Managing Mainframe Costs in the Financial Services Sector

Mainframe software charges are usually based on peak usage. Whatever an organisations' peak workload in a given month is what they pay for.

Managing those workload peaks in the Financial Services sector can be a major challenge. Workload can suddenly spike during key times of the year. In the Banking sector consumer transactions increase dramatically during Christmas and Sales periods. Similarly in the Insurance sector, quotes rise when new car registrations are released and seasonal trending means that workload can peak during winter months. It is vital that financial services organisations are able to protect themselves against large spikes in usage and therefore unforeseen large Monthly Licence Charge (MLC) software costs.

Demystifying Mainframe Pricing

The key is to gain a real understanding of:

1. How Mainframe pricing works
2. When, where and why your workload peaks occur
3. How those peaks can be reduced

1 – How Mainframe pricing works

Every month an IBM workload report produced using the Sub-Capacity Reporting Tool ("SCRT") is produced and sent to IBM. This is used to determine the peak workload and therefore what charges will be applied for the software used. The small print in the SCRT report says:

For recurring charge (MLC) products, the data supplied in this report will be used to adjust the billable MSUs in the inventory for all MLC Products listed under the MLC Product Name column on this report. In accordance with our agreement, IBM will treat a change in product licensed capacity as an order. If the MSUs have changed since the last report, software billing based on inventory MSUs will increase or decrease accordingly

For One Time Charge (IPLA) products, the data supplied in this report will be used to bill those IPLA products listed under the IPLA Product Name column in this report which exceed your entitled capacity. In accordance with our agreement, IBM will treat the use of a product in excess of it's entitled capacity as an order and you will be billed for the amount in excess of your entitlement.

This means that peaks not only drive the Monthly Licence Charge (MLC) software costs, but also that any IPLA, also called One-time Charge Products (OTC), that have entitlements currently lower than the peak grow their entitlement to the new peak. IPLA products are then charged monthly at the new entitlement peak regardless of the actual usage in the coming months.
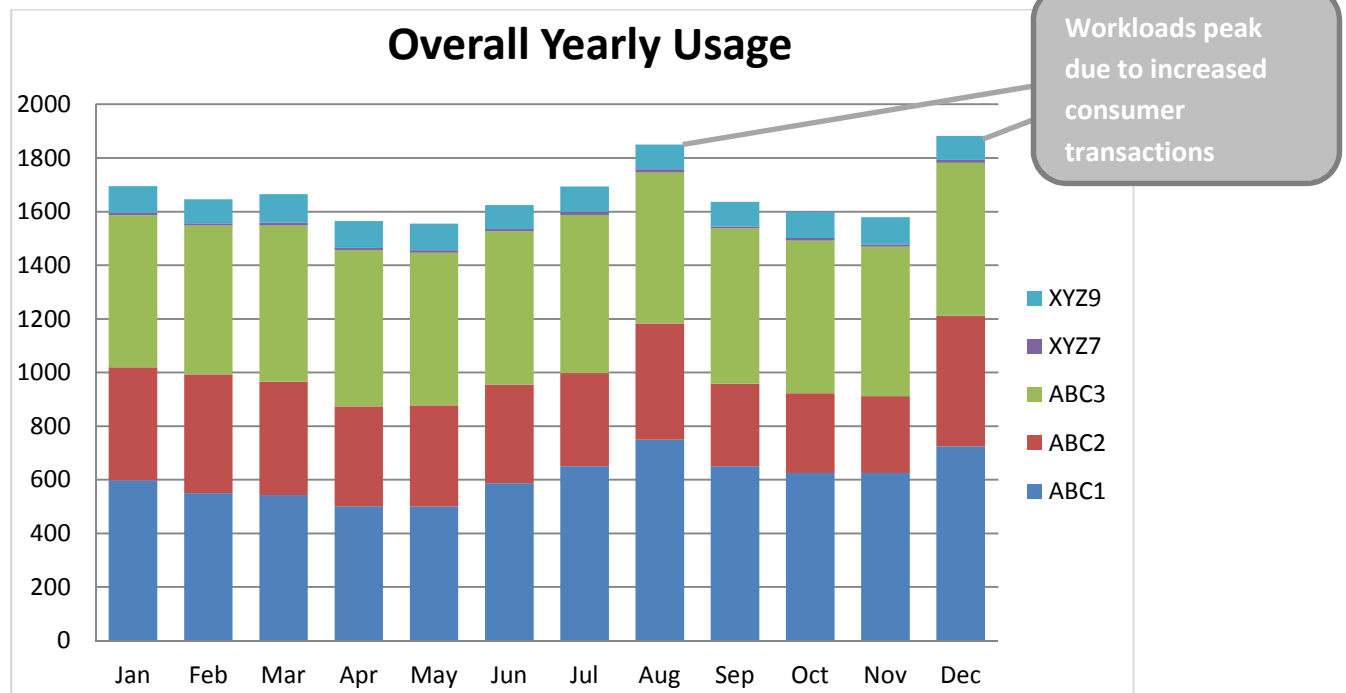
2 – When, where and why your workload peaks occur

It is not always obvious when, where or why workload peaks occur. Sometimes organisations are not aware that workload is peaking at a certain time or date, or if they are then they are not always aware of the reasons for the peak. By analysing SCRT and SMF data it is possible to get an understanding of when and where workload peaks are occurring. This analysis can often be difficult, especially in a complex environment with many z/OS logical partitions (LPARs) and many Mainframe customers do not undertake it on a regular basis. Triton's bespoke analysis tools can rapidly pinpoint the timing and location of each environment's monthly peak, providing rapid insight into monthly MLC costs.

