

Name: Eric Woyant, Sarah Barlow, Braden Litastro

Grade: A-

Course: Computer Science 380 Spring 2013

## Assignment: Laboratory #1 - Group presentation

### Presentation Assignment Evaluation

1. Did the student(s) adhere to the honor code?  Yes  No  Not Clear
2. Was the presentation delivered at the correct date and time?  Yes  No
3. Was the presentation properly organized?  Yes  No
4. Was the presentation's organization adequately communicated?  Yes  No
5. Did the presentation accurately describe the assignment?  Yes  No
6. Does the presentation contain all of the necessary components?  Yes  No  
If no, then what components are missing?

- Introduction and motivation
- Description of technique
- Technical diagrams
- Program screenshots
- Mathematical notation
- Tables of data
- Data visualizations
- Results analysis
- Related work
- Conclusion
- Future work
- Other components of a presentation

Some issues  
could be clarified  
with technical  
diagrams &  
screenshots.

7. Does the presentation contain spelling mistakes?  Yes  No
8. Does the presentation have errors in grammar?  Yes  No
9. Does the presentation exhibit poor formatting?  Yes  No
10. Did the presenter(s) evidence a strong knowledge of the topic?  Yes  No
11. Did the presenter(s) convey an enthusiastic attitude?  Yes  No
12. Did the presenter(s) regularly make eye contact?  Yes  No
13. Did the presenter(s) adequately answer questions?  Yes  No
14. Did the presenter(s) adhere to the established time limits?  Yes  No
15. Overall, how well did the presenter(s) enhance your understanding of the laboratory and enable you to develop new ideas about data management?  
 Exceptional  Good  Average  Poor

### Additional Comments

Used the IDC estimates as part of this study  
Does make a distinction between production & storage.  
USC study in 2008 - interesting. Connected to the  
positions needed to manage the data - good idea.  
Overall, good format for the slides. Mentioned the  
importance of cloud storage. Drupal data management tools.  
At the middle, did give a definition of data - earlier?  
Focus on efficiency - compression. SQL & XML.  
More about computer security.

Name: Braden Licastro, Eric Weyant, Sarah Barlow Grade: A  
Course: Computer Science 380, Principles of Database Systems, Spring 2013

## Laboratory Assignment #1

### Laboratory Assignment Evaluation

1. Did the student(s) adhere to the honor code?  Yes  No  Not Clear
2. Was the assignment turned in on time?  Yes  One Week Late  Very Late
3. Was the laboratory notebook used as the cover sheet?  Yes  No
4. Does the assignment contain all of the required deliverables?  Yes  No  
If no, then what deliverables are missing?

- Source code
- System output
- Documentation
- Build system
- Version control repository
- Tables of data
- Data visualizations
- Results analysis
- Description of algorithm or technique
- Technical diagrams
- Program screenshots
- Equations and/or other mathematical notation
- Reflective discussion of the encountered challenges
- Other types of deliverables

✓ The format of  
the assignment  
is very professional.

5. Are there mistakes in the specification, design, or testing plan?  Yes  No
6. Are there errors in the implementation of the program or test suite?  Yes  No
7. Does the writing contain spelling mistakes?  Yes  No
8. Does the writing have errors in grammar?  Yes  No
9. Does the writing have incorrect punctuation?  Yes  No
10. Does the writing exhibit factual errors?  Yes  No
11. Does the assignment exhibit poor formatting?  Yes  No
12. If asked to work in a group, did the students operate effectively as a team?  
 Yes  No  Maybe  Not Applicable
13. Did the student(s) ask the instructor for assistance with the assignment?  Yes  No
14. Should the student(s) meet with the instructor to discuss this assignment?  
 Yes  No  Maybe

### Additional Comments

You have turned in a good laboratory assignment -  
the tutorial about installing Drupal is quite nice.  
Please see me if you have any questions or comments.

LAB 1

# Investigating Data Management Tools and Technologies

---

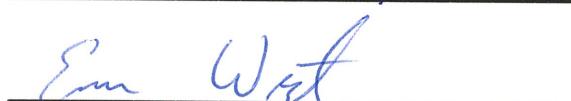
*CMPSC 380 Professor Kapfhammer*

**Braden Licastro**

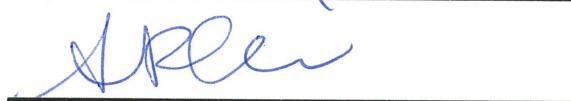


✓ very  
good  
cover  
sheet.

