

You can pick up where you left off. Just join a new session and we'll reset your deadlines.

Join a session

Back to Week 1

Lessons

Prev

Next

Specialization Introduction:
Excel to MySQL: Analytic
Techniques for Business

Introduction to Managing
Big Data with MySQL

Welcome to Managing Big Data with MySQL	1 min
What You Will Learn in This Course	10 min
Course Overview	10 min
IMPORTANT FOR EVERY LEARNER	10 min
Special Thanks!	5 min
Feedback Survey Information	10 min

Lesson 1: Problems
Databases Solve

Lesson 2: Database Design
Tools

Lesson 3: Building Your Own
Entity-Relationship
Diagrams

Lesson 4: Building Your Own
Relational Schemas

Lesson 5: Test Your
Understanding

Have a question? Discuss this lecture in the week forums.

Interactive Transcript

English

0:07 Sequel. >> S-Q-L. >> Sequel. >> S-Q-L. >> Sequel. >> Sequel. >> Sequel. >> Clearly, nobody seems to agree how to pronounce the acronym for the database language you are going to learn in this course. The most common pronunciations are either SQL or Sequel, but there is truly no agreement about which pronunciation is more accepted. A blogger named Patrick Gillespie amusingly emailed a man named Don Chamberlin to ask about the official pronunciation of the language's name. Don Chamberlin co authored SQL or sequel with Raymond Boyce at IBM in the 1970s. So if anybody should know what the language is called, it's Don.

0:47 Don responded with, Hi Pat. Since the language was originally named Sequel, many people continued to pronounce the name that way after it was shortened to S-Q-L. Both pronunciations are widely used and recognized. As for which is more official, I guess the authority would be the ISO standard, which is spelled and presumably pronounced SQL. Thanks for your interest, Don Chamberlin. So even the co author of the language can't say with confidence how the name of the language is pronounced. The best he can say is that it's presumably pronounced, not definitely pronounced or even usually pronounced SQL. To make things more confusion some of the most popular database systems in the world such as Oracle say explicitly in their documentation that the language is pronounced See-Quill. Others, such as MySQL say explicitly that their name is My Ess-Cue-Ell. So I wasn't exaggerating when I said there is no agreement about how to pronounce the name of the language. That means you can use whichever pronunciation you like better. Personally I like using both pronunciations, sometimes even in the same sentence. I will go back and forth between the two pronunciations throughout the course.

1:52 So why should we care about SQL or Sequel? Companies have to keep track of a lot of data, and the only practical way to do that is to store the data is some type of database. If you are becoming a data analyst for the business world, somehow you are going to have to figure out how to get data out of these databases in order to do any kind of analysis. This course is designed to help you learn how to do that.

2:14 In many technology driven companies, the only way to retrieve data from a database is to write your own SQL queries, which are structured lines of computer code that extract data. In other larger and more traditional companies there are people in the company who could extract the data for you, but you have to put in a formal request for them to write and execute and query. And it could take days or weeks for them to fill your request especially if many other people are asking for queries too. For this reason, companies are very eager to hire analysts who can not only analyze data but who know how to get the data in the first place. As some evidence of this, academic researchers of the Teradata University Network were in a study to determine the state of business intelligence and analytics. They asked over 400 recruiters from technical companies to answer the question. When I recruit for business intelligence, business analyst roles, it is important that the students have the following coursework or knowledge. Then the recruiters had to rank a bunch of possible responses. The number 2 answer companies gave, was SQL and Query skills. The number 3 answer companies gave was Basic Analytics. Therefore, according to this survey, the ability to pull data out of a database is even more important than analytical skills for getting a business analyst job.

3:21 We are dedicating an entire course in the specialization to helping you acquire this ability.

3:26 By the way, if you're curious about what the number 1 response companies gave in the survey was, it was very interestingly, Communication Skills. That's why we dedicated course three of this specialization to Data Visualization and Communication.

3:41 I mentioned earlier that SQL is the language used to retriever data from specific types of databases called relational databases. Relational databases are the gold standard for data bases that store highly organized and structured business data. Almost every single company in the world has at least one relational database. There are also new classes of databases specifically designed for data that are collected in an extremely fast rate, such as GPS or biosensory data, or unstructured data that does not fit well into highly prescribed formats, such as tweets or texts. However, although these databases are certainly likely to gain popularity in the future, right now they only represent a very small fraction of the database market. As you can see from this graph produced by the International Data Corporation, a global market intelligence firm, annual sales from the current database market are over \$40 billion, and are expected to top \$50 billion in the next few years. The green in the bars on this chart illustrates that almost all of these billions of dollars are spent on relational database technologies. Therefore, although there are definitely exciting things to learn and know about other databases, we chose to focus on how to manage big corporate data using relational databases in this specialization.

