A. COURSE DESCRIPTION
This course covers the concept of pipelines – assemblies of basic bioinformatics tools and data sources to solve complex data processing problems. The pipeline concept will be introduced with simple UNIX command line methods, and then extended to the use of preconfigured commercial and extensible open-source workflow management systems. Reproducibility of analysis, collection of analytic provenance information, and database integration will also be covered. (3 credit hours, On Demand)

B. PRE-OR CO-REQUISITES
Prerequisite: BINF 6203/8203.

C. OBJECTIVES OF THE COURSE
Having successfully completed this course, the student will be able to:
- Build simple analysis pipelines at the UNIX command line
- Build workflows within commercially available preconfigured software
- Work with public data services
- Build workflows within extensible open source workflow management systems
- Integrate publicly available UNIX software tools into workflow management systems as components
- Integrate simple databases into workflows
- Understand issues of reproducibility and data provenance tracking that are addressed by using workflow management systems

D. INSTRUCTIONAL METHOD
The course is presented in a hands-on laboratory format, with demonstrations and problem-solving activities making up the majority of the course activities. Brief lectures will be used to explain concepts prior to the start of hands-on activities.

E. MEANS OF STUDENT EVALUATION
Students will be evaluated on their ability to build and execute workflows following protocols presented in the class (80% of the grade) and to answer factual questions about resources and concepts presented (20% of the grade).

F. SPECIFY POLICIES THAT APPLY TO THIS COURSE:
The following policies apply to students in BINF 6215:

1. UNIVERSITY INTEGRITY
All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online at: [http://www.legal.uncc.edu/policies/ps-105.html](http://www.legal.uncc.edu/policies/ps-105.html). A set of links to various...
resources on plagiarism and how to avoid it is available at the UNCC Library website: http://library.uncc.edu/display/?dept=instruction&format=open&page=920.

2. ATTENDANCE
Attendance at lecture is required, although exceptions will be made for reasons such as illness or family emergency. Excessive absences will result in a reduced classroom participation score at the instructor’s discretion, and will negatively impact the overall course grade.

3. GRADING POLICY
Grades will be assigned on the following scale:
A=90-100%
B=80-90%
C=65-80%
U=0-65%

4. ADDITIONAL POLICIES.
The use of cell phones, beepers, or other communication devices is disruptive, and is therefore prohibited during class. Except in emergencies, those using such devices must leave the classroom for the remainder of the class period. Students are permitted to use computers during class for note-taking and other class-related work only. Those using computers during class for work not related to that class must leave the classroom for the remainder of the class period.

G. PROBABLE TEXTBOOKS OR RESOURCES

H. TOPICAL OUTLINE OF COURSE CONTENT
- Basic concept of pipelines
- Simple pipeline implementation with UNIX scripts and commands
- Workflow management in an integrated genomics environment
- Construction of genomics workflows from integrated components
- Construction of workflows in an open-source environment
- Using scientific software in an open source WMS
- Using web services in an open source WMS
- Workflow construction in an open source WMS
- Using databases in an open source WMS
- Tracking data provenance in an open source WMS