#SHAREorg

SHARE
Technology · Connections · Results

# Fit for Purpose Platform Positioning and Performance Architecture

Joe Temple
IBM

Monday, February 4, 11AM-12PM
Session Number 12927

Insert Custom Session QR if Desired.

SHARE in San Francisco 2013

# Fit for Purpose Categorized Workload Types

## Mixed Workload – Type 1

- Scales up
- Updates to shared data and work queues
- Complex virtualization
- Business Intelligence with heavy data sharing and ad hoc queries

## Parallel Data Structures – Type 3

- Scales well on clusters
- XML parsing
- Buisness intelligence with Structured Queries
- HPC applications

*Application Function   Data Structure   **Usage Pattern   SLA   Integration   Scale***

## Highly Threaded – Type 2

- Scales well on large SMP
- Web application servers
- Single instance of an ERP system
- Some partitioned databases

## Small Discrete – Type 4

- Limited scaling needs
- HTTP servers
- File and print
- FTP servers
- Small end user apps

Black are design factors    Blue are local factors

Insert Custom Session QR if Desired.

# These do not define workload

- Languages
  - c/c++,COBOL,FORTRAN, JAVA, etc.
- Middleware
  - Oracle, DB2, Websphere, MQ, Tuxedo,CICS, Encina,etc.
- Workload Type
  - OLTP, Analytics, Business Applications, Infrastructure
  - *Mixed/Consolidated, Highly Threaded, Parallel, Small Discrete*

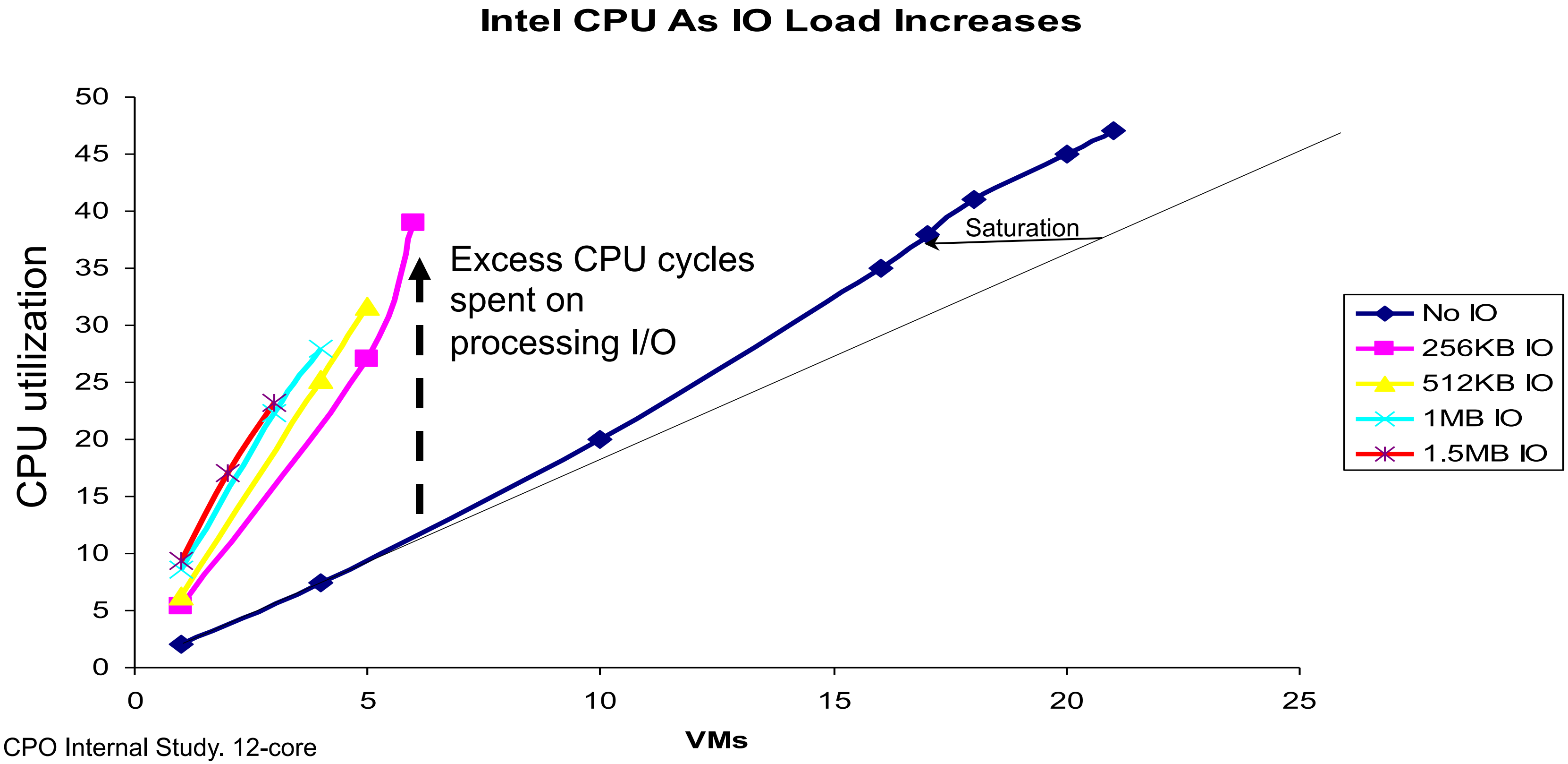
*Workload types can be used for positioning machines, but are not enough to guide platform selection*

