ESD 40.011
Activity 2
Advanced Relational Database Concepts
P. Jackson
Edited by Ying Xu for SQLiteStudio 3.2.1

Contents

Educational Objectives 1
Basic Concepts 1
Getting Started 1
Exercise 1: 2
Exercise 2 4
Exercise 3 6
Review 7

Educational Objectives

1. Use a data-driven approach to throughput improvement in a factory
2. Understand when you need to refer to a table name when selecting a field
3. Understand how to group data by a specific condition
4. Understand aggregation queries and aggregate operators
5. Understand how to join the contents of multiple tables or views with "Joins"

Basic Concepts

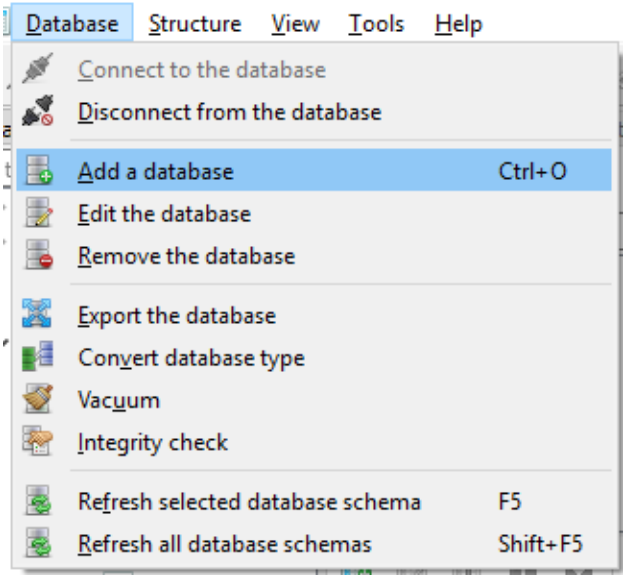
You must have completed Activity 1 (Intro to SQL) before attempting this activity.

1. You may need the following concepts:
    a. Throughput (the flow of material from input to output over some period of time; often expressed as a rate, such as units per hour)
    b. Yield loss (the loss of throughput attributable to scrap)
    c. Speed loss (the loss of throughput attributable to slower operations)
    d. Line transitions (the loss of throughput attributable to equipment changeovers from producing one product to producing another)

Getting Started

1. To receive credit for this activity, complete the activity response document and turn it in. Blanks in this document have matching entries in the activity response document.
2. Open SQLite Studio, version 3.2.1 or higher.

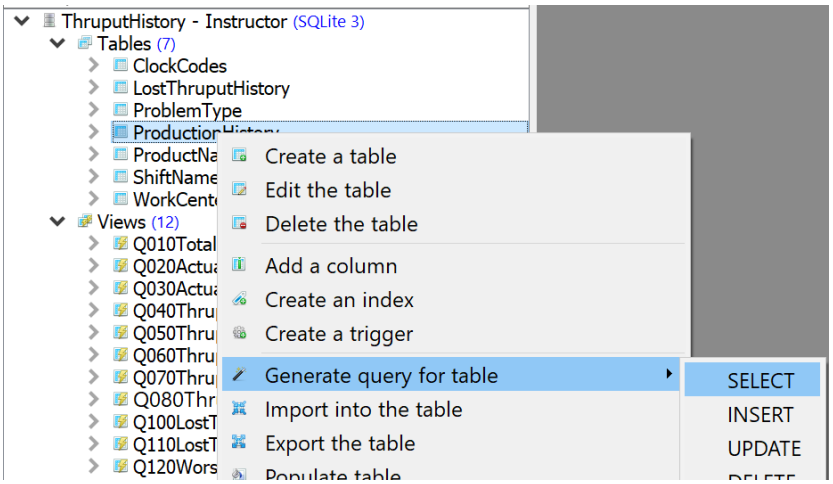
3. Remove any other databases such as the Chinook database from the list of databases in the left hand panel. Note, removing the database from the list does not delete the file from your computer. You can add the database to the list as easily as removing it.
4. Download the file “ThruputHistory.sqlite” from the course website.
5. Open the file by selecting the option “Add a database” under the “Databases” tab.



Exercise 1:

The “ProductionHistory” table is a log of each product’s production by work center, week, and shift over a period of time. We wish to analyze the data in this table to see where we should concentrate our throughput improvement efforts. The table contains five fields related to throughput. Using these fields, we can compute the total time spent on the production and the target time that should have been spent by each product, work center, week, and shift.

1. We start with a trick to get a listing of all the fields in the table. Right-click the ProductionHistory table and select “Generate query for table…SELECT”



This should result in a generate query like this:

Query History

```
SELECT ProductID,
       WeekNo,
       WCType,
       WorkCenterID,
       ShiftID,
       Demand,
       TransitionMinutesActual,
       ProductionMinutesActual,
       ThruputTarget,
       TransitionTarget
  FROM ProductionHistory;
```

Select all the text in the query and copy it into the clipboard.

