

**Revision Date February 1, 2023**

**Course Title: Big Data for Business Applications**

**Semester: Spring 2023**

**Course Prefix, Number: ISM6562**

**Credit Hours: 3**

<i>Instructor:</i>	Kaushik Dutta	<i>E-Mail:</i>	<a href="mailto:duttak@usf.edu">duttak@usf.edu</a>
<i>TA/GA:</i>	Varun Krishna Ramakrishnan	<i>E-Mail:</i>	varunkrishna@usf.edu
<i>Term:</i>	Spring 2023	<i>Dates:</i>	2/6-5/4/23.
<i>Delivery Method:</i>	Face to face in class (2-3 classes may be online based on pedagogical requirements)	<i>Location:</i>	BSN 230
<i>Office Hours:</i>	Wednesday 11 AM – 12.30 PM Please notify the instructor 12 hours before if you would attend the office hours		

#### **University Course Description:**

The course will cover data management and data analytics for businesses using various big data technologies such as No-SQL database, distributed file system, Map-Reduce, and Spark.

#### **Course Description:**

With the advent of social media and IoT (internet of things), the data volume in organizations has increased rapidly in recent years. With the increased in data volume to terabytes and even petabytes, traditional database and analytical techniques are not sufficient. In this course students will learn various big data technologies and how they can be used for data management and data analytics purposes to handle such massive dataset. The first half of the course will focus on big data storage technologies such as No-SQL database and distributed file system. The second half of the course will focus on big data computational platforms such as Hadoop map-reduce and Spark. The course will cover in-depth spark programming on big data platform.

#### **Course Objectives:**

- Identify various components of big data technologies;
- Explain how big Data technologies can be used;
- Design and develop analytics applications on big data.
- Apply the knowledge of big data technologies to design and scale analytics applications

**Learning Outcomes:**

Upon completion of this course students should be able to:

- Design and manage large business applications with millions of users
- Design analytic applications using several tens to hundreds of terabytes data
- Develop business applications running on top of several tens to hundreds of terabytes data

**Prerequisites:**

- Completion of Data Mining and Analytical Methods for Business/Statistical Programming/Statistical Data Mining course
- Relational Database
- Java or Python Programming knowledge
- Web application development
- Willingness to work hard with Big data technology stack

**Instructor Contact Information and Communication**

Canvas Email is generally the best way to get in touch with me for class-related things, and I will make every attempt to answer emails within 24 business hours. Please do not use @usf.edu for class-related communications.

Instructor: Kaushik Dutta

Office: School of Information Systems and Management (CIS Building)

Office Hours: Wednesday 11 AM – 12.30 PM

**First Day Attendance Policy**

A student who is not present on the first day of the class will automatically be dropped from the class.

**Attending the class Virtually or Physically**

Regular attendance in class is mandatory to do well in the course. There is no virtual attendance of the class. The class WILL NOT be streamed on any platform, and recordings of lectures will not be available. 2-3 classes may be delivered online. However, even on those days, the class will not be recorded through MS Teams. Suppose you miss 25% of the total class lectures for whatever reason (family emergency, religious reason, medical situation). In that case, it is very unlikely that you will be able to make up for that missed class. Please work with the instructor to do a late drop if it happens.

**Course Materials:**

Students do not need to purchase any software or book for this class. However, the instructor will direct students to various online materials, books, manuals, videos, and software to supplement the class lecture.

## Book

1. Frank Kane's Taming Big Data with Apache Spark and Python - eBook freely available from the library

1. **Frank Kane's Taming big data with Apache Spark and Python** : real-world examples to help you analyze large datasets with Apache Spark

Frank Kane.

**Author:** Kane, **Frank**

**Published:** Birmingham, UK : Packt Publishing, 2017.



eBook

Ebook Central (EBL) [USF ONLINE ACCESS](#) Click to View

USF ONLINE RESOURCE General Collection

add print email txt export



2. Ullman – Mining of Massive Datasets - <http://www.mmds.org/>

3. BOOK

PySpark Cookbook : Over 60 Recipes for Implementing Big Data Processing and Analytics Using Apache Spark and Python.