# Workload Definition

- A workload consists of a workload type ***plus local factors:***
  - Usage Pattern
    - Load Variability
  - Scale
    - Size of load
  - Service Level
    - Response or turnaround expectation at load
  - Desired Efficiency
    - Target utilization level
  - Integration
    - Connections and shared data impacts

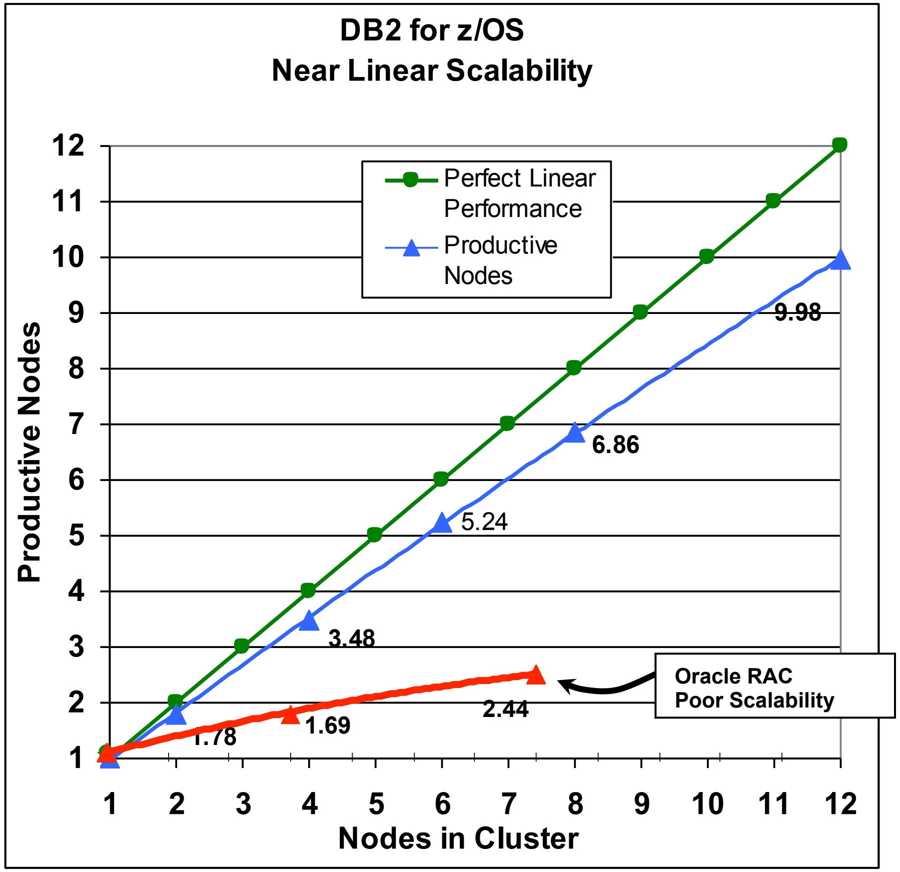
# x86 Performance Degrades As I/O Demand Increases

- Run multiple virtual machines on x86 server
- Each virtual machine has an average I/O rate
- x86 processor utilization is consumed as I/O rate increases



Source: CPO Internal Study. 12-core Westmere EP with KVM. FB at 22 tps with varying IO per transaction.

