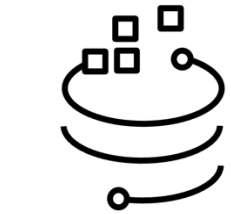
IBM Storage Scale on ARM CPU Technology News and Roadmap

IBM Storage Scale User Group
Heidelberg
March 19, 2025

Ingo Meents
IT Architect
Storage Scale Development

Disclaimer



- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Why ARM?

- Advanced RISC Machine
- Processor design licensed from ARM limited
- Simple RISC architecture, 32 and 64 bit
- Efficiency: embedded, mobile devices
- Growing into HPC, AI, ML

https://www.arm.com/markets/computing-infrastructure/high-performance-computing

TOP 500 list
Fugaku super computer

https://www.top500.org/system/179807/

European
Processor Initiative

https://www.european-processor-initiative.eu/

Grace-CPU
DPU

https://www.nvidia.com/de-de/data-center/grace-cpu/

AWS
Graviton 2 and 3

https://aws.amazon.com/de/ec2/graviton/

Top 500 – ARM systems

Fugaku super computer

https://www.top500.org/system/179807/

List	Rank	System	Vendor	Total Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
11/2024	6	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
06/2024	4	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
11/2023	4	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
06/2023	2	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
11/2022	2	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
06/2022	2	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
11/2021	1	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
06/2021	1	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
11/2020	1	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,630,848	442.01	537.21	29,899.23
06/2020	1	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Fujitsu	7,299,072	415.53	513.85	28,334.50

Alps

https://www.top500.org/system/180259/

List	Rank	System	Vendor	Total Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
11/2024	7	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS	HPE	2,121,600	434.90	574.84	7,124.00
06/2024	6	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11	HPE	1,305,600	270.00	353.75	5,194.00

Note: Systems exploiting ARM processors, not necessarily Scale

Top 500: Nov 24, Processor generation

	Processor Generation	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
1	AMD Zen-3 (Milan)	71	14.2	2,800,456,750	4,043,863,695	25,527,824
2	Xeon Gold 62xx (Cascade Lake)	70	14	266,621,090	506,973,014	6,673,752
3	AMD Zen-2 (Rome)	61	12.2	699,207,000	976,208,256	11,495,184
4	Xeon Platinum (Sapphire Rapids)	57	11.4	1,865,189,440	2,792,332,744	8,558,232
5	Xeon Gold (Skylake)	42	8.4	132,229,890	226,964,260	3,169,028
6	Xeon Platinum 83xx (Ice Lake)	35	7	421,598,380	607,804,680	4,969,236
7	Xeon Platinum 82xx (Cascade Lake)	32	6.4	124,531,650	222,585,260	2,618,512
8	AMD Zen-4 (Genoa)	30	6	2,341,574,600	3,594,181,890	15,665,648
9	Intel Xeon E5 (Broadwell)	17	3.4	61,314,194	74,758,062	1,912,180
10	Xeon Platinum (Skylake)	13	2.6	92,817,998	133,786,221	1,616,212
11	NVIDIA Grace	9	1.8	768,594,000	1,031,628,340	3,796,848
12	Intel Xeon E5 (Haswell)	9	1.8	47,788,060	65,587,808	1,381,020
13	Fujitsu A64FX	9	1.8	534,794,900	646,034,527	9,183,744

ARM Neoverse Family

Group of 64-bit ARM processor cores intended for datacenter, edge computing, and high-performance computing

Neoverse Series	Intended Usage	Level	Instruction Set	Examples
Neoverse N-series (scale out performance)	data center usage	N1	ARMv8.2-A	Ampere Altra (2-socket 80 cores) AWS Graviton2 (64 cores) Huawei Kunpeng 920
		N2	ARMv9.0-A	Alibaba Yitian 710
Neoverse E-series (efficient throughout)	edge computing	E1	ARMv8.2-A	
		E2	ARMv9.0-A	
Neoverse V-series (max performance)	high performance computing	V1	ARMv8.4-A	AWS Graviton3 (64 cores) Center for Dev of Advanced Computing (C-DAC) AUM
		V2	ARMv9.0-A	Nvidia Grace (144 cores) AWS Graviton 4 Google Axion
A64FX, Fujitsu	HPC		Armv8.2-A + SVE	Supercomputer Fugaku

V3 announced in fall 2024

This is a general list of where ARM can be found, how it can be categorized and some examples. This is not a Scale support list.