**Eric Weyant**



**Sarah Barlow**



Every second 2.9 million emails are sent, every minute twenty hours of video are uploaded to YouTube, 50 million tweets are tweeted each day and every month a total of 700 billion hours are spent on Facebook (World of Data) Copious amounts of data are produced daily. As of 2011, the total amount of data in the digital universe was approximately 1.8 zettabytes according to the International Data Corporation, with ninety percent of the world's total data being created within the last two years. *Forbes* Magazine estimated that in 2012, 2.7 zettabytes of data existed. This is enough digital data to fill a stack of DVDs stretching to the moon and back or enough fully loaded 32GB iPads to build a twenty foot wall around South America. So what is all of this data and where is it coming from? Generated data is not limited in form. Traditional conceptions of data as numerical and textual have been eclipsed by images, audio files, and video. In 2012, 68% of the data produced was generated by consumers streaming digital television, interacting with social media and sending or receiving photographs and video with digital devices (Merian) We glean further data from GPS satellites, medical equipment and public transportation logs to name a few. But the amount of data that we produce is further compounded by the generation of metadata, or the data about data. Metadata is the record of the who, what, where and when of data production. Every time the creation and transmission of data takes place, for example the sending a text message, data is collected; the number of characters in the text, the text itself, the sender, the receiver, the date and time it was sent, etc., creating a seemingly infinite amount of data. An EMC-sponsored data study found that the amount of data produced by people writing emails, taking photos and downloading music and movies is minuscule compared to the amount of data generated about them (Merian) Pinpointing an exact amount of existing data is a monumentally difficult task for this reason.

After we create all of this data, the question remains, where exactly do we store it all? Not all data that we produce will be stored, the metadata resulting from streaming a video from YouTube or watching a movie over cable or satellite does not necessarily get saved. A study conducted by research scientists at the University of Southern California determined that between 1986 and 2007, 295 exabytes of data were stored worldwide, enough data to cover the entirety of China in thirteen layers of books containing that data information (Hilbert and Lopez) The study examined sixty categories of ICT, information and communication technologies, (twenty-one analog and thirty-nine digital) over a period of two decades, the most comprehensive study of data storage conducted to that time. To determine total storage capacity, researchers multiplied the number of installed technological devices with their respective performances (Hilbert and Lopez) If we follow the storage capacity formula, which states storage capacity doubles every three years, for the year 2012, 600 exabytes of data is currently being stored (Kar) This enormous amount of information is being saved on hard drives, optical drives, digital tapes, and other forms too numerous to mention.

It is interesting to note, however, despite storing all of this data, less than 0.5% of data is currently being analyzed. It has been estimated that up to a quarter of this data has the potential to "contain information that might be valuable if analyzed." (Purcell) The problem stems from the lack of "tagging," as it is estimated that only 3% of all useful information has been tagged. Tagging provides context for the data and is essentially metadata. The large amount of data we are accumulating that goes unanalyzed or unused has caused many analysts to wonder if we are in fact becoming data "hoarders." The practice of saving data until we are forced to get rid of it is at the center of this controversy. This

could be early versions of documents and files supporting a work process that aren't needed after a project's completion, an old CS 111 lab assignment that went horribly wrong or that paper you wrote in the fifth grade. Removing useless data, for a large company could result in the saving of thousands of dollars and more space to save more data.

Projections for 2013, predict that Internet data production alone will double by the end of this year. The IDC has estimated that within the next decade, the amount of data that we produce will increase fifty times. A tracker of the amount of data produced since January 2011 is currently available on the IDC website. By 2020, there will be 5,200 GB of data for every person on Earth. Additionally, metadata is growing twice as fast as the digital universe as a whole utilizing as much as a petabyte of information for a gigabyte of stored information. Complicating the matter is the notion that the number of IT professionals to manage all of the data that we produce is estimated to only increase by 1.5 times.

## Works Cited

Hilbert, Martin and Priscilla Lopez, "The World's Technological Capacity to Store, Communicate, and Compute Information," *Science* 332, no. 6025 1 April 2011: pp. 60-65.

Kar, Saroj, "Where's the World's Data Being Stored and Who Owns It?," 21 July 2012 :  
<http://siliconangle.com/blog/2011/07/21/wheres-the-worlds-data-being-stored-and-who-owns-it/>.

Merian, Lucas, "World's data will grow by 50X in the next decade, IDC study predicts," 28 June 2011:  
[http://www.computerworld.com/s/article/9217988/World\\_s\\_data\\_will\\_grow\\_by\\_50X\\_in\\_next\\_decade\\_IDC\\_study\\_predict](http://www.computerworld.com/s/article/9217988/World_s_data_will_grow_by_50X_in_next_decade_IDC_study_predict).

