PERFORMANCE TESTING Of BATCH JOBS

By
Rupesh Garg
Performance Testing Practise

© 2010 Wipro Ltd - Confidential

Contents

© 2010 Wipro Ltd - Confidential

Contents

4.0 PERFORMANCE TUNING BATCH JOBS
4.1 REDUCING THROUGHPUT
4.2 SLOW THROUGHPUT
4.3 INFREQUENT DIPS IN CPU UTILIZATION DELAYING BATCH
4.4 OUT OF MEMORY ISSUES
4.5 SQL QUERIES TIMING OUT
5.0 CONCLUSION
6.0 REFERENCES

Introduction

- **Batch Processing :**
  – Batch jobs are meant for processing huge volumes of data without human intervention and are generally scheduled to run during off hours so that resources are utilized to the best. Majority of the batch jobs are Mainframe based. But they can also be written using UNIX scripts and DOS commands. Data for the batch job is generally present in the database. Other methods include, fetching data by script from some file or passing data as command line arguments.

Importance of Batch Performance testing

WIPRO
Applying Thought

- Performance of batch jobs is equally or in many cases, more important than Web based jobs due to factors such as:
- Batch jobs should be tuned to utilize all the resources to the maximum during off hours.
- Irrespective of the volume of the data, performance of the batch should be good enough to complete the processing within the allowed time frame.
- Execution of batch job should never affect the performance of other processes.

5   © 2010 Wipro Ltd - Confidential

Batch performance Testing Process

WIPRO
Applying Thought

Compared to other applications, batch processes follow a different approach cum methodology for performance testing, points for which are listed below:

- •Requirements Collection
- •Performance Testing Setup
- •General Performance issues

6   © 2010 Wipro Ltd - Confidential

Requirements Collection

- There are many factors to be observed and collected before starting Performance test for a batch job.
- Timings of the batch run - When is the batch expected to run in production?
- Time duration (hours) for which batch would run.
- Architectural design of a batch job. The different methods and DB calls involved would need to be known. For eg: Server mappings to each of the Methods or POJO or Web Service or DB calls.
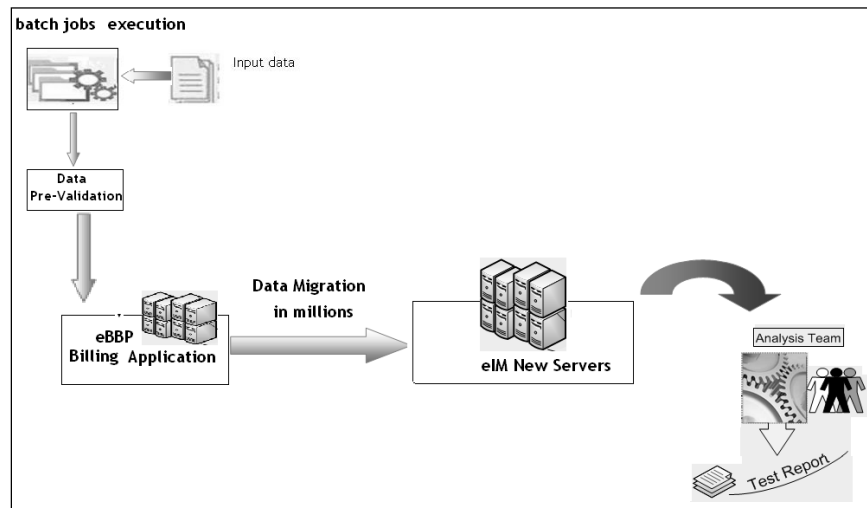
7   © 2010 Wipro Ltd · Confidential

Requirements Collection

Sample Architectural Diagram

batch jobs execution

Input data

Data Pre-Validation

Data Migration in millions

eBBP Billing Application

eIM New Servers

Analysis Team

Test Report

8   © 2010 Wipro Ltd · Confidential

Performance Testing Setup

- An environment similar to production environment has to be set up to performance test the batch job. This will include a similar system configuration set up, all the expected concurrent processes running and a similar data setup as in production.
- In most of the cases since replicating production like environment is expensive a relatively less configured system setup is made for testing and proportionate changes are made in the expected throughput during the test

© 2010 Wipro Ltd - Confidential

Counters Monitored

- All the system resources have to be efficiently and optimally utilized to obtain best performance from batch process. So, all the system level statistics has to be monitored. In addition to this all the application level statistics also will be monitored as required.
- The main monitoring counters include:
  - **Throughput**
  - **CPU & Memory Utilization**
  - **DB Query Response Time**
  - **Application level statistics**

© 2010 Wipro Ltd - Confidential

Counters Monitored - Throughput

- No. of records processed per minute (second) statistics has to be collected. A script can be executed or a DB query lookup can be made to collect to collect throughput metrics of the batch process over the given time period.