Hardware Examples

NVIDIA ARM HPC Developer Kit Server - Ampere® Altra® Max ARM Server

- Single socket Ampere® Altra® Max or Altra® Processor
- Up to 2 x NVIDIA ® A100 PCIe Gen4 GPU cards
- Up to 2 x NVIDIA ® BlueField-2 DPUs
- 8-Channel RDIMM/LRDIMM DDR4, 16 x DIMMs

Our development & test platform

QuantaGrid S74G-2U

- NVIDIA GH200 Grace™ Hopper™ Superchip
- NVIDIA Grace™ 72 Arm® Neoverse V2 cores
- 1 Processor
- NVIDIA® NVLink®-C2C 900GB/s
- 3 PCIe 5.0 x16 FHFL Dual Width slots

Grace Hopper

Blue Field3 DPU

- Up to 16 Armv8.2+ A78 Hercules cores (64-bit)
- 16GB on-board DDR5

DPU

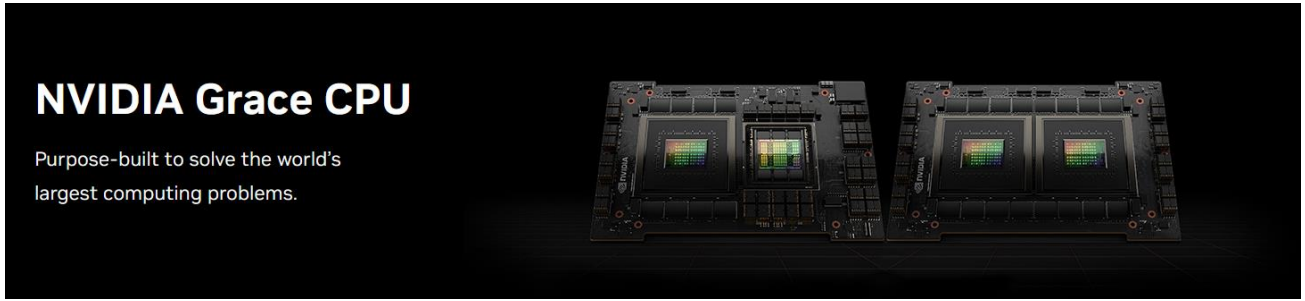
CPU Fujitsu A64FA

https://www.fujitsu.com/global/products/computing/servers/supercomputer/a64fx/

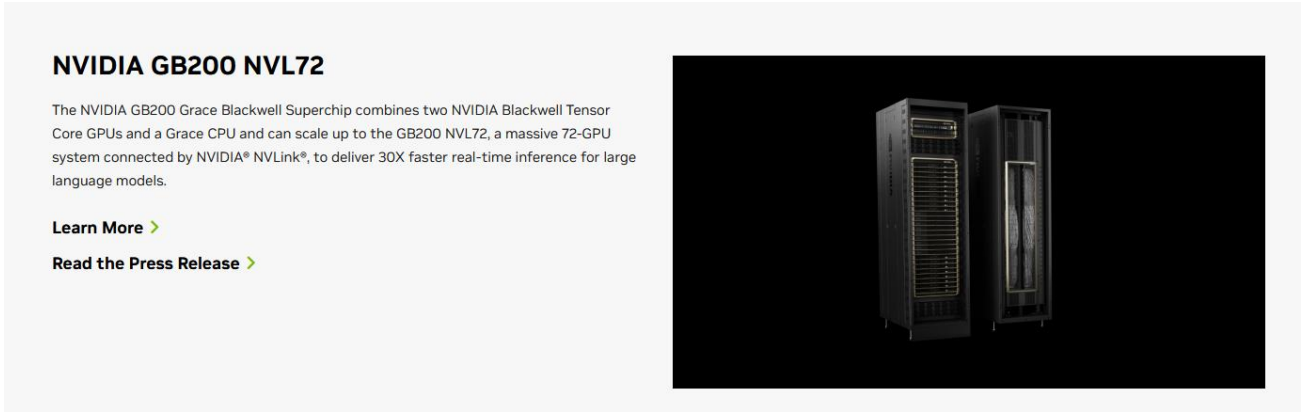