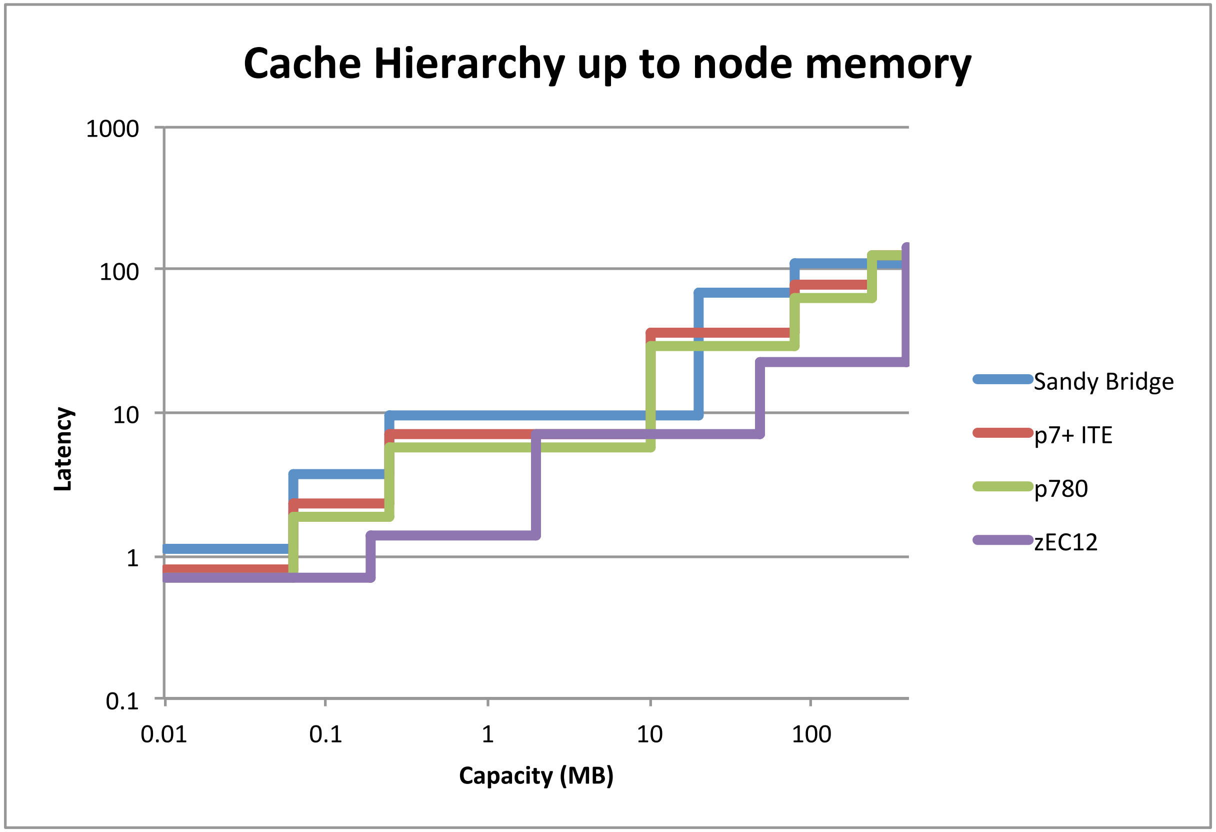
# Scaling Matters



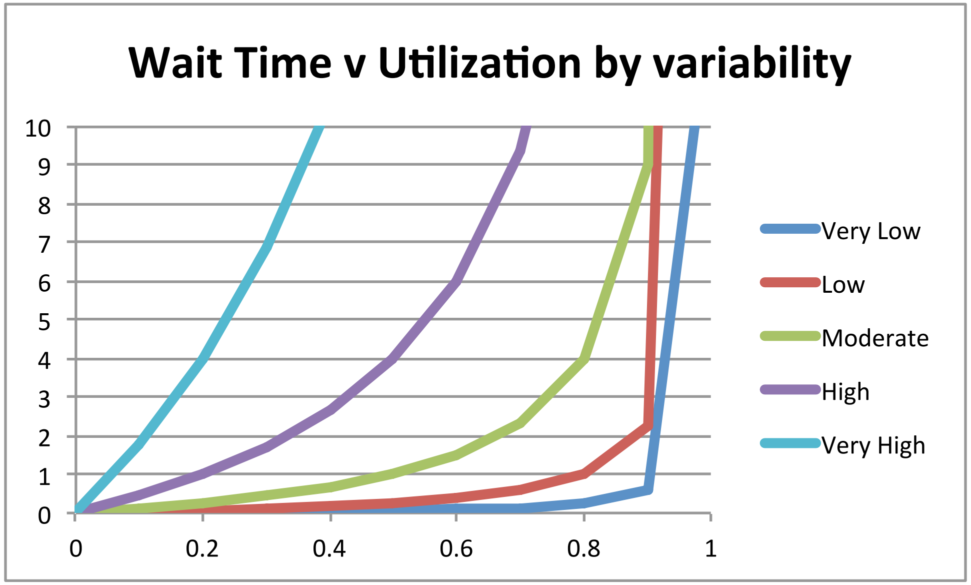
- Oracle RAC is inefficient by design
  - Network based lock and buffer management
  - Scaling RAC requires complex tuning and partitioning
  - Application partition awareness makes it difficult to add or remove nodes
- Published studies demonstrate difficult or poor scalability
  - Dell (shown in chart): Poor scalability despite using InfiniBand for RAC interconnect
  - CERN: Four month team effort to tune RAC, change database, change application
  - Insight Technology: Even a simple application on two node RAC requires complex tuning and partitioning to

Oracle RAC characteristics as shown in Dell RAC InfiniBand Study http://www.dell.com/downloads/global/power/ps2q07-20070279-Mahmood.pdf
CERN (European Organization for Nuclear Research) http://www.oracleracsig.org/pls/apex/RAC_SIG.download_my_file?p_file=1001900
Insight Technology http://www.insight-tec.com/en/mailmagazine/vol136.html