Drabas, Tomasz.; Lee, Denny.

Birmingham : Packt Publishing Ltd; 2018



BOOK

**PySpark Cookbook : Over 60 Recipes for Implementing Big Data Processing and Analytics Using Apache Spark and Python.**

Drabas, Tomasz.; Lee, Denny.

Birmingham : Packt Publishing Ltd; 2018

Available at USF ONLINE RESOURCE(SFPDA) General Collection 2 >

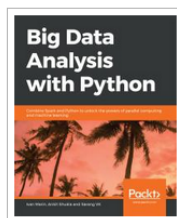
Available Online >

4. Big Data Analysis with Python: Combine Spark and Python to Unlock the Power of Parallel Computing and Machine Learning

by Ivan Marin, Ankit Shukla, and Sarang VK

Publisher

Packt Publishing, Limited



**Big Data Analysis with Python : Combine Spark and Python to Unlock the Powers of Parallel Computing and Machine Learning**

Ivan Marin, Ankit Shukla, and Sarang VK

Availability

Get all pages, require free third-party software, You have an active loan that expires in 0 day(s) and 23 hour(s).

Book Details

TITLE  
Big Data Analysis with Python

### How to Succeed in this course

- Access course on regular basis
- Read all mandatory reading materials
- Practice the examples and assignments
- Combine knowledge from other courses such as Data Mining, Analytical Methods for Business, Statistical Programming, Statistical Data Mining

### Grading Scale and Basis for Grades:

The grading will be based on the following components in the class –

Cassandra Quiz (Individual in-class)	10%
Assignments (Individual@Home)	10%
Group Project (Group@Home)	25%
Quizzes (Individual in-class only)	15%
Final Exam (Individual in-class only)	40%

The students will be given an “Incomplete” grade only as per the university policy without any exception.

The final grading will be based roughly on the following scale –

[95-100]	A/A+
[90-95)	A-
[85-90)	B+
[80-85)	B
[75-80)	B-
[70-75)	C+
[65-70)	C
[60-65)	C-
Below 60%	F

Please note that this is a tentative grading scale and can change based on the students' performance and the class's overall performance. However, the faculty reserves the right to change the grading scale as he deems appropriate based on the overall performance of the class.

### Course Outline and Schedule:

Week	Discussion Topic
1	Introduction to Big Data Application Big data Storage Big data computation
2	No-SQL Database Basics and concepts Various No-SQL database systems Cassandra
3	Cassandra, MongoDB and HBase
4	Distributed File Systems – HDFS, Cloudera Platform
6	Distributed Computing – Map-Reduce, Hadoop Map-Reduce, MongoDB Map-Reduce, Hive, Kudu, Impala
7	Data Robot Platform, Spark Platform
8	Spark basics programming – Resilient distributed data, Databricks
9	Data Cleaning, Data Visualization
9	Advanced spark programming
10	Spark SQL – Dataframe, Dataset, Spark Machine Learning Libraries
11	Catchup class
12	Final Exam – Spark Programming

### Group Project: Big data analytics project

The final project will be a group project, with each group comprising of 4 members. Each group is expected to take a publicly available dataset, and do the analysis, visualization, and model creation on the dataset using Jupyter Notebook and HDFS system on Cloudera platform or DataBricks platform.

### Standard University Policies

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at:

<https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx>

### COURSE POLICIES

**Late Work Policy:**

There are no make-up opportunities for exams. Topical assignments turned in late will be assessed a penalty of 20% for each for each day the assignment is late. Assignments will not be accepted if late by more than 5 days.

**Extra Credit Policy:**

There are no opportunities for extra credit in this course. Students' focus should be on the primary work in the course.

**Grades of "Incomplete":**

An "I" grade may be awarded to a student when 1) arrangements are made prior to the end of the semester, 2) in the judgment of the instructor a valid reason is offered for granting an Incomplete, 3) a clear path to a standard grade is agreed to by the instructor and the student which will result in successful completion of course requirements by the end of the succeeding semester. Offer specifics about your policy on incomplete grades. "I" grades not removed by the end of the next semester will be changed to "IF".

