PowerAI 使用指南
——Basic Guide

IBM System Lab Services(GCG)

通过经验，交流和反复试验，我们学会了感知我们周围的世界







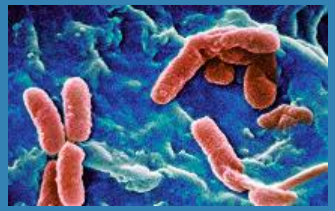
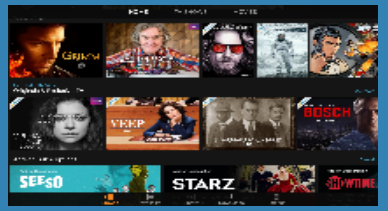

深度学习/人工智能词典

- 人工智能> 机器学习 > 深度学习
- 深度学习= 训练 (密集计算型，数据中心计算) + 推断 (嵌入式、边缘计算... 更加贴近用户端)
- 训练 = 类似神经元方式“灵感”，由数百万的数据点馈送... 重复驱动器的加权和连接
- 平台 = 软件框架 + 支持的库 + 计算
- 计算 = 加速器 +高带宽

- 期望成果: 在推断模型感知任务中具有更高的准确度

深度学习: 跨界跨领域应用

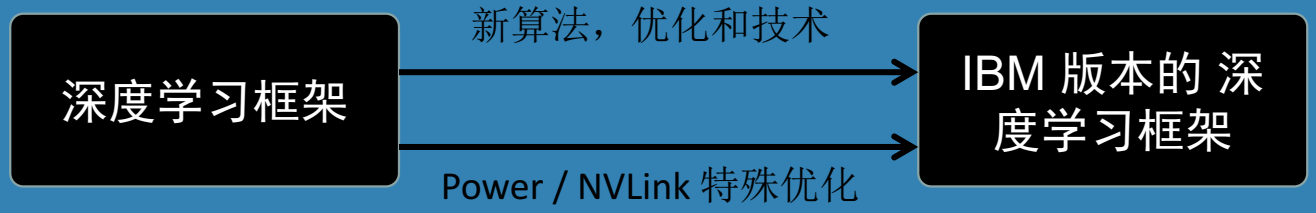


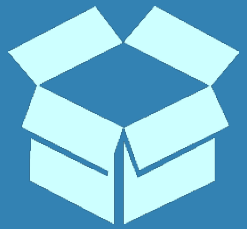
汽车和交通	国防，安全，公共安全	互联网，手机，零售消费者	生物医学	娱乐传媒
• 自动驾驶 • 行人检测 • 事故避免	• 视频监控 • 图像分析 • 面部识别和检测	• 图片标记 • 语音识别 • 自然语言处理 • 建议和情绪分析	• 药物探索 • 诊断帮助 • 癌细胞检测	• 字幕 • 搜索 • 建议 • 实时翻译

Power 人工智能/深度学习策略

- 拥抱扩展开源：增加创新，优化，新算法
- 增加系统级优化。利用NVLink Power平台，更好的集群（横向扩展）性能，简化入门过程
- 使用NVLink构建差异化的GPU加速系统解决方案



