Course Approval Form

For approval of new courses and deletions or modifications to an existing course.

registrar.gmu.edu/facultystaff/curriculum

Action Requested:

- [X] Create new course
- [ ] Inactivate existing course
- [ ] Modify existing course (check all that apply)
  - [ ] Title
  - [ ] Credits
  - [ ] Repeat Status
  - [ ] Grade Type
  - [ ] Prereq/coreq
  - [ ] Schedule Type
  - [ ] Restrictions
  - [ ] Other:

Course Level:

- [X] Undergraduate
- [ ] Graduate

College/School: College of Science
Department:
Submitted by: Nektaria Tryfona
Ext: 4884 **Email:** ntryfona@gmu.edu

Subject Code: COS **Number:** 401
(Do not list multiple codes or numbers. Each course proposal must have a separate form.)

Effective Term:

- [ ] Fall
- [X] Spring
- [ ] Summer

Year 2016

Title:
Current Banner (30 characters max including spaces): Big Data Management
New: **Big Data Management**

Credits: (check one)

- [x] Fixed: 3 or
- [ ] Variable: to

Repeat Status: (check one)

- [X] Not Repeatable (NR)
- [ ] Repeatable within degree (RD)
- [ ] Repeatable within term (RT)

Maximum credits allowed:

πGrade Mode: (check one)

- [X] Regular (A, B, C, etc.)
- [ ] Satisfactory/No Credit
- [ ] Special (A, B C, etc. +IP)

Schedule Type: (check one) LEC can include LAB or RCT

- [X] Lecture (LEC)
- [ ] Lab (LAB)
- [ ] Recitation (RCT)
- [ ] Internship (INT)
- [ ] Independent Study (IND)
- [ ] Seminar (SEM)
- [ ] Studio (STU)

Prerequisite(s):
CDS 302: Scientific Data and Databases

Corequisite(s):

Instructional Mode:

- [X] 100% face-to-face
- [ ] Hybrid: ≤ 50% electronically delivered
- [ ] 100% electronically delivered

Restrictions Enforced by System: Major, College, Degree, Program, etc. Include Code.

Are there equivalent course(s)?

- [ ] Yes
- [X] No

If yes, please list

Catalog Copy for NEW Courses Only (Consult University Catalog for models)

Description (No more than 60 words, use verb phrases and present tense)	**Notes** (List additional information for the course)
Discusses and explores advanced data management concepts such as big data, noSQL databases, cluster computing, linked data and Web data sources (APIs) by means of theoretical study, case studies and a project assignment	
Indicate number of contact hours: Hours of Lecture or Seminar per week: 2 Hours of Lab or Studio:	
When Offered: (check all that apply) [ ] Fall [ ] Summer [X] Spring	

Approval Signatures

Department Approval Date

College/School Approval Date

If this course includes subject matter currently dealt with by any other units, the originating department must circulate this proposal for review by those units and obtain the necessary signatures prior to submission. Failure to do so will delay action on this proposal.

Unit Name	Unit Approval Name	Unit Approver's Signature	Date

For Graduate Courses Only

Graduate Council Member

Provost Office

Graduate Council Approval Date

For Registrar Office's Use Only: Banner_____Catalog_____

revised 11/8/11

Course Proposal Submitted to the Curriculum Committee of the College of Science

1. COURSE NUMBER AND TITLE:

COS 401: Big Data Management

Course Prerequisites:

CDS 302: Scientific Data and Databases

Catalog Description:

Discusses and explores advanced data management concepts such as big data, noSQL databases, cluster computing, linked data and Web data sources (APIs) by means of theoretical and case studies.

2. COURSE JUSTIFICATION:

Course Objectives:

The goal of this course is to enable students to develop a good understanding of the principles and techniques with respect to new trends in data management specifically related to "big data", distributed computing and the Semantic Web. Considering for example "*big data as high volume (a lot of data), high velocity (streams) and/or high variety (different sources) information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization*."

To address this specific challenge, data management and processing has changed in recent years and many new technologies have been introduced. In this course, students will learn new principles and techniques for managing such complex datasets.

Course Necessity:

This course will introduce COS students big data management techniques.

Big data management is the organization, administration and governance of large volumes of both structured and unstructured data.

Managing big data is an essential skill for scientists needed to ensure a high level of data quality and accessibility for business intelligence and big data analytics applications.

As Corporations, government agencies and other organizations employ big data management strategies to help them contend with fast-growing pools of data, typically involving many terabytes or even petabytes of information saved in a variety of file formats, there is a great need to offer students this expertise.

Course Relationship to Existing Programs:

The course can be an elective in the Computation and Data Science minor.

Course Relationship to Existing Courses:

3. APPROVAL HISTORY:

4. SCHEDULING AND PROPOSED INSTRUCTORS:

Semester of Initial Offering:

Proposed Instructors: Dr. Dieter Pfoser

5. TENTATIVE SYLLABUS:

(attached)

College of Science

Exploratory Hall
4400 University Drive, MSN 5C3
Email: cos@gmu.edu Web: cos.gmu.edu

COS 401
Big Data Management

1. General Information

Instructor: Dr. Dieter Pfoser
When: tbd
Course website: Blackboard
Credits: 3.0
Prerequisites: None

Instructor's Office Hours: tbd with students

2. Course Objectives

Discusses and explores advanced data management concepts such as big data, noSQL databases, cluster computing, linked data and Web data sources (APIs) by means of theoretical study, case studies and a project assignment.

The goal of this course is to enable students to develop a good understanding of the principles and techniques with respect to new trends in data management specifically related to "big data", distributed computing and the Semantic Web. Considering for example "*big data as high volume (a lot of data), high velocity (streams) and/or high variety (different sources) information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization*."
To address this specific challenge, data management and processing has changed in recent years and many new technologies have been introduced. In this course, students will learn new principles and techniques for managing such complex datasets.

3. Course outline (tentative)

In this course we will cover the following topics (please note that the topics and their order are subjected to change at the discretion of the instructor, any changes will be announced in class):

Lec. #	Topic
1	Introduction and course overview – emerging trends and challenges
2	Traditional data management - relational databases
3	Traditional data management – object relational databases
4	noSQL databases overview
5	noSQL databases, MongoDB case study
6	Distributed data management – cloud computing
7	Map/Reduce concept
8	Hadoop case study
9	Introduction to Linked Data
10	Linked data case studies
11	Web data sources, APIs
12	Sourcing Web data, case studies
13	Course summary and conclusions
14	Project presentations