Revision Date 02/01/2022

General Information

## Meeting Times and Location

Session dates: 02/1/2022 - 5/17/2022; Meets Tuesday at 7:10 pm to 9:40 pm.

The class will meet online (Webex at URL <https://umbc.webex.com/meet/yuehang>) every Tuesday. Assignments, course materials and lecture slides will be posted on Blackboard.

|  |  |  |
| --- | --- | --- |
| Instructor | Email | Office Location & Hours |
| **Hang Yue** | yuehang@umbc.edu | Virtual, phone and web conference sessions by appointment |
| **Note: teaching Online, if students want to stay in classroom for group meeting** on the main campus at Information Technology 233**, please wearing a mask is strongly recommended in the classroom. Hope that you have been fully vaccinated.**  **https://highpoint-prd.ps.umbc.edu/app/catalog/classsection/UMBC1/2222/2273** | | |

## Description

This course is specifically designed to support the range of complex data challenges Data Practitioners face today from optimizing relational database systems to managing big data. Students will get an overview of relational database management systems, SQL programming, and emerging Big Data NoSQL database technologies. Also, we include the topics of advanced databases.

## Prerequisite

Students must be enrolled in the Data Science Program. Other students may be admitted with instructor permission.

## Course Learning Objectives

Upon completion, students will understand:

* the fundamentals of relational and big data database systems
* key concepts related to database design and administration
* how to acquire and store both structured and unstructured data
* how to write SQL queries to transform and process data
* fundamentals of data warehouse design
* key concepts related to big data technologies
* understand pros and cons of the different database options

## Text Books

1. Principles of Database Management: (PDBM) The Practical Guide to Storing, Managing and Analyzing Big and Small Data, Lemahieu, Vanden Broucke, and Baesens ISBN 9781107186125
2. TSQL T-SQL Fundamentals, 3rd Edition Itzik Ben-Gan ISBN 9781509302000
3. MDM Modern Database Management, 13th Edition, Hoffer, Venkataraman & Topi ISBN 780134773650. This is only available as an e-book at present. You can order the e-book from <http://www.mypearsonstore.com/bookstore/modern-database-management-9780134773650>

## Course Format and Assignments

The students will complete serval class assignments/labs/homework, a midterm exam and a final team project. This course incorporates a variety of hands-on labs and practical exercises to engage students and prepare them for challenges they may encounter in the workplace. Topics include relational databases, writing SQL queries, big data technology, and some advanced databases.

## Course Communication and Content

I will use UMBC’s Blackboard application at <https://blackboard.umbc.edu/> as the primary means for distributing course content, feedback, and communication. Occasionally, for larger downloads, I will provide them on a flash drive.

## Grading Criteria

Students are expected to participate in class discussions.

|  |  |
| --- | --- |
| Course work | Grade Distribution |
| Homework / Labs | 30% |
| Midterm Exam | 35% |
| Final Project | 35% |

Final Grade will be computed as follows:

90-100% A (>93.0 A; 90.0 – 93.0 A-)

80- 89% B (87.0 – 89.99 B+; 83.0 – 86.99 B; 80.0 – 82.99 B-)

70- 79% C (77.0 – 79.99 C+; 73.0 – 76.99 C; 70.0 – 72.99 C-)

60- 69% D

<60% F

## Course Policies

UMBC provides a range of writing assistance, which can be found in the following:

* The Writing Center: <http://lrc.umbc.edu/tutor/writing-center/>
* Research Guides & Tutorials: <http://lib.guides.umbc.edu/tutorial>

Failure to follow guidelines for each assignment, including the required format, style, length, and submission may result in at least one-letter-grade reduction on the assignment depending on the type or number of transgressions.

Late/incomplete assignments will be accepted if an extension has been agreed to in advance. Emergency situations will be handled on a case-by-case basis with appropriate justification or documentation.

Incomplete grades are granted only for extenuating circumstances and your request is made before the last week of class.

## Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC’s scholarly community in which everyone’s academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping other to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, but is not limited to failure, suspension or dismissal. Homework responses from other courses should not be reused in this class.