Workload peaks can be caused by any number of external and internal forces:

High/fluctuating transaction volumes

The fluctuating nature of financial services transaction volumes can dramatically affect workload peaks throughout the year.

Overall Yearly Usage

Workloads peak due to increased consumer transactions

Legend: XYZ9, XYZ7, ABC3, ABC2, ABC1

Badly performing/slow running applications

Badly performing or slow running applications are another source of woe when it comes to pushing up Mainframe costs. The graph below shows an example of 4-Hour-Rolling-Average (4HRA) MSU usage by software Product. If the LPAR peak is 1900 MSUs this is what software products will be charged at. For example, if your CICS application has not been tuned as well as it could be or you have an issue with performance and it is using up more MSUs than it should then the entire peak will be raised and the associated software costs will increase to the peak. There are potentially significant savings to be made by ensuring that the system is running as efficiently as possible.

Usage by Product
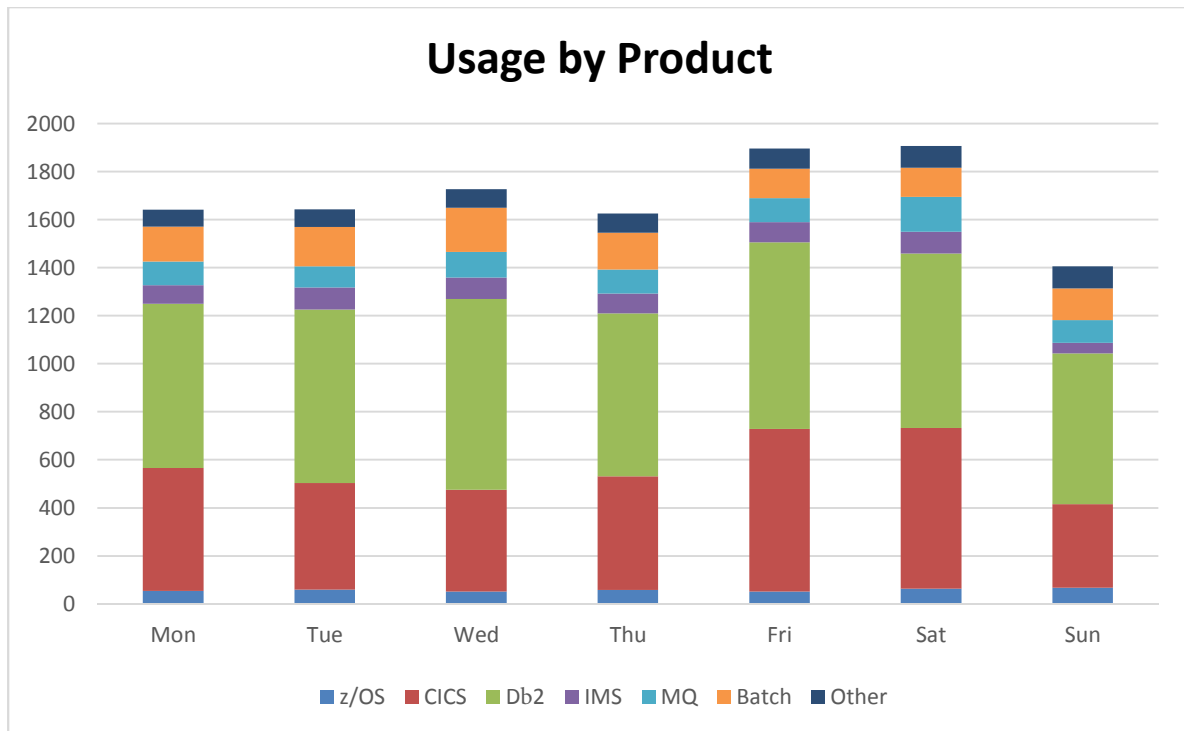
When looking at batch processing, an unmanaged workload mix can greatly affect whether batch processing is contributing to a large chunk of Mainframe software costs. By unmanaged workload mix we mean that sometimes, without realising it, organisations can be running non-essential batch processing jobs during the prime shift and pushing the workload peak up significantly. Every month an IBM workload report is produced using the Sub-Capacity Reporting Tool (SCRT) and sent to IBM. This is used to determine the peak workload and therefore what charges will be applied for the software used. By carefully analysing the SCRT report and related SMF data it is possible to gain a clearer view of where peaks are occurring. Moving batch workload to a different time may make it possible to bring peaks down and reduce the 4HRA and thus reduce costs.

