Mainframes:
z/VM and Linux for zSeries

Malcolm Beattie
EMEA Enterprise Server Group
IBM UK

beattiem@uk.ibm.com

19 Nov 2002

Introduction

- Introduction
- Hardware
- z/VM
- Linux for zSeries
- History
- Questions

Hardware

- Specs
- Performance
- Reliability
- Dynamic changes

Hardware specs

- Models
  - Plenty of very old models
  - S/390 (31-bit): ..., G5, G6 (“Generation 6”)
  - Current models are 64-bit
  - One frame or two (separate disk subsystem)
  - z900: small model with 12 processor MCM
  - z900: large model with 20 processor MCM
  - z800: “baby” z with 5 processor MCM

The MCM (Multi Chip Module)

- The MCM is the heart of the system
- 127mm x 127mm
- 35 chips on the 20 PU MCM (PU = “processor”)
- 2.5 billion transistors, 101 layers, 4224 pins...
- L1 cache: 256K+256K (I+D) for each PU
- L2 cache: 32MB (binodal)
- PUs run at 770 MHz or 970 MHz

Hardware specs: I/O

- MCM has 4 x MBA (Memory Bus Adapters)
  - each MBA has 6 x 1 Gbyte/s STI links for I/O
- Physical hardware limits
  - 64 GB RAM
  - 256 channels (each has a CHPID)
  - FICON or FCP: fibrechannel
    - 100 or 200MByte/s: max 96 FICON/FCP channels
  - ESCON: fibre (17 Mbyte/sec)

Hardware specs: I/O contd

- OSA-Express: Gigabit ethernet
  - 1 Gbit/s (or 2 Gbps?) (max 24 channels)
- ICB-3 links: for clustering (z/OS only for now)
  - 1 Gbyte/s (each takes an STI link)
- PCICA: crypto co-processors for SSL
  - max 12 CPUs (each uses up a CHPID)

Hardware reliability

- Each PU has two cores running all instructions in parallel and comparing
  - Any repeatable error causes transparent “sparing” to a free PU
- All memory (L1, L2, main, key) has error detection/correction
  - memory errors cause sparing where necessary
- All I/O paths are ECC protected
- All channels are spared...

Dynamic hardware changes

- Most hardware changes are concurrent
  - concurrent = non-disruptive = dynamic
  - zSeries docs explicitly state when any change or upgrade is disruptive: the norm is concurrency
- Concurrent:
  - channels, disk, network
- Disruptive:
  - CPU, memory, crypto (except by arrangement)

VM

- Virtualisation
- Reliability
- Manageability
- “Better than real”
- Monitoring
- Dynamic hardware changes

VM: Virtualisation

- VM stands for Virtual Machine
- VM is a hypervisor
- Its nucleus/kernel is CP (Control Program)
- CP provides virtual machines
  - also called guests
- in each of which a S/390 or z/Arch O/S runs
  - e.g. Linux, CMS, z/OS, MVS, VSE or VM itself
- they all think they have their own machine
- VM uses hardware assists for performance

Reliability

- You can run tens to thousands of O/Ses under one hypervisor
  - Test Plan Charlie ran 41400 Linux guests
  - Test Plan Omega ran 90000+ Linux guests
- so you want the hypervisor to be reliable
- VM is reliable
- 30 years of development
- used for mission critical servers

VM manageability

- VM traditionally used for hosting CMS
- CMS is a single-user O/S...
- ...but VM can run tens of thousands of them
- VM has plenty of tools for managing guests
- Very flexible, very powerful, very robust

VM “better than real” features

- VM mostly virtualises real hardware
- but it also can provide “better than real” virtual features to guests
  - GuestLAN: emulates real LAN and NICs
  - Tracing: at instruction/data/trap level
  - Monitoring: a hardware monitor for every guest
  - Spool: a combination of /tmp, email and scratch

VM monitoring

- S/390 or zSeries hardware tracks lots of performance and event monitoring info
- VM tracks lots of information too
- VM has some software to query all this
- add-on software makes this even better

VM: dynamic hardware changes

- VM can handle dynamic hardware changes
- Channels and devices can be added, removed and changed concurrently
- E.g.: end-to-end concurrent disk addition
  - wheel new disk system onto floor
  - connect up channels
  - configure online to LPAR
  - vary online to VM
  - attach to Linux guest
  - Add dynamically in Linux guest
  - Use LVM to add/resize filesystem

Linux for zSeries

- Just another port: Linux is Linux is Linux...
- Application porting
  - Endian: big (like POWER and SPARC)
  - zSeries: 64-bit (“ESAME”) and 31-bit (“ESA”)
  - S/390: 31-bit (“ESA”)
  - datatype sizes
    - ESA: 32-bit int/long, 31-bit pointers
    - ESAME: 32-bit int, 64-bit long/pointer (i.e. LP64)
- Inherits advantages from hardware & z/VM

z/Linux inherits advantages...

- from hardware
  - reliability
  - performance
- from z/VM
  - dynamic hardware changes
  - inter-guest comms: high-speed, flexible, secure
  - monitoring, tracing
  - low-overhead addition of guests for
    - hot-standby, devel, test, QA, fall-back, ...

z/Linux and z/VM work together

- z/Linux and z/VM work well together
  - Sharing
    - Disks, Filesystems, Network ports, CPU, memory, ...
  - Cloning
    - Create a fresh new Linux guest in < 30 secs
  - Dynamic changes
    - to hardware (virtual and real) and resources
- Some benefits available without z/VM too
- Demo (if local network allows IPsec)

History

- Hardware/architecture
  - S/360 about 40 years old
  - z/Arch mostly backward compatible with
    - ESA/390, ..., S/370, S/360
- VM
  - 30 years old this year