**Email:**

The primary means of communication between instructor and students will be Canvas email (not @usf.edu). "Blast emails" will occasionally be sent by the instructor to all students via Canvas. Students can feel free to email their instructor with questions at any time. Please anticipate a response time of 24 hours to email queries.

**Canvas:**

Canvas will be used in this course to disseminate materials turn in weekly assignments and return graded assignments. If you need help learning how to perform various tasks related to this course or other courses being offered in Canvas, please consult the Canvas help guides. You may also contact USF's IT department at (813) 974-1222 or [help@usf.edu](mailto:help@usf.edu).

**Classroom Recording:**

Audio and/or video recordings of lectures are allowable as per Florida law only for personal use. Such recording should not include any audio or video of other students in the class. Such recordings should not be publicly available and distributable as per the Florida law.

**Academic Integrity:**



# USFMuma

**College of Business**

School of Information Systems and Management

**UNIVERSITY of SOUTH FLORIDA**

Academic integrity is the foundation of the University of South Florida System's commitment to the academic honesty and personal integrity of its university community. Academic integrity is grounded in certain fundamental values, which include honesty, respect, and fairness. Broadly defined, academic honesty is the completion of all academic endeavors and claims of scholarly knowledge as representative of one's own efforts. The final decision on an academic integrity violation and related academic sanction at any USF System institution shall affect and be applied to the academic status of the student throughout the USF System, unless otherwise determined by the independently accredited institution.

### **Disruption to Academic Process:**

Disruptive students in the academic setting hinder the educational process. Disruption of the academic process is defined as the act, words, or general conduct of a student in a classroom or other academic environment which in the reasonable estimation of the instructor: (a) directs attention away from the academic matters at hand, such as noisy distractions, persistent, disrespectful or abusive interruption of lecture, exam, academic discussion, or general University operations, or (b) presents a danger to the health, safety, or well-being of self or other persons.

### **Student Academic Grievance Procedures:**

The purpose of these procedures is to provide all undergraduate and graduate students taking courses within the University of South Florida System an opportunity for objective review of facts and events pertinent to the cause of the academic grievance. An "academic grievance" is a claim that a specific academic decision or action that affects that student's academic record or status has violated published policies and procedures, or has been applied to the grievant in a manner different from that used for other students.

### **Disability Access:**

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at:

<https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx>

### **Sexual Misconduct/Sexual Harassment Reporting:**

USF is committed to providing an environment free from sex discrimination, including sexual harassment and sexual violence ([USF System Policy 0-004](#)). The USF Center for Victim Advocacy and Violence Prevention is a confidential resource where you can talk about incidents of sexual harassment and gender-based crimes including sexual assault, stalking, and



# USF Muma

**College of Business**

School of Information Systems and Management

**UNIVERSITY of SOUTH FLORIDA**

domestic/relationship violence. This confidential resource can help you without having to report your situation to either the Office of Student Rights and Responsibilities (OSSR) or the Office of Diversity, Inclusion, and Equal Opportunity (DIEO), unless you request that they make a report. Please be aware that in compliance with Title IX and under the USF System Policy, educators must report incidents of sexual harassment and gender-based crimes including sexual assault, stalking, and domestic/relationship violence. If you disclose any of these situations in class, in papers, or to me personally, I am required to report it to OSSR or DIEO for investigation. Contact the USF Center for Victim Advocacy and Violence Prevention: (813) 974-5757.

### **Religious Observances:**

All students have a right to expect that the University will reasonably accommodate their religious observances, practices, and beliefs. If you observe religious holidays, you should plan your allowed absences to include those dates. However, at the end it is student's responsibility to arrange for any make-up.