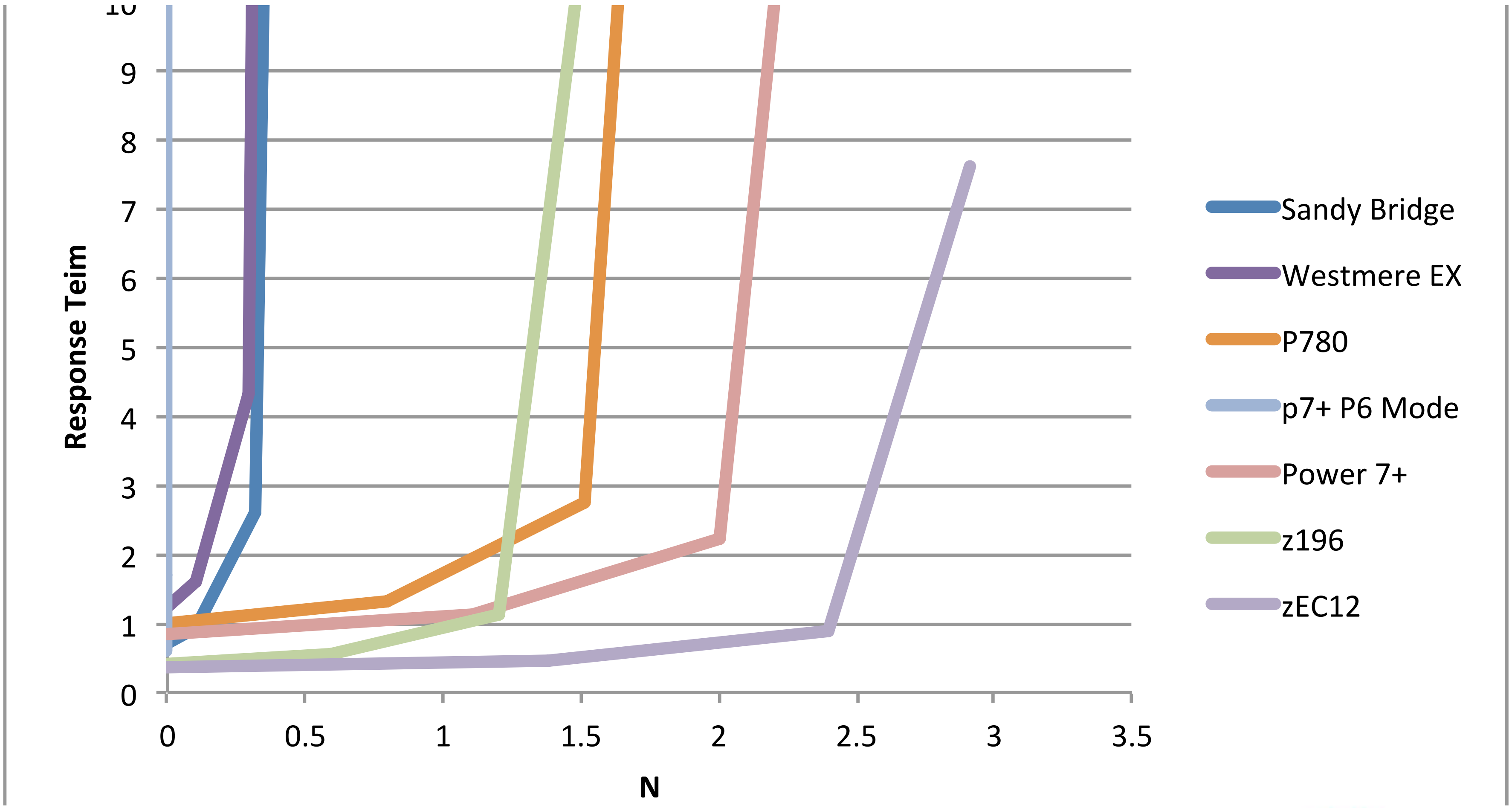
# Cache Working Set Matters



# Queuing and Load Variability Matter



# Response time and Consolidation matter



Each machine partition is three dedicated cores, number of loads varies

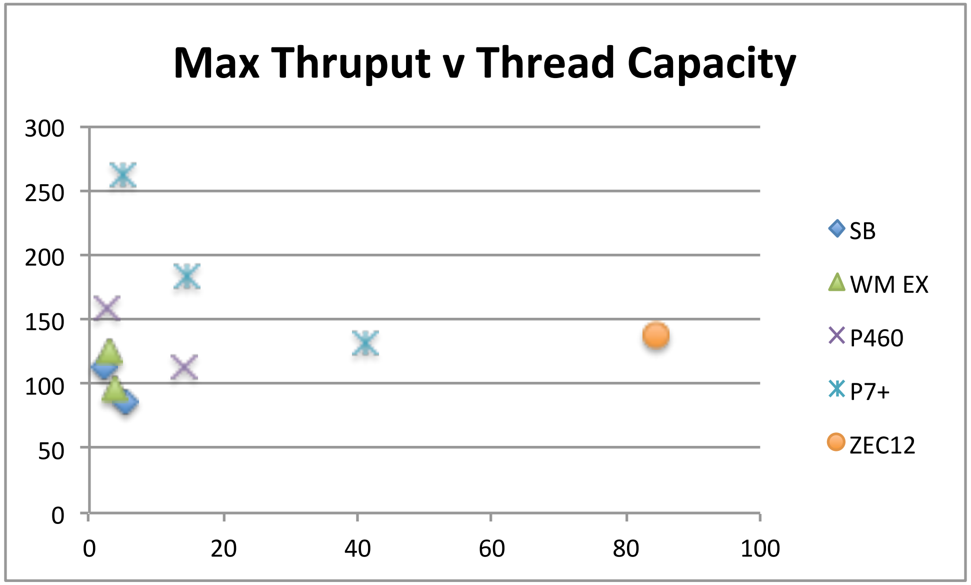
# Modeling and benchmarks

- There is enough data in the machine specification to make an architectural performance model
- We know that distributed on line loads usually have high variability
- However, the resulting model has relatively  low precision
  - It is better to use measurements
  - Traditional measurement of maximum throughput metrics will not help enhance the model.  It simply replaces it with another low precision model.
  - We should measure single thread speed
    - Single thread Saturation
    - Scaling with increased thread count (related to saturation)
    - Interval data of the usage and possibly throughput pattern

# Performance architecture involves requirements as well as comparisons

- How is response time defined?
  - Completion of a single thread of work?
  - Completion of many threads of work?
- What response time is required and what  fraction of the peaks need to be “covered”?
  - There is a trade off between peak coverage, cost and utilization efficiency.
  - Feasibility can become and issue if the SLA is too “tight”.
- Is “aggregate throughput” meaningful to users or is the preferred metric the number of loads contained in the machine while meeting the SLA?

