

Non-Stop Computing

Mark A. Dyer

CHTMA13@OUTLOOK.COM



SLIDE 1 of 312

**JUST KIDDING ... LET'S
LEARN AND HAVE FUN!**



SLATE
RETRACT

Storage of the past:

Here's How Hard it Was to Move a 5MB IBM Hard Drive in 1956



By Editorial Staff [Twitter](#) [Email](#)

Posted on September 15, 2015



History

- Mainframe
 - International Business Machines and the BUNCH: Burroughs, UNIVAC, NCR, Control Data Corporation, and Honeywell.
- Minicomputers
 - Digital Equipment Corporation, Data General, Wang Laboratories, Apollo Computer, and Prime Computer.
- Microcomputers
 - Intel, Zilog ^[1], Motorola, RCA, National Semi Conductor and others

The old joke: **Microsoft vs. General Motors**

- At a computer expo, Bill Gates reportedly compared the computer industry with the auto industry and stated "If GM had kept up with technology like the computer industry has, we would all be driving \$25.00 cars that got 1,000 miles to the gallon."

Terms / Nomenclature

- “minicomputers” or “departmental computing”
 - In a 1970 survey, the [*New York Times*](#) suggested a consensus definition of a minicomputer as a machine costing less than \$25,000, with an input-output device such as a teleprinter and at least [4K](#) words of memory, that is capable of running programs in a higher level language, such as [Fortran](#) or Basic.
 - Minicomputer versus microcomputer



 **TANDEM**COMPUTERS

Tandem Computers: A great idea

- Tandem Computers
 - Tandem Computers, Inc. was the dominant manufacturer of fault-tolerant computer systems for ATM networks, banks, stock exchanges, telephone switching centers, and other exacting, time sensitive needs.
- Founded by James “Jimmy” Treybig ^[1].



Treybig Leadership

- Gaye Clemson who hasn't worked for Tandem Computers in 20 years, still comments:
 - “It doesn't seem to matter how short or long it was, people who worked there, almost to a person, say it was the best place they ever worked in their professional careers”

Tandem Computers Summary

- Formed in 1974.
- In 1997, Tandem Computers, Inc. was acquired by Compaq Computer Corp.
- In 2001, Compaq and HP announced their plan to merge and consolidate their similar product lines.
- Still exists today **“HP Nonstop systems For industries that never stop”** See:
(<http://h17007.www1.hp.com/us/en/enterprise/servers/integrity/nonstop.aspx>)

HPE Solutions Today ...



[Hybrid IT with Cloud](#) [Mobile & IoT](#) [IT for Data & Analytics](#) [Services](#) [How to Buy](#) [Contact](#) [🔍](#) [🛒](#) | [☰ Menu](#)

HPE Integrity NonStop

Fault tolerant solutions, for businesses that never stop.

[Overview](#) [Portfolio](#) [Case Studies](#) [Tech Specs](#) [Resources](#)

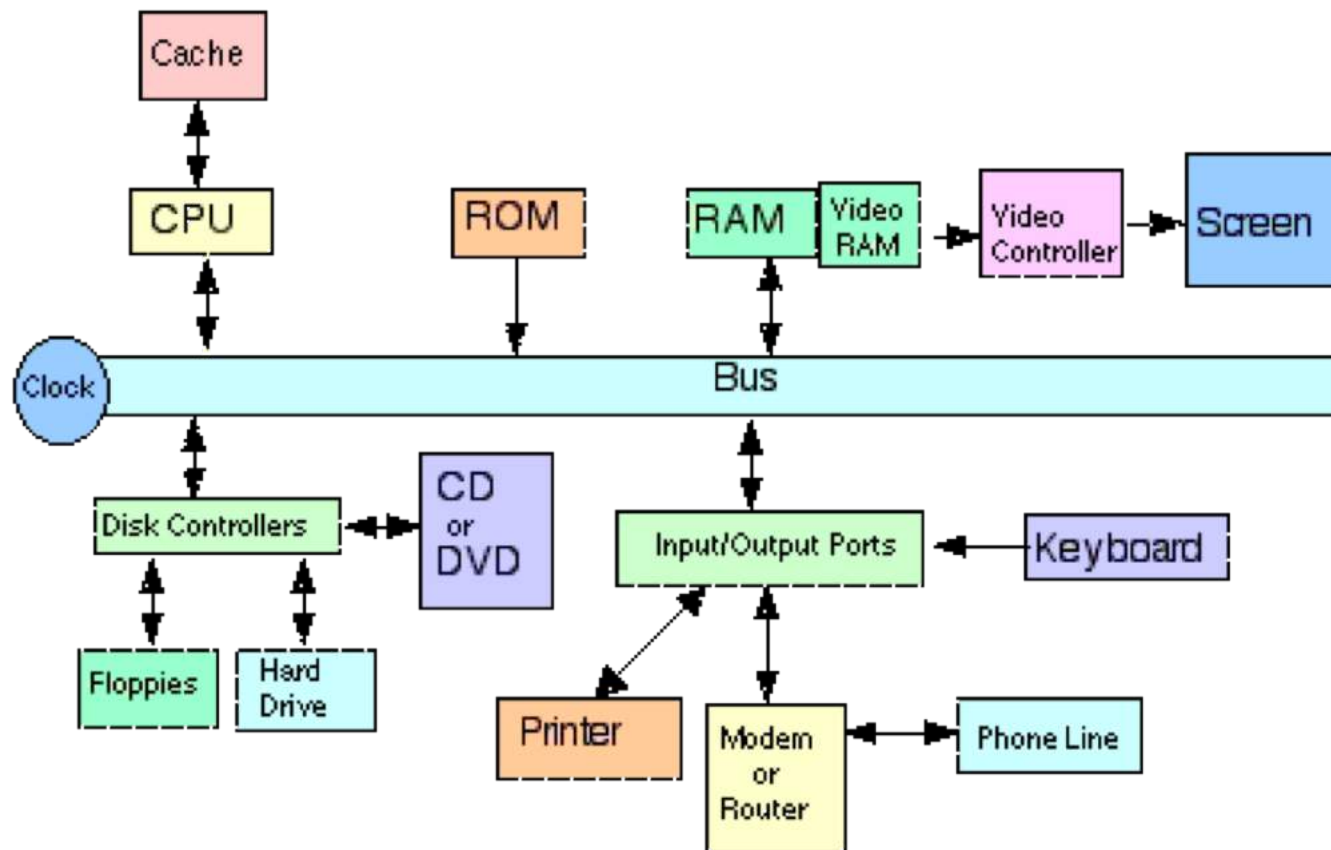
[VIEW BEST PRACTICES GUIDE](#)



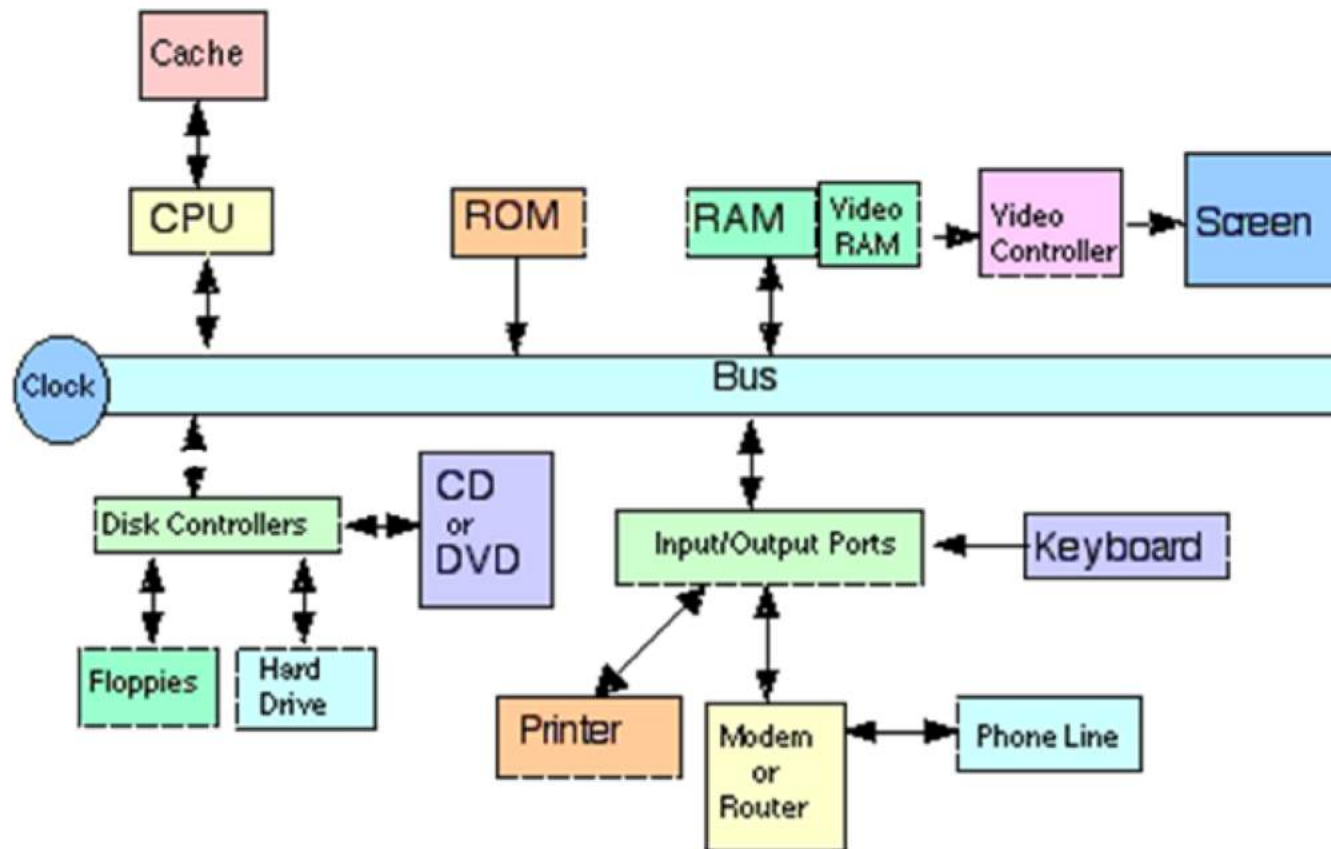
Because customers never wait

What did computing look like in 1974?