Merian, Lucas, "By 2020 there will be 5200 GB of data for every person on Earth," 11 Dec 2012:  
[http://www.computerworld.com/s/article/9234563/By\\_2020\\_there\\_will\\_be\\_5\\_200\\_GB\\_of\\_data\\_for\\_every\\_person\\_on\\_Earth?taxonomyId=19&pageNumber=2](http://www.computerworld.com/s/article/9234563/By_2020_there_will_be_5_200_GB_of_data_for_every_person_on_Earth?taxonomyId=19&pageNumber=2).

Purcell, Andrew, "Less than 1% of global data analyzed," 9 Jan 2013:  
<http://www.isgtw.org/spotlight/less-1-global-data-analyzed>.

Savitz, Eric, "Big Data: The Hidden Opportunity," *Forbes Magazine*, 1 May 2012:  
<http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/>

"The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011," March 2008, IDC.

"The World of Data We're Creating on the Internet," 11 Oct 2010: <http://www.good.is>. 

Great - F  
am glad to  
see that you  
furnished a  
references page

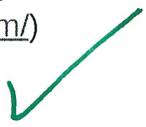
2) A listing of the tools or features that you would like to see in a data management tool. If the desired tools already exist, then please reference their Web site addresses and clearly explain what the tools do. If the tools and/or features do not already exist, then you should clearly describe the functionality that you are requesting.

### Tools/Features of Data Management Tools

The tool/ features of data management tools have been broken down into two groups:

- features that belong in nearly every data management tool or features that are very popular. For example, the following features can all be found in SQLite or phpMyAdmin, which can be downloaded at <http://www.sqlite.org/download.html> and [http://www.phpmyadmin.net/home\\_page/downloads.php](http://www.phpmyadmin.net/home_page/downloads.php), respectively.
  1. Store/retrieve consistent data – One of the main purposes of a data management tool is the ability to store and retrieve data by using a specific search or query. Without this feature data storage would be useless.
  2. Modify/delete - More main features of a data management tool is the ability to modify and even delete data once you store it into the database. This simplifies the process of retrieving data, modifying it, and then storing it.
  3. Analyze data(patterns) - After storing and eventually modifying data, a data management tool needs to be able to allow the user to analyze this data which can be done by showing patterns.
  4. Performance/efficiency - Obviously the performance of a data management tool is very important to the user and data management systems can employ different techniques in order to improve efficiency. For example, a technique known as clustering can be used in order to “cluster” or group data that is related closer in memory in order to improve the time in sorting.
  5. Atomicity - An “all or nothing” rule that states that if any part of a transfer fails then the entire transfer fails. This is important because if this is not the cause data could be lost if a transfer failed. For example, if an individual was transferring money from their savings account to their checking account and the transfer failed halfway, without atomicity, that money could just become corrupted and be gone forever.



6. User friendly/simplicity - User friendliness is important because if a database is difficult to use then its performance is irrelevant as more time will be wasted in trying to figure out the program. A user friendly system will also be chosen over one that is not.
  7. Sorting - Data can be nearly useless if it is not sorted. In order to see how data compares to other data sorting is a necessity in a data management tool. For instance, if managing a group of people's birth years, it would be useful to sort these into either descending or ascending order.
  8. Data abstraction - In order to simplify use for different types of users, abstraction creates a layer of security by preventing some users from accessing certain areas of data. A regular user should not have access to the same files as a programmer or a database manager.
  9. Safety/security (encryption) - Data management tools should provide safety and security to the user. When using a data management tool the user should never feel that their data is at risk.
  10. Sharing - When working with a database, users often work in groups and therefore it is necessary to be able to share the work with multiple users easily and quickly.
  11. Reliability - Reliability is important because a data management system cannot be used if the data is not correct. So even if it can give an answer to a query immediately, it is useless to the user if it is not the correct answer.
- tools that already exist and can be seen in a present data management tool. For example, cloud storage is a feature that exists in numerous data management tools, however, there are many databases that do not possess this feature.
  - 12. Scalability - This is important because a database is needed to be able to scale itself so that if given a larger task it is able to complete it or will be able to complete it if the system is enlarged itself. For example, if a database system cannot handle a growing amount of work, but it is upgraded and given a faster CPU and more RAM, it should be able to make use of this and become more efficient and handle the growing work. This feature can be found in Twitter (found at <https://twitter.com/>) which uses its own framework ontop of MySQL in order to scale its

databases to handle the vast amounts of incoming tweets daily.