2. Create a new view, title it “Q010TotalMinutesActualDetail” and paste the query from the clipboard. That should save you some typing. Now edit the query in the view so that it lists only the fields shown below. Be sure to include the calculated field “TotalMinutesActual”.

View name: Q010TotalMinutesActualDetail

```
SELECT ProductID,
       WeekNo,
       WCType,
       WorkCenterID,
       ShiftID,
       Demand,
       TransitionMinutesActual,
       ProductionMinutesActual,
       TransitionMinutesActual+ProductionMinutesActual as TotalMinutesActual
  FROM ProductionHistory
```

3. Save the view and look at the data tab:

	ProductID	WeekNo	WCType	WorkCenterID	ShiftID	Demand	TransitionMinutesActual	ProductionMinutesActual	TotalMinutesActual
1	P013	1	0	B2	3	3200	30.00	1001.00	1031
2	P014	1	0	B1	1	3700	30.00	1127.00	1157
3	P002	1	2	A2	3	3700	60.00	2133.00	2193
4	P012	1	1	L4	1	3200	240.00	2297.00	2537
5	P011	1	1	L2	1	2500	213.00	2372.00	2585

Interpreting the first line, we see that Product “P013” started production in WorkCenter “B2” on the third shift; the number of units in the production order was 3200. The length of time to setup (“transition”) the equipment to produce this product was 30 minutes. Then the production run took 1001 minutes (almost 17

hours). The total time the equipment was engaged in this production run was 1031 minutes.

4. We cannot tell whether these production runs are efficient or not without referring to some benchmark. The ProductionHistory table includes two fields ThruputTarget and TransitionTarget which give realistic but ideal targets for each of the products in each workcenter. ThruputTarget is the ideal production rate (units per hour) and TransitionTarget is the number of minutes the setup or transition should take under ideal conditions. We use these data to estimate a target time for the job. Create a view called "Q020ActualAgainstTarget" using this query:

View name: Q020ActualAgainstTarget

```
SELECT ProductID,
       WeekNo,
       WCType,
       WorkCenterID,
       ShiftID,
       Demand,
       TransitionMinutesActual,
       ProductionMinutesActual,
       TransitionMinutesActual+ProductionMinutesActual as TotalMinutesActual,
       TransitionTarget as TransitionMinutesTarget,
       Demand/ThruputTarget*60 as ProductionMinutesTarget,
       TransitionTarget+Demand/ThruputTarget*60 as TotalMinutesTarget
  FROM ProductionHistory
```

5. For the P013 job on B2 we just looked at, the one that took 1031 minutes, what should have been the time to perform that job under ideal conditions? _____

6. It would be nice to know what the types of workcenters are. This factory has only three types (see the WorkCenterTypes table). For this we will use a LEFT JOIN and we will alias the view name of the underlying view. Create the following view:

View name: Q030ActualAgainstTargetWithWCDescription

```
SELECT Q020.*,WCDescription FROM Q020ActualAgainstTarget as Q020 LEFT JOIN WorkCenterTypes
ON Q020.WCType=WorkCenterTypes.WCType
```

Note that we use "Q020" as an alias for the long view name and we match two tables based on the common field "WCType". We used a LEFT JOIN because we do not want to risk losing any of the throughput data simply because it is missing a WCType.

7. What is the description of the workcenter type for the B2 workcenter we looked at?_____

Exercise 2

1. Now we are ready to do some analysis. Let's sum up the total actual production time and the total target production time over all products, workcenters, and shifts. Create the following query and title it "Q040ThruputTotalsPlantWide":

View name: Q040ThruputTotalsPlantWide

```
SELECT
SUM(TotalMinutesActual) as SumTotalMinutesActual,
SUM(TotalMinutesTarget) as SumTotalMinutesTarget
FROM Q030ActualAgainstTargetWithWCDescription
```

2. Look at the resulting data of the view:

Total rows loaded: 1

	SumTotalMinutesActual	SumTotalMinutesTarget
1	1635262	1165200

Observe that all of the rows of the underlying view ("Q030") are summarized into one single row. Observe that the total time actually used is about 1.64 million minutes but the ideal number is quite a bit lower, 1.17 million minutes. This tells us that there is opportunity for throughput improvement.

3. Let's do our analysis using relative comparisons between actual and target. Create the following view:

View name: Q050ThruputPerformancePlantWide

```
SELECT
(SUM(TotalMinutesActual)-SUM(TotalMinutesTarget))/SUM(TotalMinutesTarget)*100 as PerCentImprovementPossible
FROM Q030ActualAgainstTargetWithWCDescription
```

4. To the nearest whole number, what is the percentage of throughput improvement possible in this plant?_____
5. We can't make improvements everywhere all at once. We have to find a place to start. Let's look for the type of workcenter where we have the greatest opportunity for improvement. For this we use the GROUP BY clause. Create a view and call it "Q060ThruputPerformanceByWCType":

View name: Q060ThruputPerformanceByWCType

```
SELECT WCDescription,
(SUM(TotalMinutesActual)-SUM(TotalMinutesTarget))/SUM(TotalMinutesTarget)*100 as PerCentImprovementPossible
FROM Q030ActualAgainstTargetWithWCDescription
GROUP BY WCType
```

Instead of one single row, this will aggregate the data based on WCType. There will be one row for each WCType. Recall there are 3 workcenter types so there will be three rows in the result. What will be the content of those rows? We will show the WCDescription, because that is the same for each row in the original view, for each WCType, and we will perform aggregation calculations for the rest.