Typical “micro” computer bus. Note the single points of failure.



Typical “micro” computer bus. Note the single point of failure.

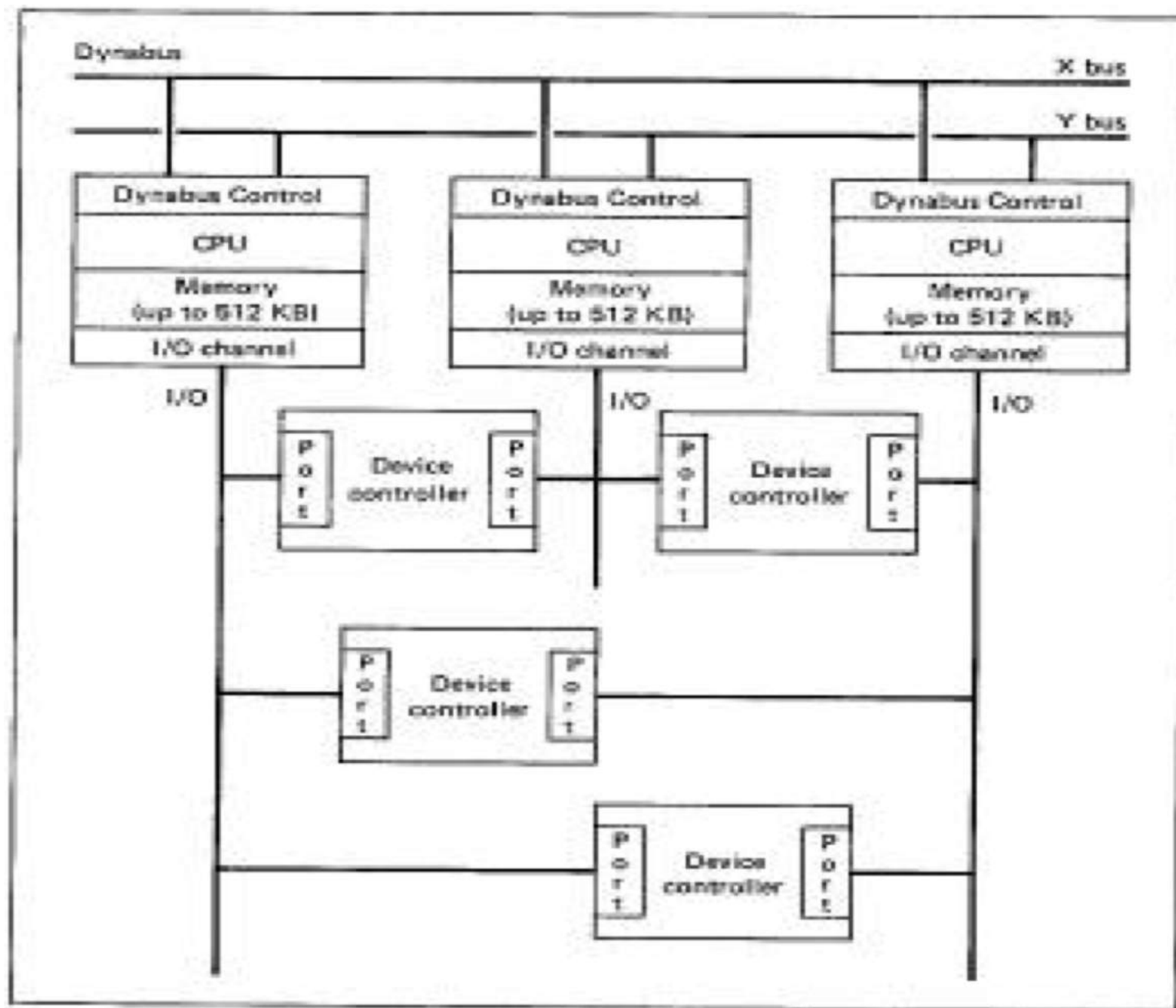


Fig. 2. Tandem 16 system architecture.

Note the redundancies.

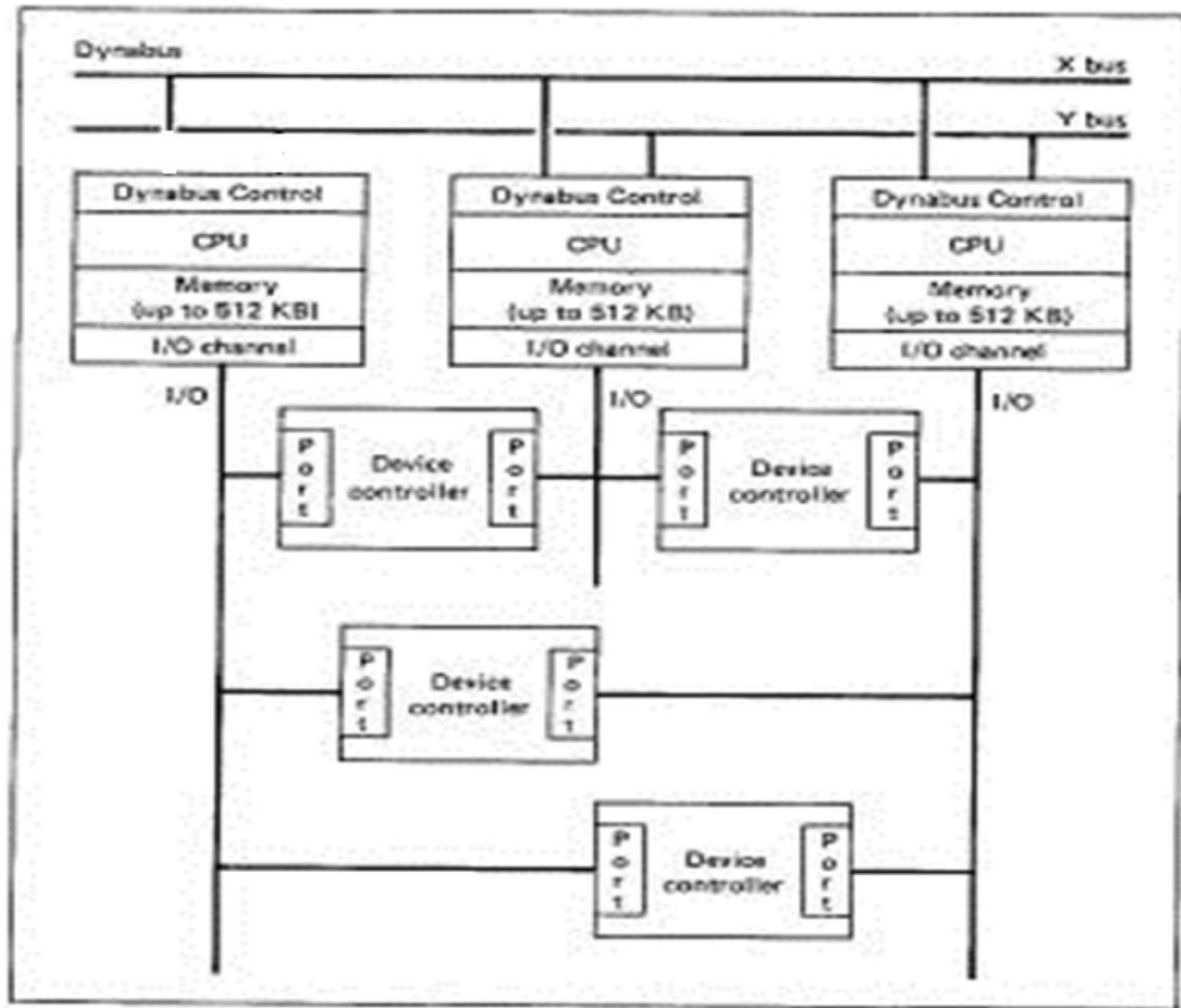


Fig. 2. Tandem 16 system architecture.

Note the redundancies.

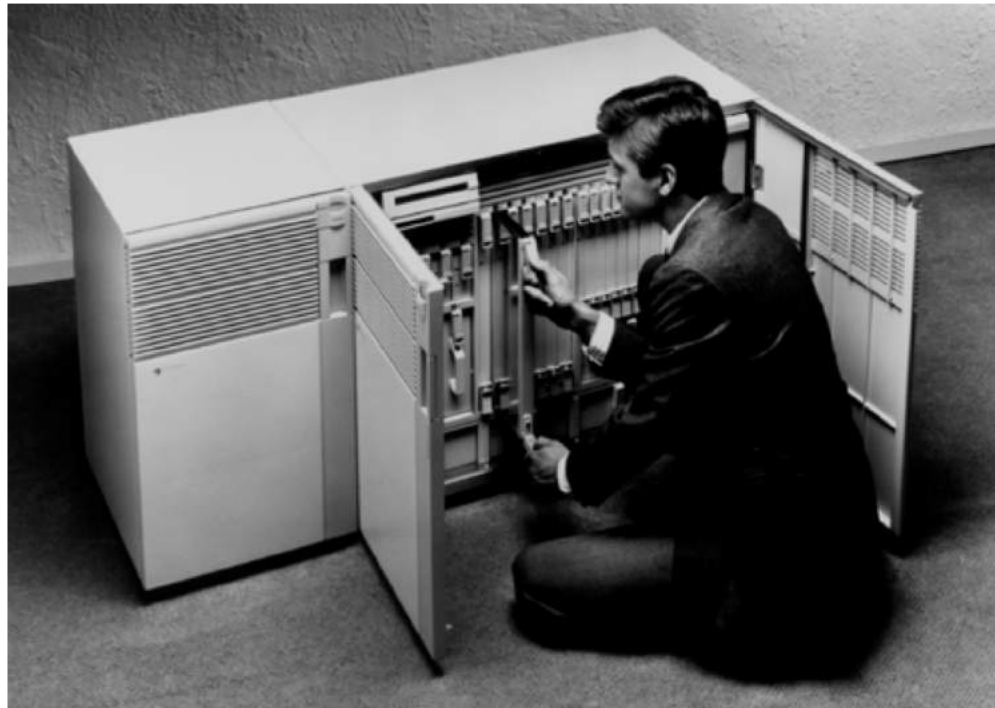
Comparison in 1970s and 80s

- Conventional Config
 - Single CPU (single core), single IO bus and single disk drive.
 - Large configurations had multiple CPUs (typically less than 8).
 - This period was the move from large mainframe to “minicomputers” or “departmental computing”^[1]
 - In August 1981, the first IBM PC went on sale
- Tandem Configuration
 - Between 2 and 16 processors, with their own memory, I/O buses, and dual connections to their custom inter-CPU computer bus, Dynabus. The modules were constructed with dual paths so that any single failure would always leave at least one bus (both I/O and Dynabus), free for use by the other modules