# There is a design tradeoff between throughput and capacity



Here: Throughput is Clock * SMT muttiplier/threads per core * total threads
Thread Capacity is Clock * SMT muttiplier/threads per core * cache/thread

It is best to replace is Clock * SMT muttiplier/threads per core by measured thread speed

Throughput can't be faster that thread speed * Threads
Thread capacity is how much work can stack on a single thread which is related to both the thread speed and the cache available.

# Virtual Machine Density and the Tradeoff

As VM’s per core of the workload increases
the importance of aggregate throughput decreases

As the size of a virtual machine increases
The importance of its internal throughput rate increases.

Increased density favors favors z;
increased VM size favors Power

Intel is favored when resources can be aggregate without scaling penalties.
Power and z are favored when resources can be shared without scaling penalties.

# Do you need a deep dive to understand workload fit?

- Workload fit involves more than determining the workload type and a throughput ratio rule of thumb.
  - Operational considerations will change the relative capacity of machines
  - Throughput ratios do not generally take operational tradeoffs into consideration
- An Performance Architecture workshop can provide such a deep dive.
  - The objectives of the workshop are to build a model which produces characteristic curves
    - Response time v Throughput
    - Response time v Load Count or VM count
    - Response time v utilization
    - Throughput v utilization
    - Scaling
- The workshop can work with machine specs and assumed usage patterns in lieu of data but collection of data will yield better results

# Fit for purpose thinking comes down to: Know the legacy, workload, and costs

***Know the current IT Environment***

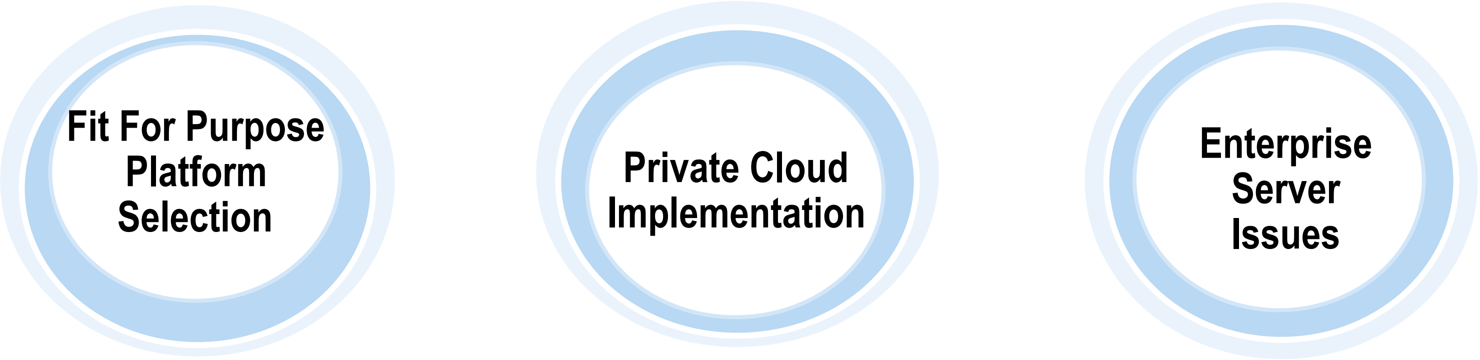
***Understand the workload***

***Examine costs***

Workload analysis gets technical fast, and real cost analysis is a deep dive.

# Eagle Engagements

- **Free of Charge** total cost of ownership study that helps customers evaluate the lowest cost option among alternative approaches. The study usually requires one day for an on-site visit and is **specifically tailored to a customer's enterprise.**
- The study can be focused on at least one of the areas below :



Fit For Purpose Platform Selection

Private Cloud Implementation

Enterprise Server Issues

- We conduct Eagle studies for System z, POWER, and PureSystems accounts
- Over 300 customer studies since the formation of the TCO Eagle team in 2007
- **Engage our Eagle-Eyed Experts!**
  - Start by requesting your IBM Contact to send an email to eagletco@us.ibm.com
  - For deep workload analysis workshop use the same link and ask for Joe Temple
  - Will be ramping up capability for workload deep dives in the coming quarters.