13. Portability/compatibility - If the database can be ported to a different environment (such as another operating system) or it is compatible to another environment, then it is much more useful as the user is able to manage their data from different computing environments. This feature can be found in Firefox (which can be found at <http://www.mozilla.org/en-US/>) which is available on nearly every operating system available.
14. Crash tolerance - As with any computer tool or software it is desirable to not have to deal with crashes. However, having a crash free system is nearly impossible and the next best thing is having a good crash tolerance. Or in other words, the ability to recover from an unexpected crash without any data loss or other problems. This feature can be found in Thunderbird (which can be found at <https://www.mozilla.org/en-US/thunderbird/>) as it autosaves your emails so that if it were to crash you will not lose the current state of your email when restarting the service.
15. Visualization - This roughly means representing data in visual way that allows it to be communicated easily and effectively. For example, placing data into a chart can dramatically improve the way that data is presented. This feature can be found in LibreOffice Calc (which can be found at <http://www.libreoffice.org/>) as it allows the user to select data and convey it into a table or graph very easily.
16. Redundancy - A major issue with data management is data redundancy. Having multiple copies of files is redundant and can take up storage for no reason. Thus, a data management tool should have the ability to take care of redundant files and determine whether it necessary to have two copies of the same file. This feature can be found in Duplicate Cleaner (which can be found at <http://www.digitalvolcano.co.uk/content/duplicate-cleaner>) as it scans a system and checks for duplicate files and deletes them.
17. Compression - Databases can be very large and storage consuming and compressing them can be very effective in order to save space. Compression is also useful for other features such as sharing. This feature can be found in SQLite (which can be found at <http://www.sqlite.org/download.html>). As the name suggests, this is a

lightweight piece of software compared to a tool that is large and take up more storage space.

18. Isolation - Isolation, or multiple users accessing same file simultaneously is important for group purposes like sharing. For example, this lab is being done using Google docs so that our entire group can access and edit each part of the lab. As just mentioned, this feature can be found in Google Docs (which can be found at [docs.google.com/](https://docs.google.com/)).
19. Computation - Computation is important as it can save vast amounts of time as data is often needed to be computed to be useful to the user. For instance, data such as a company's revenue needs to be computed with math operations in order for it to be useful to the company. This feature can be found in Microsoft Excel (which can be found at <http://office.microsoft.com/en-us/excel/>) as users are able to select amounts of data in various "cells" and perform mathematical operations.
20. Change/access logging - In order to keep track of changes made to a database a log is needed to be kept. This is important in order to know who is making changes or accessing the database. For example, if the data system crashes after changes are made the log will show what was being accessed at the time of the crash. This feature can be found in FileZilla (which can be found at <http://filezilla-project.org/>) as it provides the user with logging to file.
21. Cloud storage - Being able to store data into the cloud will eliminate the problems of limited storage. Cloud storage is available from Dropbox (which can be found at <https://www.dropbox.com/>) as users have the ability to store their files into the cloud and access them from any location.
22. Open source - This feature is not necessarily necessary, however, it is a feature that we would want in a data management system because it allows the user to have full control of the data management software. An example of an open source data management system is Drupal (which can be found at <http://drupal.org/>) as it is maintained and developed by a community of 630,000+ users and developers.

**File Management and Access****1. Firefox:**

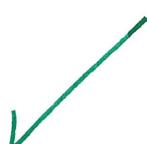
Web Address:	<a href="http://www.mozilla.org/en-US/firefox/fx/">http://www.mozilla.org/en-US/firefox/fx/</a>
Features:	Auto complete search and URL recommendations, simplified user interface, app tabs, new tab thumbnails, switch to tab, Firefox Sync, session restore, spell check, bookmark tagging, download manager, RSS reader, offline browsing, speed boost, crash protection, do not track, private browsing, anti-virus integration, add-ons, integrated web developer tools, password manager, auto complete forms, pop-up blocker, integrated search with suggestions, HTML5/CSS3 support, wide range of font types supported
Data Managed:	Cookies, Usernames, Passwords, Form Data, History, Bookmarks, Web Content, Add-ons

**2. Thunderbird:**

Web Address:	<a href="https://www.mozilla.org/en-US/thunderbird/">https://www.mozilla.org/en-US/thunderbird/</a>
Features:	Tabbed email navigation, Easy to use address book, integrated chat software, inline web search, advanced search tools, message archives, activity tracking, combined email folders, do not track, junk mail filter, phishing protection
Data Managed:	Passwords, Usernames, Emails, Attachments, Large Files, Chat Accounts, Archives, Searches

**3. FileZilla:**

Web Address:	<a href="http://filezilla-project.org/">http://filezilla-project.org/</a>
Features:	FTP, FTPS and SFTP support, IPv6 support, large file transfer support, bookmarks, search filters, remote file editing, remote search, transfer speed limits, cross platform
Data Managed:	Passwords, Account information, bookmarks, ftp and local data



Category	Definition	Example
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100



### Image Management

#### 8. Gallery CMS:

Web Address:	<a href="http://www.gallerycms.com/">http://www.gallerycms.com/</a>
Features:	Simple installation, simplistic interface, supports XML/JSON feeds, upload multiple images, drag and drop support, user and feed management
Data Managed:	Text, images, video, user accounts, passwords, data feeds