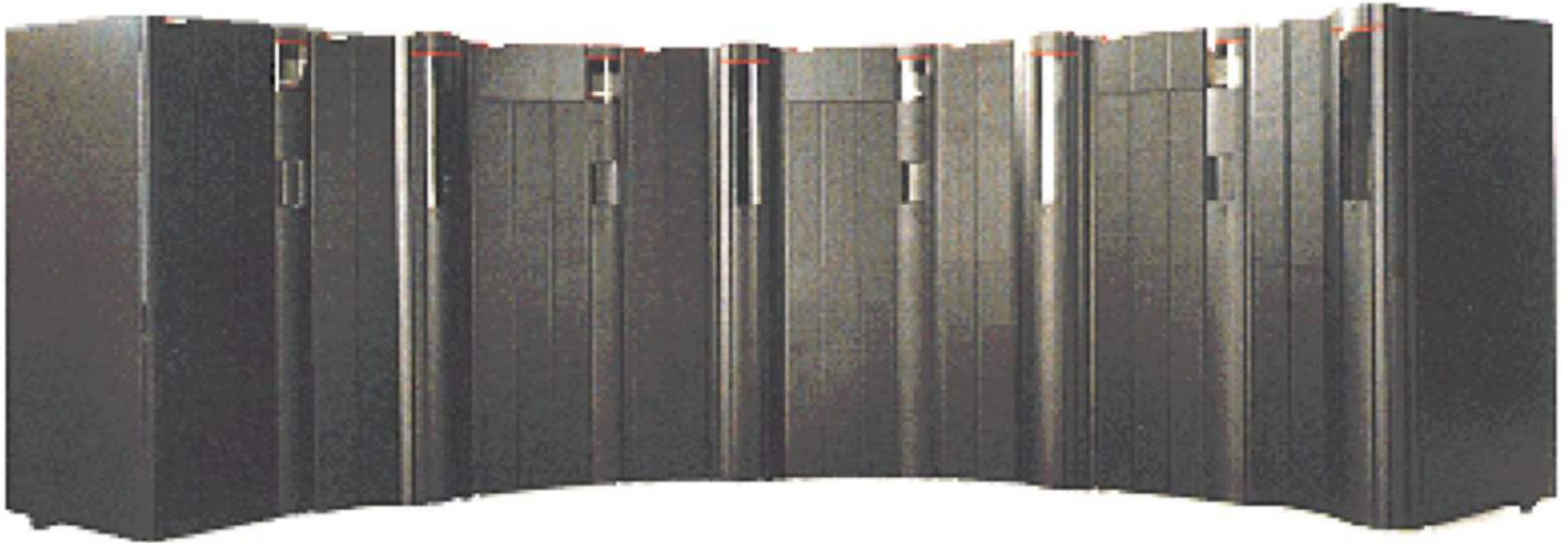
Tandem's first NonStop system: 1976.



The Tandem Integrity S2 was a fault-tolerant, high-performance UNIX system ...



Himalaya K10000: 1995



Himalaya K20000, could consist of up to 4080 processors with 16,711 Tb storage

Tandem Computers Timeline

- In 2005, the current product line of HP Integrity Nonstop servers, based on [Intel Itanium](#) microprocessors, was introduced.
- In 2014, the first systems "Nonstop X" running on the Intel X86 chip were introduced.

Nonstop Hardware

- InfiniBand™ is an industry-standard specification that defines an input/output architecture that leverages switched, point-to-point channels with data transfers up to 120 gigabits per second.
- Cluster I/O Modules (CLIMs)

HP Nonstop Systems

- Servicing over 1.4 billion credit and debit cards worldwide with a charge volume over \$3.5 trillion in 2011
- Supporting 8 of the top 25 global retailers to acquire their transactions.
- Processing over \$122 billion of fuel card transactions in the U.S. in 2011.

HP Nonstop Systems

- Servicing over 375M subscribers in advanced Telco network applications, such as Home Location Register (HLR), Home Subscriber Service (HSS), and other network applications.
- Powering mission-critical applications at 100% of the top 10 global manufacturers.
- Supporting several of the world's leading medical institutions.

What do all these share?



Their designs haven't changed much in years ... Why?



Because they WORK!



Geezer Stories

- TANDEM engineer told what happened after a California earthquake.
 - Laying on it's back ... still running, but on its back!
- Took one of our TANDEMs down
 - ... A decade since it last was cold started.

Data Processing terms ...

- Checkpointing
- Failover
- Process Pair
- Inter-process Communication
- Multiprocessing
- Uptime
- MTBF

Checkpointing

- *Checkpointing* is the saving of program state, so that it may be reconstructed later in time.
- Checkpointing provides the backbone for rollback recovery (fault-tolerance), playback debugging, process migration and job swapping.

Failover

- Failover is the constant capability to automatically and seamlessly switch to a highly reliable backup. This can be operated in a redundant manner or in a standby operational mode upon the failure of a primary server, application, system or other primary system component.

Interprocess Communication

- **Interprocess communication (IPC)** is a set of programming interfaces that allow a programmer to coordinate activities among different program processes that can run concurrently in an operating system. This allows a program to handle many user requests at the same time.

Multiprocessing

- **Multiprocessing :**
 - Multi-processing is the ability of a system to support more than one processor at the same time.
 - Applications in a multi-processing system are broken to smaller routines that run independently.
 - The operating system allocates these threads to the processors improving performance of the system.
- **Loosely Coupled Multiprocessor:**
 - A loosely coupled multiprocessor system is a type of multiprocessing where the individual processors are configured with their own memory and are capable of executing user and operating system instructions independent of each other.
 - Such systems are connected via high-speed communication networks.
 - Loosely coupled multiprocessor systems are also known as distributed memory, as the processors do not share physical memory and have their own IO channels.

Reliability terms ...

- Uptime:
- Uptime is a computer industry term for the time during which a computer is operational.
- MTBF:
- MTBF (mean time between failures) is a measure of how reliable a hardware product or component is.

Nonstop Operating System

- The Tandem Nonstop series ran a custom [operating system](#), initially called **T/TOS (Tandem Operating System)** but soon named **Guardian**. It supported a "Nonstop" programming paradigm that allowed a program to be completely fault tolerant.

Nonstop Operating System

- Several other companies introduced failover technologies but only Guardian supported completely fail-safe transaction processing.
- Hardware redundancies and software redundancies.

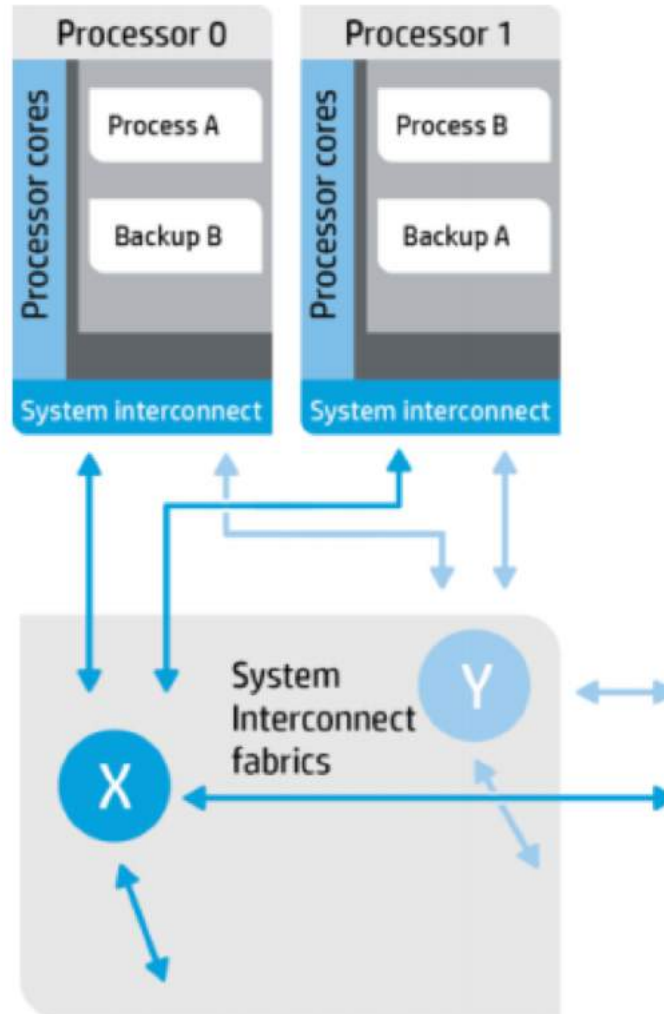
Architected for continuous availability

- Upon detecting a failure of the primary process (due to either hardware or software fault), the backup process takes over as the primary and starts processing using the context it received through checkpointing. The backup processes require only minimal processing to consume checkpointing data.
- The primary benefit is that the backup process has the context, and upon abrupt termination of the primary process, can continue processing using that context.

