

coursera

Explore

Q

What do you want to learn?

For Enterprise

php

php

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

Lesson 1: Problems Databases Solve

Problems with Having a Lot of Data Used by a Lot of People7 min

How Relational Databases Help Solve Those Problems8 min

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

0:03

Why do we need databases in all this complicated database query language? Why can't businesses keep things simple and just store things in spreadsheets? It seems like that would make things a lot easier on everyone.

0:14

Well, let's examine that premise that keeping company data in [spreadsheets would make things easier on everyone](#).

0:20

Imagine you are working with Daniel Egger as he stars up Egger's Roast Coffee, the business he described in Course 1 of the specialization Business Metrics for Data Driven Companies. Your first week of business, Daniel personally handled orders from some great companies that he dutifully kept track of in this spreadsheet. You entered all the information that it seemed important to track and things look good.

0:41

Now, fast forward a couple of years. Edger's Roast Coffee's done extraordinarily well, and Daniel has become coffee king of North Carolina.

0:49

There are huge numbers of orders coming in every day, and now many people are entering those orders into the spreadsheet you made the first week, and many people are keeping track of and modifying the orders as customers call to ask questions.

1:01

Actually I just said that very calmly, but the truth is what I just said puts misleadingly positive outlook on what's probably happening in the ordering department of Egger's Roast Coffee. What's really happening is that multiple people are trying to enter those orders into a spreadsheet, but that doesn't always happen because now Egger's Roast Coffee has over 1 million orders. So the one spreadsheet you made when the company started has been broken into six separate spreadsheets so that Excel doesn't crash every time you try to enter an order.

1:28

Breaking the spreadsheet into six has helped prevent Excel from crashing. But now every time a customer calls to ask a question about an order or an invoice, It takes a couple of minutes before the employee can find the correct spreadsheet that has the order. Especially because each spreadsheet is still very large so it takes Excel a while to open each one.

1:45

The more annoying part is that multiple people are trying to use the spreadsheet at the same time. So when a customer calls with question, you have to wait until somebody else finishes with the spreadsheet before you can open it yourself. As a consequence, you all start coming up with elaborate ploys to keep the spreadsheet open on your computers so that you are the one that has it loaded when a customer calls.

2:06

Fortunately, everybody at Egger's Roast Coffee realizes that it's bad business practice to have employees get kicked off the ordering spreadsheets when they need them. So the company agrees to move their spreadsheets to an online service that let's multiple people edit the spreadsheet at the same time. Cool, but now new problems are arising. First of all, the spreadsheets are still pretty big. So they still take a while to open and search. Second, sometimes multiple people end up interacting with a spreadsheet at the same time, and they end up writing over each other's entries. Third, there's starting to be a lot of inconsistencies in the data which are hard to catch if you don't have multiple spreadsheets open at once. And only a few people have computers that are powerful enough to open multiple spreadsheets of that size at once anyway.

2:52

As some examples of these inconsistencies, here are some excerpts from the master spreadsheets.

2:58

Is the customer Archway's contact Trudy, or Mark? And if it's Trudy, which email address is in the spreadsheet is her true email?

3:06

Are Trudy's email addresses different because someone entered the second one by mistake, or because Trudy has actually changed her email?

3:13

One of the reasons you might worry that Trudy's second email was a mistake, is because lots of mistakes have been occurring recently. Every time a returning customer makes an order, an Egger's Coffee employee has to type in all that information about the returning customer again, even though it's already been entered in the spreadsheet multiple times. Since the employees are doing their best to keep the customer happy, they end up entering the information really quickly, and as a side effect, are more likely to make a mistake.

3:38

Other issues that have gotten annoying are that when a mistake is made in one place in the spreadsheet it often means that an employee has to change multiple fields in the spreadsheet to correct the mistake. So when it turned out that Lina wasn't actually the Egger's roast coffee employee to make this sale, somebody had to update this column that had Lina's name but that person also had to update these columns with Lina's contact information and sale number. It would have saved a lot of time if that were automatic.

4:04

There were formulas in these columns at one point that made updating automatic, but now there are so many different spreadsheets and mistakes in the spreadsheets, that the formulas often output the wrong thing, so nobody trusts them. And the problems don't stop there. Sometimes, it would be useful to keep some information in a row, but delete other information. Here for example, Lulu's decided to cancel their order. So you delete the order information in an effort to make sure the formulas that tally up the sum of your sales are correct. But you would still like to keep Lulu's contact information in the company's records so that you can try to woo them back to placing an order in the future.

4:39

The problem is that you have a row with a lot of empty data, an employee looking over the spreadsheet trying to reconcile records for an audit, might think the row is a big mistake and either get worried or erase it all together.

4:51

Speaking of audits, the company ended up switching invoice systems along the way, and some orders are being made on the old invoice system while others are being made with the new system. The employee who was responsible for reconciling the companies records needs to separate the sales made on two systems, but it's difficult to do so because yet another separate analysis spreadsheet is used for the analysis. And it's hard to keep that analysis spreadsheet up to date when there are new orders being added all the time. Further, Egger's Roast Coffee would like to start collecting more information about its customers to improve marketing efforts. However, the number of columns that need to be added to each spreadsheet to accomplish these marketing analysis would more than triple the size of the spreadsheets. That would make the spreadsheets even slower than they are now.

5:33

So, as you can see, starting by entering Egger's Roast Coffee records into a spreadsheet is easy and made a lot of sense. But as the company grew, spreadsheets became increasingly inefficient and inaccurate. At this point keeping the company's data and spreadsheets would really hinder the company. If Egger's Roast Coffee ever gets to the point in the future where they're making millions of sales per year, the spreadsheet system will become a huge bottleneck in the company's ability to maintain customer service, keep track of its finances, and make business decisions. For Egger's Roast Coffee to stay competitive, they need a data storage solution that makes their information easily retrievable, easy to update, and accessible and modifiable by multiple people at the same time. In addition, especially because so many people will be accessing the data, there needs to be a way to make sure the data are consistent across all the stored records and that data aren't being deleted or lost inadvertently.

6:24

The storage method also needs to be space efficient, so the data aren't being repeated and taking up extra space.

6:30

Of course, the storage needs to be fast as well. Data need to be entered, search, and retrieved quickly.

6:38

Finally, another thing we didn't talk about yet, is that the company's data also needs to be very secure. When you're storing financial information, adding a password to a spreadsheet is not going to be enough to keep that information safe and protected. That long list of problems that arise when a lot of people are using a lot of data are the exact problems databases are meant to solve. Watch the next video to find out how relational databases in particular, manage to do that.

Downloads

Lecture Video

mp4

Subtitles (English)

WebVTT

Transcript (English)

txt

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