### Image Manipulation

#### 9. GNU IMage MAnipulation Program:

Web Address:	<a href="http://www.gimp.org/">http://www.gimp.org/</a>
Features:	Customizable interface, photo enhancement, digital retouching, hardware support, wide array of file formats, extendable
Data Managed:	Plug-ins, Images

### Audio Manipulation

#### 10. Audacity:

Web Address:	<a href="http://audacity.sourceforge.net/">http://audacity.sourceforge.net/</a>
Features:	Cross platform, multiple input/output, timed recording, high resolution recording, device toolbar, simultaneous multi channel recording, easy to read level meters import/export wide range of file types, basic to advanced editing features, automatic crash recovery, advanced analysis tools, and downloadable effects
Data Managed:	Audio, Plug-ins and Effects

## I. Table of Contents

1 Introduction.....	2
1.1 Purpose .....	2
1.2 Scope .....	2
1.3 Server Requirements .....	2
2 Installation Manual.....	2
2.1 Pre-requisites.....	2
2.2 Pre-installation Tasks.....	3
2.3 Installation Procedure .....	3
2.3.1 Initial File Setup .....	3
2.3.2 Database Setup .....	4
2.3.3 Running installation scripts.....	5
2.4 Post-installation.....	9
3 Additional Information .....	10
3.1 Document Revisions.....	10
3.2 Reference Documents.....	10

## 1 Introduction

### 1.1 Purpose

This document is provided to serve as installation guidelines for the Drupal 7 Content Management System.

The objective of this document is to set up a functioning Drupal 7 base installation.



An inadequate, incomplete or nonexistent install procedure may lead to website instability and data loss. Check prerequisites and pre-install checklist before proceeding.

### 1.2 Scope

This document is intended to direct and assist a novice user with the detailed set-up of the Drupal 7 Content Management System.

It will not cover detailed aspects such as module and theme installation, Drupal Core configuration, or updates to themes, modules, or the Drupal Core environment.

### 1.3 Server Requirements

To install and run Drupal 7, the following requirements must be met: [1]

1. Apache version 2.0 or greater
2. PHP 5.2.0 or greater
3. MySQL 5.0 or greater, PostgreSQL 8.3 or greater, or SQLite 3.4.2 or greater

Failure to meet these requirements will cause installation failures or a non-operable installation of Drupal 7.

## 2 Installation Manual

### 2.1 Pre-requisites

Before beginning the installation of Drupal 7, verify that the following are available and configured for use:

- Web server (Must meet server requirements outlined in section 1.3)
- FTP Software
- Stable internet connection
- Any archive software (WinRAR, 7-zip, etc)
- Any text editor
- Web browser with internet access

## Drupal Installation Manual

### 2.2 Pre-installation Tasks

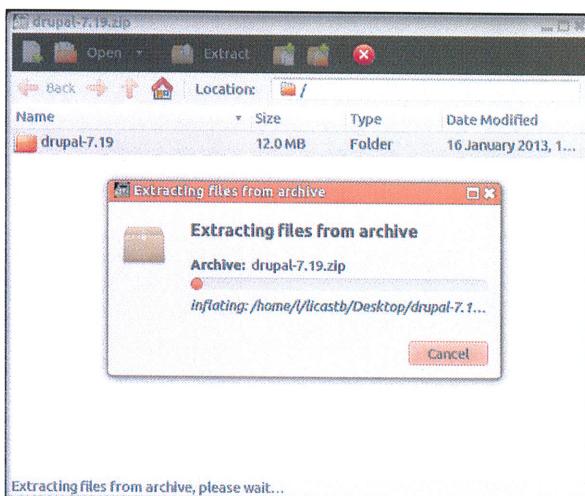
Before installation, visit [www.drupal.com/start](http://www.drupal.com/start) and download the latest release of Drupal 7. It will be available for download as a tar.gz or a zip file. Either will work, but some archive managers only support the zip format. Refer to your software manual to determine which file format(s) are supported.

For the purposes of this installation it is going to be assumed that there is an active domain and an empty directory available and properly configured to host this website.