Architected for continuous availability

- Nonstop Operating System is designed and built for the highest availability.
- Nonstop OS supports the process-pair technology used to create pairs of software fault-tolerant processes that run on different processors to perform critical system tasks such as disk volume management.

Nonstop process-pair technology:



Nonstop Operating System

- While conventional systems of the era, including [mainframes](#), had mean time between failure on the order of a few days, the Nonstop system was designed to fail 100 times less, with "uptimes" measured in years.

Nonstop Operating System

- Nevertheless the Nonstop was deliberately designed to be price-competitive with conventional systems, with a simple 2-CPU system priced at just over two times that of a competing single-processor mainframe, as opposed to four or more times of most competing solutions.

Today, under HP ...

- The HP Integrity NonStop computers are a line of [fault-tolerant](#) server computers based on the [Intel Itanium](#) processor platform, and optimized for transaction processing. Average availability levels of 99.999% have been observed.

Today, under HP ...

- NonStop systems feature a [massively parallel processing](#) (MPP) architecture and provide linear scalability. Each CPU (systems can be expanded up to over 4000 CPUs) runs its own copy of the OS. This is a [shared nothing architecture](#) — a "share nothing" arrangement also known as [loosely coupled multiprocessing](#), and no "diminishing returns" occur as more processors are added.

What do you mean “diminishing returns”?

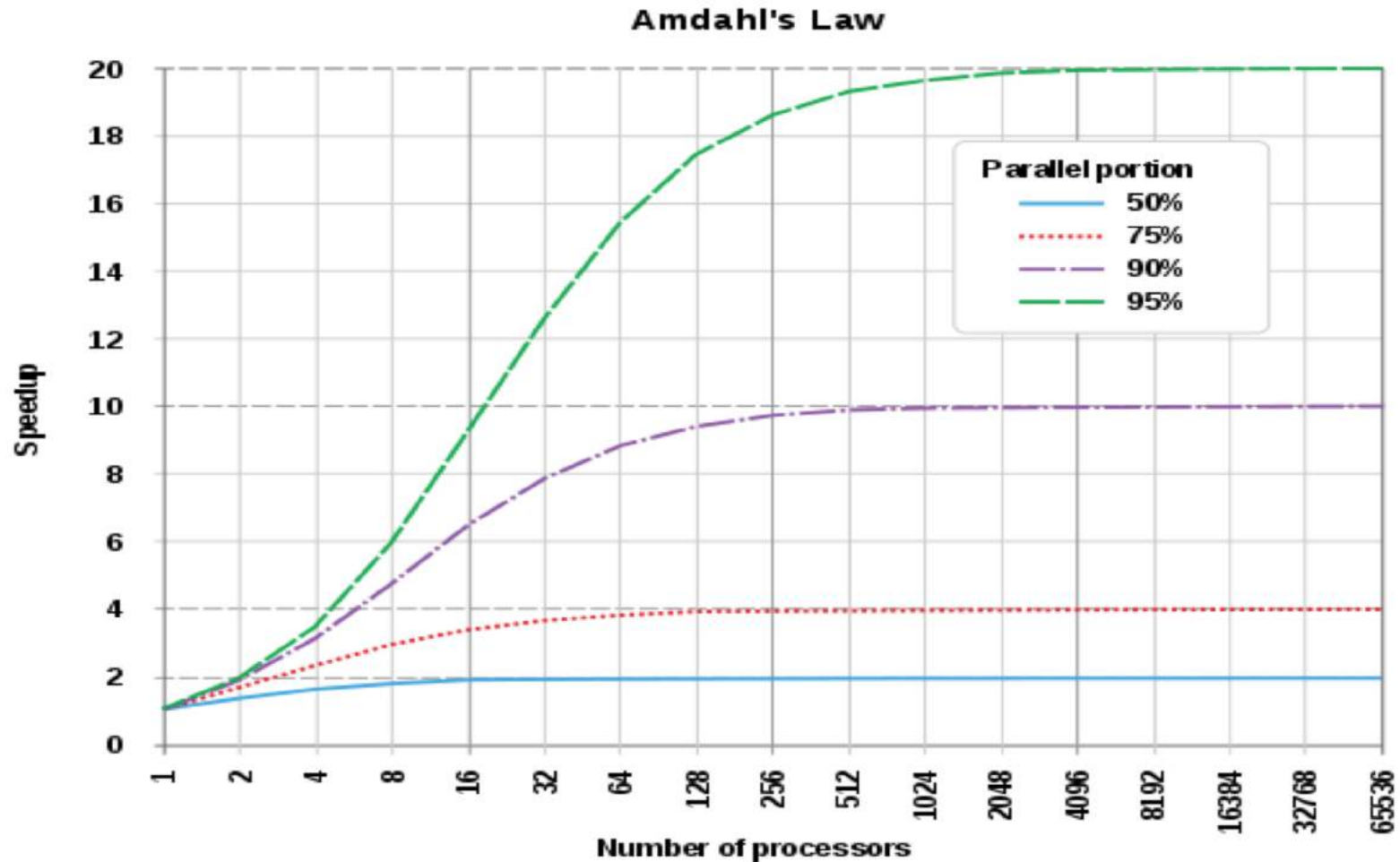
- **What is Amdahl's Law?**

- Amdahl's law is a formula used to find the maximum improvement possible by improving a particular part of a system. In parallel computing, Amdahl's law is mainly used to predict the theoretical maximum speedup for program processing using multiple processors.
- It is named after Gene Amdahl, a computer architect from IBM and the Amdahl Corporation.”

What do you mean “diminishing returns”?

- “Amdahl's law is often used in [parallel computing](#) to predict the theoretical speedup when using multiple processors. For example, if a program needs 20 hours using a single processor core, and a particular part of the program which takes one hour to execute cannot be parallelized, while the remaining 19 hours ($p = 0.95$) of execution time can be parallelized, then regardless of how many processors are devoted to a parallelized execution of this program, the minimum execution time cannot be less than that critical one hour. Hence, the theoretical speedup is limited to at most 20 times ($1/(1 - p) = 20$). For this reason, parallel computing with many processors is useful only for highly parallelizable programs.

What do you mean “diminishing returns”?



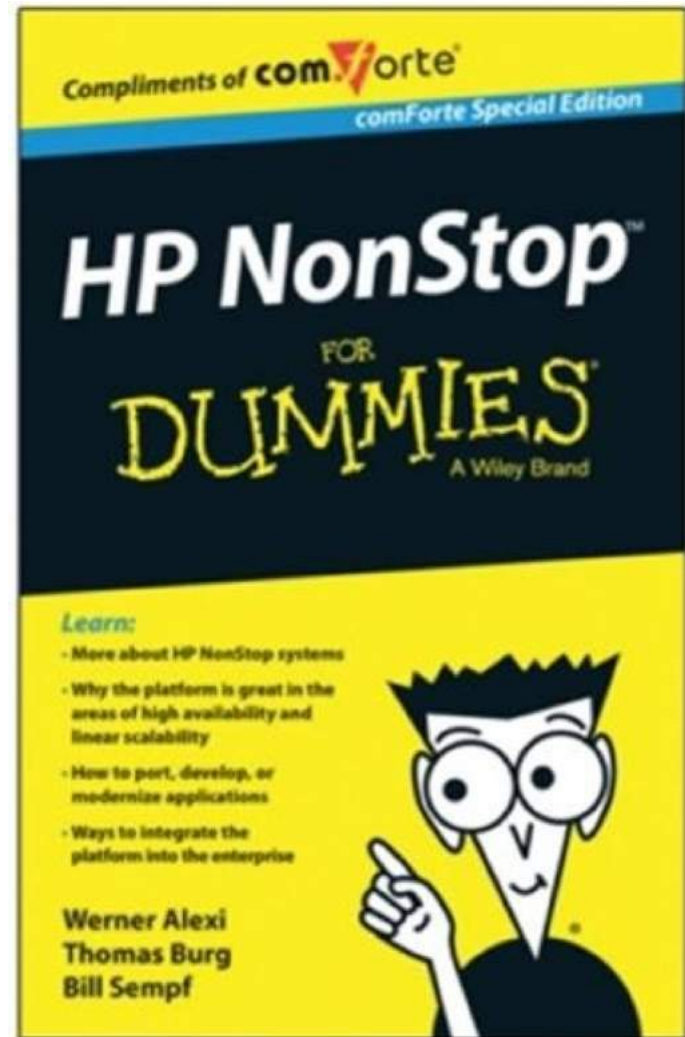
The Always On Operating System

- The HPE Nonstop operating system (OS) is unique. It delivers fault tolerance through a share-nothing parallel processing architecture with multiple levels of error detection. Fault isolation and workload takeover capabilities provide application and database availability on both a local and global scale. [HP Ad]

The Always On Operating System

- As a result, HPE Nonstop has unparalleled ability to detect, isolate, and recover from hardware and software failures without affecting critical applications and their users.
[HP Ad]

Tandem Software



Learn:

- More about HP NonStop systems
- Why the platform is great in the areas of high availability and linear scalability
- How to port, develop, or modernize applications
- Ways to integrate the platform into the enterprise

Nonstop Software: TACL

- TACL: HP Tandem Advanced Command Language.
 - TACL is a software application that provides an interface to the HP Nonstop operating system. You can use TACL either as an interactive command interface or as an interpreted programming language to construct programs.

Nonstop Software: TACL

- Similar to:
 - Bash
 - Ksh
 - csh

Nonstop Software: TAL

- TAL: Transaction Application Language ... used to be called "**Tandem** Application Language".
 - The Transaction Application Language (TAL) is a high-level, block-structured language that works efficiently with the system hardware to provide optimal object program performance ...

Nonstop Software: TAL

- Typical “ALGOL” like procedural language
 - **Transaction Application Language** or **TAL** (originally "Tandem Application Language") is a block-structured, procedural language optimized for use on [Tandem](#) hardware. TAL resembles a cross between [C](#) and [Pascal](#). It was the original [system programming language](#) for the Tandem [CISC](#) machines, which had no [assembler](#).