PowerAI介绍:
快速开始深度学习



预编译主要的深度学习框架软件包

易于安装 & 开始使用企业级支持的深度学习

利用NVLink优化性能

高性能计算基础设施使得这些变得可行

PowerAI 平台

深度学习框架

Caffe | NVCaffe | IBMCaffe | Torch

TensorFlow | DL4J | Theano

支持的函数库

OpenBLAS | Bazel | Distributed Frameworks | NCCL | DIGITS

加速服务器和基础架构

NVLink 服务器集群

高速并行文件系统

扩展到云

很快到来

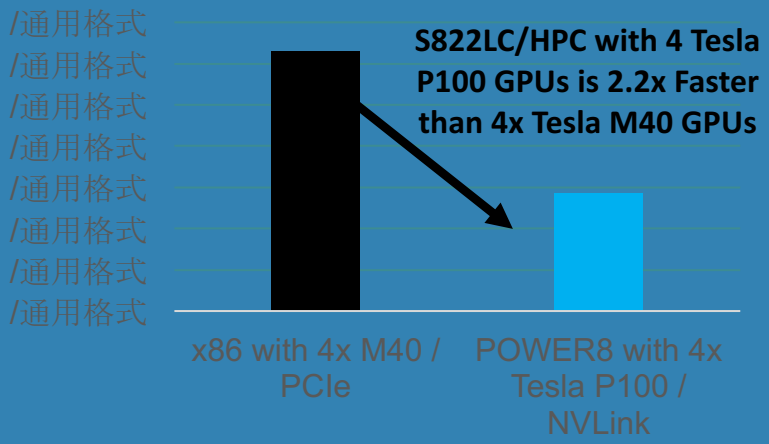
PowerAI 简化访问和安装

- 测试二进制构建的常见深度学习框架，以方便实施
- 简单完整安装的过程记录在ibm.biz/powerai
- 未来的重点是优化POWER的特定包：OpenBLAS，NVIDIA Caffe，TensorFlow和Torch

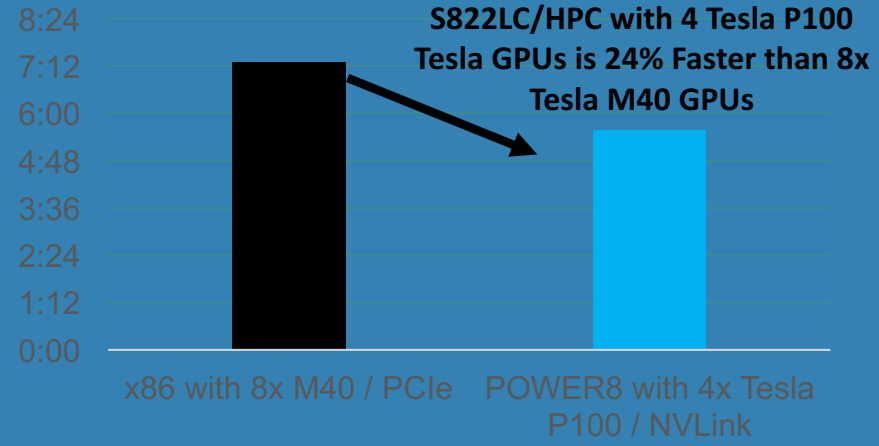
	PowerAI
OS	Ubuntu 16.04
CUDA	8.0
cuDNN	5.1
Built w/ MASS	Yes
OpenBLAS	0.2.19
Caffe	1.0 rc3
NVIDIA Caffe	0.14.5
IBM Caffe	1.0 rc3
NVIDIA DIGITS	v5.0.0
Torch	7
Theano	0.8.2
TensorFlow	v1.0.1 and v0.12.0
GPU	4 x P100
Base System	S822LC/HPC

PowerAI on S822LC for HPC: 2.2x Faster

Training time (minutes): AlexNet and Caffe to top-1, 50% Accuracy (Lower is better)



BVLC Caffe vs IBM Caffe / VGGNet Time to Top-1 50% accuracy: (Lower is better)