AWS Graviton-Prozessor in Amazon EC2

https://aws.amazon.com/de/ec2/graviton/

ARM @ Nvidia



Grace Grace Super Chip
Grace Hopper Super Chip



Grace Blackwell Super Chip
announced in March
at GTC24

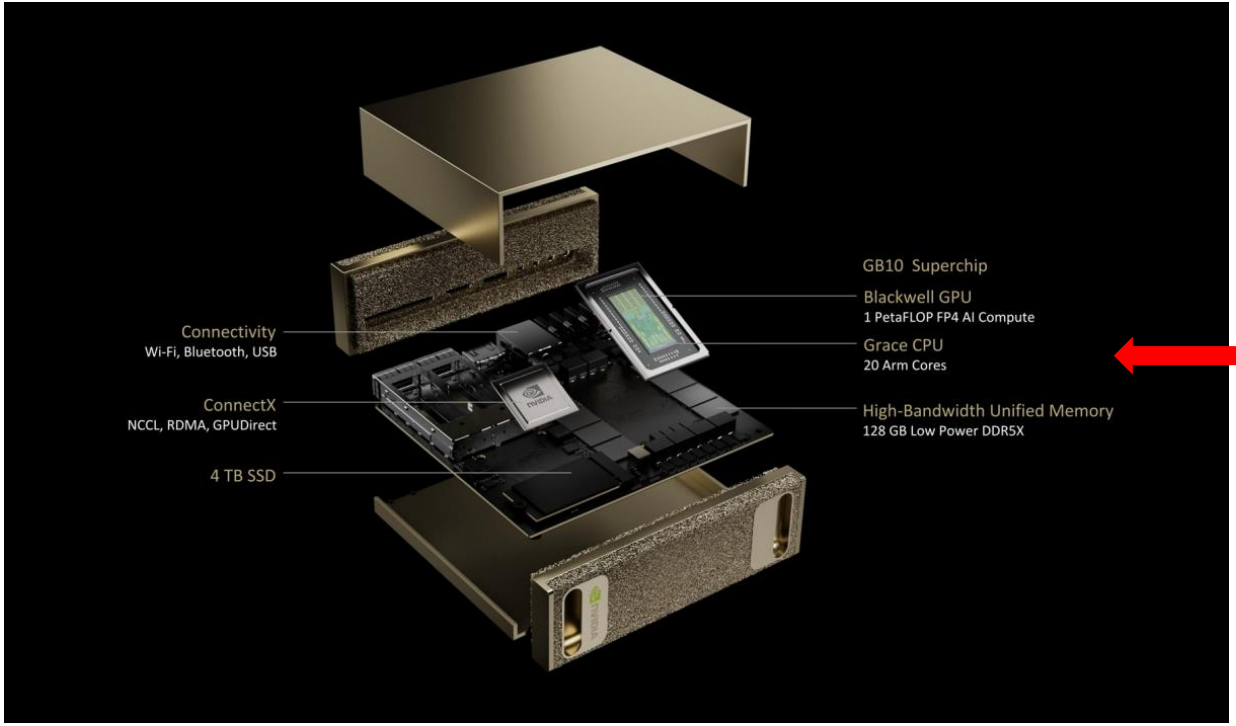
Grace Blackwell & Grace Hopper

https://www.nvidia.com

Grace = ARM CPU where our clients runs
Hopper or Blackwell = GPU where we can put data with GDS

Nvidia Project DIGITS

CES announcement 2025



ARM GPU

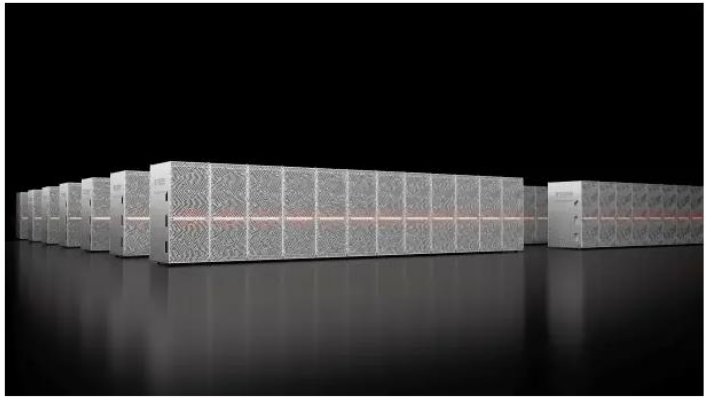
https://www.nvidia.com/en-us/project-digits/

ARM @ Jülich

Jupiter becomes the first German exaflops supercomputer

FZ Jülich explains how Jupiter 2024 combines the EU processor Rhea1, US technology from Nvidia and British ARM calculation cores.

