Software Evolution

Date: Thursday 28th May 2009

Time: 14:00 – 16:00

Please answer any THREE Questions from the FOUR questions provided

This is a CLOSED book examination

The use of electronic calculators is NOT permitted
1. a) Researchers have classified the software changes observed to occur into four kinds: corrective, adaptive, perfective and preventive. Which type or types of software change are being carried out in the following situations? Give a brief justification for your answer in each case.

i) A key program used by your company is continually subject to problem reports from users. The program is written in COBOL and your one remaining COBOL expert is about to retire. As lead maintenance officer, you decide to use a software tool to translate the code to Java. You then ask one of your most talented object-oriented programmers to look at the translated code and to ensure that sensible design patterns are used throughout.

ii) After a series of complaints from users that important transactions now run very slowly, you decide to migrate your company’s main sales system from the legacy CODASYL database on which it currently operates to a more modern relational database system that can take advantage of RAID arrays (parallel arrays of disks, giving fast performance). One of your developers is initially sceptical, but later comments on how much simpler and easier to understand the database queries have become in the new system.

iii) Your bosses have made a strategic decision that all purchasing of parts for your company’s products will now be done over the internet, using business-to-business e-commerce technology. You are given the job of ensuring that the warehouse inventory system can interface with the new parts purchasing system. While working on the conversion you realise that it will be possible to generate several new reports for management on usage and productivity of the warehouse, which they have been requesting for some time. You gain funding for these new reports and put your team to work to implement them.

(6 marks)
(Question 1 continues from the previous page)

b) Draw a data dependency graph for the following fragment of code:

```java
public static int wc(String file) throws Exception {
    int lc = 0;
    int wc = 0;
    int cc = 0;
    FileReader fr = new FileReader(file);
    boolean ws;
    String whitespace = " \t\n\r";
    int ch = 0;
    while ((ch = fr.read()) != -1) {
        cc++;
        if (ch == '\n') {
            lc++;
        } else {
            if (whitespace.indexOf(ch) == -1) {
                if (ws) {
                    ++wc;
                } else {
                    ws = true;
                }
            } else {
                ws = false;
            }
        }
        if (cc != 0) {
            ++lc;
        }
    }
    return wc;
}
```

(10 marks)

c) Either the process of generating the DDG or its final form should have indicated a number of problems or inelegancies within the code given in b). Describe any two of these, and point out the features of the DDG that highlight them. (4 marks)
a) The F&W Company manufactures flunges and worbles for the emerging hi-tech omchip industry. You have recently taken a job as team leader for the main sales and production systems for this company, having never worked in the omchip industry before. Describe the top-down code reading strategy, and explain the difficulties you are likely to encounter in applying this strategy when first attempting to understand the source code of these systems. What action could you take to make your job easier in future? (5 marks)

b) In dealing with the first bug report assigned to you, you narrow the problem down to the following fragment of code:

```
WorkQueue queueForToday = new WorkQueue();
WorkQueue queueForLater = new WorkQueue();
FWDate today = FWDate.today();
Worker[] workers = Factory.onDuty(today);
int numWorkers = workers.length;
for (int i=0; i<numWorkers; i++) {
    workers[i] = new Worker(queueForToday);
    workers[i].start();
}

Task[] flungeTasks = Factory.tasksToComplete(Flunge);
for (Task task : flungeTasks) {
    Worker i = bestWorker(workers, task);
    if (i == 0)
        queueForLater.addWork(i, task);
    else
        queueForToday.addWork(i, task);
}

Task[] worbleTasks = Factory.tasksToComplete(Worble);
for (Task task : worbleTasks) {
    Worker i = bestWorker(workers, task);
    if (i == 0)
        queueForLater.addWork(i, task);
    else
        queueForToday.addWork(i, task);
}

Iterator workersIter = workers.iterator();
while (workersIter.hasNext()) {
    Worker w = workersIter.next();
    if (w.fullLoad())
        queueForToday.addWork(w, Worker.NO_MORE_WORK);
}
```

What domain knowledge (i.e. information about the business operations of the F&W company) can you guess at from this fragment of code? In your answer, you should give at least 3 hypotheses about the workings of the F&W factory, justifying each by pointing out the features of the code that led you to infer them. (6 marks)

(Question 2 continues on the following page)
c) Point out at least 3 examples of the standard idioms covered in the course that you can see in the fragment of code given in b). For each, give the line numbers at which the idiom appears (start and end), and a hypothesis as to the high-level behaviour being implemented by the idiom. (6 marks)

d) What can we infer about the strategic priorities of the F&W management from the separate handling of flunge tasks and worble tasks in the code given in b)? (3 marks)

3. FirstShark Finance offers loans and insurance to small- to medium-sized businesses. This is a market that has become increasingly competitive in recent years, with the need to introduce new and innovative products becoming ever more pressing. The IT division of FirstShark has struggled to keep up with the need for change, due to its reliance on two legacy systems, dating from the time of FirstShark’s creation as a merger of two separate companies, more than a decade ago. The systems are old, complicated and difficult to change without jeopardising core revenue-generating processes.

The recent business climate has put even more pressures on the company. Government have been introducing ever more stringent regulatory and legal requirements with severe financial penalties for non-compliance; senior management at FirstShark predict that more changes of this kind are on the way. In addition, the company has suffered from poor decision making in the past regarding loans and insurance sales to unreliable customers. New risk management procedures are being designed for the company, and software support for these is now urgently required if the company is to avoid bankruptcy within the next decade.

