□□ CMS News Letter □□□□□□□□□□□□Core Micro Systems □□□

報道発表「Prime STOR ZFS HPC」のご報告

□□□□□□□□□□□□□□□□□□□□□□□□ Vol.010 2010/11/29 □□

いつもお世話になっております。
コアマイクロシステムズの吉政でございます。

本日は、24日に発表した「Prime STOR ZFS HPC」の報道発表を事前にご連絡したくメール致しました。

「Prime STOR ZFS HPC」はNexentaStor搭載の弊社フラグシップ製品の一つであり、ZFS／RAIDZ並列負荷分散方式を採用した10GB／Sを誇る超高速HPC／映像メディア向け低価格ネットワークストレージ製品になっています。是非ご注目頂きたいです。

それでは、今号も宜しくお願いいたします！！！

□□ INDEX □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[News & Topics]

◇11月24日報道発表のご報告

ZFS／RAIDZ並列負荷分散方式を採用した
10GB／Sを誇る超高速HPC／映像メディア向け低価格ネットワークストレージ
Prime STOR ZFS HPCを発売開始

◇10月1日報道発表のご報告

強力なストレージ仮想化機能により、オンデマンドでの性能強化と
柔軟な容量拡張を実現した
「Prime STOR ZFS ENTERPRISE」を10月1日より発売開始

[Web更新情報]

第2回クラウンドコンピューティングEXPO
サンウェイテクノロジー社殿のブースにてSuperSSDを展示いたしました。

インターネットと運用技術シンポジウム（IOTS2010） に協賛します。

Super RAID導入事例：某大学　映像関連学部を公開しました。

ドキュメントセンターの更新
・スーパーストレージ（半導体ストレージ関連）
・NexentaStor詳細資料

[テクニカルコラム：かんたん♪ ＺＦＳ]
第十一回：DTrace その 2

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

◆◆>[News & Topics]<◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆

◇11月24日報道発表のご報告

ZFS／RAIDZ並列負荷分散方式を採用した
10GB／Sを誇る超高速HPC／映像メディア向け低価格ネットワークストレージ
Prime STOR ZFS HPCを発売開始

＜報道発表原文＞
http://www.cmsinc.co.jp/news/2010/news2010112401.html

＜解説＞

HPCや映像関係のお客様の要件が厳しくなる中で他社と一歩差をつけるHPCおよび映像メディア向けの高い拡張性と低価格を両立した製品になります。実機のデモも可能です。また、OEMでの提供も可能になりますので、お気軽にお申し付けください。

＜掲載結果＞
キーマンズネット、Exciteニュース、ZDnet、朝日新聞（ネット版）、CNETなどの主要メディアに掲載されました。

・キーマンズネットの掲載情報は以下をご覧ください。
  http://www.keyman.or.jp/3w/prd/80/20032180/

◇10月1日報道発表のご報告

強力なストレージ仮想化機能により、オンデマンドでの性能強化と
柔軟な容量拡張を実現した
「Prime STOR ZFS ENTERPRISE」を10月1日より発売開始

＜報道発表原文＞
http://www.cmsinc.co.jp/news/2010/news2010093001.html

＜掲載結果＞
キーマンズネット、インターネットコム、毎日新聞（ネット）、
Yahoo!ニュース、infoseekニュース、livedoorニュース、MSN、
@Niftyニュース、Exciteニュースなど主要9メディアに掲載されました。

・キーマンズネットの掲載情報は以下をご覧ください。
  http://www.keyman.or.jp/3w/prd/05/20031505/
・インターネットコムの掲載情報は以下をご覧ください。
  http://japan.internet.com/webtech/20101001/9.html

◆◆>[Web更新情報]<◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆

第2回クラウンドコンピューティングEXPO
サンウェイテクノロジー社殿のブースにてSuperSSDを展示いたしました。
⇒ http://www.cmsinc.co.jp/news/2010/news2010110801.html

インターネットと運用技術シンポジウム（IOTS2010） に協賛します。
⇒ http://iot.ipsj.or.jp/iots/2010

Super RAID導入事例：某大学　映像関連学部を公開しました。
⇒ http://www.cmsinc.co.jp/casestudy/casestudy_05.html