Reading time: 3 min. Save in Pocket 8



The exaflops supercomputer Jupiter at FZ Jülich (Image: Nvidia/Eviden)

Jupiter is equipped from 2024 and consists of two parts: Jupiter Booster and Jupiter Cluster. The latter consists of more than 1300 nodes, in which the European ARM processor SiPearl Rhea1 is stuck.

Booster power

In Jupiter Booster, around 6000 knots with four Nvidia Grace Hopper GH200s deliver about 1 FP64 flops. For AI algorithms that get by with only 8 bit accuracy, up to 93 eFlops are even ready – and a gigantic 2.3 petabyte of super-fast HBM3e storage.

Quadruple module Nvidia Quad GH200 with four Grace and Hopper chips each, which are networked with each other.
(Image: Nvidia)

The new quad GH200 modules from Nvidia are used in Jupiter. Each H100 chip (hopper) controls 96 GByte HBM3e-RAM with up to 4 TByte/s. For the 72 ARM cores of each Grace chips, 120 GB of LPDDR5X RAM with up to 500 GB/s are available. The H100 GPUs can also use this memory.

Jupiter is also equipped with 21 petabytes of fast mass storage, with 40 IBM Elastic Storage Servers 3500 with NVMe SSDs. Together, they should provide data transfer rates of up to 4 TByte/s when reading and 3 TByte/s when writing.

https://www.heise.de/news/Jupiter-wird-erster-deutscher-Exaflops-Supercomputer-9529768.html
(translated by Firefiox)

https://www.fz-juelich.de/en/ias/jsc/jupiter/tech
(update 2025, ESS 6000)

Quad Nvidia ConnectX-7 Cedar module

BullSequana XH3515-HMQ
At Super Computing, Hamburg, 2024

Quad Nvidia ConnectX-7 Cedar module

4 CX-7, GH200
Liquid cooled

Packages

SE packages

```
-rwxr-xr-x 1 root root 128181750 Apr 19 01:38 Storage_Scale_Data_Access-5.2.0.0-aarch64-Linux-install
-rw-r--r-- 1 root root        90 Apr 19 01:38 Storage_Scale_Data_Access-5.2.0.0-aarch64-Linux-install.md5
-rwxr-xr-x 1 root root 137226406 Apr 19 01:38 Storage_Scale_Data_Management-5.2.0.0-aarch64-Linux-install
-rw-r--r-- 1 root root        94 Apr 19 01:38 Storage_Scale_Data_Management-5.2.0.0-aarch64-Linux-install.md5
```

RHEL

```
gpfs_rpms/gpfs.license.dm-5.2.0-0.aarch64.rpm
gpfs_rpms/gpfs.crypto-5.2.0-0.aarch64.rpm
gpfs_rpms/gpfs.compression-5.2.0-0.aarch64.rpm
gpfs_rpms/gpfs.gskit-8.0.55-19.1.aarch64.rpm
gpfs_rpms/gpfs.gpl-5.2.0-0.noarch.rpm
gpfs_rpms/gpfs.base-5.2.0-0.aarch64.rpm
```

Ubuntu

```
gpfs_debs/gpfs.compression_5.2.0-0_arm64.deb
gpfs_debs/gpfs.gpl_5.2.0-0_all.deb
gpfs_debs/gpfs.gskit_8.0.55-19.1_arm64.deb
gpfs_debs/gpfs.crypto_5.2.0-0_arm64.deb
gpfs_debs/gpfs.license.dm_5.2.0-0_arm64.deb
gpfs_debs/gpfs.base_5.2.0-0_arm64.deb
```

Included in the “new” licences
- data access edition
- data management edition

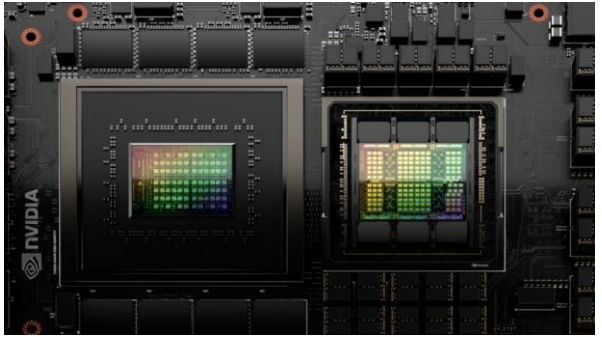
Not available in standard, advanced and Erasure code editions