### **Incomplete Grade Policy:**

An Incomplete grade ("I") is exceptional and granted at the instructor's discretion only when students are unable to complete course requirements due to illness or other circumstances beyond their control. The course instructor and student must complete and sign the "I" Grade Contract Form that describes the work to be completed, the date it is due, and the grade the student would earn factoring in a zero for all incomplete assignments. The due date can be negotiated and extended by student/instructor as long as it does not exceed two semesters for undergraduate courses and one semester for graduate courses from the original date grades were due for that course. An "I" grade not cleared within the two semesters for undergraduate courses and one semester for graduate courses (including summer semester) will revert to the grade noted on the contract.

### **Online Proctoring**

Quizzes and examinations will be conducted using online proctoring tools. Keeping the audio and video (microphone and camera) on during such exams and quizzes is a must. If the student is not willing to use these the student is asked not to register for this course. Any student may elect to drop or withdraw from this course before the end of the drop/add period. Online exams and quizzes within this course may require online proctoring. Therefore, students will be required to have a webcam (USB or internal) with a microphone when taking an exam or quiz. Students understand that this remote recording device is purchased and controlled by the student and that recordings from any private residence must be done with the permission of any person residing in the residence. To avoid any concerns in this regard, students should



select private spaces for the testing. The University library and other academic sites at the University offer secure private settings for recordings and students with concerns may discuss location of an appropriate space for the recordings with their instructor or advisor. Students must ensure that any recordings do not invade any third-party privacy rights and accept all responsibility and liability for violations of any third-party privacy concerns. Setup information will be provided prior to taking the proctored exam. For additional information about online proctoring, you can visit the online proctoring student FAQ at <https://www.usf.edu/innovative-education/student-resources/index.aspx>

### **Turnitin.com**

In this course, turnitin.com may be utilized. Turnitin is an automated system which instructors may use to quickly and easily compare each student's assignment with billions of web sites, as well as an enormous database of student papers that grows with each submission. Accordingly, you will be expected to submit all assignments in both hard copy and electronic format. After the assignment is processed, as instructor I receive a report from turnitin.com that states if and how another author's work was used in the assignment. For a more detailed look at this process visit <http://www.turnitin.com>. Essays are due at turnitin.com the same day as in class.

### **Netiquette Guidelines:**

1. Act professionally in the way you communicate. Treat your instructors and peers with respect, the same way you would do in a face-to-face environment. Respect other people's ideas and be constructive when explaining your views about points you may not agree with.
2. Be sensitive. Be respectful and sensitive when sharing your ideas and opinions. There will be people in your class with different linguistic backgrounds, political and religious beliefs or other general differences.
3. Proofread and check spelling. Doing this before sending an email or posting a thread on a discussion board will allow you to make sure your message is clear and thoughtful. Avoid the use of all capital letters, it can be perceived as if you are shouting, and it is more difficult to read.
4. Keep your communications focused and stay on topic. Complete your ideas before changing the subject. By keeping the message on focus you allow the readers to easily get your idea or answers they are looking for.
5. Be clear with your message. Avoid using humor or sarcasm. Since people can't see your expressions or hear your tone of voice, meaning can be misinterpreted.

### **Email and Discussion Board Guidelines:**

1. Use the subject line effectively by using a meaningful line of what your email or discussion is about.

2. Keep your emails and postings related to the course content. You should not post anything personal on a discussion board, unless is requested by the instructor.
3. Any personal, course or confidential issues should be directly communicated to the instructor via email. The discussion boards are public spaces; therefore, any personal issues should not be posted there.
4. Posts and discussions unrelated to class content are prohibited. Warning will be given for such posting. Repeat offenders will lose the access to the online class materials and thus risk failing the class.

#### **End of Semester Student Evaluations:**

All classes at USF make use of an online system for students to provide feedback to the University regarding the course. These surveys will be made available at the end of the semester, and the University will notify you by email when the response window opens. Your participation is highly encouraged and valued.

#### **WhatsApp, GroupMe, and Student-to-Student Communication:**

While students may use digital communication tools (WhatsApp, GroupMe, etc.) to communicate with fellow students, it is important to remember that academic integrity policies still apply in these environments. Informing others about the contents of tests is prohibited by [the official regulation](#), as is receiving unauthorized information about an examination. Students are expected and required to immediately report instances of such violations to the instructor.

