Exercise 1

FIGURE 1 The File Structure for Problems 1-3

PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	16833460.00
25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	12500000.00
25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	32512420.00
25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	21563234.00
27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	10314545.00
29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	25559999.00
31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	56850000.00

1. **How many records does the file contain? How many fields are there per record?**

   The file contains seven records (21-5Z through 31-7P) and each of the records is composed of five fields (PROJECT_CODE through PROJECT_BID_PRICE.)

2. **What problem would you encounter if you wanted to produce a listing by city? How would you solve this problem by altering the file structure?**

   The city names are contained within the MANAGER_ADDRESS attribute and decomposing this character (string) field at the application level is cumbersome at best. (Queries become much more difficult to write and take longer to execute when internal string searches must be conducted.) If the ability to produce city listings is important, it is best to store the city name as a separate attribute.

3. **What data redundancies do you detect? How could those redundancies lead to anomalies?**

   Note that the manager named Holly B. Parker occurs three times, indicating that she manages three projects coded 21-5Z, 25-9T, and 29-2D, respectively. (The occurrences indicate that there is a 1:M relationship between PROJECT and MANAGER: each project is managed by only one manager but, apparently, a manager may manage more than one project.) Ms. Parker's phone number and address also occur three times. If Ms. Parker moves and/or changes her phone number, these changes must be made more than once *and they must all be made correctly... without missing a single occurrence*. If any occurrence is missed during the change, the data are "different" for the same person. After some time, it may become difficult to determine what the correct data are. In addition, multiple occurrences invite misspellings and digit transpositions, thus producing the same anomalies. The same problems exist for the multiple occurrences of George F. Dorts.

FIGURE 2 The File Structure for Problems 4-6

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CODE	JOB_CHG_HOUR	PROJ_HOURS	EMP_PHONE
1	Hurricane	101	John D. Newson	EE	85.00	13.3	653-234-3245
1	Hurricane	105	David F. Schwann	CT	60.00	16.2	653-234-1123
1	Hurricane	110	Anne R. Ramoras	CT	60.00	14.3	615-233-5568
2	Coast	101	John D. Newson	EE	85.00	19.8	653-234-3254
2	Coast	108	June H. Sattlemeir	EE	85.00	17.5	905-554-7812
3	Satellite	110	Anne R. Ramoras	CT	62.00	11.6	615-233-5568
3	Satellite	105	David F. Schwann	CT	26.00	23.4	653-234-1123
3	Satelite	123	Mary D. Chen	EE	85.00	19.1	615-233-5432
3	Satellite	112	Allecia R. Smith	BE	85.00	20.7	615-678-6879

4. **Identify and discuss the serious data redundancy problems exhibited by the file structure.**

Given the file's poor structure, the stage is set for multiple anomalies. For example, if the charge for JOB_CODE = EE changes from $85.00 to $90.00, that change must be made twice. Also, if employee June H. Sattlemeier is deleted from the file, you also lose information about the existence of her JOB_CODE = EE, its hourly charge of $85.00, and the PROJ_HOURS = 17.5. The loss of the PROJ_HOURS value will ultimately mean that the Coast project costs are not being charged properly, thus causing a loss of PROJ_HOURS*JOB_CHG_HOUR = 17.5 x $85.00 = $1,487.50 to the company.

Incidentally, note that the file contains different JOB_CHG_HOUR values for the same CT job code, thus illustrating the effect of changes in the hourly charge rate over time. The file structure appears to represent transactions that charge project hours to each project. However, the structure of this file makes it difficult to avoid update anomalies and it is not possible to determine whether a charge change is *accurately* reflected in each record. Ideally, a change in the hourly charge rate would be made in only one place and this change would then be passed on to the transaction based on the hourly charge. Such a structural change would ensure the historical accuracy of the transactions.

You might want to emphasize that the recommended changes require a lot of work in a file system.

5. **Looking at the EMP_NAME and EMP_PHONE contents, what change(s) would you recommend?**

A good recommendation would be to make the data more *atomic*. That is, break up the data componnts whenever possible. For example, separate the EMP_NAME into its componenst EMP_FNAME, EMP_INITIAL, and EMP_LNAME. This change will make it much easier to organize employee data through the employee name component. Similarly, the EMP_PHONE data should be decomposed into EMP_AREACODE and EMP_PHONE. For example, breaking up the phone number 653-234-3245 into the area code 653 and the phone number 234-3245 will make it much easier to organize the phone numbers by area code. (If you want to print an employee phone directory, the more atomic employee name data will make the job much easier.)

6. Given your answer to Problem 4, what new files should you create to help eliminate the data redundancies?

The data sources are probably the PROJECT, EMPLOYEE, JOB, and CHARGE. The PROJECT file should contain project characteristics such as the project name, the project manager/coordinator, the project budget, and so on. The EMPLOYEE file might contain the employee names, phone number, address, and so on. The JOB file would contain the billing charge per hour for each of the job types – a database designer, an applications developer, and an accountant would generate different billing charges per hour. The CHARGE file would be used to keep track of the number of hours by job type that will be billed for each employee who worked on the project.