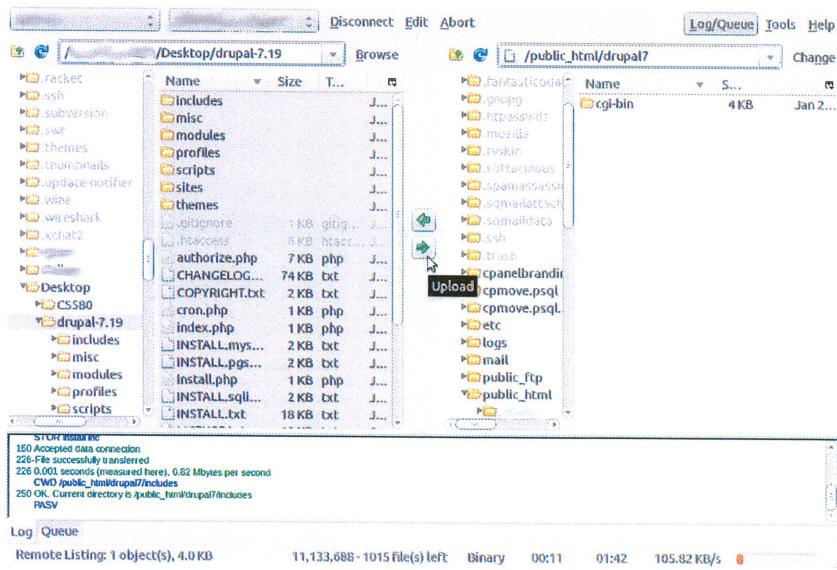
### 2.3 Installation Procedure

#### 2.3.1 Initial File Setup

1. Extract the installation files from the archive using your archive manager.

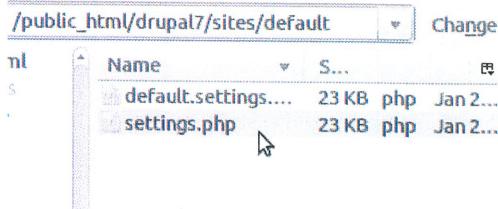


2. Move the extracted files into a directory within your web servers public HTML directory or document root. [1]

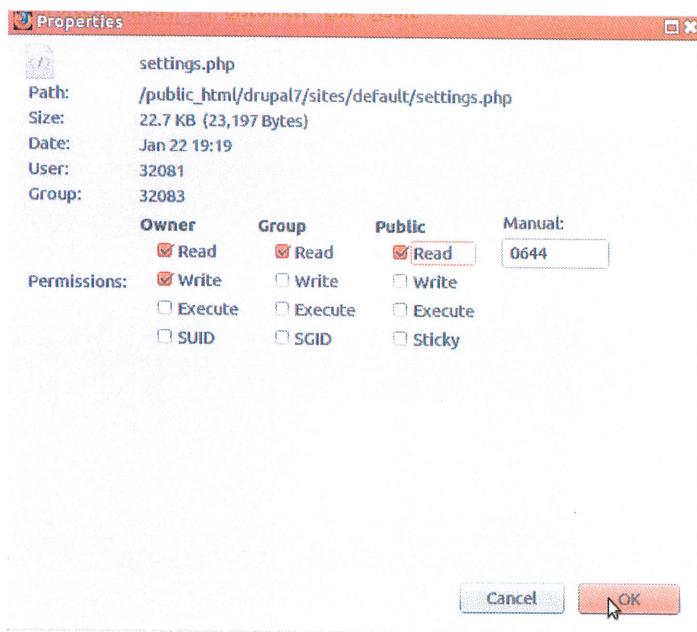


## Drupal Installation Manual

3. Create a configuration file by copying the example, named default.settings.php (located in the sites/default directory of the extracted files). Rename the copied file to settings.php. [1]



4. Give the web server write privileges (644) for the settings file. [1]



The file work has been completed for this tutorial. Feel free to modify the .htaccess file for your needs, install/remove modules and themes, etc. This is all we will be covering in this tutorial.

### 2.3.2 Database Setup

To set up a database for your Drupal 7 installation, log in to your database management software of choice. For this installation walk-through we will be using the integrated Cpanel database tools.

1. Create the database and password using your database management tool.

A screenshot of a 'Create New Database' interface. A text input field labeled 'New Database:' contains the value 'd7db'. To the right of the input field is a green circular icon with a white checkmark. Below the input field is a large blue 'Create Database' button.

## Drupal Installation Manual

2. Next you will have to set database access permissions. Open your user permissions dialogue and give the database user access for the following:

*SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, INDEX, ALTER*

Manage User Privileges

User: temp  
Database: d7db

ALL PRIVILEGES	
<input checked="" type="checkbox"/> ALTER	<input checked="" type="checkbox"/> CREATE
<input type="checkbox"/> CREATE ROUTINE	<input type="checkbox"/> CREATE TEMPORARY TABLES
<input type="checkbox"/> CREATE VIEW	<input checked="" type="checkbox"/> DELETE
<input checked="" type="checkbox"/> DROP	<input type="checkbox"/> EXECUTE
<input checked="" type="checkbox"/> INDEX	<input checked="" type="checkbox"/> INSERT
<input type="checkbox"/> LOCK TABLES	<input type="checkbox"/> REFERENCES
<input checked="" type="checkbox"/> SELECT	<input type="checkbox"/> SHOW VIEW
<input type="checkbox"/> TRIGGER	<input checked="" type="checkbox"/> UPDATE

You have successfully created a database and user at this point. If the privileges have been set successfully move on to section 2.3.3 – running the install scripts.