Refer to the UMBC policy at: <http://catalog.umbc.edu/content.php?catoid=17&navoid=879#academic-integrity>.

#### **Accessibility and Disability Accommodations, Guidance and Resources**

Accommodations for students with disabilities are provided for all students with a qualified disability under the Americans with Disabilities Act (ADA & ADAAA) and Section 504 of the Rehabilitation Act who request and are eligible for accommodations. The Office of Student Disability Services (SDS) is the UMBC department designated to coordinate accommodations that would create equal access for students when barriers to participation exist in University courses, programs, or activities.

If you have a documented disability and need to request academic accommodations in your courses, please refer to the SDS website at [**sds.umbc.edu**](http://sds.umbc.edu/) for registration information and office procedures.

SDS email: disAbility@umbc.edu  
SDS phone: (410) 455-2459

If you will be using SDS approved accommodations in this class, please contact me (instructor) to discuss implementation of the accommodations. During remote instruction requirements due to COVID, communication and flexibility will be essential for success.

#### **Sexual Assault, Sexual Harassment, and Gender Based Violence and Discrimination**

UMBC’s [**Policy on Sexual Misconduct, Sexual Harassment and Gender Discrimination**](https://oei.umbc.edu/policy-on-sexual-misconduct-sexual-harassment-and-gender-discrimination/) and Federal Title IX law prohibit discrimination and harassment on the basis of sex in University programs and activities.  Any student who is impacted by sexual harassment, sexual assault, domestic violence, dating violence, stalking, sexual exploitation, gender discrimination, pregnancy discrimination, gender-based harassment or retaliation should contact the University’s Title IX Coordinator to make a report and/or access support and resources:

Mikhel A. Kushner, Title IX Coordinator (she/her/hers)  
410-455-1250 (direct line), kushner@umbc.edu

*You can access support and resources even if you do not want to take any further action.*You will not be forced to file a formal complaint or police report. Please be aware that the University may take action on its own if essential to protect the safety of the community.

If you are interested in or thinking about making a report, please see the [**Online Reporting Form**](https://umbc-advocate.symplicity.com/titleix_report/index.php/pid364290?). Please note that, while University options to respond may be limited, there is an anonymous reporting option via the online form and every effort will be made to address concerns reported anonymously.

**Notice that Faculty are Responsible Employees with Mandatory Reporting Obligations:**

All faculty members are considered *Responsible Employees*, per [**UMBC’s Policy on Sexual Misconduct, Sexual Harassment, and Gender Discrimination.**](https://oei.umbc.edu/policy-on-sexual-misconduct-sexual-harassment-and-gender-discrimination/) Faculty are therefore required to report possible violations of the [**Policy**](https://oei.umbc.edu/policy-on-sexual-misconduct-sexual-harassment-and-gender-discrimination/) to the Title IX Coordinator, even if a student discloses something they experienced before attending UMBC.

While faculty members want you to be able to share information related to your life experiences through discussion and written work, students should understand that faculty are required to report Sexual Misconduct to the Title IX Coordinator so that the University can inform students of their [**rights, resources and support**](https://oei.umbc.edu/files/2020/04/OEI-KYR-Resources-5.pdf).

If you need to speak with someone in confidence, who does not have an obligation to report to the Title IX Coordinator, UMBC has a number of [**Confidential Resources**](https://humanrelations.umbc.edu/sexual-misconduct/umbc-resource-page-for-sexual-misconduct-and-other-related-misconduct/#confidential) available to support you:

* The [**Counseling Center**](https://counseling.umbc.edu/): 410-455-2472 / After-Hours 410-455-3230
* [**University Health Services**](https://uhs.umbc.edu/): 410-455-2542
* Pastoral Counseling via [**Interfaith Center**](https://campuslife.umbc.edu/diversity-and-inclusion/interfaith-center/): 410-455-3657; interfaith@umbc.edu

Other Resources:

* [**Women’s Center**](https://womenscenter.umbc.edu/) (for students of all genders): 410-455-2714; womenscenter@umbc.edu.
* [**Shady Grove Student Resources**](https://oei.umbc.edu/shady-grove-title-ix-resources/), [**Maryland Resources**](https://oei.umbc.edu/maryland-resources/), [**National Resources**](https://oei.umbc.edu/national-resources/).

