

1. Welcome


This is the first topic for the course, inviting you to understand how the course works and what to expect. Generally course materials are organised in e-books and you will see one or more e-books grouped into a section for each topic of the online course.

How to read the course materials: COMP3425/COMP8410 is being delivered in *hybrid* mode for all students in S1 2021. Students have options to participate remotely (also called *online*) or physically *on-campus*. The same online materials are used for the undergraduate COMP3425 in *on-campus mode*, postgraduate COMP8410 in *on-campus mode* (and also *on-line*), and postgraduate COMP8410 in *blended mode*. In some places, advice and content is customised for one or two of these alternatives. Differentiated material will be identified in **purple as u/g only, p/g only, blended only, on-campus only, or online only**. Please be conscious of which of these categories apply to you.

ACTION: Read and follow this instruction:

Throughout the course materials you will see remarks like this, in green. These are instructions for you to follow.

ACTION 1. Read the [course outline](#) -- it is very important that you have this [course outline](#) in mind throughout the course. You are assumed to know this and a great deal of it is actually helpful. You can come back to this document at any time as it is easy to find right at the top of course home page.

 [Course Outline](#)

ACTION 2. Read the [course schedule](#) and keep it on hand through the course. You will need it to plan your days for attendance and assessment tasks.

 [Course Schedule](#)

ACTION 3. Check your knowledge of relational databases. You will need some basic understanding of relational databases and SQL in this Data Mining course.

 [Use this if you need a refresher for SQL](#)

ACTION 4. Set up your computer for practical exercises.

You will first need to use these platforms in Week 3. So you if you have problems, then do not panic, and seek help in your first lab class in Week 2.

This course has a strong practical component and you will need to use, at least PostgreSQL, Rattle, R, Protege, and OWL-Miner that relies on the Java runtime). You may also want to use additional data mining software of your choice, such as Python tools.

Firstly, if you have not done so already, you will need to register with [Streams](#). Log in with your ANU user id and ANU password. If you change your ANU password during the course, you will need to come back and re-register with [Streams](#) to use lab facilities including the VDI.

Now, you may choose to either (a) Use the ANU-supplied Horizon VDI through which all the required software is pre-installed or (b) use your own self-installed software in versions for Windows, Mac or Linux, or (c) visit the physical laboratories in the CSIT and HN buildings on campus where all software is installed or (d) use remote-access to the CSIT and HN laboratories on campus or (e) use any combination of the above. There is also an option (f) to use an Oracle Virtual Box VM, <https://cs.anu.edu.au/docs/student-computing-environment/linuxlabs/softwareaccess/> but this may be slow on low-performance personal computers and the ANU staff will not offer support. Be aware that the VDI (option (a)) offers an Ubuntu linux-based desktop environment and may be difficult on low-throughput network connections, although it has been found to be trouble-free for most students and is likely to be the least-fuss remote option. Alternatively, self-installation has the advantages that you will be interacting with your own familiar O/S and once you have invested the extra effort to install, you can keep the software for future work. Rattle has proven to be easy to install for most students, but always there are a few Mac users that never solve all their problems.

If you choose the VDI option, then follow the instructions here: <https://cs.anu.edu.au/docs/student-computing-environment/linuxlabs/VDI/>. You will need to reboot your machine part-way through, so you will need to remember to come back here afterwards. You will continue to need to continue to run [GlobalProtect](#) every time you subsequently use the Horizon VDI.

For the **PostgreSQL** database work, you can self-install PostgreSQL if you prefer, without assistance from course staff. You will be able to complete all the PostgreSQL work by suitably adapting the instructions given here. For all other options, to set up for database work you will need to follow the instructions here to use the PostgreSQL server on Partch.

 [Practical Exercise: Set Up for Data Warehouse Activities](#)

If you choose to self-install **Rattle**, then follow the instructions here. If you chose to use another access method instead then you will also need to follow these instructions, beginning with "Start Rattle" as you do not need to install first.

 [Practical Exercise: Install Rattle and get to know it](#)

The remaining course software is available in the labs and through the VDI or self-installation, but we need not be concerned with that until later in the course.

ACTION 5. Enrol in a lab class. Enrolment opens 9am on the first day of the course.



[Tutorial/Laboratory Groups](#)