Usual utilities: TEDIT

- PS TEXT EDIT (TEDIT) is a multi-screen block mode text editor designed to make editing your documents easier and more efficient. ... You interact with TEDIT by pressing function keys and keyboard keys, or by typing commands on a command line.

Usual utilities: TEDIT

- Tandem has two type of text editors. EDIT and TEDIT. EDIT is a line based text editor and TEDIT is a full screen text editor. TEDIT can be considered an improved version of EDIT. The only area where EDIT can be considered better than TEDIT is in its ability to run scripts on text files.
- Tandem uses function keys as shortcuts for various commands and so does TEDIT. The function keys definitions can be customized for the needs of a specific user.

Usual utilities

- Compilers
- SQL databases
- TCP/IP Stack
- SSL Stack
- Security Management tools
- In short ... A full environment

Programming in a Non-Stop Environment



As with most things in computer science, you are given three options: fast, cheap, and correct (you may pick two).

Sample TAL code

Sample Source File

```
!This is a source file named MYSRC.

?SOURCE $SYSTEM.SYSTEM.EXTDECS (INITIALIZER)
                                !Include system procedure

PROC myproc MAIN;              !Declare procedure MYPROC
  BEGIN
    INT var1;                  !Declare variables
    INT var2;
    INT total;

    CALL initializer;          !Handle start-up message
    var1 := 5;                 !Assign value to VAR1
    var2 := 10;                !Assign value to VAR2
    total := var1 + var2;      !Assign sum to TOTAL
  END;                         !End MYPROC
```

Sample TACL

Sample Program

This program, of type macro, purges a file:

```
?TACL MACRO
#FRAME == Establish a local environment for variables
#PUSH err == Declare a variable called "err"

== Set err to the result of the #PURGE function:
#SET err [#PURGE %1%]

== Display the results of the purge operation:
[#IF NOT err |THEN|
  #OUTPUT Purge of file %1% complete!
|ELSE|
  #OUTPUT Error [err]
]
#UNFRAME == Delete the local environment
```

A macro accepts positional arguments; %1% refers to the first argument supplied when you run the program.

Inter-process Communication

- The start-up message is a system message sent to a process when it starts.
- \$RECEIVE: A special file name through which a process receives messages from other processes.

Inter-process Communication

- Useful for setting up a client – server model
 - The server can pass to the client processes what is to be done next
 - As many clients as necessary
- Necessary for checkpointing in a nonstop environment.



Programming difficulties

Coupling

- In [software engineering](#), **coupling** is the degree of interdependence between software modules; a measure of how closely connected two routines or modules are; the strength of the relationships between modules.

Cohesion

- In [computer programming](#), **cohesion** refers to the *degree to which the elements of a [module](#) belong together*. Thus, cohesion measures the strength of relationship between pieces of functionality within a given module. For example, in highly cohesive systems functionality is strongly related.

Programming difficulties

- Coupling is usually contrasted with cohesion. Low coupling often correlates with high cohesion, and vice versa. Low coupling is often a sign of a well-structured computer system and a good design, and when combined with high cohesion, supports the general goals of high readability and maintainability.

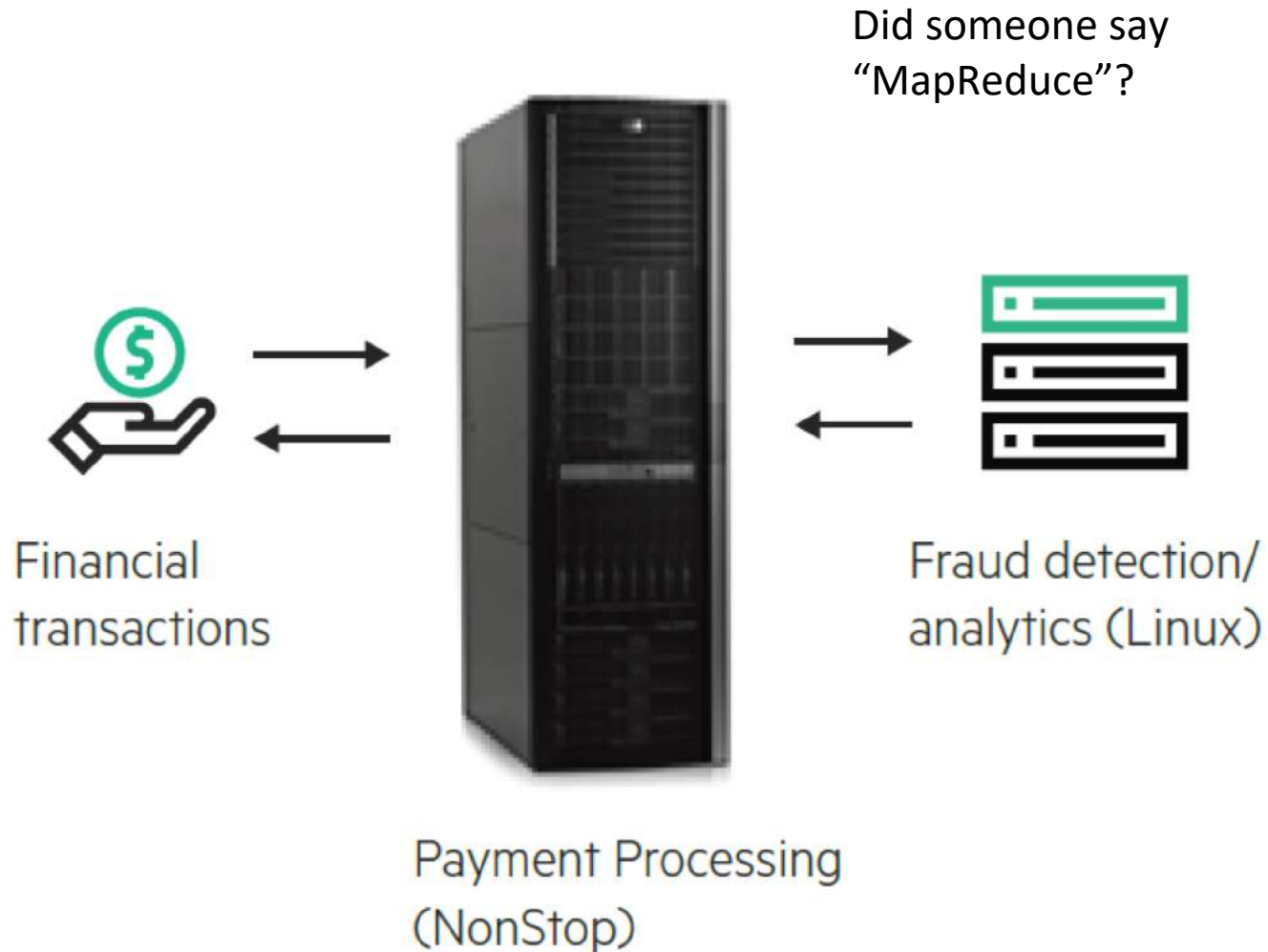
Problems

- Complexity in programming
 - Constant Checkpointing!
 - Constant IPC housekeeping.
- Cost of hardware
 - Two (or more) of EVERYTHING!
- Cost of personnel
 - JavaScript script kiddies?
 - Specialize training; Hard to find skill set.

How to integrate TANDEMS into an architecture



TANDEMS as a front end system for RAS



How to integrate TANDEM into an architecture

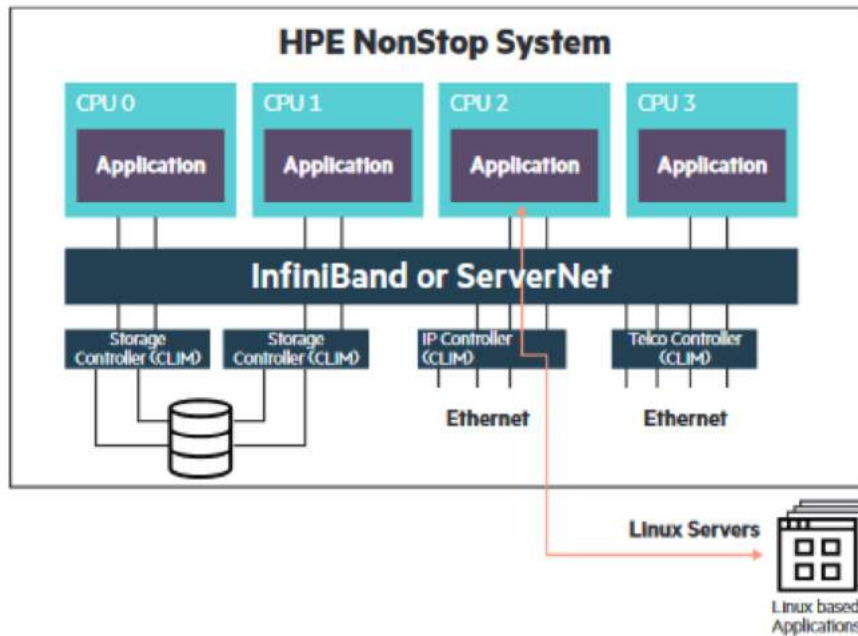


Figure 1: Traditional method of connecting NonStop and Linux applications over TCP/IP

