Enterprise IT - TCO over 5 years

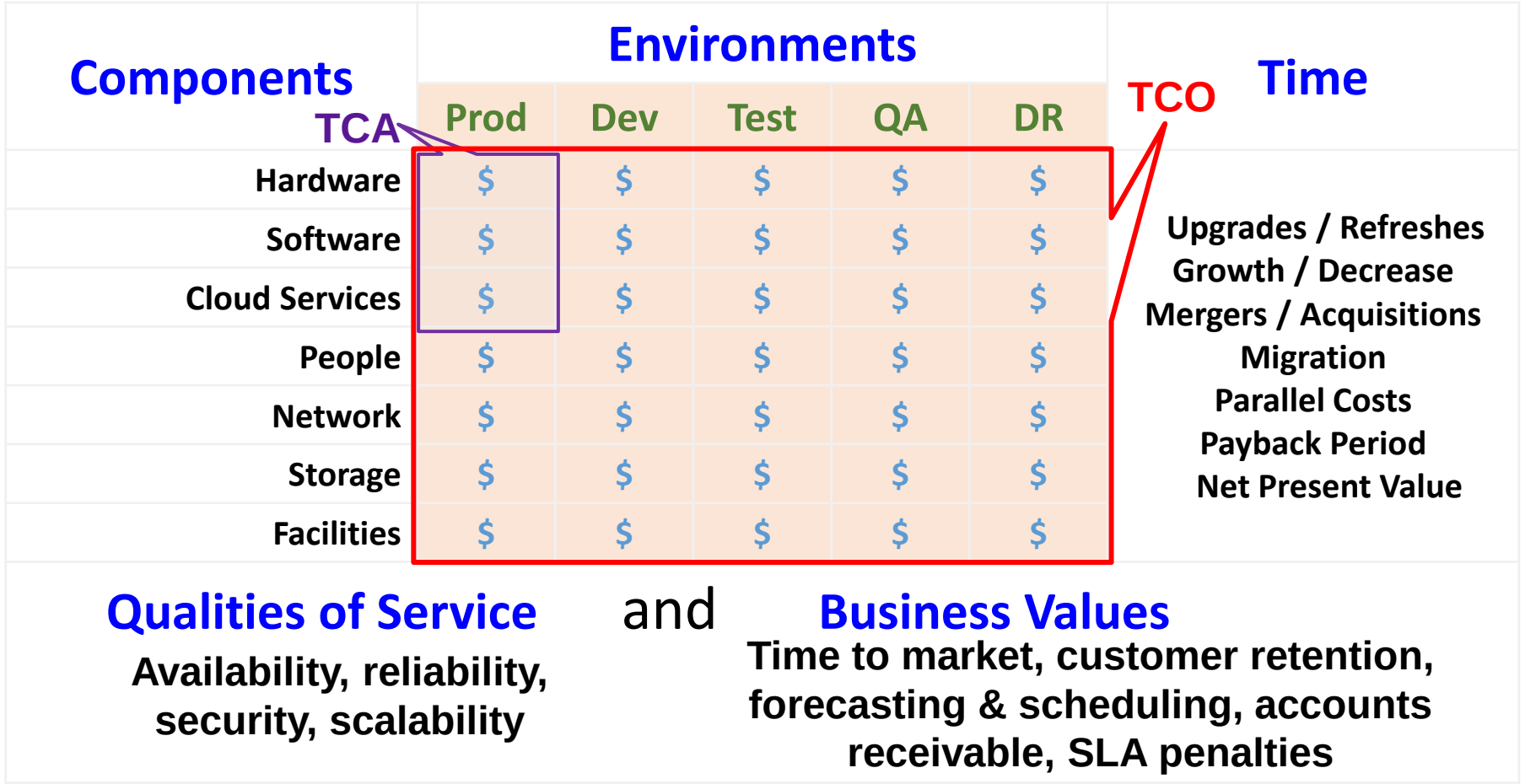
Simplified Case Study

To Software Engineering students at University of Rome "Tor Vergata"

By Giuseppe Calavaro, Ph.D.
IT Economics Consulting
IBM

TCO and Business Value – Get the complete picture

An **IT Economics study** quantifies business values, technical requirements and costs in a TCO



Total Cost of Ownership is much more than **Total Cost of Acquisition**!

TCO Assessment

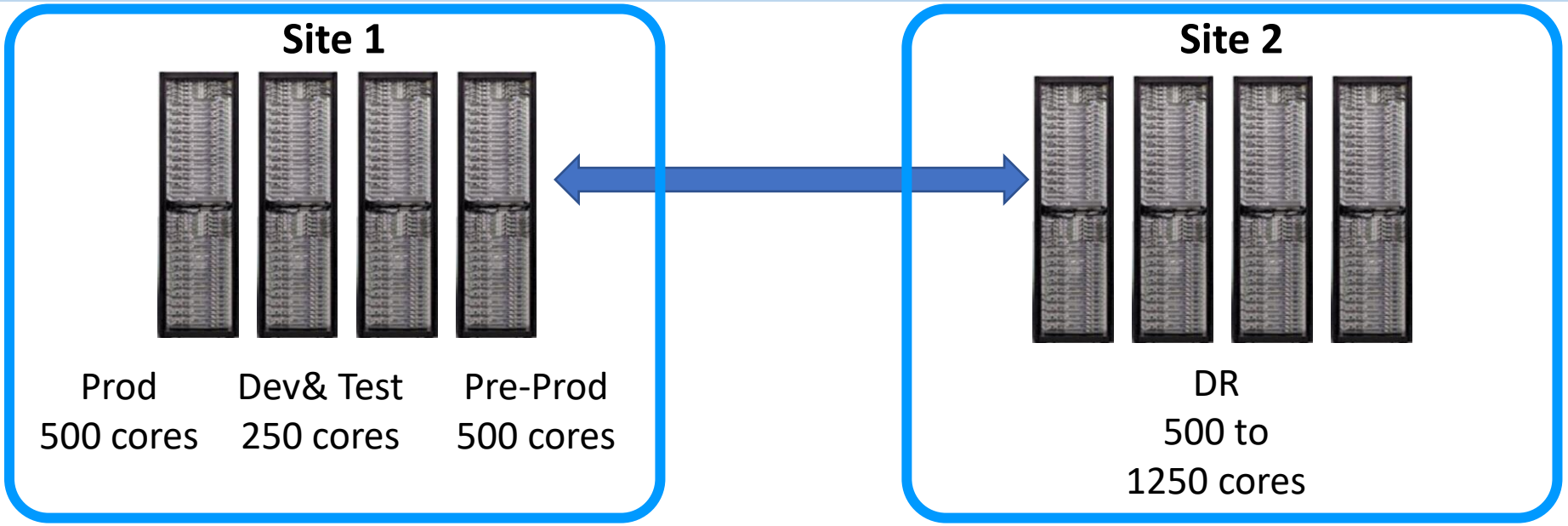
1. Analyze the UC and assess the workload
   – Averages and peaks
2. Assess the target platforms architecture and its size to support the workload
   a) Size the Production Environment considering the HA requirements
   b) Size the Dev and Test
   c) Size the Pre-Prod Environment
   d) Size the DR