11 © 2010 Wipro Ltd - Confidential

Counters Monitored -CPU & Memory Utilization

- These metrics should be always within the limit of what the batch job is expected to consume. Unlike other applications, high CPU of 90% is acceptable for batch process provided no other concurrent tasks will run and the application is finely tuned. This implies optimal utilization of resources by batch thus resulting in maximum throughput. In case, CPU utilization is higher than expected, it's always advisable to collect server level statistics which eases digging into the issue.

12

Counters Monitored - DB Query Response Time

- This is one of the most important monitoring counters to be captured while doing batch job performance testing. Most of the Batch jobs do DB updates. If the Query response time is low, any type of code or configuration changes won't be able to bring any commendable improvement in the batch performance.

© 2010 Wipro Ltd - Confidential

Counters Monitored - Application level statistics

- Need to keep track of no. of concurrent threads or sessions during the batch run. These values have to be increased or decreased as situation demands.

© 2010 Wipro Ltd - Confidential

Monitoring Tools

- For applications with Windows OS – perfmon is used to collect the System statistics. This includes, total CPU utilization, Memory utilization, process level breakdown of processor utilization, context switches, queue length etc.

15 © 2010 Wipro Ltd - Confidential

Monitoring Tools

- For UNIX and AIX there will be scripts executing commands in regular intervals to collect all the metrics during the test run. There are commands which can collect CPU, memory, network statistics, DB connections etc.

16 © 2010 Wipro Ltd - Confidential

Monitoring Tools

- Mercury J2EE diagnostics tool can be used to collect most of these statistics. It gives the response times of Web Service calls, EJB calls, query response times, breakdown of these response times across different tiers etc. It also provides info on Total memory and CPU utilization, concurrent threads, no. of database connections etc.

General Performance issues

- There are few performance issues which are frequently seen to occur during Performance testing batch process. Similar issues are found in batch processes across different applications because the behaviour and purpose these batch jobs are the same. Issues often seen are:

  - Slow throughput – Batch taking longer time than expected
  - Throughput reducing with time – No. of records being processed per minute is reducing with time.
  - High CPU utilization
  - Infrequent dips in CPU utilization – Thus delaying total batch process
  - SQL Queries timing out
  - Out of memory issues

Performance Testing Approach for Batch

WIPRO
Applying Thought

- The approach to be followed to Performance test batch jobs includes:
  – Defining scope for Performance testing
  – Feasibility Study
  – Determine test data requirements, batch execution methodology, data dependency, methods to create data, etc.
  – Identify metrics to be collected.
  – Set up an isolated batch testing environment.
  – Do necessary monitoring setup.
  – Create required data volume and run the batch.
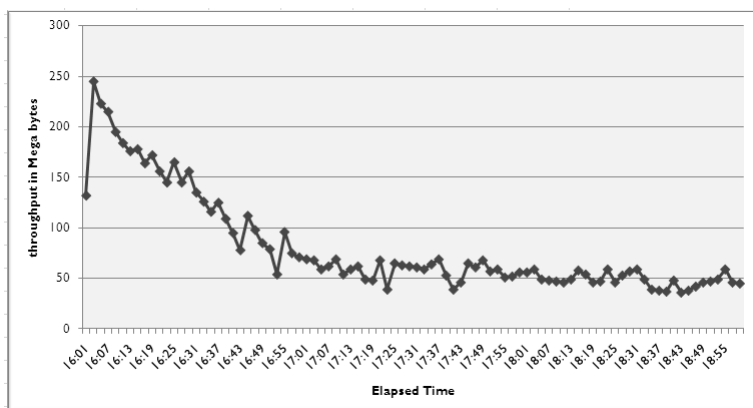 – Analyse the throughput and other system metrics and repeat the test if needed.

19 ©2010 Wipro Ltd - Confidential

Performance Tuning Batch jobs

WIPRO
Applying Thought

- **Reducing throughput**
  – Batch job works uniformly throughout the run. So it should give uniform throughput also. If the throughput is reducing with time, it can be either due to memory issues or specific queries leading to such ambiguous throughput.
  – If it's a memory issue, this throughput will be stabilized after memory reaching the required limit. Resolving such memory issues is explained in the below section.
  – In the second case where queries are the culprit, there will be few queries whose response time increases with time. This graph shows the inverse proportionality which holds good for query response
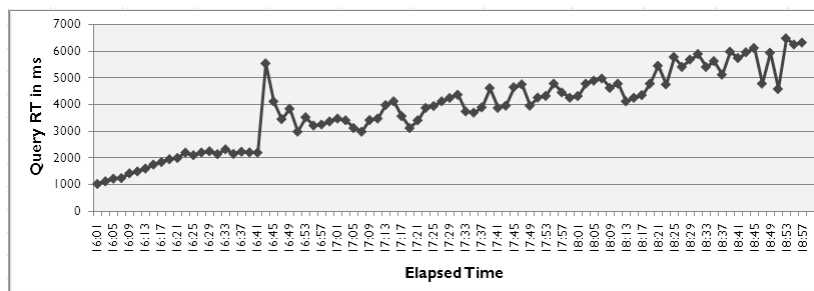
20 ©2010 Wipro Ltd - Confidential

Reducing throughput - Throughput

Reducing throughput - Query Response Time

Reducing throughput - CPU Utilization

WIPRO
Applying Thought

23   © 2010 Wipro Ltd - Confidential

SQL queries timing out

WIPRO
Applying Thought

- If the SQL queries are timing out and the database utilization statistics are very less, we can easily avoid time outs by increasing the no. of connections to database (provided the job is Multi-threaded). If database utilization is already maximum then the SQL queries and database table has to be tuned to respond faster or the worst case, Query 'time out' value has to be increased to prevent transaction failures.