To meet these needs, management have agreed to fund a major re-engineering project to create new computer systems that will meet the company’s needs. You have been put in charge of the team responsible for planning and undertaking this re-engineering project. Your task is to bring the required new functionality on-line as soon as practicable, while allowing the current business of the company to proceed unimpaired.
a) As a first step, you ask your team to determine the architecture of the current computer systems. The result is shown below:

As the diagram shows, the company currently operates using two systems, one for loans and one for insurance sales. The systems were originally completely independent, but at some time in the past, in order to provide a consistent view of the current customers across the two systems, specialised modules were created to allow each system to access customer data stored in the other.

The older Loans system is implemented in COBOL with a CODASYL database. The more recent Insurance system is implemented in a 4th generation language, specialised for access to data through its proprietary data management server technology.

How would you characterise this architecture in terms of its decomposability? (2 marks)
b) Having analysed the existing and new functionality required, your team have come up with the following target architecture:

![Architecture Diagram]

In this architecture, all data (whether regarding products, customers, or sales) is consolidated into one database, built on the most up-to-date relational technology. New GUIs are provided for the existing sales clerks, plus new components implementing the business logic relating to the sales of insurance and loan products. Furthermore, business logic for managing customers has been separated out into a further layered component, for better code reuse and ease of extension.

(Question 3 continues on the following page)
A new risk management component has been created, to encapsulate the new decision making procedures designed for the company, along with new logic protecting against non-compliance with statutory regulations. Full implementation of this component will require the storage of new types of data in the database. This new component is also the foundation for a completely new group of functionalities that will be provided by the new system: a fully-featured management information system (MIS), allowing senior management to track decision making and to monitor the success of the new decision making procedures. The MIS will also provide functionality to allow senior management to verify that company procedures and practice do not violate statutory regulations. A new business logic component and a new GUI is included in the architecture as the home for the new functionality.

Finally, a data access layer is included, to protect the business logic from changes to the underlying data structures.

Based on these considerations, discuss the suitability of each of the three migration strategies (i.e. forward migration, reverse migration and general migration) for this re-engineering project. State which of the three, after your analysis of each, you would adopt for this project. (9 marks)

c) Using your preferred migration strategy from part b) of the question, design a migration path for this system that brings the important functionality on-line soonest, while maintaining existing mission-critical functionality as necessary. For each migration step, you should:

- Indicate clearly the changes to the components and the changes to, and placement of, any gateways that have been introduced in this step. This can be done by drawing partial architecture diagrams, showing the changes, for example.

- Justify your choice of component to migrate in terms of the benefits, costs and risks of the step, and the business priorities of the company.

At each stage, your answer should clearly distinguish legacy components from target components, and forward gateways from reverse gateways. State any additional assumptions you need to make about the company, its policies or its IT infrastructure. (9 marks)
4. The ACME company produces bottled drinks. A number of suppliers provide various items: bottles, caps, and some of the ingredients needed to prepare the drinks. ACME keeps a stock of its supplies in an automated warehouse: as soon as it enters the warehouse, each item delivered by a supplier is equipped with an RFID tag, a small electronic sticker that can be detected when it passes through certain gates that are placed at various points within the warehouse. Using this technique, ACME can track both the stock levels and the progress of each item as it moves through the warehouse to make its way to the production line.

Information about these items and how they move around is stored in multiple databases, as shown in the figure below:

For simplicity we only consider items of type B (bottle) and M (short for “magic ingredient”). Database DB_B stores detailed information for each bottle. Database DB_Bmove stores all the events that pertain to bottles, for instance “B1 left room X”, “B1 entered room Y”, etc. Databases DB_M and DB_Mmove store the same type of information, but relative to items of type M.

During a normal 8 hour day, new B items arrive at an average rate of 30 per minute and they all move through 4 gates by the end of each day. And, on average, items of type M arrive at a rate of 1 per minute, and also move through the 4 gates during the day.

The four databases are frequently queried to monitor the status of the stock and of the production line.

ACME also keeps an additional database, called H (for History), into which the content of all 4 databases is periodically copied. Database H is used to compute statistics on the overall production over long period of times, for instance to determine how many units of M were used to produce a certain number of bottles.

(Question 4 continues on the following page)
The main requirements for data migration are:

- daily copies of new items from DB_B and DB_M should be available in H at the beginning of each day;
- weekly copies of new items from DB_Bmove and DB_Mmove should be available in H at the beginning of each week;
- some data transformation is needed to map the data in each of the four databases, to fit the schema of H.

The following questions should all be answered in the context of the above scenario.

a) ACME data managers need to decide on a strategy for periodically migrating data from the four primary databases to database H. In this scenario, DB_B and DB_Bmove are relational databases with trigger capability, while DB_M and DB_Mmove are legacy systems with no triggers, but with a full log of all their insertions. Bulk loading (ETL) tools are available to feed H.

For each of the three main data migration strategies (push, pull and remote/hybrid ETL), describe its suitability for use in this particular scenario. Based on this characterisation of each strategy, state which one you would choose to implement and justify your choice. (8 marks)

b) The RFID technology used by ACME is not very reliable, and the data feed into the four primary databases is not always correct. The following problems may occur: tags sometimes fail to be read, sometimes they are read more than once, and sometimes the system reports the wrong ID for a tagged item.

For each of these issues, describe the potential consequences for applications that use database H, and name a corresponding data quality dimension (from those described in the lectures) that applies to data in H because of it. (6 marks)

c) How could ACME data engineers address each of the data quality problems mention in part b)? For each problem, describe the enhancements or additions to the data architecture shown in the figure that could alleviate it, giving a brief explanation of the architecture change and how it will deal with the data quality issue. (6 marks)

END OF EXAMINATION