   Repeat this step for all Platforms being evaluated ensuring to arrive at a sizing that satisfy the same Non-Functional Requirements (or existing platform features) in equivalent ways
3. Assess the TCO for each Target Platform documenting all costs elements being considered and the assumptions considered
4. Describe the outcomes with data and graphs and references to all input data and assumptions

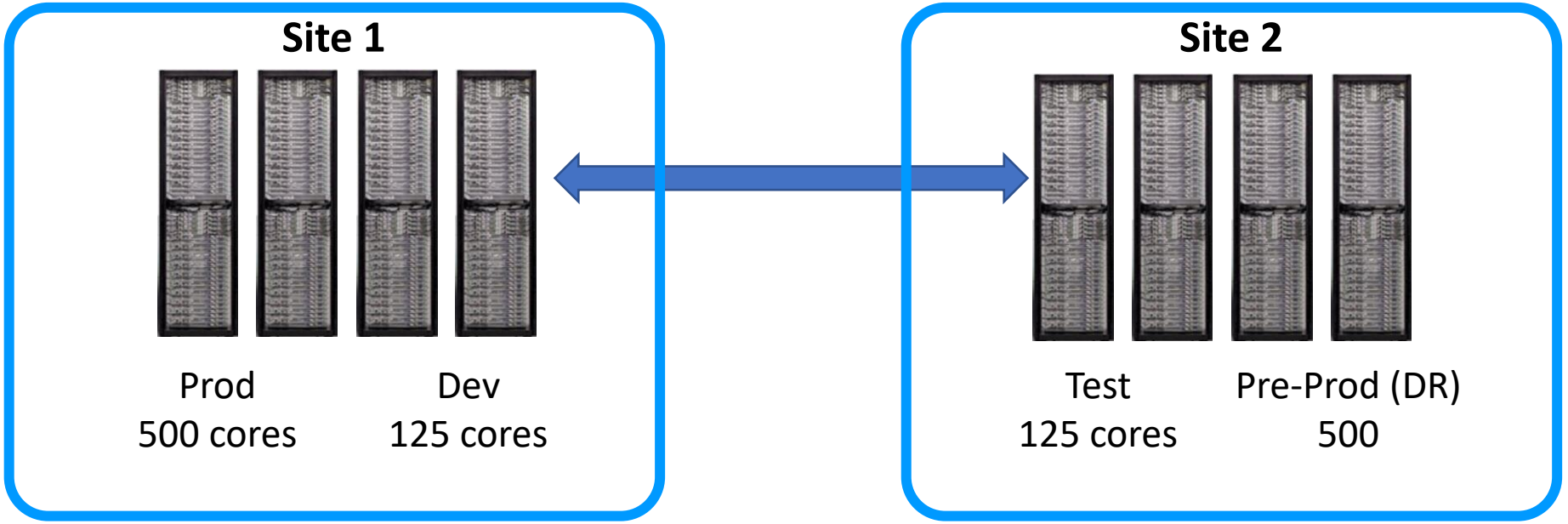
Sizing

- Calculation of required cores for a real workload requires sophisticated analysis.
  - Technical Architecture, GHz & utilization, VM & Hypervisor effect, System Overhead, Networking co-location effect, Sysplex, etc.
- In order to make a simple and realistic sizing, for this exercise a simplified approach is used. We consider the typical core ratio as found in literature.
  - Reference: *A. Micarelli, S. Proietti Conti, "Why Application Requirements Should Drive the Choice of your Infrastructure", 2021 IBM Corporation web published paper*
  - They found that "business-critical loads show most x86 Linux workloads have a core consolidation ratio ranging from 10 to 32.5 distributed cores to one IFL with an average of 17x fewer cores"
  - Download at: https://www.ibm.com/downloads/cas/2VVYYVNP
- A TCO for a specific environment should evaluate the exact sizing for the workload on the target platforms.
  - Even the same platform technology could result in slightly different sizing depending on the specific CPU, deployment Architecture, Networking, etc.
- For our example we assume a core consolidation ratio to deliver the same TPS at 10, that is a very conservative choice.