24   © 2010 Wipro Ltd - Confidential
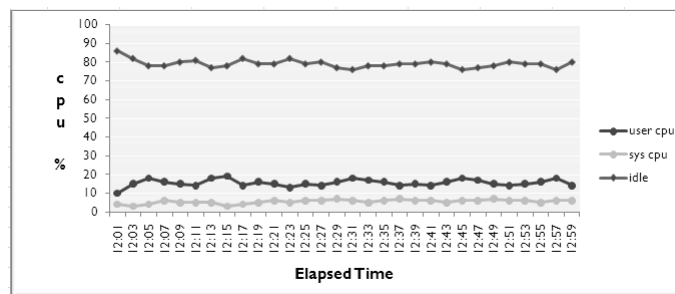
Slow throughput

- This is one of the most often faced issues during Performance testing batch jobs. This can be mainly because of any of the three reasons:
  – 1. Resources are not utilized effectively. CPU Utilization in this case will be lesser than limit. Increasing the no. of concurrent threads can easily help this situation. Threads should be increased considering current CPU and the CPU limit.
  – In the graph below, CPU utilized is only 20%. Throughput of the batch can be surely increased by better utilizing the system resources.
  – 2. Back ends are responding slow, which in turn slows down the whole batch process

25   © 2010 Wipro Ltd - Confidential

CPU Utilization

CPU utilization is high and queries are responding fast. But still if the throughput is less, then the application has to be profiled to find the methods causing high CPU with proper profiling tools like JProbe.
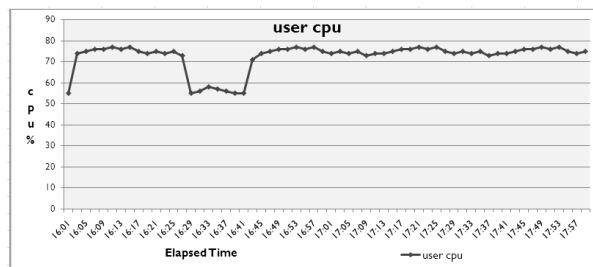
26   © 2010 Wipro Ltd - Confidential

Infrequent dips in CPU Utilization delaying batch

- If the CPU utilization goes down infrequently during the run, it will be due to slowly responding backend. Whenever backend responds slowly, it makes the processor wait for the response bringing down the CPU utilization. Data or queries taking long time have to be analysed and tuned to respond like other queries.

CPU Utilization

27 © 2010 Wipro Ltd · Confidential

Out of memory issues

- Out of memory issues in batch jobs are similar to other applications. Whenever the application hits out of memory, either the application will be in need for additional memory or there will be a memory leak in the application.
- If it's the first case then appropriate memory as required by the process has to be allocated. If it's a memory leak, the leaking objects in the batch process has to be identified and fixed using tools like Heap dump analyser, JProbe Memory debugger etc.
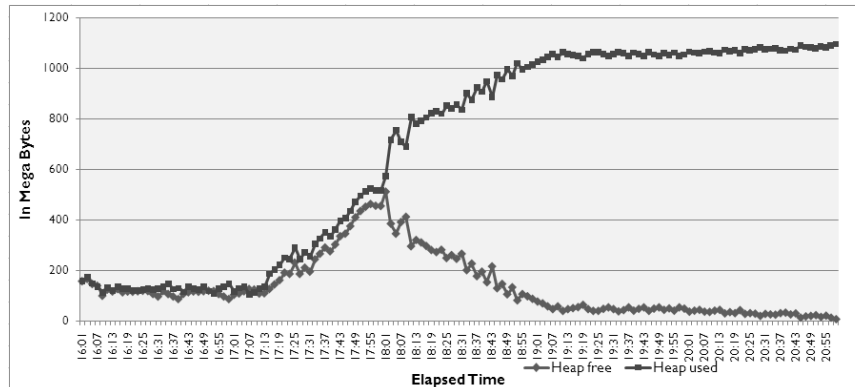
28 © 2010 Wipro Ltd · Confidential

Out of memory issues

Heap Utilization

29   © 2010 Wipro Ltd · Confidential

Conclusion

- Most of the times, applications are primary causes of performance bottlenecks which leads to inefficient use of the resources. Batch jobs performance testing much depends on the hardware utilizations, so by following performance optimization methods will always help in utilizing resources in efficient manner and thus make the jobs run at its peak efficiency.

30   © 2010 Wipro Ltd · Confidential

References

- www.oracle.com
- www.oracledba.in
- HP, LR Communities

Thank You