◇　ドキュメントセンターのご紹介　◇

弊社のドキュメントセンターをご覧になったことはありますか？
去年の10月に開設して以来、技術資料などのダウンロードを大変多くの方にご利用頂いています。もし、まだ利用された事が無い方がいらっしゃいましたら、この機会に是非お試しください。ストレージのニッチな使い方や数値データなど役立つ情報が満載です。

＜ドキュメントセンター＞
http://www.cmsinc.co.jp/doccenter/index.html

＜更新情報＞
・スーパーストレージ（半導体ストレージ関連）
・NexentaStor詳細資料

◆◆>[テクニカルコラム：かんたん♪ ＺＦＳ　]<◆◆◆◆◆◆◆◆◆◆◆◆

第十一回：DTrace その 2

◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇

こんにちは。朝晩はずいぶんと寒くなってきました。そろそろ年賀状の準備などが気になる季節ですね。みなさんいかがお過ごしでしょうか。

このコラムでは実際に ZFS を使う上で役に立つ Tips を紹介します。

今までのファイルシステムとは一味違う ZFS　を活用して面倒なファイル管理の苦労を減らしましょう！

パフォーマンス改善を行う時、負荷に対して処理能力を持て余している部分を補強してもほとんど改善は望めません。一方、処理能力を目一杯使っている部分を補強すると、大きな改善を期待することができます。

前回は、システム情報を収集するためのツールとして DTrace の基本を紹介させていただきました。今回は、いつくかのサンプルプログラムを使用したシステム情報の収集方法について紹介します。

最初はどのプロセスがどのデバイスを使用しているのかを表示するプログラムです（Solaris 動的トレースガイド P.319）

```
#!/usr/sbin/dtrace -s

#pragma D option quiet

io:::start
{
    @[args[1]->dev_statname, execname, pid] = sum(args[0]->b_bcount);
}

END
{
    printf("%10s %20s %10s %15s¥n", "DEVICE", "APP", "PID", "BYTES");
    printa("%10s %20s %10d %15@d¥n", @);
}
```

テキストエディタでこのプログラムが書かれたファイルを作成し、以下のように実行してください（ファイル名を ioproc.d）とします。

```
# chmod 700 ioproc.d
# ./ioproc.d

^C
    DEVICE                  APP        PID           BYTES
       sd1                sched          0           65536
       sd1        zpool-syspool          5        16625152
      zfs2              pageout          2        17784832
```

自動的に終了しないので Ctrl + C を押して終了させます。DEVICE は I/O デバイス名、APP はプロセスの実行名、BYTES は入出力が行われたバイト数です。

以下のプログラムはデバイス毎の IO 速度を表示します（Solaris 動的トレースガイド P.320）

```
#!/usr/sbin/dtrace -s

#pragma D option quiet

io:::start
{
    start[args[0]->b_edev, args[0]->b_blkno] = timestamp;
}

io:::done
/start[args[0]->b_edev, args[0]->b_blkno]/
{
    this->elapsed = timestamp - start[args[0]->b_edev, args[0]->b_
blkno];
```

```
    @[args[1]->dev_statname, args[1]->dev_pathname] =
    quantize((args[0]->b_bcount * 976562) / this->elapsed);
    start[args[0]->b_edev, args[0]->b_blkno] = 0;
}

END
{
    printa(" %s (%s)¥n%@d¥n", @);
}
```

テキストエディタでこのプログラムが書かれたファイルを作成し、以下のように実行してください（ファイル名を iospeed.d）とします。

```
# chmod 700 iospeed.d
# ./iospeed.d

dtrace: aggregation size lowered to 1m
^C

sd1 (/devices/pci@0,0/pci1179,1@1d/storage@2/disk@0,0:r)
value  ------------- Distribution ------------- count
    32 |                                         0
    64 |                                         3
   128 |                                         1
   256 |@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@  2257
   512 |                                         1
  1024 |                                         0


sd2 (/devices/pci@0,0/pci8086,3410@9/pci1000,9280@0/sd@1,0:a)
value  ------------- Distribution ------------- count
   128 |                                         0
   256 |                                         1
   512 |                                         0
  1024 |                                         2
  2048 |                                         0
  4096 |                                         2
  8192 |@@@@@@@@@@@@@@@@@@@                      172
 16384 |@@@@@                                    52
 32768 |@@@@@@@@@@@                              108
 65536 |@@@                                      34
131072 |                                         0
```