Examples of possible x86 Technical Architecture



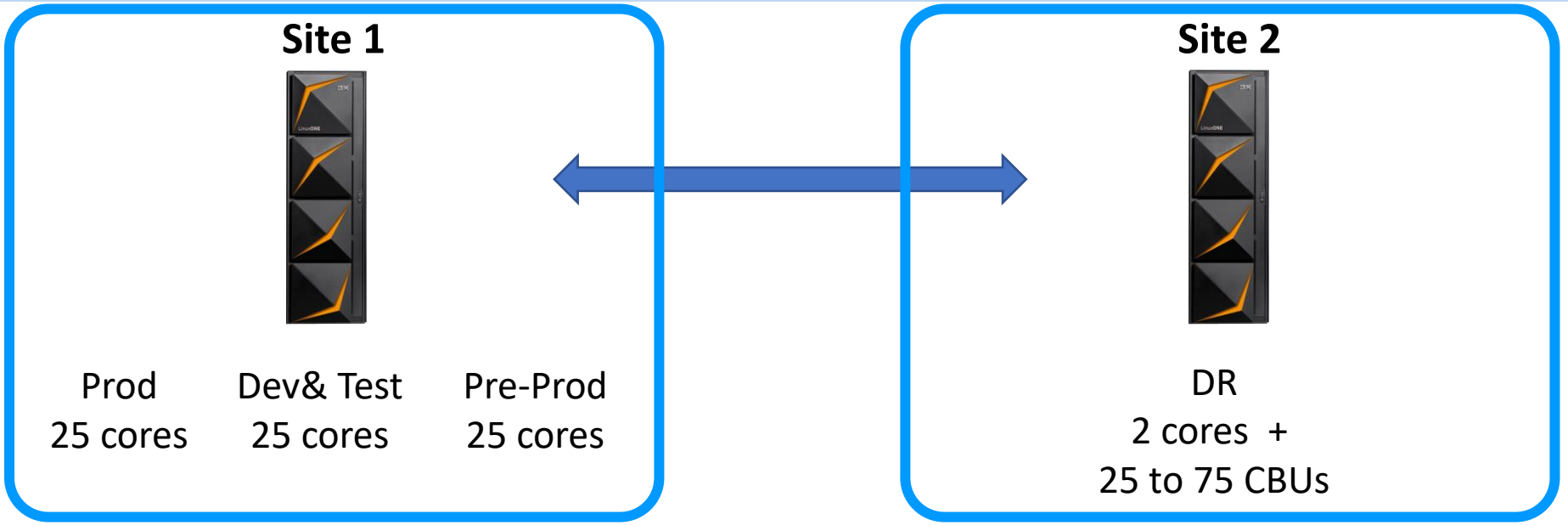
Example 1:
From 1750 cores to 2500 cores depending if DR is for Production only or for all environments



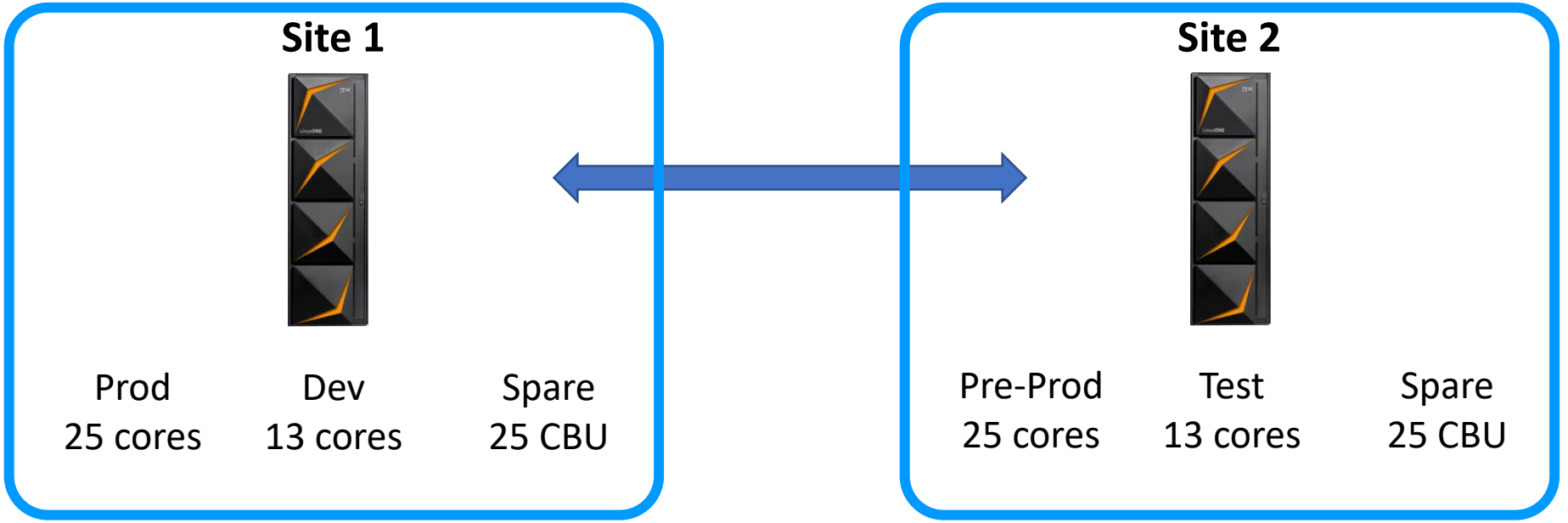
Example 2:
Tot 1250 cores.
Pre-Prod act as DR.