Child Abuse and Neglect:

Please note that Maryland law and [**UMBC policy**](https://oei.umbc.edu/child-protection/) require that I report all disclosures or suspicions of child abuse or neglect to the Department of Social Services and*/*or the police.

#### **Pregnancy**

UMBC’s [**Policy on Sexual Misconduct, Sexual Harassment and Gender Discrimination**](https://oei.umbc.edu/policy-on-sexual-misconduct-sexual-harassment-and-gender-discrimination/) expressly prohibits all forms of Discrimination and Harassment on the basis of sex, including pregnancy. [**Resources for pregnant students**](https://oei.umbc.edu/students/) are available through the University’s Office of Equity and Inclusion.  Pregnant and parenting students are encouraged to contact the Title IX Coordinator to discuss plans and assure ongoing access to their academic program with respect to a leave of absence or return following leave related to pregnancy, delivery, or the early months of parenting.

In addition, students who are pregnant may be entitled to accommodations under the ADA through the [**Student Disability Service Office**](https://sds.umbc.edu/accommodations/registering-with-sds/), and/or under Title IX through the [**Office of Equity and Inclusion**](https://oei.umbc.edu/pregnancy-and-parenting/).

#### **Religious Observances & Accommodations**

UMBC [**Policy**](https://provost.umbc.edu/files/2020/08/Religious-Holidays-2020-21.pdf) provides that students should not be penalized because of observances of their religious beliefs, students shall be given an opportunity, whenever feasible, to make up within a reasonable time any academic assignment that is missed due to individual participation in religious observances. It is the responsibility of the student to inform the instructor of any intended absences for religious observances in advance, and as early as possible.  For questions or guidance or to request an accommodation, please contact the [**Office of Equity and Inclusion**](http://oei.umbc.edu/) at [**oei@umbc.edu**](mailto:oei@umbc.edu).

#### **Hate, Bias, Discrimination and Harassment**

UMBC values safety, cultural and ethnic diversity, social responsibility, lifelong learning, equity, and civic engagement.

Consistent with these principles, [**UMBC Policy**](https://oei.umbc.edu/discrimination-policy/) prohibits discrimination and harassment in its educational programs and activities or with respect to employment terms and conditions based on race, creed, color, religion, sex, gender, pregnancy, ancestry, age, gender identity or expression, national origin, veterans status, marital status, sexual orientation, physical or mental disability, or genetic information.

Students (and faculty and staff) who experience discrimination, harassment, hate or bias or who have such matters reported to them should use the [**online reporting form**](https://umbc-advocate.symplicity.com/titleix_report/index.php/pid066642?) to report discrimination, hate or bias incidents; reporting may be *anonymous.*