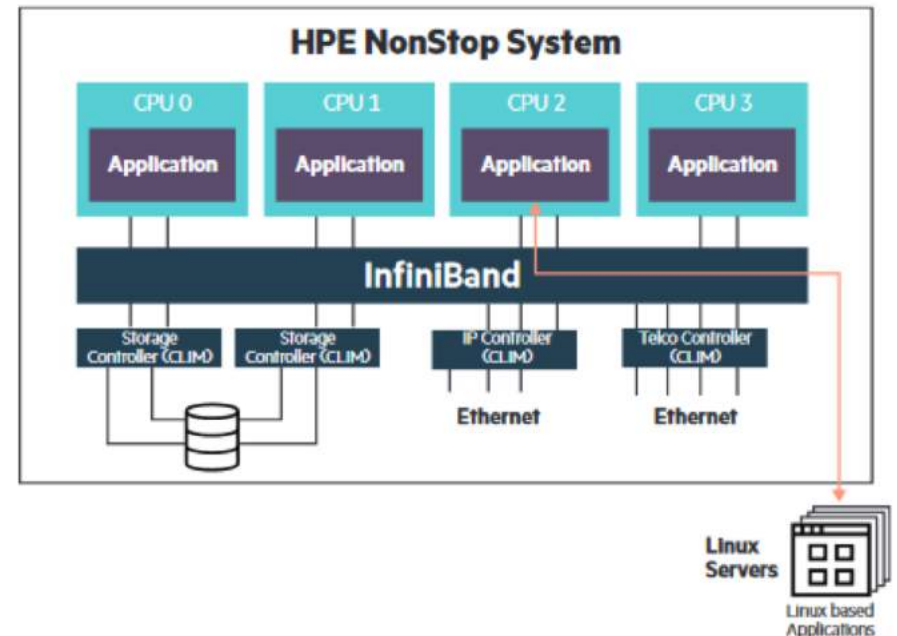
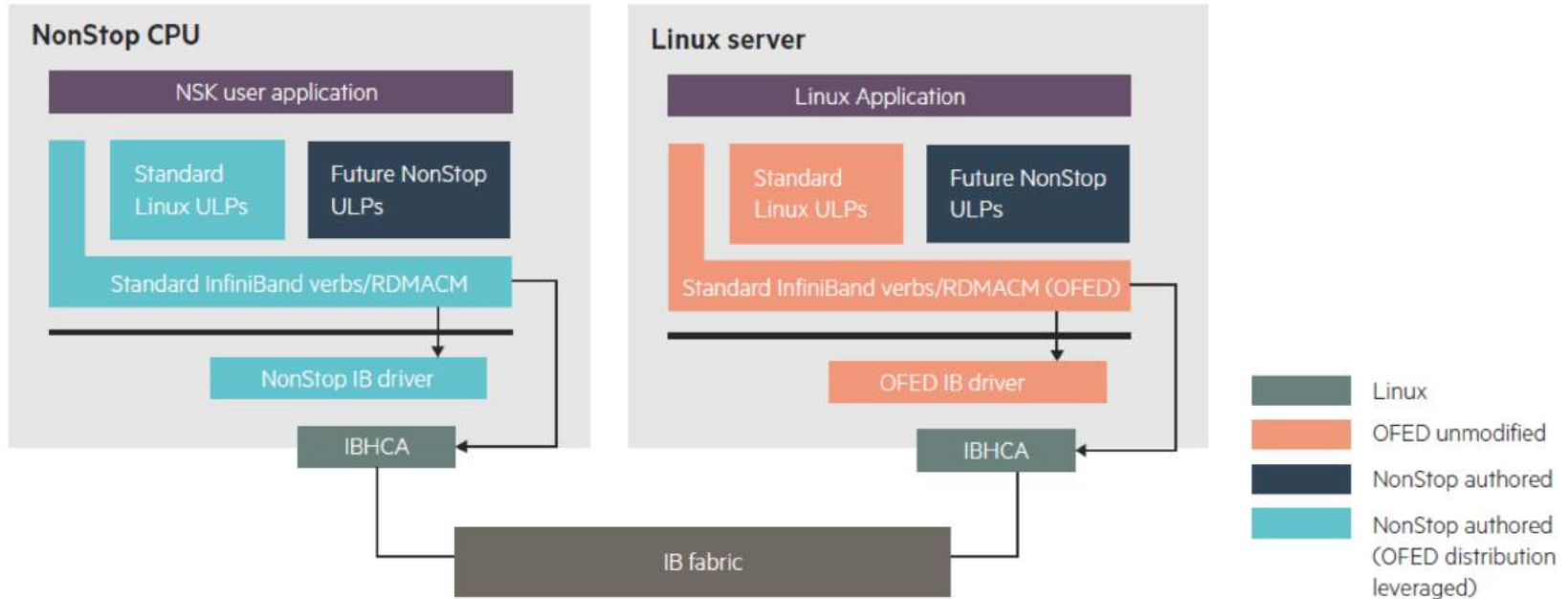
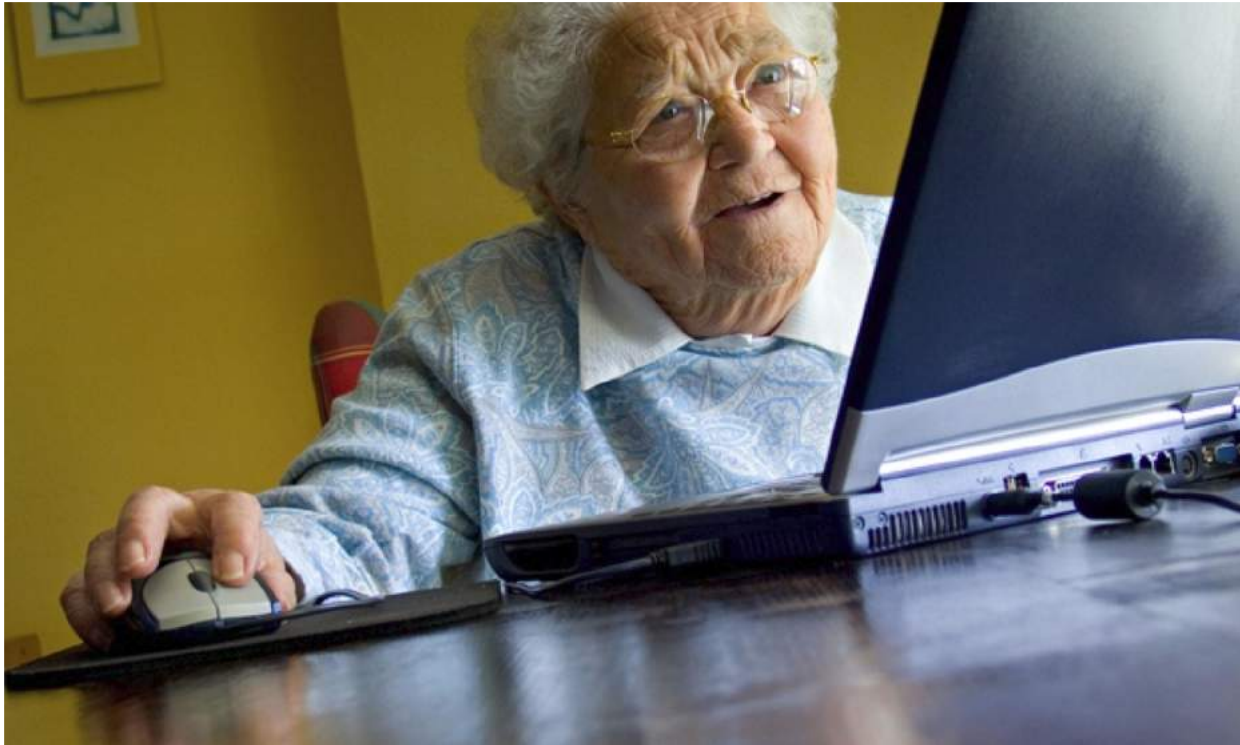


Figure 2: Connecting NonStop X and Linux applications using NSADI

How to integrate TANDEMS into an architecture



Why doesn't everyone use nonstop technology?



So .. What?

- Why aren't all computers Non-Stop?
 - Cost.
- Why aren't all processes duplicated in a process-pair?
 - Complexity
- What would be a good domain for Nonstop systems?
 - Where downtime is too expensive!

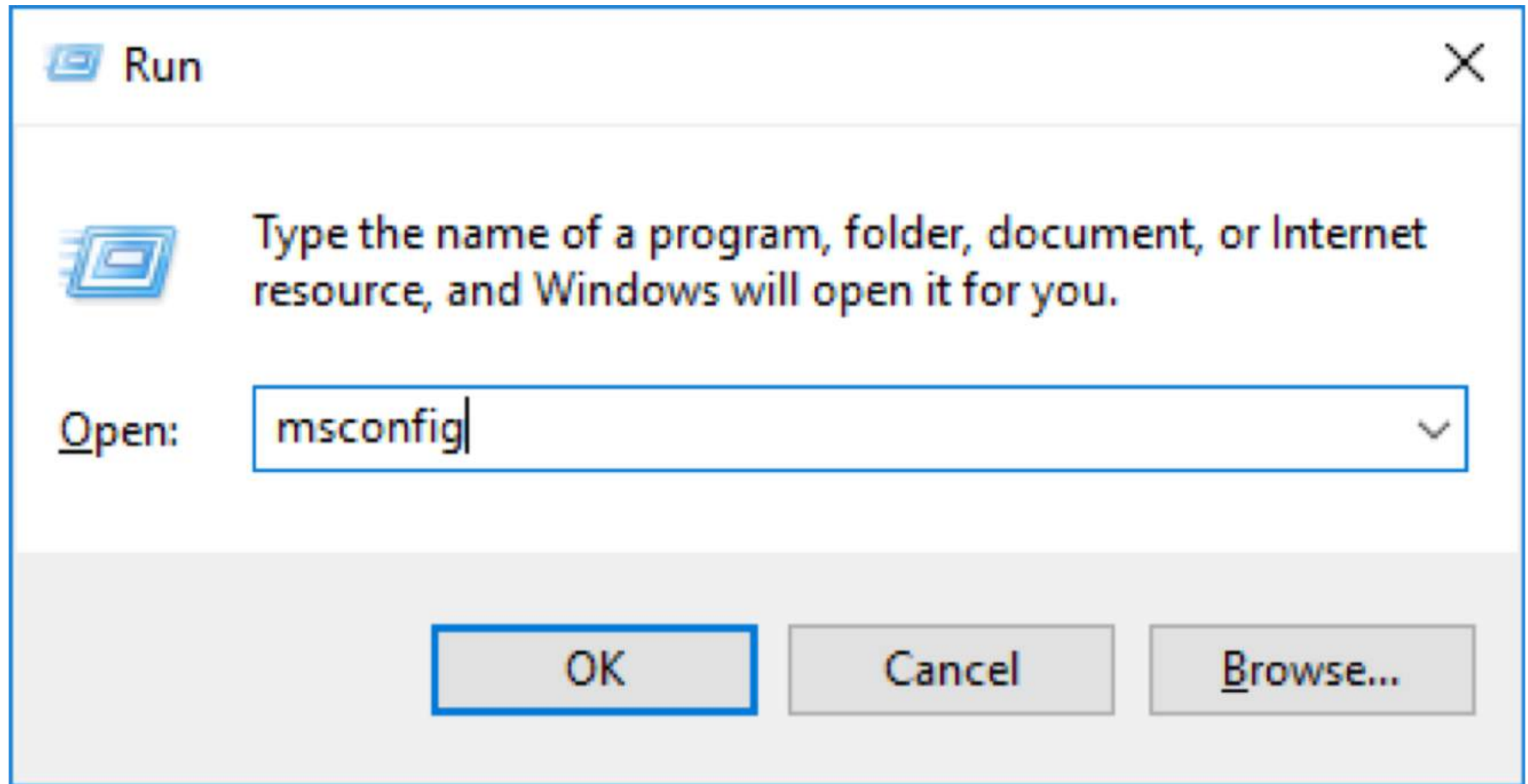
So ... How do you fix these beasts?