Business Accept no Pre-Prod env. in case of DR

Examples of possible LinuxONE Technical Architecture



Example 1:
75 cores (IFL) on main sites and 2 on Secondary site and 25 to 75 backup cores (CBU) depending if DR is for Production only or for all environments

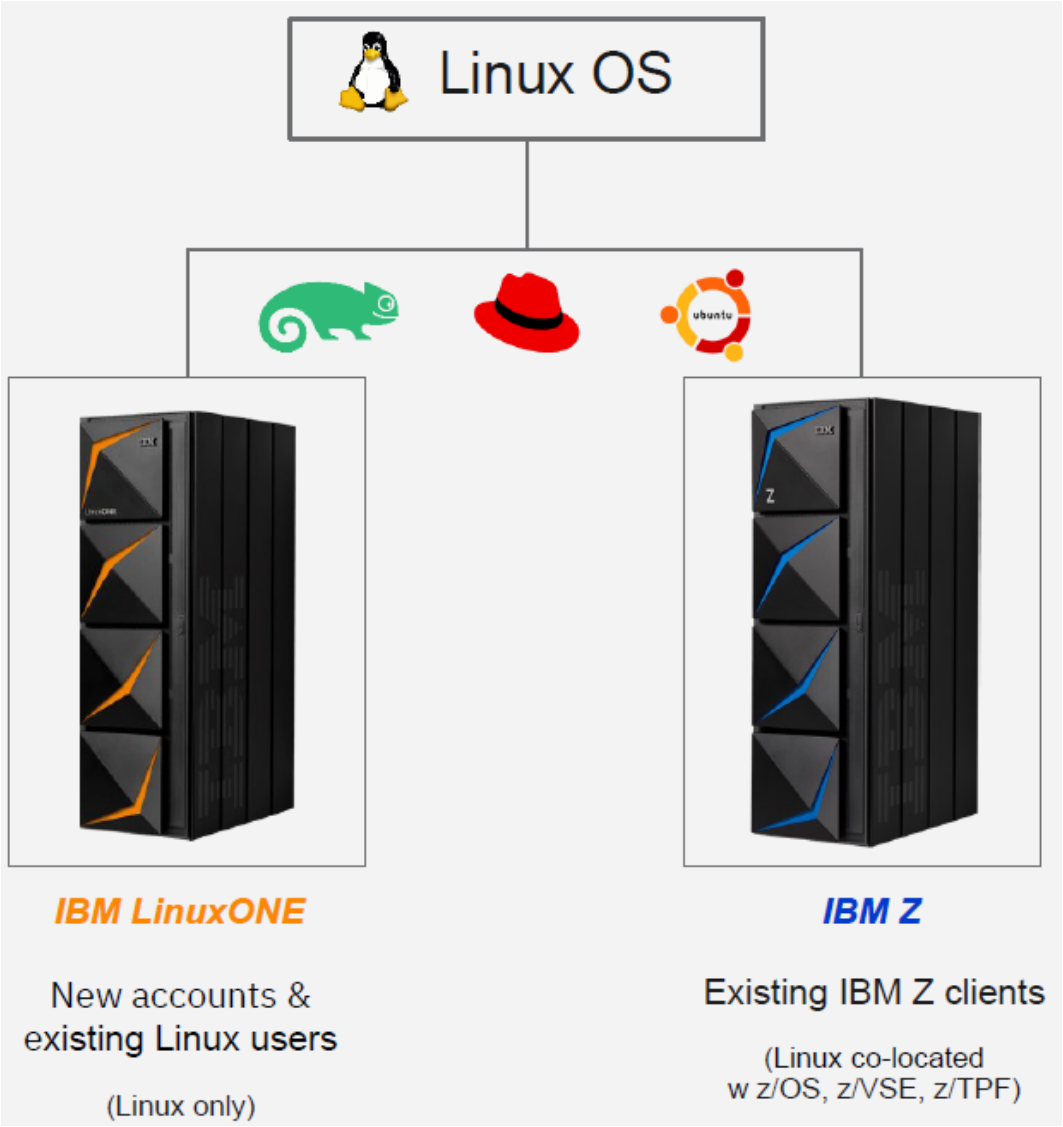


Example 2:
Tot 76 cores (IFL)
And 50 Spares (CBU)
Business will have full Prod and Pre-Prod in case of DR

Linux for z/Architecture serves two IBM brands

- Both are supporting the 3 major Enterprise Linux distribution
  - Red Hat
  - Suse
  - Ubuntu

- IBM is also contributing open source for no charge



Software Stack

- OS
- VM
- Application Server
- DB
- Data Replication Tools
- Monitoring Tools
- Security Tools

We need to know:

- Acquisition cost
  - Acquiring the licenses you get a capital asset
  - Not present in case of Subscription model
- Yearly cost
  - For maintenance/support usually is 20% or purchase price
  - Only cost in case of subscription model
- From which year you start paying the yearly cost
- Available discount
- Licensing model
  - Perpetual vs Subscription
  - Unit of measure
    - Server vs Socket vs Core vs PVU or other units
    - How are these units measured by?
    - Are there conversion factors?
- Are the same prices applied for all environments or platforms?
  - Sometime there are special prices or different conversion factors for development environments or Non-Production Environments or Platforms

Other Costs

People costs

- Environments need to be managed
  - Perform backups, restore from backups, monitor the system is running fine, act in case of problem, performance degradation or disaster, apply software patches & minor version upgrades, identify and replace hardware that is broken, replace hardware during regular hardware refresh, monitor security and act in case of attack, etc.
- Usually a Full Time Equivalent (FTE) can monitor about 30 servers.
- Consider the fully loaded cost and not the salary cost
  - On top of the salary there are also the Enterprise costs for their employees shared on each one by profile

CARBON FOOTPRINT

The carbon footprint of computing power is an important element to be evaluated. This is not only depending on the electricity consumption but by many other factors including, and not limited, to the location. Countries have different ways to get electricity that have very different carbon footprint.

Network costs

- The Enterprise networking system is complex, redundant for security and availability
- These costs are accounted by server or ports
  - In average we found that the average cost of networking is about 7000$ per server

Facilities costs

- Space
  - Each server on premise require to be allocated in the Data Center (DC)
  - The cost of the Data Center is usually reallocated by sq meter occupied by the servers, considering not only the size of the server but also the space required around
- Electricity
  - As the number of server grows the electricity bill impact become substantial
  - The electricity cost need to consider not only the consumption by the server but also by the cooling system

Example of Simplified Case Study

Description of the environment UC & required TPS

UC: Credit Card (CC) Transaction for a Bank

This conceptual transaction is made of many IT transactions.