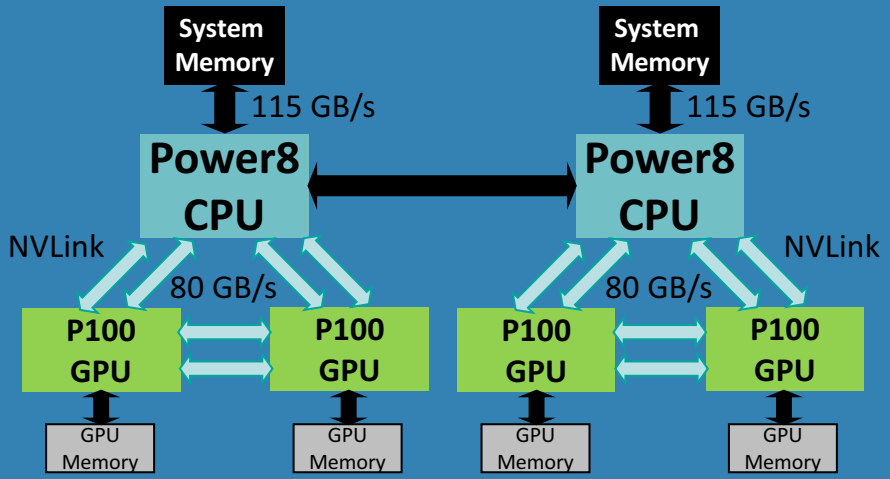


IBM S822LC 20-cores 2.86GHz 512GB memory / 4 NVIDIA Tesla P100 GPUs / Ubuntu 16.04 / CUDA 8.0.44 / cuDNN 5.1 / IBM Caffe 1.0.0-rc3 / Imagenet Data

Intel Broadwell E5-2640v4 20-core 2.6 GHz 512GB memory / 8 NVIDIA TeslaM40 GPUs / Ubuntu 16.04 / CUDA 8.0.44 / cuDNN 5.1 / BVLC Caffe 1.0.0-rc3 / Imagenet Data

PowerAI利用POWER8和P100之间的NVLink来提高系统带宽

- CPU和GPU之间的NVLink支持跟大的带宽，可以快速访问系统内存中的大型数据集
- GPU和CPU-GPU之间的两个NVLink连接导致更快的数据交换

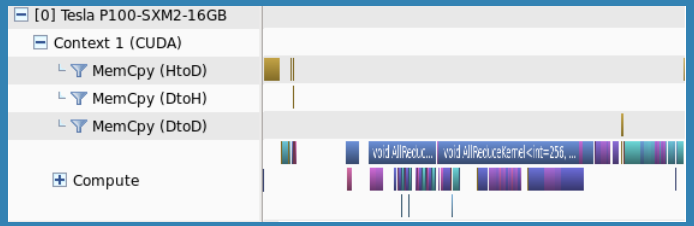


NVLink和P100 优势:
缩短通信时间，整合最快的GPU进行深度学习

- NVLink减少通信时间和开销
- GPU-GPU, Memory-GPU的数据获取更快，训练时间更短

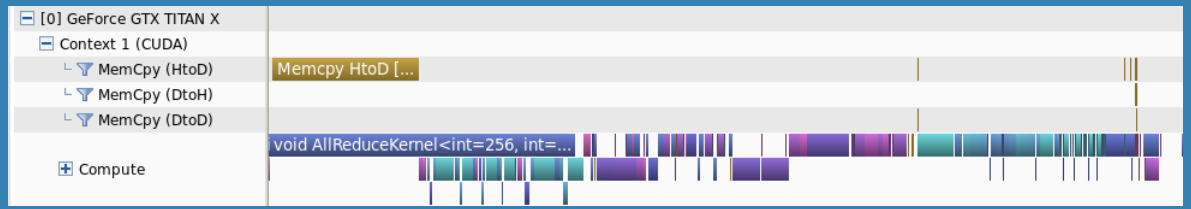
IBM 优势: 数据通信和GPU性能

POWER8 + Tesla P100+NVLink



78 ms

x86 based GPU system



170 ms

ImageNet / Alexnet: Minibatch size = 128

PowerAI的应用场景示例

- 销售主管：
- 更好地了解客户的观点，建议，提高互动性
- IT 主管:
  企业级支持，更好的性价比和共享基础设施

- 数据科学家和开发人员:
  NVLink意味着新的创新机会，更快的访问更大的内存，执行更大的训练模型和更多的迭代...完善，提高

- 研究科学家:
- 探索理解现有数据集的新方法

信息 – 深度学习

- 商业价值
  - 效率 – 减少的训练时间允许数据科学家快速迭代和改进模型
  - 准确 –更高的性能: 更多的运行 (集合), 更大的数据集, 为什么不是两者皆有?