6. Which of the three workcenter areas of the plant offers the greatest (relative) opportunity for throughput improvement?_____
7. Let's drill down. We know the area of the plant to focus on but which machine should we focus on? Let's group by WCType and WorkcenterId:

View name: Q070ThruputPerformanceByWCTypeWCId

```
SELECT WCDescription,
WorkcenterID,
(SUM(TotalMinutesActual)-SUM(TotalMinutesTarget))/SUM(TotalMinutesTarget)*100 as PerCentImprovementPossible
FROM Q030ActualAgainstTargetWithWCDescription
GROUP BY WCType,WorkCenterId
```

8. Which WorkcenterID offers the greatest relative opportunity for throughput improvement?_____

9. Sometimes the combination product and equipment is what we should focus on. Look at this:

View name: Q080ThruputPerformanceByWCTypeWCIdProductId

```
SELECT WCDescription,
WorkcenterID,
ProductId,
(SUM(TotalMinutesActual)-SUM(TotalMinutesTarget))/SUM(TotalMinutesTarget)*100 as PerCentImprovementPossible
FROM Q030ActualAgainstTargetWithWCDescription
GROUP BY WCType,WorkCenterId,ProductId
```

Observe that every time we add another field to the GROUP BY list, we add another field to the list of SELECT fields. This helps us to identify what the aggregation row refers to.

10. For the workcenter we identified above, what product on that workcenter offers the greatest relative opportunity for improvement?_____

Exercise 3

Now that we have identified the area, workcenter, and product to focus attention on, let's focus on what problems we can address. The table LostThruputHistory is a table I derived from a log of problems people on the shop floor had recorded while working on products. I did some calculations based on the log file to estimate how many minutes of thruput time were lost for each problem recorded. Let's use this information, together with the insights of the previous analysis to pick some problems to work on.

1. Create a view, title it "Q100LostThruputForL1P11" based on the following query:

View name: Q100LostThruputForL1P11

```
SELECT WeekNo,
       WCType,
       WorkCenterID,
       ProductID,
       ShiftID,
       ProblemType,
       Description,
       NumOccurrences,
       LostThroughputMinutes
  FROM LostThruputHistory
  WHERE WorkCenterId="L1" OR (WCType="1" AND ProductID="P011")
```

Here we are only interested in problems that relate specifically to WorkCenter L1 or to Product P011 in WCType 1. We should be able to find some "interesting" problems here.

2. How many problems surfaced in this category?_____
3. Let's summarize these problems as follows:

View name: Q110LostThruputForL1P11Summary

```
SELECT WCType,
    WorkcenterID,
    ProductID,
    ShiftId,
    ProblemType,
       COUNT(WeekNo) as ProblemWeekCount,
       MIN(Description) as ProblemDescription,
       SUM(NumOccurrences) as TotalOccurrences,
       SUM(LostThroughputMinutes) as TotalLostThruputMinutes
 FROM Q100LostThruputForL1P11
 GROUP BY WCType,WorkcenterID,ProductID,ShiftId, ProblemType
```

Note that for each field in the "group by" list we have a SELECT field to help identify the resulting row. The remaining SELECT fields are aggregations because there can be many records in the underlying view for each row in the aggregation. We count the number of weeks, take the *"first"* description (in alphabetic order), and sum the remaining fields.

4. Let's just pick out the three biggest problems:

View name: Q120WorstProblems

```
SELECT * FROM Q110LostThruputForL1P11Summary
ORDER BY TotalLostThruputMinutes DESC
LIMIT 3
```

5. The codes may provide some additional useful information. For example, what are the problem types? We connect the above results with tables "ProblemType" and "ShiftNames" to find out the details about the problem

View name: Q130WorstProblemsWithExplanations

```
SELECT Q120.*,
       ProblemType.ProblemTypeDescription, ShiftNames.Shift
  FROM Q120WorstProblems AS Q120
       LEFT JOIN
       ProblemType ON Q120.ProblemType = ProblemType.ProblemType
       LEFT JOIN
       ShiftNames ON Q120.ShiftID = ShiftNames.ShiftID
```

6. You are in the elevator with the plant manager. What do you tell her about opportunities to improve her throughput?_____

Review

We have used aggregation queries to analyze the data we are given. We identified the area, work center, and product which offers the greatest opportunity for throughput improvement as well as the top few problems which should be addressed immediately to improve throughput.