### 2.3.3 Running installation scripts

To run the Drupal 7 installation scripts open a web browser and navigate to the web address of your new Drupal installation. (ex. [www.example.com](http://www.example.com)) You should see a page that looks similar to the one below.

Select an installation profile

  Standard  
Install with commonly used features pre-configured.

Minimal  
Start with only a few modules enabled.

**Choose profile**

Choose language

Verify requirements  
Set up database  
Install profile  
Configure site  
Finished

We are going to use the standard settings for this installation. Click save and continue.

## Drupal Installation Manual

### Choose language



English (built-in)

[Learn how to install Drupal in other languages](#)

[Save and continue](#)

✓ Choose profile

► Choose language

[Verify requirements](#)

[Set up database](#)

[Install profile](#)

[Configure site](#)

[Finished](#)

English is the only language offered natively so we will use that. Click Save and continue to move to the next page.

### Database configuration



**Database type \***

MySQL, MariaDB, or equivalent

SQLite

The type of database your Drupal data will be stored in.

✓ Choose profile

✓ Choose language

✓ Verify requirements

► Set up database

[Install profile](#)

[Configure site](#)

[Finished](#)

**Database name \***

d7db

The name of the database your Drupal data will be stored in. It must exist on your server before Drupal can be installed.

**Database username \***

temp

**Database password**

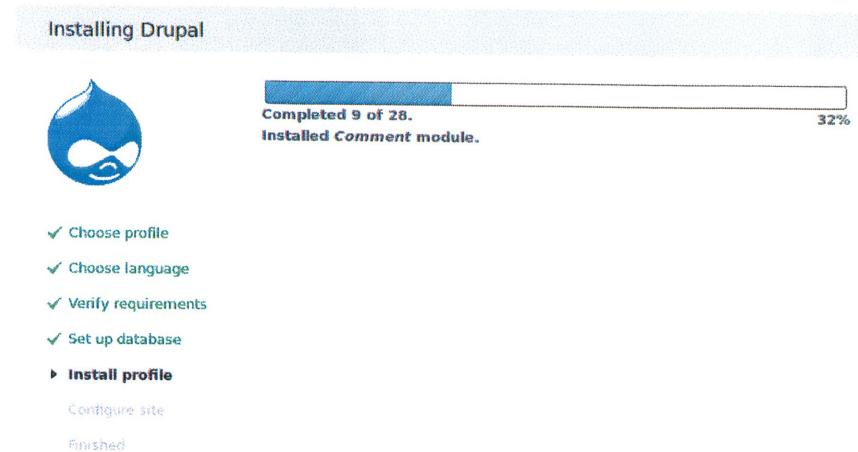
\*\*\*\*

► ADVANCED OPTIONS

[Save and continue](#)

This is where the database name, username, and password will be set up. Enter them in the appropriate fields after selecting the database type you will be installing Drupal on. Advanced options are fine as the default values in a majority of situations. Click Save and continue to move to the next step.

## Drupal Installation Manual



After clicking the Save and continue button Drupal should be able to connect to the database successfully and will proceed with the installation and default module configuration. You should see something similar to the image below.

After set up has finished you will be presented with a rather lengthy page asking for some information. I have broken the descriptions into sections to keep the monotony to a minimum.

### 1. Site Information

- I. Site name – The name that will appear in the header of the Drupal website.
- II. Site e-mail address – Email address that all emails will be sent from

**SITE INFORMATION**

**Site name \***  
Drupal 7 Installation Demo Site

**Site e-mail address \***  
noreply@example.com

Automated e-mails, such as registration information, will be sent from this address. Use an address ending in your site's domain to help prevent these e-mails from being flagged as spam.

### 2. Site Maintenance Account

- I. Username – The root account's login. Used to perform admin functions
- II. Email address – Email for the root user

## Drupal Installation Manual

**SITE MAINTENANCE ACCOUNT**

**Username \***  
rootuser  
Spaces are allowed; punctuation is not allowed except for periods, hyphens, and underscores.

