**Exercise 4. Perform Capacity Planning**

**Name: \_\_\_\_\_\_\_\_\_Zou Xuan\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_2019-Mar-9\_\_\_\_\_\_\_\_\_\_**

1. **Throughput Calculation and Projection**

**First**, identify which machine would directly serve the data entry done by the nurses and doctors.

* How many nurses and doctors does that machine have to serve?
* How many data entries would they do in a day? (entering weight, height, observations, etc)
* How much throughput in term of data entries per minute does the machine have to handle? Please note the number of hours per day and make assumption on how long would they work. Consider also peak vs off-peak behavior.

|  |  |  |
| --- | --- | --- |
|  | **Daily throughput** | **Per second/minute throughput** |
| **Screening Data Entry** | **54000** | **1.875 tps** |
| **Calculation** | 1. The machine has to serve 5\*300 nurses, 1\*300 doctors. 2. One team (5 nurses+1doctor) would enter 6 classes of maximum 30 students screening data per day. So, the daily throughput is 6\*300\*30=54000 3. They work from 8am to 5pm, lunch break 12pm-1pm. Hence the total working hours per day is 8. Throughput per minute = 54000/8/60/60=1.875 tps | |

Secondly, consider whether the assumption that you have made in the calculation would change in the subsequent 3 years, and circle your choice below (option a or b)

1. If you think that it is quite likely that the assumption will not change, then the estimation throughput that you’ve calculated can still be used for the subsequent 3 years.
2. If you think that it is more likely that the assumption changes, calculate the new throughput 3 years later and write it down at the space below

|  |  |
| --- | --- |
|  | **Per second/minute throughput** |
| **Student data entry throughput after 3 years** | **1.9897tps** |
| **Calculation: Based on historical data analysis, there number of students has roughly increment of 2% each year. We can reasonably assume the data will keep increasing 2% each year.**  **3 years later, the daily throughput will be 54000\*1.02\*1.02\*1.02=57305.232. Throughput per minute will be 57305.232/8/60/60=1.9897tps** | |

1. **Throughput Calculation and Projection**

It is reasonable to assume that SHB need to update its student record data from the latest master student records in mainframe. In this particular case, the case study cites the limitation in mainframe that the mainframe can only prepare a file which contains all of the student records (more than 10 million – refer to the table in the case study for actual figure).

One of the common design for the student download would follow the design below:

Student Record Update

FTP Server

Mainframe

SHB DB

Our task is to calculate the throughput required for “Student Record Update” batch job. In order to do this, you need to get the number of records (assume 1 student would have 1 student record) and the batch window available to run our batch job.

For this exercise, you can make your own assumption on the batch window that you think is reasonable. In real projects, we would need to orchestrate batch window for various batch job.

Once you have the number of records and the batch window, it should be quite trivial to calculate the expected throughput of the batch job in term of how many record it need to process every minute.

|  |  |
| --- | --- |
|  | **Student record/minute throughput** |
| **Mainframe Record Processing** | **487.2 per second** |
| **Calculation**  **Mainframe will start process the files at 9pm and ends before 3am, the total processing time is 6 hours per day, 2014 total student records is 10,523,298 =10,523,298**  **The throughput of mainframe record processing is 10,523,298 /6\*60\*60 =487.2 per second.** | |

Again, consider whether the assumption that you have made in the calculation would change in the subsequent 3 years, and circle your choice below (option a or b)

1. If you think that it is quite likely that the assumption will not change, then the estimation throughput that you’ve calculated can still be used for the subsequent 3 years.
2. If you think that it is more likely that the assumption changes, calculate the new throughput 3 years later.

|  |
| --- |
| **3 years later, the daily throughput becomes 6\*30\*300\*1.02\*1.022\*1.02=57305, assume there are 250 working days per year.**  **There will be 57305\*260=14,899,300**  **The throughput of mainframe record processing will be 14899300/6\*60\*60 =689.8 per second.** |

1. **Disk Requirement for the next 5 years**

* Calculate a ballpark figure (upper limit) on how much space a student screening information entered by the nurses and doctors would need based on the E-R diagram in the case study.
* Calculate the disk requirement to store the **screening data and its growth** over 5 years