#### **Laptop:**

You should have your own laptop with at least 8 GB RAM to do the work needed in the class.

#### **Title IX Policy:**

Title IX provides federal protections for discrimination based on sex, which includes discrimination based on pregnancy, sexual harassment, and interpersonal violence. In an effort to provide support and equal access, **USF has designated all faculty (TA, Adjunct, etc.) as Responsible Employees, who are required to report any disclosures of sexual harassment, sexual violence, relationship violence or stalking.** The Title IX Office makes every effort, when safe to do so, to reach out and provide resources and accommodations, and to discuss possible options for resolution. Anyone wishing to make a Title IX report or seeking accommodations may do so online, in person, via phone, or email to the Title IX Office. For information about Title IX or for a full list of resources please visit: <https://www.usf.edu/title-ix/gethelp/resources.aspx>. *If you are unsure what to do, please contact Victim Advocacy – a confidential resource that can review all your options – at 813-974-5756 or [va@admin.usf.edu](mailto:va@admin.usf.edu).*

#### **Course Hero / Chegg Policy:**

The [USF Policy on Academic Integrity](#) specifies that students may not use websites that enable cheating, such as by uploading or downloading material for this purpose. This does apply specifically to Chegg.com and CourseHero.com – almost any use of these websites (including uploading proprietary materials) constitutes a violation of the academic integrity policy.

### **Academic Accommodations**

Students with disabilities are responsible for registering with Student Accessibility Services (SAS) in order to receive academic accommodations. For additional information about academic accommodations and resources, you can visit the SAS website.

[SAS website for the Tampa and Sarasota-Manatee campuses.](#)

[SAS website for the St. Pete campus.](#)

### **Academic Support Services**

The USF Office of Student Success coordinates and promotes university-wide efforts to enhance undergraduate and graduate student success. For a comprehensive list of academic support services available to all USF students, please visit the [Office of Student Success website](#).

### **Canvas Technical Support**

*If you have technical difficulties in Canvas, you can find access to the Canvas guides and video resources in the “Canvas Help” page on the homepage of your Canvas course. You can also contact the help desk by calling 813-974-1222 in Tampa or emailing [help@usf.edu](mailto:help@usf.edu).*

[IT website for the Tampa campus.](#)

[IT website for the St. Pete campus.](#)

[IT website for the Sarasota-Manatee campus.](#)

### **Center for Victim Advocacy**

The [Center for Victim Advocacy](#) empowers survivors of crime, violence, or abuse by promoting the restoration of decision making, by advocating for their rights, and by offering support and resources. Contact information is available online.

### **Counseling Center**

The Counseling Center promotes the wellbeing of the campus community by providing culturally sensitive counseling, consultation, prevention, and training that

enhances student academic and personal success. Contact information is available online.

[Counseling Center website for the Tampa campus.](#)

[Counseling Center website for the St. Pete campus.](#)

[Counseling Center website for the Sarasota-Manatee campus.](#)

### **Campus Free Expression:**

*It is fundamental to the University of South Florida's mission to support an environment where divergent ideas, theories, and philosophies can be openly exchanged and critically evaluated. Consistent with these principles, this course may involve discussion of ideas that you find uncomfortable, disagreeable, or even offensive.*

*In the instructional setting, ideas are intended to be presented in an objective manner and not as an endorsement of what you should personally believe. Objective means that the idea(s) presented can be tested by critical peer review and rigorous debate, and that the idea(s) is supported by credible research.*

*Not all ideas can be supported by objective methods or criteria. Regardless, you may decide that certain ideas are worthy of your personal belief. In this course, however, you may be asked to engage with complex ideas and to demonstrate an understanding of the ideas. Understanding an idea does not mean that you are required to believe it or agree with it.*

### **USF Core Syllabus Policies**

USF has a set of central policies related to student recording class sessions, academic integrity and grievances, student accessibility services, academic disruption, religious observances, academic continuity, food insecurity, and sexual harassment that **apply to all courses at USF**. Be sure to review these online at: <https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx> Any USF policy will supersede any policy mentioned in this syllabus.