4 Hour Avr Peak - Product Breakdown
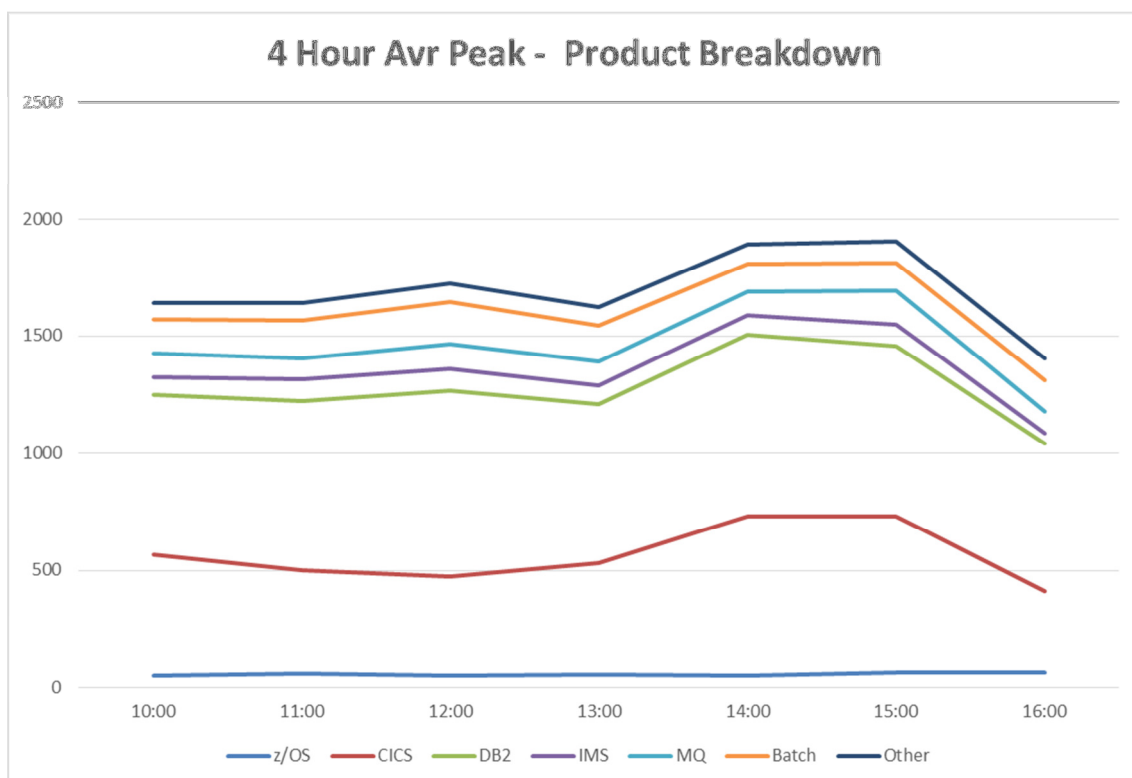
3 – How peaks can be reduced

There are three key areas to look at and Triton's zTune service covers each of these phases in detail:

Phase 1 – LPAR Optimisation
- Using 3-6 months of RMF data, comprehensive environment reports are produced that often highlight initial savings of 3-5%
- Analysis typically takes 4-6 weeks
- Possible savings of up to £1M – based on case studies

Phase 2 – Workload Optimisation
- Building on the analysis from Phase 1, Phase 2 drills further into the focus areas that are driving peaks.
- Using recent data (last 3 months) but trending data for 12+ months
- Analysis takes 3-6 months depending on environment complexity
- Typically drives an additional 5-10% saving
- Possible savings of up to £1.5M – based on case studies

Phase 3 – Workload Tuning
- Building on the previous analysis, focused workload tuning of peak CPU consumers yields further cost reductions
- Speed of implementation is set by the organisation
- Additional savings of up to £1M – based on case studies

zTune for Financial Services organisations

The fluctuating workload environment that Financial Services organisations operate in means that workload and therefore cost spikes can be a real issue for both CIOs and CFOs. Triton's zTune service implements controls that remove software cost surprises which gives robustness to both the capacity plan and the financial plan. Focused MSU reduction tuning can be translated into real cost savings.

Understanding workload profiles and taking control of peaks on a seasonal basis not only promotes good Capacity management disciplines but generates opportunities for cost reductions.

About Triton Consulting

Triton Consulting specialises in Data Management and has been an IBM Premier Business Partner since 1998. Specialising in Db2 for both the mainframe and distributed systems, Triton provides a full range of services from consultancy through to education and 24/7 Db2 support.

For more information visit www.triton.co.uk