- 为什么是IBM?
  - 超越深度学习 - 运行其他加速工作负载（模拟，分析，HPC）
  - PowerAI 平台
    - 专为生产系统
    - 易于实施 – 经过完整测试和构建框架，安装和配置详细的文档。
    - 软件包中包括应用程序，库和其他系统组件意味着更快的部署速度。
  - 首先进入市场 - 竞争对手仍在用以往的模式进行销售, IBM首先推出基于Tesla P100 GPU的系统 … 在深度学习领域取得先机。

S822LC for HPC: 2路Power8，4个GPU NVLink系统的PowerAI的具体配置

- 包含的:
- 2 POWER8 10 Core CPUs
- 4 NVIDIA P100 ”Pascal” GPUs
- 256 GB System Memory
- 2 SSD storage devices
- High-speed interconnect (IB or Ethernet, depending on infrastructure)

- 可选的:
- Up to 1 TB System Memory
- PCIe attached NVMe storage

PowerAI入门

- 安装PowerAI在你现有的S822LC的高性能计算服务器上

  http://ibm.biz/powerai

- 没有高性能服务器S822LC?
  - 参考架构/系统要求适用于使用POWER8，NVLink和Tesla P100的第一个系统（下一张幻灯片）
  - 访问IBM POWER HPC Cloud合作伙伴，以便今天在POWER8 / P100上测试这些框架
  - https://power.jarvice.com/ (Nimbix HPC Cloud)

PoweAI安装介绍

- **PowerAI可以通过两种方式安装，分别是网络和本地（ mldl-repo-local & mldl-repo-network），两种方式的安装过程都是通过配置repository进行安装，过程基本类似。**

- **下载PowerAI所需的repository package文件(.deb file)**
  from https://public.dhe.ibm.com/software/server/POWER/Linux/mldl/ubuntu/

- **安装repository package:**
  - $ sudo dpkg -i mldl-repo-*.deb
  - $ sudo apt-get update

通过apt-get 进行安装

1、把所有的框架一次全部安装

All the Deep Learning frameworks can be installed at once using the power-mldl meta-package:

$ sudo apt-get install power-mldl

2、根据需要使用的框架选择安装

The Deep Learning frameworks can be installed individually if preferred. The framework packages are: caffe-bvlc caffe-ibm caffe-nv chainer digits tensorflow theano torch

Each can be installed with:

$ sudo apt-get install <framework>

更新旧版本的PowerAI

安装新版本的repository package：

使用本地repo的话直接安装：

$ sudo dpkg -i mldl-repo-local_3.4.1_ppc64el.deb

使用网络repo的话需要先删除旧版本再安装：

$ sudo apt-get purge mldl-repo-local

$ sudo dpkg -i mldl-repo-network_3.4.0_ppc64el.deb

更新repository meta-data

$ sudo apt-get update

更新过程

更新PowerAI全部系统环境：

```
$ sudo aptitude dist-upgrade
```

仅更新PowerAI相关框架：

```
$ sudo aptitude upgrade power-mldl
```

仅更新某个PowerAI环境中的包：

```
$ sudo aptitude upgrade tensorflow
```

PowerAI上手体验

每个框架提供了一个shell脚本用来激活当前的环境变量，可以自己修改bashrc里的参数达到设置效果，但建议使用对应目录下的配置脚本来设置当前的环境变量，比如：

$source /opt/DL/<framework>/bin/<framework>-activate

在每个框架的目录下，都提供了一个基础功能测试的脚本用以验证功能是否完善:

$ <framework>-test

Caffe alternatives

Packages are provided for upstream BVLC Caffe (/opt/DL/caffe-bvlc), IBM optimized BVLC Caffe (/opt/DL/caffe-ibm), and NVIDIA's Caffe (/opt/DL/caffe-nv). The system default Caffe (/opt/DL/caffe) can be selected using Ubuntu's alternatives system:

$ sudo update-alternatives --config caffe

There are 3 choices for the alternative caffe (providing /opt/DL/caffe).

```
There are 3 choices for the alternative caffe (providing /opt/DL/caffe).

  Selection    Path                Priority   Status
------------------------------------------------------------
* 0            /opt/DL/caffe-ibm    100       auto mode
  1            /opt/DL/caffe-bvlc   50        manual mode
  2            /opt/DL/caffe-ibm    100       manual mode
  3            /opt/DL/caffe-nv     75        manual mode

Press <enter> to keep the current choice[*], or type selection number: 1
update-alternatives: using /opt/DL/caffe-bvlc to provide /opt/DL/caffe (caffe) in manual mode
```

Users can activate the system default caffe:

source /opt/DL/caffe/bin/caffe-activate

Or they can activate a specific variant. For example:

source /opt/DL/caffe-bvlc/bin/caffe-activate

Attempting to activate multiple Caffe packages in a single login session will cause unpredictable behavior.