4:49 The second part of the title of this course refers to MySQL which is a particular brand or platform of relational database. There are many relational database platforms including Oracle, Microsoft, SQL Server, DB2, SQLite, Microsoft Access, and PostgreSQL. We chose to focus on MySQL. To help explain why, I'd like to introduce you to Ryan Luecke, senior software engineer at Box.

5:16 >> Hi, my name is Ryan Luecke and I work at Box as a senior software engineer on our caching infrastructure team, which is closely related to our database engineering team.

5:27 Box is an online file sharing and collaboration service for businesses. Box lets multiple people across the country or across the world work on documents at the same time and store them in a safe place that is accessible to anyone. A good reason to highlight Box in this course is that Box's entire interface is based on relational database concepts. Every time you store a new document in a folder in your Box account, it's like a new row in a table. Every time you add a new collaborator to a document it's like adding a new relationship between tables. So in Ryan's words, >> Really, then what Box is, is a very pretty interface to get to all of the underlying database operations. >> This means that people who work at Box have a really deep understanding of what relational databases do, because their entire product is based on making sure the relational database works quickly and perfectly. And importantly for this course, Box uses MySQL as its primary database. Here is Ryan's explanation for their choice.

6:25 >> Choosing a relational database is a difficult task. There are a lot of relational databases and non-relational databases to choose from. And one reason we chose MySQL was it's reliable, it's mature, it's open source, and well understood. There's a lot of industry and community around MySQL. There are conferences around MySQL, and a lot of tools for it. And it also meets our relational needs, so it's able to do what we need it to do to store data for millions of users, and billions of pieces of content.

7:00 >> But box is by no means the only well known company to use MySQL. Here are some other companies that rely heavily on MySQL databases.

7:09 >> So Facebook uses MySQL, YouTube within Google. Yelp uses MySQL, Dropbox. GitHub which stores a lot of developer files and projects, LinkedIn, Etsy online store, Send grid, Twitter, Booking.com, Square, Pintrest, Yahoo, and lots of others.

7:36 >> Those are a lot of big name companies. Given that MySQL is open source, free for anybody to use, well understood, and used by so many companies you might want to work for, we thought it was an important database with which to give you some experience.

7:49 Each database system does use its own slightly different version of SQL language, so you will have to be prepared to look for the differences if you start working with another type of database. Don't worry, languages are very similar. Writing SQL queries is kind of like driving a car. Every time you use a different car, you might have to spend some time looking for where the buttons and adjustment knobs are. But if you know how to drive one car, you know how to drive all of them, as long as they have the same transmission type, that is. Same thing with different relational database systems. Once you know how to use SQL to interact with one database, you know how to use SQL to interact with all relational databases. You are going to get a sense of how small or large the differences in SQL languages are in this course because we decided to give you access to another type of relational database in addition to a MySQL database. The second relational database it a Teradata database. We chose Teradata because it is the leader in what is called data warehousing. A data warehouse is a copy of historical and current company data that is structured explicitly for querying and analytical reporting. Their warehouses are commonly used for things like company dashboards that allow executives to make decisions based on real time data and graphs that show how those data relate to trends from the past. Well known companies which store petabytes of data in Teradata bases include Walmart, Sam's Club, Verizon, AT&T, Bank of America, and Apple. Having experience with both a MySQL and a Teradata database will give you confidence that you can navigate different database systems and will provide you with a strong competitive edge in the data analytics marketplace.

9:17 Here's what we will cover in the next few weeks. By the end of this course, you will be able to describe the structure of relational databases, interpret and create entity relationship diagrams and relational schemas that describe the contents of specific databases. Write queries that retrieve and sort data that meet specific criteria, and retrieve such data from MySQL and Teradata databases that contain over one million rows of data. Execute practices that limit the impact of your queries on other coworkers. Summarize rows of data using aggregate function and segment aggregations according to specified variables. Combine and manipulate data from multiple tables, across a database. Retrieve records and compute calculations that are dependent on dynamic data features. And translate data analysis questions into SQL queries. Now long from now you will have a highly coveted skill set that will make you very attractive to data analyst recruiters. You will also have a powerful set of tools to help you provide real tangible value to your business. I'm very excited to learn about all the doors that will open for you as you embrace this important feature of the data driven business world. It's going to be a fun ride.

Downloads

Lecture Video	mp4
Subtitles (English)	WebVTT
Transcript (English)	txt

Would you like to [help us translate](#) the transcript and subtitles into additional languages?