## Course Syllabus (maybe updated later)

| Week | Lecture & Discussion | Reading | Exercises |
| --- | --- | --- | --- |
| Week 1  2/1/2022  Online  <https://umbc.webex.com/meet/yuehang> | Introduction  History of Databases  Database Setup  Class Survey  Online Class Setup | PDBM – Ch 1 & Ch 2 Pg. 3-33;  CH 5. Pg. 93 - 102  TSQL – Pg. 10-19  TSQL – Getting started appendix Pg. 395 | Software installment |
| Week 2  2/8/2022  Online  <https://umbc.webex.com/meet/yuehang> | Database Modeling and Lab1   * ER Model | PDBM Ch 3 pg. 38-57 | * ER tool installment and study * Lab1 Part 1of ER |
| Week 3  2/15/2022  Online  <https://umbc.webex.com/meet/yuehang> | Database Modeling and Lab2   * Relational Model * Homework 1 | PDBM Ch4 pg. 104-139  Reading: E. F. Codd Paper  PMDB Ch. 4 Pg. 79-88 | * Lab1 Part2 of ER |
| Week 4  2/22/2022  Online  <https://umbc.webex.com/meet/yuehang> | Basic SQL Programming and Exercise 1  Data Management  Homework2 | TSQL Ch. 1, Ch. 2, Ch. 3 |  |
| Week 5  3/1/2022  Online  <https://umbc.webex.com/meet/yuehang> | Advanced SQL Programming  Homework3 and Exercise 2 | TSQL – Ch. 4, Ch. 5, Ch. 6  Ch. 11, Flow Elements Pg. 367-370 |  |
| Week 6  3/8/2022  Online  <https://umbc.webex.com/meet/yuehang> | Datawarehouse/ETL  and Final Project Assignment | PDBM Chapter 17   * Sec 17.1 Decisions, page 552 * Sec 17.2, Data Warehouse Definition, page 553 * 17.3 Star & Snowflake Schema, p.554- 560 * 17.4 ETL page 565 - 567 * Sec 17.5 Data Marts p 567 - 569 * Sec 17.9 OLAP & OLAP queries p 574 - 583 |  |
| Week 7  3/15/2022  Online  <https://umbc.webex.com/meet/yuehang> | Midterm Exam – take home exam |  |  |
| Week 8  3/22/2022  Online  <https://umbc.webex.com/meet/yuehang> | No class, spring break |  |  |
| Week 9  3/29/2022  Online  <https://umbc.webex.com/meet/yuehang> | AWS and Mid Exam Review |  |  |
| Week 10  4/5/2022  Online  <https://umbc.webex.com/meet/yuehang> | AWS glue Exercise 3 |  |  |
| Week 11  4/12/2022  Online  <https://umbc.webex.com/meet/yuehang> | Big Data and HBase Lab 3 |  |  |
| Week 12  4/19/2022  Online  <https://umbc.webex.com/meet/yuehang> | BI and Data Visualization |  |  |
| Week 13  4/26/2022  Online  <https://umbc.webex.com/meet/yuehang> | Tableau Exercise 4 |  |  |
| Week 14  5/3/2022  Online  <https://umbc.webex.com/meet/yuehang> | Final Team Project |  |  |
| Week 15  5/10/2022  Online  <https://umbc.webex.com/meet/yuehang> | Final Team Project |  |  |
| Week 16  5/17/2022  Online  <https://umbc.webex.com/meet/yuehang> | Data science on database research for the application and development of future databases |  |  |

# Exam Schedule

| Date | Subject |  |
| --- | --- | --- |
| Week 7, 3/15/2022 | Midterm Exam |  |
| Week 14-15, 5/3 to 5/10, 2022 | Final Exam – Team Project |  |

# Additional Information and Resources

## Learning Resources

For additional guidance, Lynda.com offers excellent tutorials and walk throughs on SQL, Python, Hadoop and database design. Please refer to this page for information on how to access <https://wiki.umbc.edu/display/faq/Lynda.com>.

## Recommended Software

* To Install SQL Server for Windows:

SQL Server 2019 Express

<https://www.microsoft.com/en-us/sql-server/sql-server-2019>

* Management Studio

<https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-2017>

* To Install SQL Server for Mac:

<https://learnsqlteam.com/2020/05/15/installing-mssql-on-mac/>

* **To Install “databricks community edition”**

<https://databricks.com/product/faq/community-edition>

* **To Install Tableau (Student free license)**

<https://www.tableau.com/academic/students>

* **To Install AWS free Tier**

https://aws.amazon.com/free/?all-free-tier.sort-by=item.additionalFields.SortRank&all-free-tier.sort-order=asc&awsf.Free%20Tier%20Types=\*all&awsf.Free%20Tier%20Categories=\*all

Data 604: Setup SQL Tools in details

Update 9/3/2021, Hang Yue, Adjunct Professor

Revised 2/6/2019, Patty Stanton, Adjunct Professor

Revised 5/26/2020, Jyotsna Potarazu, Adjunct Professor

There are a variety of database management system and software tools to administer those systems. To follow along with the Class lectures, we will use SQL Server 2017 and SQL Server Management Studio.