Let's assume that an x86 core is able to perform 20 TPS of the these conceptual transactions

- Every CC transaction requires:
- Checking the CC is valid,
- Checking the POS is valid,
- Checking the money amount is valid
  - That the specific CC is within single transaction limits, daily limits and monthly limits
  - If the CC is pre-loaded CC, that there is sufficient money in the basket
- Check if the transaction is suspect
  - Run AI to warn or stop suspect transactions
- If the POS requires a second identification step, perform that step and check complete successfully,
- At the end of the transaction, issue OK or abort
- Store all the information about the transaction whatever is the exit status

Sizing

		Workload (TPS)	Core per Server	Max Server Faults	HA Ratio															
		1.000	16	1	32%															
	Typical core Ratio	TPS by Core	# of Cores for workload TPS	HA cores	Prod cores	Dev % of Workload	Dev cores	Test % of Workload	Test cores	PreProd % of Workload	PreProd cores	DR active cores % of Prod+PreProd	DR cores	DR CBU (spare) cores	Tot # of Active cores	x86 to z Active Core ratio	Tot # of Active cores no DR	x86 to z Core ratio no DR	Tot # of cores	x86 to z Core ratio
x86	10	20	50	16	66	50%	25	50%	25	100%	50	100%	116		282	17,6	166	10,4	282	10,8
LinuxONE	1	200	5		5	50%	3,0	50%	3,0	100%	5			10	16		16		26	

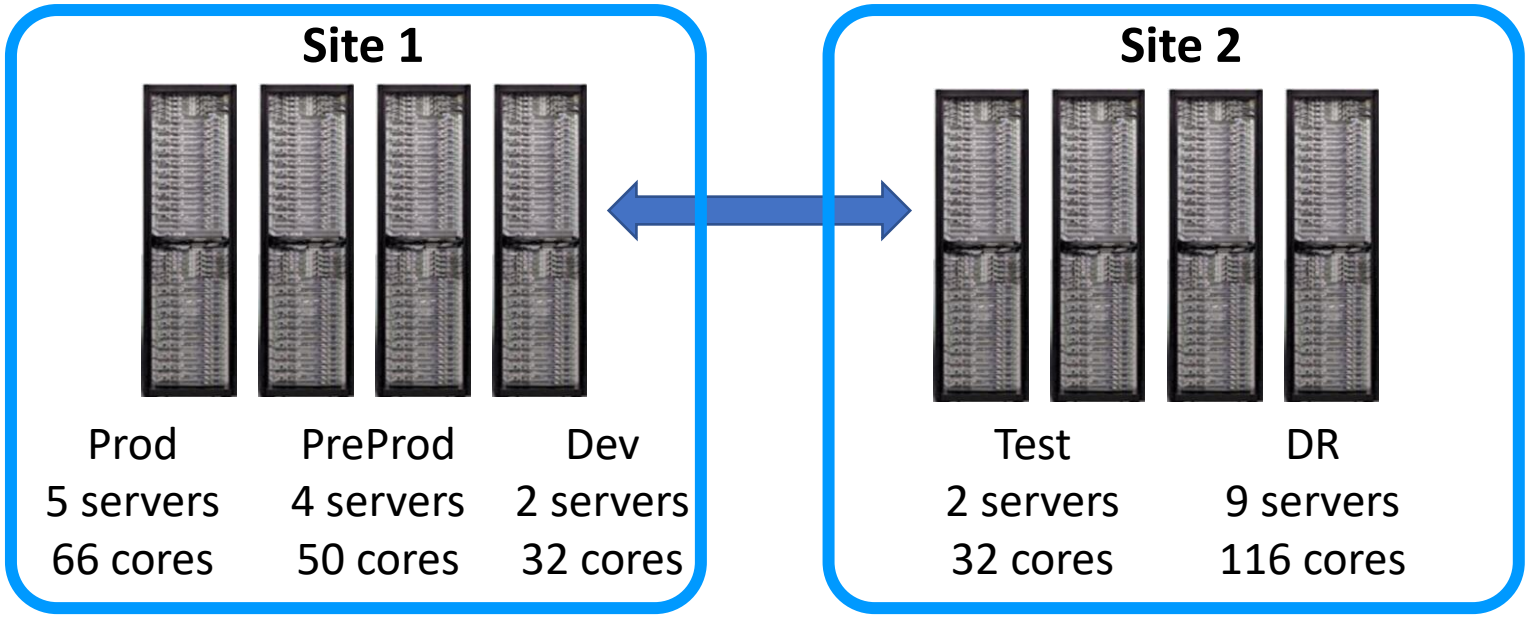
- The input provided is that the systems must support up to **1000 TPS**
  - Our input is that a x86 core is able to process 20 TPS*
  - Based on literature and what said previously, a LinuxONE core (aka IFL) is able to perform 10x TPS, therefore 200 TPS

**A precise sizing would require to know the x86 server and its config. Yet, as the differences of these servers would not impact the overall view, for simplification we adopt this approach*