実行すると計測を開始します。そして Ctrl + C が押されるまで情報を収集し、デバイス毎の統計情報を表示します。 iostat でも同じ情報を収集することができますが、このプログラムの方がより分かりやすい形で結果が表示されます。sd1 と sd2 を比較すると sd1 は 256 ～ 512 [KB/sec]、sd2 は 8 ～ 32 [MB/sec] の速度で読み書きが行われていることが分かります。

例えば、これが sd1 から読み出しして sd2 へ書くような処理だった場合、sd1 が律速であることが分かります。このような処理のパフォーマンスを改善するには sd1 を早いデバイスにすると効果的です。

次のプログラムは write() システムコールが復帰するまでの時間を、 プロセス毎に表示するプログラムです。

```
#!/usr/sbin/dtrace -s

syscall::write:entry
{
    self->ts = timestamp;
}
```

```
syscall::write:return
/self->ts/
{
    @time[execname] = quantize(timestamp - self->ts);
    self->ts = 0;
}
```

テキストエディタでこのプログラムが書かれたファイルを作成し、以下のように実行してください（ファイル名を writetm.d）とします。

```
# chmod 700 writetm.d
# ./writetm.d

dtrace: script './writetm.d' matched 2 probes
dtrace: aggregation size lowered to 2m
^C

  tar
           value  ------------- Distribution ------------- count
             512 |                                         0
            1024 |                                         20
            2048 |@@@@@@                                   23505
            4096 |@@@@@@@@                                 31122
            8192 |@@@@@@@@@@@@@@                           51358
           16384 |@@@@@@@@@                                34861
           32768 |@@                                       7059
           65536 |                                         1531
          131072 |                                         381
          262144 |                                         135
          524288 |                                         39
         1048576 |                                         3
         2097152 |                                         17
         4194304 |                                         46
         8388608 |                                         27
        16777216 |                                         36
        33554432 |                                         50
        67108864 |                                         40
       134217728 |                                         6
       268435456 |                                         0
```

実行すると計測を開始します。そして Ctrl + C が押されるまで情報を収集し、プロセス毎の統計情報を表示します。 ここでは tar についての結果のみを抜粋しています。

プログラムの write の部分を read に変更すれば、read() システムコールの所要時間を得ることもできます。 timestamp は D の組み込み変数でナノ秒単位のカウンタです。このプログラムのように相対的な時間を計測する場合に使用することができます。

今回のコラムでは、DTrace を使用したシステム情報の収集例を紹介させていただきました。次回は ARC の情報収集について紹介します。

（つづく）

■ 参考文献

(1) サンマイクロシステムズ、"SunOS リファレンスマニュアル 1 : ユーザコマンド"、Part No. 819-1210-13、2007 年 7 月

(2) サンマイクロシステムズ、"SunOS リファレンスマニュアル 1M : システム管理コマンド"、Part No. 819-1211-13、2007 年 7 月

(3) サンマイクロシステムズ、"Solaris 動的トレースガイド", Part No. 819-0395-12、2008 年 10 月

(4) サンマイクロシステムズ、"Dtrace ユーザガイド", Part No. 819-6259-10、2006 年 5 月

（高田　浩生）

□□
※　CMS BizStor Updateのバックナンバーはパートナーログインにアップされています。

編集後記

今年も残るところ1か月です。残務整理に入るのも大事ですが、来年の計画を立てることも大事ですよね。何ごとも「先手必勝」、「先んずれば人を制す。」だと改めて思います。その前に、「己を知れば百戦危うからず」ですよね。自分の現状を知った上での計画でなければ、単なる妄想に終わってしまいそうです。基本的なことですが本当に大事だと思います。（吉政）

コアマイクロシステムズでは、お客様事例を募集しております。事例公開にご協力いただきましたお客様の社名やお取り組みを弊社のホームページや販促資料に記載し、皆様のビジネスの露出度向上に些少でも貢献出来ればと思います。お気軽に弊社担当までお申し付けください。

―◇本メールマガジンの登録解除方法◇―――――――――――――――――
このメールに対する登録解除は★そのまま返信★でその旨をお伝えください。

本メールマガジンの送付先追加と送付先変更方法
このメールマガジンに★そのまま返信★で下記の項目をCut & Pasteして項目を埋め、返信して下さい。

変更の場合、旧送付先：
（１）メールアドレス：
（２）氏名　　：
（３）会社名　　：
（４）部署名　　：
（５）役職名　　：

発行人：コアマイクロシステムズ株式会社　吉政　忠志