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Year 1** | **Year 2** | **Year 3** | **Year 4** | **Year 5** |
| **Disk Requirement** | **22.87 GB** | **36 GB** | **49.39 GB** | **63.05 GB** | **76.98 GB** |

**Calculation**

|  |
| --- |
| **1.**   1. **Student table: 10+10+4+12+1+1+8+9+30+8=93 bytes (1 record per student)** 2. **Drug Allergy: 6\*5=30 bytes (assume 2 records per student)** 3. **Basic measurement: 6\*3+2\*4=26 bytes (1 record per student)** 4. **Appointment request=301=10+4=14 bytes (1 record per student)** 5. **Health Assessment: 10+2\*4=18 bytes (1 record per student)** 6. **Medical history: 4 bytes (1 record per student)** 7. **Immunization Not Performed: 10+4+2+10+60=86 bytes (assume 3 records per student)** 8. **Immunization Performed: one record 10+4+2+4+60=80 bytes (6 record per student)** 9. **School and Drug Code table no need, as it’s configured for all students.** 10. **One student will roughly contain 93+30\*2+26+14+18+4+86\*3+80\*6=953 bytes.**   **2.**   1. **13,500,000 students per year, assume one student need 1 KB space.** 2. **Year 1 will need (13,500,000\*1 KB=12.87 GB) + (10,523,298(historical data) \*1 GB)=10 GB=22.87 GB . Below assume 2% student increasement each year** 3. **Year 2 =12.87\*1.02+ Year1 =36 GB** 4. **Year 3= 12.87\*1.02\*1.02+ Year2 =49.39** 5. **Year 4= 12.87\*1.02\*1.02\*1.02+ Year3 =63.05** 6. **Year 5= 12.87\*1.02\*1.02\*1.02\*1.02+ 1 Year4 =76.98** |

1. **Hardware/Software Requirement**

Calculate the number of hardware and software required for the project. Use the throughput from section 1 and 2 to guide your estimates on the server size and refer to your physical design to put down the hardware/software requirement.

|  |  |  |
| --- | --- | --- |
| **Hardware/Software** | **Quantity** | **Note/Justification** |
| **client PCs**  **Load Balancer server**  **Application servers**  **Database servers**  **DNS Server**  **Web Server**  **FTP Server** | **900**  **1**  **2**  **2**  **1**  **1**  **1** | **Each team 5 nurses+1 doctor, assume 2 staffs use 1 PC. 300 teams would require 300\*6/2=900**  **To provide High availability**  **To provide High availability**  **To provide High availability** |

1. **Procurement Plan**

Create a procurement lead time matrix for SHB. Put your assumption for the project. You can use the typical procurement lead time needed for the organization you worked for.

|  |  |  |
| --- | --- | --- |
| **Procurement Plan for** | Client PC |  |
|  |  |  |
| **Process** | **Source** |  |
| Internal procurement process (raise request to PO) | Internal organization | (2 working days) |
| Lead time from PO to delivery | Hardware vendor | (4 weeks) |
| Hardware installation time | Hardware vendor | (1 week) |
| Software installation time | Software vendor/in house IT | (1week) |
| Production setup time | In house IT | (1 day) |
| Testing | In house IT/user | (0.5 day) |
|  |  |  |
|  |  |  |
| **Planning buffer** | | 1 week |
| **TOTAL** | | 7 weeks 3.5 days |
|  |  |  |
| Estimated downtime/reduced SLA | | 1 day |
| Lead time for downtime approval | | 1 day |

1. **Metrics**

Identify 3 application metrics and 3 system metrics to be monitored that you think will be helpful in providing input for future capacity planning.

Propose tools to monitor them (e.g. : HP OpenView, IBM Tivoli, Windows Performance Monitor, Unix shell tools, Application self-logging, etc)

|  |
| --- |
| **Application Metrics** |
| **1.Screening data entered per hour** |
| **2.Number of disconnected mode data per hour** |
| **3.Number of Reporting request per day** |

|  |  |
| --- | --- |
| **System Metrics** | **Tools** |
| **1. CPU usage** | IBM Tivoli |
| **2.RAM usage** | IBM Tivoli |
| **3. I/O wait** | IBM Tivoli |