- The computers we are used to now are tightly built together.
- A system with discrete parts can be easily serviced by removing failing components.
 - Changing flat tires while moving
- This approach is not the norm.
 - Usually not worth the cost / complexity / trouble!

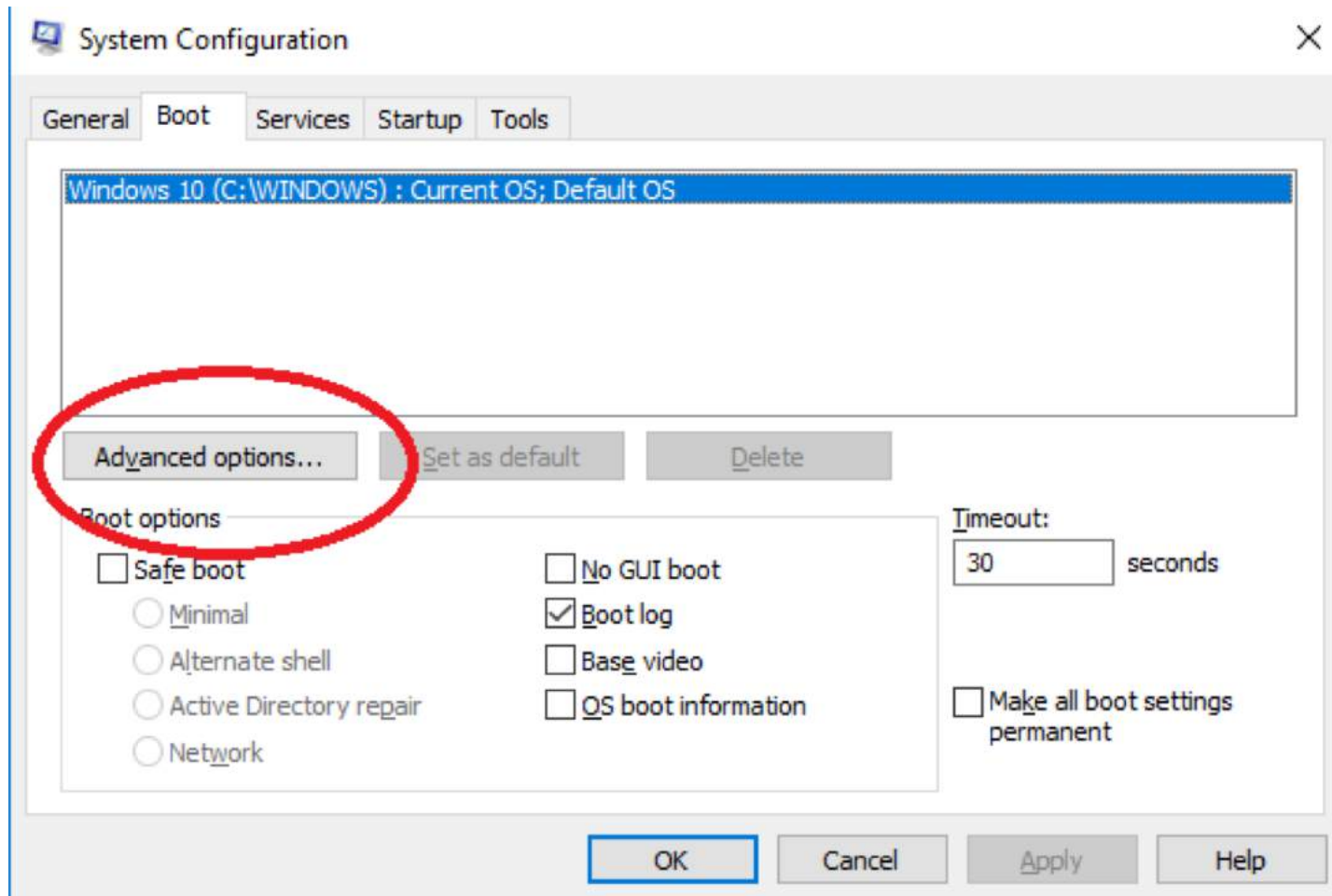
Let's look behind the curtain ...



Simple Hardware Example (1):




Simple Hardware Example (2):



Simple Hardware Example (3):

BOOT Advanced Options

☒ Number of processors:  ☐ Maximum memory:

4 0

☐ PCI Lock

☐ Debug

Global debug settings

☒ Debug port: ☒ Baud rate:

COM1: 115200

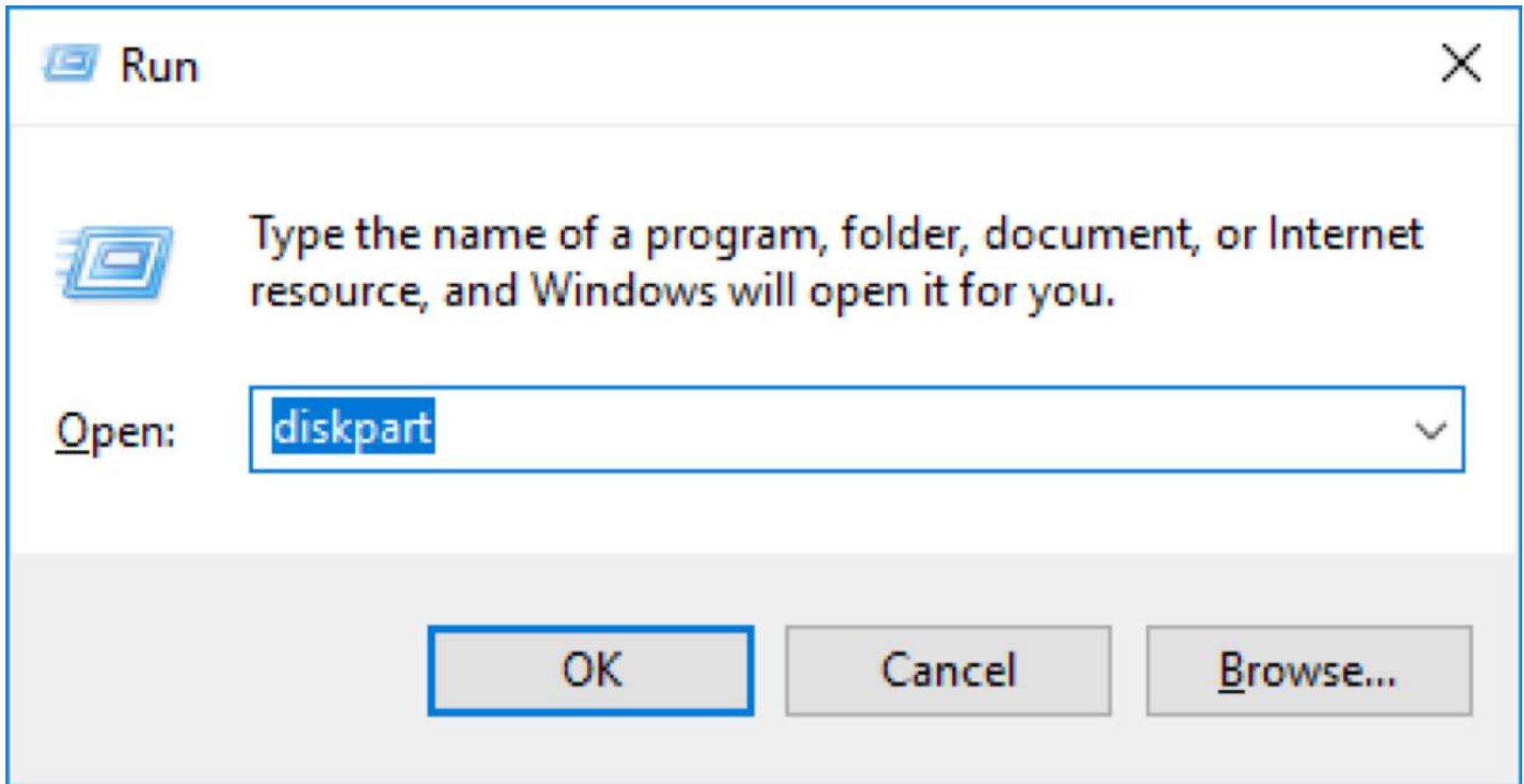
☐ Channel:

0

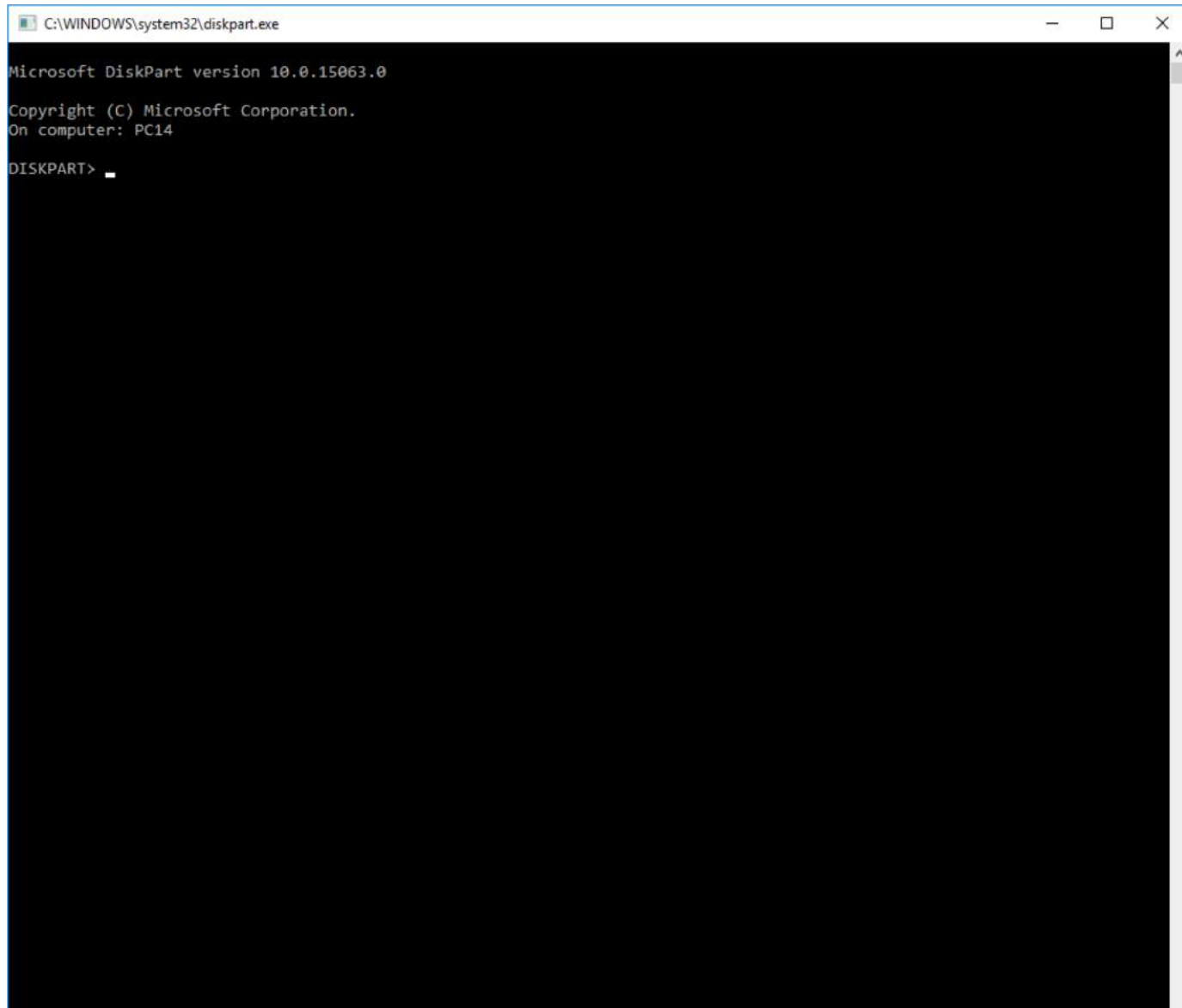
USB target name:

OK Cancel

Simple Hardware Example (4):



Simple Hardware Example (5):

A screenshot of a Windows command prompt window titled "C:\WINDOWS\system32\diskpart.exe". The window has a black background with white text. The text inside the window reads: "Microsoft DiskPart version 10.0.15063.0", "Copyright (C) Microsoft Corporation.", "On computer: PC14", and "DISKPART> _".

```
C:\WINDOWS\system32\diskpart.exe

Microsoft DiskPart version 10.0.15063.0

Copyright (C) Microsoft Corporation.
On computer: PC14

DISKPART> _
```

Simple Hardware Example (6):

```
C:\WINDOWS\system32\diskpart.exe
On computer: PC14

DISKPART> help

Microsoft DiskPart version 10.0.15063.0

ACTIVE          - Mark the selected partition as active.
ADD             - Add a mirror to a simple volume.
ASSIGN          - Assign a drive letter or mount point to the selected volume.
ATTRIBUTES     - Manipulate volume or disk attributes.
ATTACH         - Attaches a virtual disk file.
AUTOMOUNT      - Enable and disable automatic mounting of basic volumes.
BREAK          - Break a mirror set.
CLEAN           - Clear the configuration information, or all information, off the
                disk.
COMPACT        - Attempts to reduce the physical size of the file.
CONVERT        - Convert between different disk formats.
CREATE         - Create a volume, partition or virtual disk.
DELETE         - Delete an object.
DETAIL         - Provide details about an object.
DETACH         - Detaches a virtual disk file.
EXIT           - Exit DiskPart.
EXTEND         - Extend a volume.
EXPAND         - Expands the maximum size available on a virtual disk.
FILESYSTEMS    - Display current and supported file systems on the volume.
FORMAT         - Format the volume or partition.
GPT            - Assign attributes to the selected GPT partition.
HELP           - Display a list of commands.
IMPORT         - Import a disk group.
INACTIVE       - Mark the selected partition as inactive.
LIST           - Display a list of objects.
MERGE          - Merges a child disk with its parents.
ONLINE         - Online an object that is currently marked as offline.
OFFLINE        - Offline an object that is currently marked as online.
RECOVER        - Refreshes the state of all disks in the selected pack.
                Attempts recovery on disks in the invalid pack, and
                resynchronizes mirrored volumes and RAID5 volumes
                that have stale plex or parity data.
REM            - Does nothing. This is used to comment scripts.
REMOVE         - Remove a drive letter or mount point assignment.
REPAIR         - Repair a RAID-5 volume with a failed member.
RESCAN         - Rescan the computer looking for disks and volumes.
RETAIN         - Place a retained partition under a simple volume.
SAN            - Display or set the SAN policy for the currently booted OS.
SELECT         - Shift the focus to an object.
SETID          - Change the partition type.
SHRINK         - Reduce the size of the selected volume.
UNIQUEID       - Displays or sets the GUID partition table (GPT) identifier or
                master boot record (MBR) signature of a disk.

DISKPART>
```

Simple Hardware Example (7):

```
C:\WINDOWS\system32\diskpart.exe

Copyright (C) Microsoft Corporation.
On computer: PC14

DISKPART> LIST VOLUME

Volume ###  Ltr  Label          Fs      Type          Size      Status       Info
-----
Volume 0      D             DVD-ROM        0 B      No Media
Volume 1      D             SYSTEM RESE   NTFS     Partition    100 MB     Healthy      System
Volume 2      C      PC14           NTFS     Partition   1382 GB    Healthy      Boot
Volume 3      D      PQSERVICE     NTFS     Partition    15 GB     Healthy      Hidden
Volume 4      F             Removable     0 B      No Media
Volume 5      G             Removable     0 B      No Media
Volume 6      H             Removable     0 B      No Media
Volume 7      I             Removable     0 B      No Media
Volume 8      J             Removable     0 B      No Media
Volume 9      L      GoFlex_L@1E    NTFS     Partition    931 GB     Healthy
Volume 10     M      Seagate BKU    NTFS     Partition   4657 GB    Healthy
Volume 11     K      PC11_D_K@D8    NTFS     Partition    596 GB     Healthy

DISKPART> SELECT VOLUME 11

Volume 11 is the selected volume.

DISKPART> _
```


Simple Hardware Example (8):

```
C:\WINDOWS\system32\diskpart.exe
-
□
X

INACTIVE - Mark the selected partition as inactive.
LIST      - Display a list of objects.
MERGE     - Merges a child disk with its parents.
ONLINE    - Online an object that is currently marked as offline.
OFFLINE   - Offline an object that is currently marked as online.
RECOVER   - Refreshes the state of all disks in the selected pack.
            Attempts recovery on disks in the invalid pack, and
            resynchronizes mirrored volumes and RAID5 volumes
            that have stale plex or parity data.
REM       - Does nothing. This is used to comment scripts.
REMOVE    - Remove a drive letter or mount point assignment.
REPAIR    - Repair a RAID-5 volume with a failed member.
RESCAN    - Rescan the computer looking for disks and volumes.
RETAIN    - Place a retained partition under a simple volume.
SAN       - Display or set the SAN policy for the currently booted OS.
SELECT    - Shift the focus to an object.
SETID     - Change the partition type.
SHRINK    - Reduce the size of the selected volume.
UNIQUEID  - Displays or sets the GUID partition table (GPT) identifier or
            master boot record (MBR) signature of a disk.

DISKPART> SELECT VOLUME 11

Volume 11 is the selected volume.

DISKPART> OFFLINE
```


Simple Hardware Example (9):

- LINUX CPU Control Commands

To check information about your CPU, run this command from the Terminal:

```
cat /proc/cpuinfo
```

To disable a CPU core, run now this command:

```
echo 0 | sudo tee  
/sys/devices/system/cpu/cpu1/online
```

To enable it, use this command:

```
echo 1 | sudo tee  
/sys/devices/system/cpu/cpu1/online
```

Simple Hardware Example (10):

- LINUX Drive Control Commands

To unmount a LINUX disk, use the Umount command.

DISKPART uses a GUI to mount / unmount and more.

Final Thoughts

- Good People.
- Good Ideas.
- Good Designs.
- Good Practices.
- Good Questions?
- **Good Night!**