Technical Architectures for the two cases

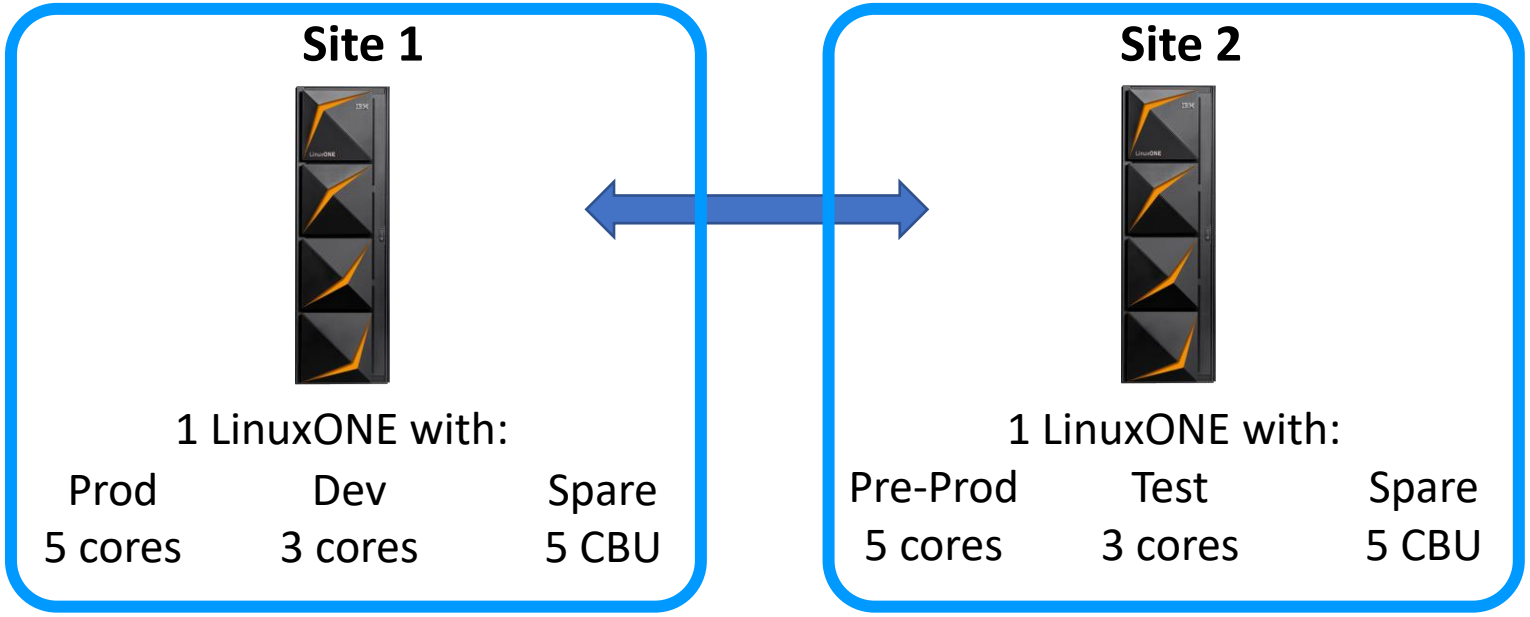
Inputs:

Workload=
1000 TPS



Case 1:
x86 Server:
Rack server w
Xeon Gold 6146 12C
3.20GHz
(2 Chips, 24 Cores)

22 servers
282 cores



Case 2:
LinuxONE Server

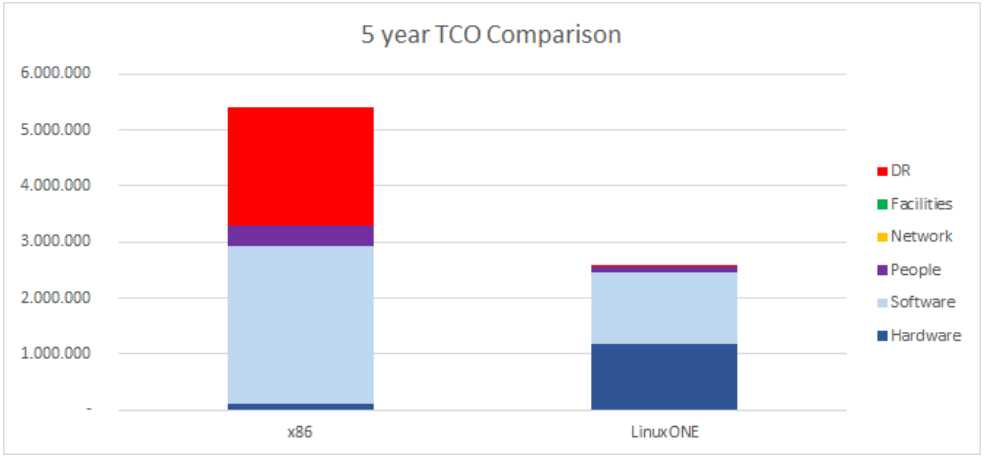
2 servers
16 cores (IFL)
and 10 spares (CBU)

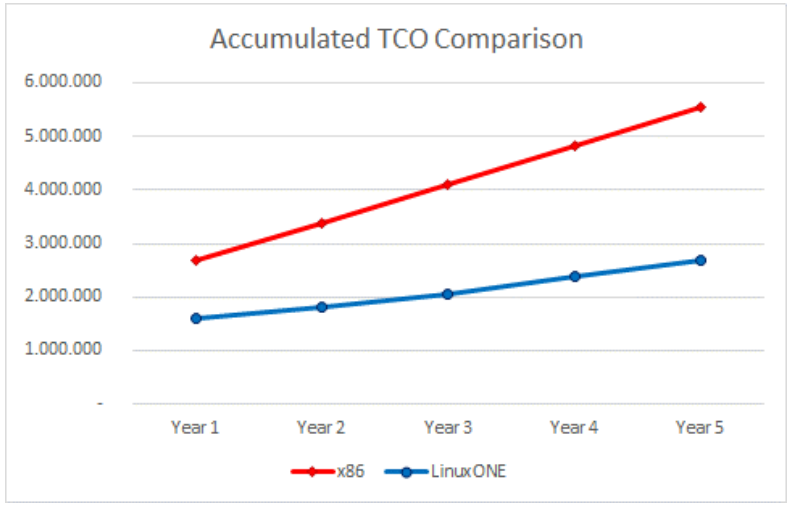
TCO Comparison on 5 years

	x86	Environments					
	Rack server w Xeon Gold 6244 8C 3.60GHz (2 Chips, 16 Cores)	Prod	Dev	Test	PreProd/QA	DR	Tot by Comp
Components	Hardware	39.200	15.680	15.680	31.360	70.560	**172.480**
	Software	1.116.450	433.025	433.025	850.850	1.952.100	**4.785.450**
	People	366.667					**366.667**
	Network	34.300	13.720	13.720	27.440	61.740	**150.920**
	Space	13.000	2.253	2.253	4.507	13.000	**35.013**
	Electricity	10.950	4.380	4.380	8.760	19.710	**48.180**
	Total by Env	1.580.567	469.058	469.058	922.917	2.117.110	**5.558.710**

