生花

but you'll be able to write queries ,but you'll write them in sort of physical operator form
rather than ,and you know in SQL and then translate them

但你还是能去写查询语句的，但你会以物理运算符的形式去将SQL翻译并进行查询，而不是用你所知道的SQL语句去进行查询

14.29–14.31 虚生花

because it's not something we're doing just yet

因为这不是我们正在做的事情

14.31–14.33

so the name of the system is BusTub

因此，该系统的名字是BusTub

14.33–14.36

I will explain offline what that means

我会在线下解释它的含义

14.36–14.39

but we had a nice logo logo made

但我们给它做了一个非常好看的Logo

14.39–14.45

and again what announced onGithub, we'll post a link on Piazza in week two

在第二周，我们会在Github上上传内容并在Piazza上发布链接

14.45–14.47

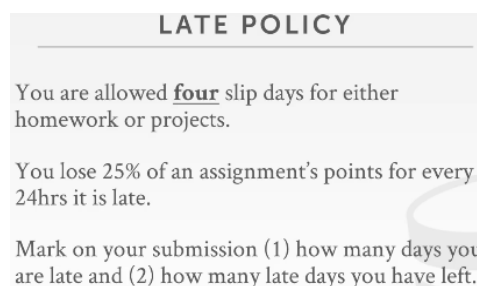
well right now it's the first project

现在，它是第一个项目

14.47–14.50

so I'm pretty excited about having this the everyone work on this this semester

So我对这学期为这个项目作出贡献的每个人而感到兴奋



14.50–14.57

all right for the late policy for the homework and projects ,every student is allowed four slip days,

这里我来讲下关于作业和项目晚交的延期政策，每个学生允许有四天的宽限期（这段时间内交作业不扣分）

14.57–15.07

so basically for any homework or project you say I you know I found my daily you can decrement affront from your account

一般来讲，对于任何作业或者项目，如果延期，我们会对你进行扣分

15.07–15.17

so the be on a on each homework and the project submission ,you just say how many late days you use how many ladies you left so sort of allow yourself to keep track of what ,you know have how many late days you have

对于每个作业和项目来说，请在你的提交记录上标明你延期的天数以及你所剩下的延期天数
15.17–15.28

so after you run out of slip days, then it'll be you lose 25% for every on the and the sum of total points every time, it's it's it's for every 24 hours that it's late

当你用完了宽限期，如果你作业没交，那么每24小时就会扣掉你项目或者作业的25%的分数
15.28–15.32

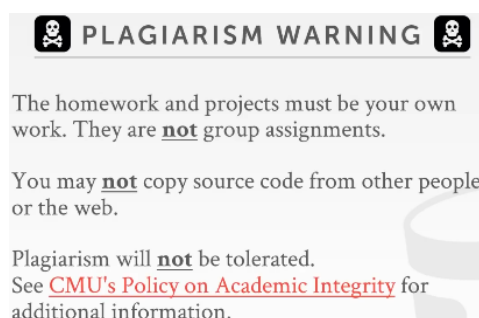
so again this wouldn't keep attracting us as we go along in the semester

随着学期的进行，我们不会一直关注这个

15.32–15.38

obviously if there's medical munitions or other issues that come up please contact me and we can accommodate you

如果生病或者其他问题，请与我联系，我们可以给你提供备选方案



15.38–15.45

So again as I said before now all these projects and the homework should be done individually

这里再说一次，如我之前所说，所有的项目和作业必须独立完成

15.45–15.47

,they're not group assignments

它们并不是团队作业

15.47–15.51

you're not allowed to work it out together and submit a single submission together

我不允许你们组队完成它，然后一起单独交一份

15.51–15.53

you should be doing everything individually

你应该独立完成每一件事

15.53–16.01

now some code for some projects may be still online ~~from from wish~~ or from previous years

某些前几年项目的代码可能网上现在还存在

16.01–16.06

don't take that we're gonna run it through boss the players plagiarism checker,

请不要抄袭，我们会对它们进行抄袭检查

16.06–16.12

if we catch you running their shitty code ,unfortunately you know we got to report me to Warner Hall

如果我们抓到你抄袭他们的狗屎代码，那么很不幸，你懂的这件事会通报给我

16.12–16.16

so don't do that because this ,you know you know it's stupid it'll fuck your life

所以请不要抄袭，因为这很愚蠢，这也会毁了你的生活

16.16–16.20

and it makes everything harder ,right just don't don't plagiarize

它会让所有事情变得很艰难，所以请不要抄袭

16.20–16.30

okay and yeah, if you're unsure, check the academic policy or policy for academic integrity ,or contact me if you're unsure about what to do ,and this includes also for the extra credit

如果你不确定，请查看下学术政策，如果你不确定该做什么，请联系我，当然这也包括额外加分的部分

16.30–16.37

and just because it's extra credit and it's optional doesn't mean that you can't also get caught for plagiarism

因为它是额外加分并且是可选的，但这并不意味着你因为抄袭而不会陷入困境

16.37–16.38

so don't do that as well

因此请不要这么做

16.38–16.44

and I'll remind you every single time we put a new project out every single time ,you know we we tell why the extra credit okay

每次我们放出一个新项目时，我就会告诉你们如何得到额外加分



16.44–16.51

all right, so if you want to go beyond the kind of things we're talking about this this course, if you really like databases which I do

除了我们所讨论这门课的内容以外，你还想学更多的话，如果你真的就像我那样热爱数据库的话

16.51–16.54 虚生花

you have no idea how much I love databases

你可能并不知道我是多么热爱数据库

16.54–16.59

If you want to go beyond the course material

如果你想去学习课程资料以外的东西

16.59–17.04

there's two sort of ways to get involved in database research or other database topics going on at Carnegie Mellon

