Course Objectives: An overview of the principles of data integration, the fundamental basis for developing useful and flexible business intelligence platforms. Modern data integration needs differ from traditional approaches in four main dimensions that parallel differences between big data and traditional data: volume, velocity, variety, and veracity.

Learning Outcomes: Students will need to demonstrate an understanding of how to integrate data using relational, dimensional, and key-value (big data) data models.

Prerequisites: JAVA, SQL, and database design experience. Graduate student in the Lindner College of Business or permission from the instructor.


Managing Data in Motion: Data Integration Best Practice Techniques and Technologies by April Reeve, Elsevier, 2013.

Course Details:

- Students are expected to come prepared for every class period by reading the assigned reading material ahead of time.
- Class lectures will focus on highlighting the textbook topics with additional insight using supplemental material.
- The course involves hands-on development activities using Oracle, Cloudera, and Hadoop.
- Class participation (and attendance) is critical to learning - absence for a class does not relieve a student from responsibility for the learning the subject matter or completing the assignments. Due dates will remain as listed for all assignments unless otherwise noted by the instructor.

Grading Details:

- Students are expected to take the tests as scheduled.
- Make-up tests will not be given unless requests are made early in the semester/flex and under only extraordinary circumstances.
- Assignments will be due on the date/time specified. Late submissions will not be accepted.
- All assignments are due at the start of the class session on the due date and should be turned in directly to the instructor.
- All assignments should contain a cover page with the students’ names and the assignment title clearly labeled.
A | 93-100 | Excellent | 4.0000
A-| 90-92.9 |         | 3.6667
B+| 87-89.9 |         | 3.3333
B | 83-86.9 | Good    | 3.0000
B-| 80-82.9 |         | 2.6667
C+| 77-79.9 |         | 2.3333
C | 70-76.9 |         | 2.0000
F | 0-69.9  | Fail    | 0.0000
P |         | Pass    | N/A
U |         | Unsatisfactory | N/A
T |         | Audit   | N/A
I |         | Incomplete | 0.0000
I/F|         | Failure  | 0.0000
W |         | Withdrawal (Official) | N/A

**Grading:**

This course uses graduate-level grading and is worth 2.0 graduate-level credits.

---

**Academic Integrity:**

As with all Lindner College of Business efforts, this course will uphold the highest ethical standards, critical to building character. Ensuring your integrity is vital and your responsibility. LCB instructors are required to report ANY incident of academic misconduct (e.g., cheating, plagiarism) to the college review process, which could result in severe consequences, including potential dismissal from the college.

All academic programs at the Lindner College of Business will apply a “Two Strikes Policy” regarding Academic Integrity. Any student who has been found responsible for two cases of academic misconduct may be dismissed from the College.

The “Two Strikes Policy” supplements the UC Student Code of Conduct (see: [http://www.uc.edu/conduct/Code_of_Conduct.html](http://www.uc.edu/conduct/Code_of_Conduct.html)). All cases of academic misconduct (e.g., cheating, plagiarism, falsification) will be formally reported by faculty. Students will be afforded due process for allegations, as outlined in the policy. If a student is found guilty of academic misconduct in two instances, the student may be dismissed from the Lindner College of Business.

**Inclement Weather and Class Cancelations:**

I will cancel classes due to weather only if the University cancels all classes. Information concerning weather closings can be obtained through the University’s website. Please see the inclement weather policy located at ([http://www.uc.edu/news/NR.aspx?id=12766](http://www.uc.edu/news/NR.aspx?id=12766)) for further information.
Title IX:
Title IX is a federal civil rights law that prohibits discrimination on the basis of your actual or perceived sex, gender, gender identity, gender expression, or sexual orientation. Title IX also covers sexual violence, dating or domestic violence, and stalking. If you disclose a Title IX issue to me, I am required forward that information to the Title IX Office. They will follow up with you about how the University can take steps to address the impact on you and the community and make you aware of your rights and resources. Their priority is to make sure you are safe and successful here. You are not required to talk with the Title IX Office. If you would like to make a report of sex or gender-based discrimination, harassment or violence, or if you would like to know more about your rights and resources on campus, you can consult the website www.uc.edu/titleix or contact the office at 556-3349.