Tuning recommendations

Recommended settings for optimal Deep Learning performance on the S822LC for High Performance Computing are:

Enable Performance Governor

```
$ sudo apt-get install linux-tools-common linux-tools-generic linux-cloud-tools-generic cpufrequtils lsb-release
$ sudo cpupower -c all frequency-set -g performance
```

Enable GPU persistence mode

Use nvidia-persistenced (http://docs.nvidia.com/deploy/driver-persistence/index.html) or

```
$ sudo nvidia-smi -pm ENABLED
```

Set GPU memory and graphics clocks

```
$ sudo nvidia-smi -ac 715,1480
```

For TensorFlow, set the SMT mode

```
$ sudo ppc64_cpu --smt=2
```

MNIST 数据集训练过程——数据准备

1、进入caffe目录下

cd $CAFFE_HOME

2、下载Mnist 数据集

./data/mnist/get_mnist.sh

3、将其转换成Lmdb数据库格式，执行完此shell脚本后，会在./examples、mnist下增加两个新目录，mnist_test_lmdb和mnist_train_lmdb

./examples/mnist/create_mnist.sh

Mnist 数据集训练过程——训练模型

训练模型：
./examples/mnist/train_lenet.sh

打开train_lenet.sh可以看到，事实上通过caffe 命令，加上solver文件作为参数实现的训练

```
#!/usr/bin/env sh
./build/tools/caffe train --solver=examples/mnist/lenet_solver.prototxt
```

$CAFFE_ROOT/examples/mnist/lenet-solver.prototxt

```
# The train/test net protocol buffer definition
net: "examples/mnist/lenet_train_test.prototxt" #网络模型文件路径
test_iter: 100 #test的迭代次数，要和test layer中的batch_size结合起来理解
test_interval: 500 #训练时每迭代500次测试一次
base_lr: 0.01  #学习率
momentum: 0.9  #动量
weight_decay: 0.0005 #权重衰减
lr_policy: "inv" #学习率策略
gamma: 0.0001
power: 0.75
display: 100 #每迭代100次显示
max_iter: 10000 #最大迭代次数。这个数设置太小，会导致没有收敛，精确度很低。设置太大，会导致震荡，浪费时间
snapshot: 5000 #snapshot用于设置训练多少次后进行保存
snapshot_prefix: "examples/mnist/mnist-model/mymnist"
solver_mode: GPU
```