# Students with Windows Machines

## Local Installation

If you have Windows Machines, you can install Microsoft SQL Server Express <https://www.microsoft.com/en-us/download/details.aspx?id=55994> and

Install SQL Server 2019 Express Edition

<https://www.microsoft.com/en-us/sql-server/sql-server-2019>

Management Studio <https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-2017> locally on your machine. This is the recommended option if you are able as it will give you the most experience working with a database management system.

# Students with Advanced SQL Skills

Students who have advanced SQL skills can opt to use tools of their preference. They will be responsible for looking up and translating any difference in syntax if using a DBMS other than SQL Server.

# Students with MACs

To Install SQL Server for Mac

<https://learnsqlteam.com/2020/05/15/installing-mssql-on-mac/>

## VirtualBox

If you have a Mac and at least 4 GB of memory and at least 25 GB of available hard disk space, you can download a virtualbox sandbox that has SQL Express and Management Studio preloaded.

Instructions for installing VirtualBox and the sandbox are given below

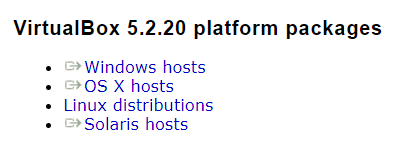
# VIRTUALBOX Instructions

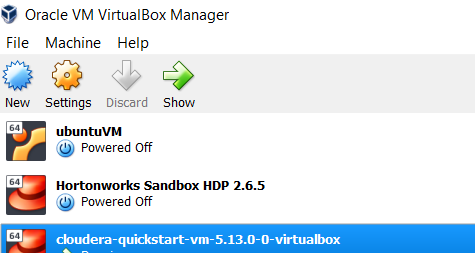
## Description

Oracle VirtualBox is an open source software that allows you to run virtualized versions of common computer and operating systems such as Linux or Windows. It is a good tool to learn how to install and interact with applications and technology in a low-cost development environment. Caution must be used in that if you do utilize any of the tools for production such as Windows, you need to ensure you comply with the vendor’s licensing requirements.

## Installation Instructions

* Download from <https://www.virtualbox.org/wiki/Downloads>
  + Click on applicable link for the VirtualBox 5.2.24 platform packages (or most recent edition) to download the installation file for your operating system (**Windows host** for windows computer, **OS X hosts** for apple computer).



* + Save and install the program according to your operating system. For Windows, run the VirtualBox-5.2.20-125813-Win.exe file to launch the install program. For Apple, the file ends with dmg extension
  + Note for those with Apple computers, after you install VirtualBox, you will need to go into **System Preferences**, then **Security & Privacy** and enable Oracle VirtualBox.
  + For Intel based laptops, especially Lenovo and gaming/high performance laptop, virtualization might be disabled. You will need to go into your BIOS and enable it. Please do a web search based on your laptop vendor to get the instructions to enable as instructions may vary depending on your laptop. I can help you in class if you are unable to figure it out.
  + Additional information can be found at <https://www.virtualbox.org/manual/UserManual.html>.
* In Windows 10, you can type in the search ‘VirtualBox’ or ‘Oracle VirtualBox’ to find and launch the Oracle VM VirtualBox Manager.  
   
* Stop here for now.

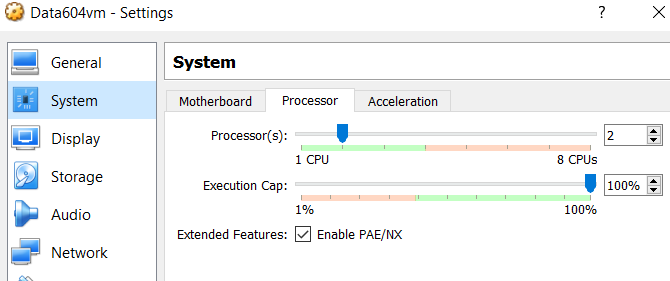
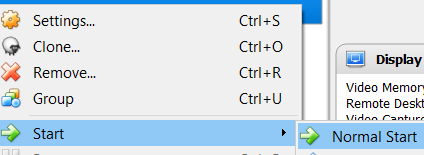
# SQL Server Sandbox