	LinuxONE	Environments					
		Prod	Dev	Test	PreProd/QA	DR	Tot by Comp
Components	Hardware	367.500	220.500	220.500	367.500	42.000	**1.218.000**
	Software	398.250	238.950	238.950	398.250	-	**1.274.400**
	People	100.000					**100.000**
	Network	8.575	5.145	5.145	8.575	-	**27.440**
	Space	22.750	13.650	13.650	22.750	-	**72.800**
	Electricity	1.729	1.037	1.037	1.729	-	**5.533**
	Total by Env	898.804	479.282	479.282	798.804	42.000	**2.698.173**

- Above schemas summarize the TCO for the different environments on 5 years, for each cost component
  - The savings with the LinuxONE platform is 51% vs the x86 equivalent platform
- Below diagrams highlights the component contribution to total cost and cost per year that make the accumulated TCO





x86 TCO Assumptions

- Hardware
  - Rack server w Xeon Gold 5115 10C 2.40GHz (2 Chips, 20 Cores)
  - Acquisition cost: 10K€ - 30% discount including 3 years Support
  - Support for following years is 20% of purchase price
- Network
  - 7000€ per Server -30% disc
    Maint @ 10% of PP from y2
- People
  - One FTE covers 30 servers
  - Average yearly fully loaded costs=100K€
- Space
  - Fully loaded cost of Sq. meter=2.800€
- Electricity
  - Cost per kWh=0,10€
- Software
  - OS: subscription 2000€/year – 20% disc per socket
  - VM: license @5000€/socket - 20% disc
    maint @ 20% of purchase price from y2
  - Application Server: license @100€/PVU -75% disc (70 PVU per core)
    maint @ 20% of purchase price from y2
  - DB: license @40.000€ for 2 cores -75% disc
    maint @ 20% of purchase price from y2
  - Data Replication Tools: subscription 3000€/year – 75% disc per core
  - Monitoring Tools: license @5.000€/server -40% disc. Maint @ 20% of PP from y2
  - Security Tools: license @5.000€/server -40% disc. Maint @ 20% of from y2

LinuxONE TCO Assumptions

- Hardware
  - ❑LinuxONE servers configured as the example @ 725K€ for 3 years – 30% disc
  - ❑Maint @ 10% PP from y4
- Network
  - ❑14000€ per LinuxONE Server -30% disc
  - ❑Maint @ 10% of PP from y2
- People
  - ❑One FTE covers 10 servers
  - ❑Average yearly fully loaded costs=120K€
- Space
  - ❑Fully loaded cost of Sq. meter=2.600€
- Electricity
  - ❑Cost per kWh=0,10€

- Software
  - ❑OS: subscription 6000€/year – 20% disc per socket
  - ❑VM: included in HW
  - ❑Application Server: license @100€/PVU -50% disc (120 PVU per core) maint @ 20% of purchase price from y2
  - ❑DB: license @40.000€ per cores -50% disc maint @ 20% of purchase price from y2
  - ❑Data Replication Tools: subscription 3000€/year –50% disc per core
  - ❑Monitoring Tools: license @5.000€/server -40% disc. Maint @ 20% of PP from y2
  - ❑Security Tools: license @5.000€/server -40% disc. Maint @ 20% of from y2

Demo of the Excel Tool

Some real-world considerations on Tech Architecture

Usually, in the Enterprise environments there are

- **Many workloads** that the Enterprise has to manage
  - The example here is considering a single isolated workload
- **An existing infrastructure** to leverage
  - The example here is considering a new system on new dedicated servers
- When you have many workloads running on the same cores you can leverage **statistical multiplexing** utilization models to **size by the average** TPS requirements instead of by the peak
  - The example here is considering for both cases only the peak TPS
- A **refresh cycle** for existing platforms that, if well planned and supported by the technology, it will allow to leverage the move of workloads among available platforms in an easy and optimized way
  - The example here is considering new servers in all sites

These consideration lead to:

- On Z servers you can have up to **hundred of parallel workloads running on the same cores**, thanks to optimized firmware integrated VM
  - You can **size your Workload by the average TPS** instead of the peak
- On Z servers you can **activate cores** on existing environments
  - Having machines with up to 190 cores, instead of purchasing a new server your can **purchase cores** and this will cost much less than purchasing new machines
- You can have Z servers in your private cloud, public cloud or both, boosting the advantages of a **hybrid cloud environment** to further reduce costs while gaining Enterprise grade NFR satisfaction such as security, availability, etc. on the cloud advantages.

In order to simplify the case study and your assignments, the TCO calculation is not taking in consideration the above considerations and related advantages of Z platforms, as well as the economic value of the **additional security features that have a high-cost impact in case would be required to replicate on x86**.