Counseling Services:
Students have access to counseling and mental health care through the University Health Services (UHS), which can provide both psychotherapy and psychiatric services. In addition, Counseling and Psychological Services (CAPS) can provide professional counseling upon request; students may receive five free counseling sessions through CAPS without insurance. Students are encouraged to seek assistance for anxiety, depression, trauma/assault, adjustment to college life, interpersonal/relational difficulty, sexuality, family conflict, grief and loss, disordered eating and body image, alcohol and substance abuse, anger management, identity development and issues related to diversity, concerns associated with sexual orientation and spirituality concerns, as well as any other issue of concerns. After hours, students may call UHS at 513-556-2564 or CAPS Cares at 513-556-0648. For urgent physician consultation after-hours students may call 513-584-7777.

Project:
During the course, students will be put into teams of 3. Each team will be given access to a Hadoop cluster and/or virtual machine. In this project, teams will acquire and integrate new data into existing sales data to develop an insightful analysis and business sales strategy. Your goal is to present a thoughtful proposition, analysis, and recommendation to the executive sales team for how to improve profitability for a web-based sales and delivery company.

For the project, the teams will be required to collect one outside source of data (derived from a web-based source) to integrate into the sales data used in the class assignments. The outside data source being integrated into the sales data must contain more than 100 tuples and more than 10 attributes when combined together (when represented in a dimensional format). The data must be used to develop a business-solution that informs the team and improves their sales strategy with insights that are not available in the original sales data alone.

Your team must hand in a 10-page (double-spaced) analysis that supports your logic and sales plan, and will present your plan to your class the following week. Your team will be graded on the business potential, novelty, and practicality of your solution. You work should be of professional quality and contain no spelling, grammar, or logic errors. Team members will be awarded points for the project based on their contributions. More details about the project can be found on the project statement worksheet. Additional grading details are provided on the rubric and team member contribution form.
### Course Schedule:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percent of Grade</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1</td>
<td>25%</td>
<td>Sat 1/27/18</td>
</tr>
<tr>
<td>Exam 2</td>
<td>25%</td>
<td>Sat 2/17/18</td>
</tr>
<tr>
<td>Assignments (4 total)</td>
<td>20%</td>
<td>About 1 per week</td>
</tr>
<tr>
<td>1 Page Exec Summary</td>
<td>5%</td>
<td>Sat 2/3/18</td>
</tr>
<tr>
<td>Project</td>
<td>25%</td>
<td>Sat 2/17/18</td>
</tr>
</tbody>
</table>

*Quiz and Test dates may be subject to change

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Topic</th>
<th>Reading</th>
<th>Assignment Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/13 Sat</td>
<td>1</td>
<td>Data Integration and Big Data Overview</td>
<td>DIM 1-3, 18,19,21,22</td>
<td></td>
</tr>
<tr>
<td>1/20 Sat</td>
<td>2</td>
<td>Dimensional, Relational, and Key-Value Models</td>
<td>BDI 2, 3</td>
<td>A1: Aligning Different Data Schemas and Sources</td>
</tr>
<tr>
<td>1/27 Sat</td>
<td>3</td>
<td>Schema Alignment and Record Linkage</td>
<td>CAP 7, BDI 3</td>
<td>A2: Hadoop Setup/Using HDFS Exam 1 (Theory of Data Integration)</td>
</tr>
<tr>
<td>2/3 Sat</td>
<td>4</td>
<td>Introduction to Hadoop and Hive</td>
<td>CAP 1</td>
<td>A3: Executing a MapReduce Job 1 Page Executive Project Summary</td>
</tr>
<tr>
<td>2/10 Sat</td>
<td>5</td>
<td>Data Clusters and Map Reduce</td>
<td>CAP 2, 11</td>
<td>A4: Managing Data with Hive</td>
</tr>
<tr>
<td>2/17 Sat</td>
<td>6</td>
<td>Acquiring and Integrating Data with Hive</td>
<td>CAP 3, 7</td>
<td>10-Page Analysis Due Exam 2 (Using Hadoop and Big Data Solutions)</td>
</tr>
<tr>
<td>2/24 Sat</td>
<td>7</td>
<td>30-Minute Project Presentations</td>
<td></td>
<td>Project Presentations</td>
</tr>
</tbody>
</table>

*Assignment dates may be subject to change