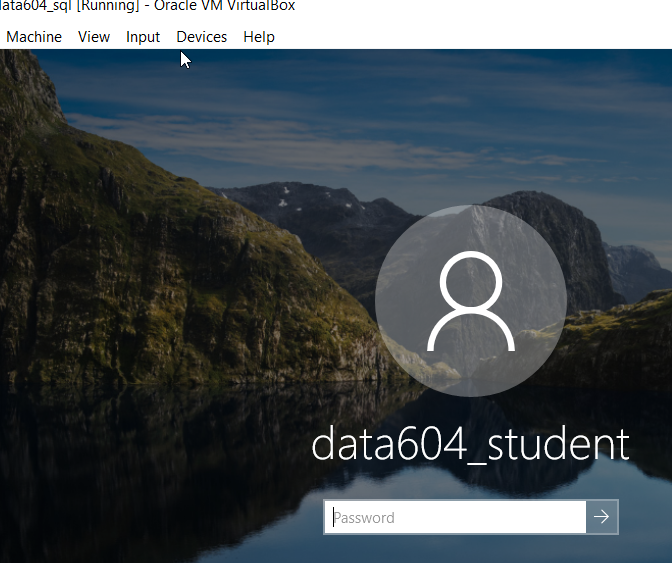
## Description

I have created an educational version of a Windows sandbox with SQL Server Express and Management Studio installed. You can directly download from <https://umbc.box.com/s/cko1yk9ba69brdv97434gop6h21l8ms7>. This is a very large file that will take a while to download. Once it is finished, you will need to unzip it to a folder of your choice. If you are unable to download, please let me know and I can provide you a copy on a USB drive. If you are loading from the USB, make sure that you copy the sandbox folder to a drive on your laptop. We will be using the VirtualBox platform in this class as it works well with both the Apple and Windows based operating systems.

The sandbox is a little over 8GB in size and you will need at least 25GB of space on your computer to run and work with the sandbox. If you downloaded the file, extract the file from the zip file. You will see a file called Data604\_sql with an ova extension. This is the virtual image.

## Setting up the Sandbox:

* Once you have unzipped the sandbox, double-click on it and Virtual Box will load a screen to import the virtual appliance
* Click **Import** to continue.
* Now this critical step as you will select the Cloudera virtual hard drive in this step which will import your virtual machine.
* You will now see your virtual machine listed in the powered off state: 
* Let’s double check the machine settings and configure the CPU before we power on. Make sure the machine is selected and click on the Settings icon. 
* Select **System** from the left menu. Confirm in the Motherboard tab that the base memory set to 2gb. You can adjust here as needed.
* Next, select the **Processor** tab. The machine settings default to one CPU (these are CPU cores- if you have a dual core CPU, there will be two CPUs listed for each core). For better performance, you can additional CPUs or leave at one. 
* You don’t need to change anything else but feel free to explore the other settings. If your laptop is GPU enabled, you can add 3d acceleration to the virtual machine in the Display settings but it is not needed for this sandbox. If the future, you can adjust these settings but you will need to make sure the virtual machine is in a Powered Off state to change them.
* **Right-click** on your node, and from the pop-up menu select **Start** and then **Normal Start** 
* It will take a few minutes to start and could take much longer depending on your machine.

Once it starts, you will see a second VirtualBox window: 

Your virtual machine is now installed. If you need to switch between keyboard capture modes (inside or outside the virtual machine), press the right mouse and right control key.

The windows user account, password, and security prompts are:

windows account data604\_student

pw d@t@rul3s

name of 1st school attended umbc

first pets name: True Grit

Childhood nick name True Grit

Additionally, this sandbox is not licensed in any form or manner for production use. Please only use it for educational purposes. Any use other than educational is illegal.