**E-mail address \***  
noreply@example.com

**Password \***  
\*\*\*\*\* Password strength: Weak

**Confirm password \***  
\*\*\*\*\* Passwords match: yes

To make your password stronger:  
• Make it at least 6 characters  
• Add uppercase letters  
• Add numbers  
• Add punctuation

### 3. Server Settings

- I. Default Country – Choose the country that will be selected by default when new users register. Also usually the geographic location of the site.
- II. Default time zone – Default time zone for new users and date stamps

**SERVER SETTINGS**

**Default country**  
United States  
Select the default country for the site.

**Default time zone**  
America/New York: Tuesday, January 22, 2013 - 22:11 -0500  
By default, dates in this site will be displayed in the chosen time zone.

### 4. Update notifications are personal preference. After selecting the options you want, click Save and continue as shown in the image below.

**UPDATE NOTIFICATIONS**

Check for updates automatically  
 Receive e-mail notifications

The system will notify you when updates and important security releases are available for installed components. Anonymous information about your site is sent to [Drupal.org](#).

**Save and continue**

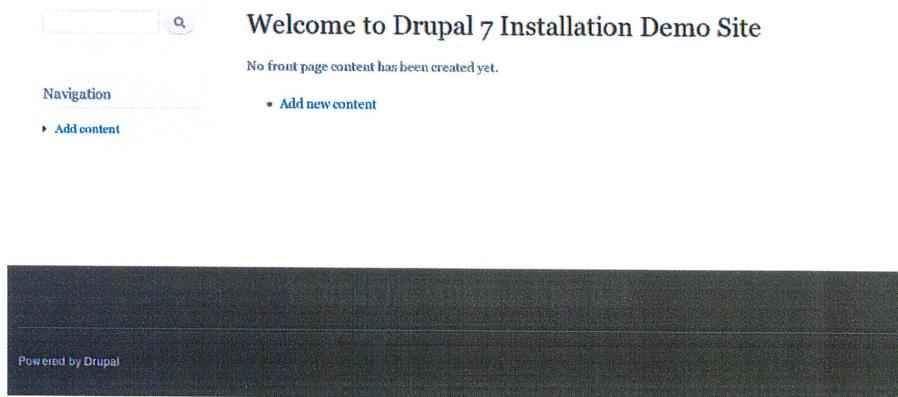
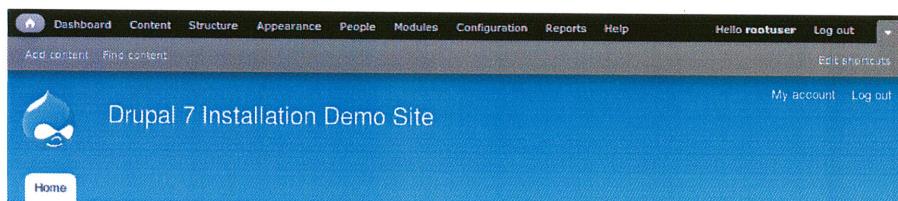
## Drupal Installation Manual

5. If everything was properly filled out and configured, you should see the success message listed below.



### 2.4 Post-installation

If everything went well and you clicked the “View your new site” button on the previous screen, your new website should look something like this:



At this point you can rest assured that everything is installed, and if there are no errors displayed it has been set up and configured correctly. From here you can post content, manage your website, etc. Log out and you will see the site as the public would.

Congratulations and job well done!

## 3 Additional Information

### 3.1 Document Revisions

Date * Revision Number	Description
1.20.2013 * 1.0	Document initially created.
1.22.2013 * 1.5	Changed formatting and added database directions.
1.23.2013 * 1.5.1	Fixed auto correct typo and several minor grammatical errors.

### 3.2 Reference Documents

- [1] Sea, Eric. "Quick Install for Beginners." *Drupal*. Drupal.org, 13 Aug. 2012. Web. 22 Jan. 2013. <<http://drupal.org/documentation/install/beginners>>.

Wow! Very  
good installation  
document.

## Lab 1: Investigating Data Management Tools and Technologies

Braden Licastro  
Eric Weyant  
Sarah Barlow

## World Data Production and Storage

As of 2012, 2.7 zettabytes of data had been produced in the world

- Enough fully loaded 32GB iPads to build a twenty foot wall around South America

Currently 600+ exabytes of information in storage

- Storage capacity formula states storage capacity doubles every three years

IDC estimates that data production will have increased 50x by the year 2020

## Features of Data Management Tools

- Store/retrieve data
- Modify/delete
- Analyze data /sorting
- Performance
- Atomicity
- Data abstraction
- Safety/security
- Reliability
- Crash tolerance
- Redundancy
- User friendly
- Cloud storage

## Best Data Management Tool

- Drupal - Flexible and secure web based content management system
- Open source, extendable, native SSL integration, quick and easy setup
- Supports virtually any data type and size
- Interface fast, easy to navigate, takes advantage of other web services

## Expectations

- SQL
- XML
- Database/Query Efficiency
- Database Security
- Data Management Techniques