Ready? Let's train!

```
root@nova-eabf7820-3853-4ee8-90d5-565be95cbca3:/home/opuser/caffe-master-20150813#  ./examples/mnist/train_lenet.sh
I0518 06:54:23.101384 24768 caffe.cpp:118] Use GPU with device ID 0
I0518 06:54:23.262435 24768 caffe.cpp:126] Starting Optimization
I0518 06:54:23.262526 24768 solver.cpp:36] Initializing solver from parameters:
test_iter: 100
test_interval: 500
base_lr: 0.01
display: 100
max_iter: 10000
lr_policy: "inv"
gamma: 0.0001
power: 0.75
momentum: 0.9
weight_decay: 0.0005
snapshot: 5000
snapshot_prefix: "examples/mnist/mnist-model/mymnist"
solver_mode: GPU
net: "examples/mnist/lenet_train_test.prototxt"
I0518 06:54:23.262779 24768 solver.cpp:74] Creating training net from net file: examples/mnist/lenet_train_test.prototxt
I0518 06:54:23.263417 24768 net.cpp:289] The NetState phase (0) differed from the phase (1) specified by a rule in layer mnist
I0518 06:54:23.263450 24768 net.cpp:289] The NetState phase (0) differed from the phase (1) specified by a rule in layer accuracy
I0518 06:54:23.263478 24768 net.cpp:44] Initializing net from parameters:
```

```
I0612 15:10:25.185286  1370 solver.cpp:456]     Test net output #0: accuracy = 0.9887
I0612 15:10:25.185308  1370 solver.cpp:456]     Test net output #1: loss = 0.036992 (* 1 = 0.036992 loss)
I0612 15:10:25.186771  1370 solver.cpp:251] Iteration 9500, loss = 0.0040812
I0612 15:10:25.186789  1370 solver.cpp:267]     Train net output #0: loss = 0.00408101 (* 1 = 0.00408101 loss)
I0612 15:10:25.186800  1370 sgd_solver.cpp:106] Iteration 9500, lr = 0.00606002
I0612 15:10:25.187003  1370 solver.cpp:287]     Time: 0.265889s/100iters
I0612 15:10:25.356017  1370 solver.cpp:251] Iteration 9600, loss = 0.00198696
I0612 15:10:25.356040  1370 solver.cpp:267]     Train net output #0: loss = 0.00198677 (* 1 = 0.00198677 loss)
I0612 15:10:25.356050  1370 sgd_solver.cpp:106] Iteration 9600, lr = 0.00603682
I0612 15:10:25.356214  1370 solver.cpp:287]     Time: 0.169183s/100iters
I0612 15:10:25.506953  1370 solver.cpp:251] Iteration 9700, loss = 0.00341411
I0612 15:10:25.506973  1370 solver.cpp:267]     Train net output #0: loss = 0.00341392 (* 1 = 0.00341392 loss)
I0612 15:10:25.506983  1370 sgd_solver.cpp:106] Iteration 9700, lr = 0.00601382
I0612 15:10:25.507148  1370 solver.cpp:287]     Time: 0.150913s/100iters
I0612 15:10:25.659124  1370 solver.cpp:251] Iteration 9800, loss = 0.00911176
I0612 15:10:25.659211  1370 solver.cpp:267]     Train net output #0: loss = 0.00911158 (* 1 = 0.00911158 loss)
I0612 15:10:25.659222  1370 sgd_solver.cpp:106] Iteration 9800, lr = 0.00599102
I0612 15:10:25.659389  1370 solver.cpp:287]     Time: 0.152221s/100iters
I0612 15:10:25.814558  1370 solver.cpp:251] Iteration 9900, loss = 0.00373757
I0612 15:10:25.814580  1370 solver.cpp:267]     Train net output #0: loss = 0.00373739 (* 1 = 0.00373739 loss)
I0612 15:10:25.814590  1370 sgd_solver.cpp:106] Iteration 9900, lr = 0.00596843
I0612 15:10:25.814755  1370 solver.cpp:287]     Time: 0.155347s/100iters
I0612 15:10:25.964187  1370 solver.cpp:506] Snapshotting to binary proto file examples/mnist/lenet_iter_10000.caffemodel
I0612 15:10:25.970006  1370 sgd_solver.cpp:273] Snapshotting solver state to binary proto file examples/mnist/lenet_iter_10000.solverstate
I0612 15:10:25.974470  1370 solver.cpp:357] Iteration 10000, loss = 0.00264752
I0612 15:10:25.974485  1370 solver.cpp:377] Iteration 10000, Testing net (#0)
I0612 15:10:26.090816  1370 solver.cpp:456]     Test net output #0: accuracy = 0.9912
I0612 15:10:26.090837  1370 solver.cpp:456]     Test net output #1: loss = 0.0286396 (* 1 = 0.0286396 loss)
I0612 15:10:26.090849  1370 solver.cpp:362] Optimization Done.
I0612 15:10:26.090864  1370 caffe.cpp:292] Optimization Done.

real    0m20.143s
user    0m30.636s
sys     0m5.188s
```

法律声明

商标信息

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.

特别通知

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.

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.