Homework assignment

- See attached document with Assignment description
- Output is a presentation similar to slides 10 to 16
- Study the TCO tool in teams but homework is personal
- Homework will require a lot of time for the internet search, study and understanding the tool, evaluate the sensitivity of input changes and prepare the output
- Exam Test will be similar to Homework, yet time will be limited to 2 hours

Conclusions and message to bring home

- Review the Simplified Case Study slide
- Review and make your opinion on slide:
  - *Some real-world considerations on Tech Architecture*
- On Hardware Technical Architecture
  - A dedicated course would be required to learn how to define and optimize such architectures to get the most at lower cost
  - The intention here is to learn how to analyze a given architecture but you should have gor some ideas to validate them
- On Software Stack
  - This is the most important cost contributing element
  - Licensing models and way to count changes often and by vendor
- TCO can change drastically with minor changes of some parameters while remain substantially the same with changes of other parameters...
  - Your homework is to reflect and provide considerations on these aspects

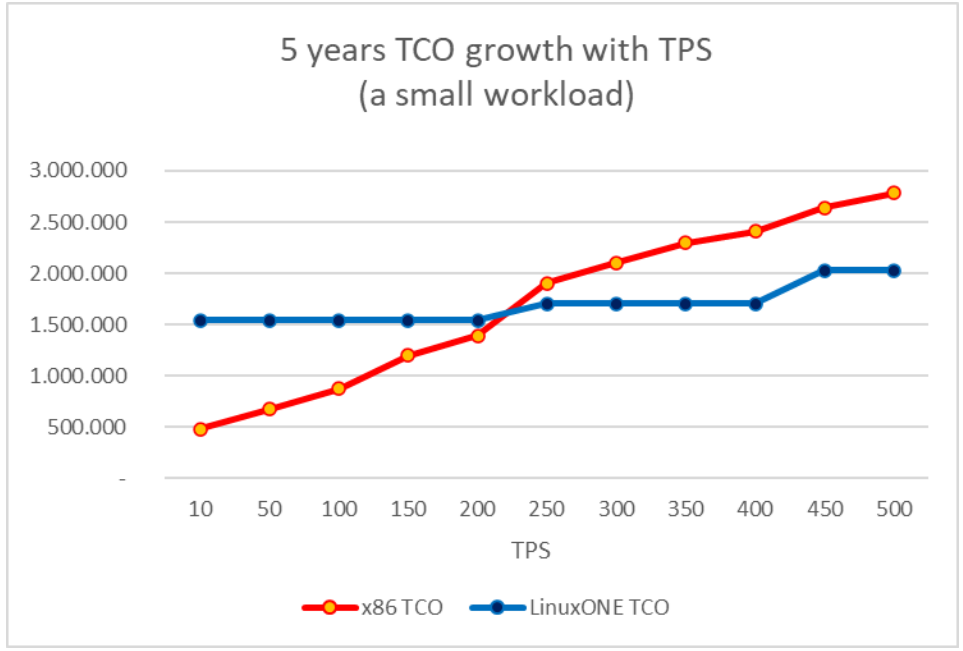
Disclaimer

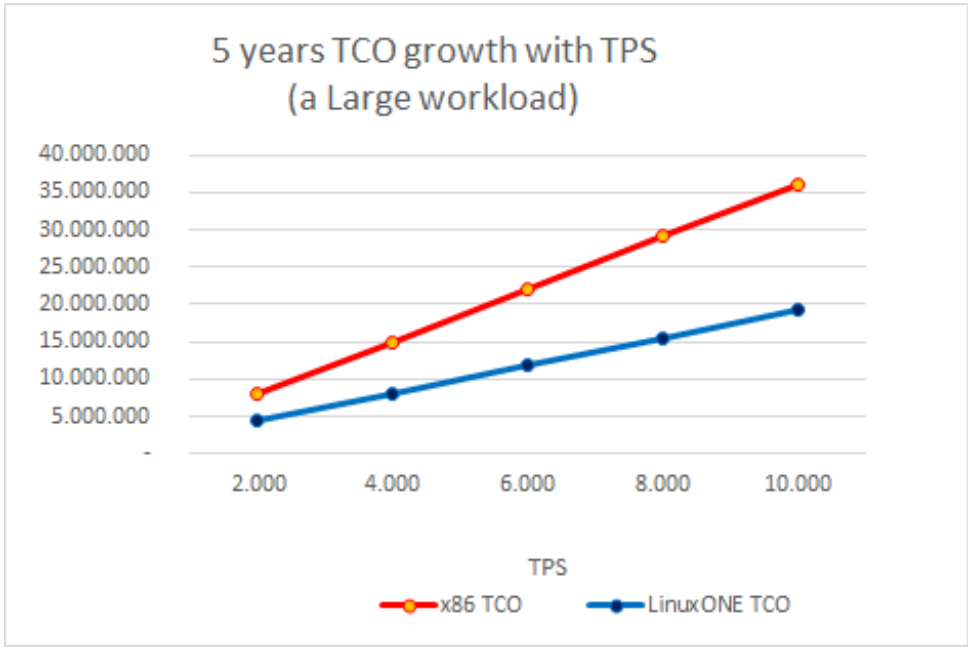
- This is an example for education purpose
- The values are just examples to show the different possible values, metrics, and different units used to charge the elements of an IT solution
- The views and information herein are those of the presenter; they do not necessarily reflect the views of IBM or any other entities in the IT field

Backup

Example of how TCO changes with TPS

- Example based on the assumptions already stated





Accounting basics: Capex vs Opex

- **Capex** (Capital Expenditure or Expense)
  - money an enterprise spends to buy, maintain, or improve its fixed assets, such as buildings, vehicles, equipment, or land
- **Opex** (Operating Expenditure or Expense)
  - ongoing expenses that are inherent to the operation of the asset. Opex includes items like electricity or cleaning.
- Example:
  - the purchase of a photocopier involves capex, and the annual paper, toner, power and maintenance costs represents opex

- An ongoing question for the accounting of any company is whether certain costs incurred should be *capitalized* or *expensed*.
- For tax purposes, capex is a cost that cannot be deducted in the year in which it is paid or incurred and must be capitalized.
- The capital expenditure costs are then amortized or depreciated over the life of the asset in question.