A Look At Grace Hopper


```
[root@fscc-ghv1-1 ~]# lscpu
Architecture:            aarch64
  CPU op-mode(s):        64-bit
  Byte Order:            Little Endian
CPU(s):                  72
  On-line CPU(s) list:   0-71
Vendor ID:               ARM
  BIOS Vendor ID:        NVIDIA
  BIOS Model name:       Grace A02
  Model:                 0
  Thread(s) per core:    1
  Core(s) per socket:    72
  Socket(s):             1
  Stepping:              r0p0
  Frequency boost:       disabled
  CPU max MHz:           3474.0000
  CPU min MHz:           81.0000
  BogoMIPS:              2000.00
  Flags:                 fp asimd evtstrm aes pmull sha1 sha2 crc32 atomics fphp asimdhp cpuid asimdrdm jscvt fcma lrcpc dcpop
                         sha3 sm3 sm4 asimddp sha512 sve asimdfhm dit uscat ilrcpc flagm ssbs sb dcpodp sve2 sveaes svepmull sv
                         ebitperm svesha3 svesm4 flagm2 frint svei8mm svebf16 i8mm bf16 dgh
Caches (sum of all):
  L1d:                   4.5 MiB (72 instances)
  L1i:                   4.5 MiB (72 instances)
  L2:                    72 MiB (72 instances)
  L3:                    114 MiB (1 instance)

[root@fscc-ghv1-1 ~]# uname -a
Linux fscc-ghv1-1 5.14.0-284.30.1.el9_2.aarch64+64k #1 SMP PREEMPT_DYNAMIC Fri Aug 25 10:26:28 EDT 2023 aarch64 aarch64 aarch64 GNU/Linux

[root@fscc-ghv1-1 ~]# cat /etc/os-release | grep -i pretty
PRETTY_NAME="Red Hat Enterprise Linux 9.2 (Plow)"
```



http://www.nvidia.com

GDS on Grace Hopper


```
[root@fscc-ghv1-1 ~]# nvidia-smi
Mon Apr 22 16:36:02 2024
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 550.54.14              Driver Version: 550.54.14      CUDA Version: 12.4     |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  NVIDIA GH200 480GB             Off |   00000009:01:00.0 Off |                    0 |
| N/A   29C    P0             66W /  900W |       1MiB /  97871MiB |      6%      Default |
|                                         |                        |             Disabled |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI        PID   Type   Process name                              GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|  No running processes found                                                             |
+-----------------------------------------------------------------------------------------+

[root@fscc-ghv1-1 ~]# ofed_info -s
MLNX_OFED_LINUX-24.01-0.3.3.1:


[root@fscc-ghv1-1 ~]# gdsio -f /gpfs/gpfs1/foo -x0 -I1 -s1G -i4k -d0 -w8     # functional test against dummy storage
IoType: WRITE XferType: GPUD Threads: 8 DataSetSize: 1047476/1048576(KiB) IOSize: 4(KiB) Throughput: 0.905979 GiB/sec,
Avg_Latency: 33.545450 usecs ops: 261869 total_time 1.102620 secs
```

Some Major Porting Steps

- Adjust memory layout for ARM64 (shared segment, pagepool, token manager area, GNR area)
- Updates to “legacy” assembler code (memory barriers, atomic operations)
- Assembler optimizations for GPFS Mutexes (kernel & userspace, stats, debug/tracing)
- Hardware Interfaces (timer counter, memory management)
- Build system needs to account for new architecture (makefiles, packaging)
- Test tools need to be re-compiled / changed (unit, function, system & performance tests)
- Ksh – move to distro korn shell

Storage Scale on ARM - Timeline

- **5.1.9 / 5.1.9.1 Tech Preview (Nov 2023)**
  - RHEL 9.2. only
  - Some functional restrictions (client only, GDS, encryption, GUI, AFM, etc.)
- **5.2.0 Scale on ARM GA** (April 2024)
  - RHEL 9.3, Ubuntu 22.04
  - Client only, 4K & 64K kernels, most base scale functionality
    - Grace Hopper included
- **5.2.1 Scale on ARM 5.2.1** (August 2024)
    - NSD server functionality
- **5.2.2.1**
    - RHEL 9.5 support – missed Ubuntu 24.04 support
- **Roadmap items**
  - ECE, AFM, BlueField-3 support
- **Currently not in plan**
  - Protocols, BDA / HDFS, CNSA, TCT, HSM

Any questions?

meents@de.ibm.com

Special thanks to the team: Swen Schillig, Olaf Weiser, Ralph Würthner, Anthony Wong, Richard Zollo

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment. .