那么这里有两种方式去参与数据库研究或者其他在卡内基梅隆大学的数据库课题

17.04–17.13

so the CMU database group has our weekly meetings on Mondays at 4:30 in the gates, gates building own a floor

在每周一下午的4:30, CMU数据库讨论组会在GHC 8102举行每周会议

17.13-17.23

and this is other students visitors from companies ,and people in Pittsburgh from abroad coming and giving talks about what you know the the kind of researcher kind of work that they're doing

这里会有其他学校的学生, 一些大公司的来宾或者一些来自国外的人来参加这场会议, 然后这些研究人员会对他们的工作进行演讲

17.23-17.28

If you want to get involved in the development of a sort of advanced system

如果你想参与高级系统的开发

17.28-17.35

we have our team meetings on Tuesdays at 12 o'clock also in the gates building

我们在每周二中午十二点的GHC 8115房间举办我们的团队会议

17.35-17.39

so we're building in addition to bustub sort of the academic system

我们在为开发我们的学术系统BusTub而努力

17.39-17.45

we have a new sort of full-featured database management system that we've been doing for several years now

我们有一套功能完善的新数据库管理系统, 为此我们开发了好几年

17.45-17.52

that again if you want to get involved in this kind of stuff ,you should come check that out and I'll send a reminder on this Piazza

如果你想加入这项工作, 那你可以看下我之后在Piazza上发的提醒信息

17.52-18.03

it will say also to if you want to take the advanced class (15-721) in the spring ,all those projects are based on on this this new this other system we're building

如果你想在春季选15-721这门高级课程, 那么它里面的所有项目都会基于我们所开发的这个新系统

18.03-18.11

so if you want to get that and sort of learn learn how that system works and get involved in the early days of this thing ,no by all means come to this

如果你想知道系统是如何工作的并想参与到早期开发中, 那就啥也别说参与进来吧

Databases

18.11-18.21

okay ,so with that that that's that's it for the course I mean please if you have questions about things post posted coach them on Piazza and I'll respond

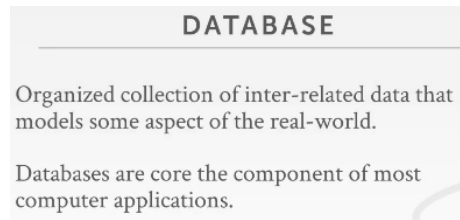
好了, 这就是这门课关于学术诚信所涉及的内容, 如果你还有任何疑问, 请在Piazza上贴出来, 我会进行回复

=====

18.21–18.25

all right ,so now let's talk about Databases

好了，现在让我们来讨论下数据库



18.21–18.34

the Databases are super important in in real life

数据库在现实生活中非常重要

18.34–18.36

because they're used everywhere

因为它们无处不在

18.36–18.45

,so many but every single complex are energy and any computer applications you can think of ,at the end of the day deep down inside of it there's going to be a database
你能想象到的任何电脑应用程序，它们内部都会有一个数据库

18.45–18.55

right ,if it's a mobile phone application, if it's running your desktop, if it's a website right, if it's some kind of complex computer simulation, the end of the day there's always a database

如果它是一个手机应用或者桌面程序，或者是一个网站，又或者是某种复杂的电脑模拟程序，在它们里面始终会有一个数据库

18.55–19.00

,everyone has database problems many things can then just be reduced down to database problems

这些应用都有数据库层面的问题，许多事情都能最后简化为数据库层面的问题

19.00–19.09

so a database the definition I like to use is that it's a it's a collection of data that's related to gathers in some way

所谓数据库，我喜欢这样去定义它，即它是以某种方式去进行关联的数据集合

19.09–19.12

that's trying to model some aspect of the real world

这就可以试着对现实世界的某些方面进行建模

19.12–19.17

right ,it's not just a bunch of loose files you have randomly sitting on your laptop

它们并不是那些随机分布在你电脑上的那些零散文件

19.17–19.22

right and that in some ways that the database but it's not really useful one because you can't ask questions about it

某种意义上，这算是一种数据库，但并不是那么有用，因为你无法对它进行一些查询操作

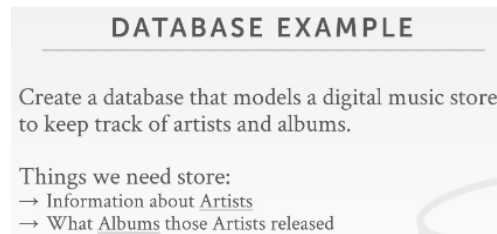
19.22–19.26

so this data is usually related together or have some common theme to them

这些数据通常会关联在一起或者它们之间有某些共同主题

19.26–19.30

and it's trying to you know model some aspect of something that's going on in reality
可以试着去对现实一些事的某些方面进行建模



19.30– 19.41 虚生花

so the example I always like to use for this class is that say we want to have a digital music store by something like Spotify or the iTunes Store right

这个例子是我最喜欢在这个课堂上用的，假设我们想要有一个类似于Spotify或者iTunes商店那样的电子音乐商店

19.41–19.46

and so we're so backing this application will be a database

在这些应用程序背后做支持的就是数据库

19.46–19.51

that's going to keep track of the various Artists we have and their Albums

它用来保存许多艺术家以及他们的专辑信息

19.51–20.00 虚生花

right ,so the what we would put in this database would be basic information about their Artists ,and then information about what Albums that those Artists have released

在这个数据库中我们所放入的是艺术家们的基本信息，以及他们所发行的专辑信息

20.00–20.05

all right so that's that cool right there that's that's that's a database

这很Cool对吧。这就是数据库！