

2005 linuxos erőműve

Egyesek nagy teljesítményű, kétprocesszoros rendszert akartak építeni. Mások csendes gépre vágytak, ami zenei célokra is megfelel. Úgy döntöttünk, mindkét igényt kielégítjük.

A 2005-ös évet nyugodtan hívhatjuk az energiakezelés évének. A processzorgyártók számtalan tanulmányban taglalják a wattokkal való takarékoskodást, és a februárban megrendezett *Linux-World* konferencián és kiállításon is komoly figyelmet kapott az energiakezelés.

Talán az iparág aggódni kezdett a globális felmelegedés miatt? Talán az informatikai cégek vezetői több kardhalat akarnak enni, és csökkenteni akarják a gyárak higanykibocsátását? Aligha. Napjaink kiszolgálóiban egyre több és forróbb processzor üzemel, és a felhasználók légkondicionáló rendszerei képtelenek megbirkózni ekkora terheléssel. A NASA-nak vízűtéssel kellett ellátnia 10240 processzort tartalmazó, *Columbia* nevű fűrtjét, ahogy arról a januári számban írtunk is.

Minden megtakarított watt 3,6 kJ-nyi, illetve, hogy pontosak legyünk, 3,4 BTU-nyi hőt jelent, amivel könnyebbé válik a felhasználó dolga. A penge kiszolgálókkal és az akár négy processzort is tartalmazó egy egység magas rendszerekkel teli adatközpontokban mindez a hő összeadódik.

Az asztali *Linuxok* möhön falják fel a linuxos kiszolgálók sok milliárd dolláros piacának maradványait, így az energiafogyasztás az asztali gépeknél is egyre nagyobb figyelmet kap.

A ventilátorok hangosak. Ha a processzorokat jobb energiakezeléssel látjuk el, akkor kevesebb hőt termelnek, és kevesebb ventilátorra lesz szükségünk, illetve a meglévőket

alacsonyabb fordulatszámon tudjuk működtetni. Mi ventilátorok helyett mást választottunk, mint még lesz róla szó.

Mi sem természetesebb, hogy az energiafogyasztás a hordozható számítógépek és az egyéb mobil eszközök esetében is fontos. A hálózati tápellátás nélküli üzemidő növelésének kérdésével későbbi írásainkban fogunk foglalkozni.

Alaplap: a rendszer szíve

Kedveljük a *Tyan* alaplapjait, ahogy az egyedi linuxos rendszereket építő cégek is. A négy darab *Opteron* processzor befogadására képes *Tyan Thunder K8QS Pro* kicsit később jött ki ahhoz, hogy a múlt év linuxos erőművének részévé válhasson. Az *AMD 8000*-es sorozatú lapkakészletre épül. Amikor lapkakészletet mondunk, egy picit más hardverelemre gondolunk, mint amit az *Intel* alapú gépekben láthatunk. Az *AMD64* alapú rendszerekben minden processzor saját, beépített memóriavezérlővel és saját memóriabankkal rendelkezik, ezek közt *HyperTransport* kapcsolat áll fenn. Az *AMD64*-es, többprocesszoros gépek valójában kisebb *NUMA (Non-Uniform Memory Architecture, nem egységes memória-architektúra)* gépek, és ezeknél magában a lapkakészletben nincs is memóriavezérlő.

Tavaly egy *Celestica A8440*-es szekrénybe szerelhető váz szolgált a linuxos erőmű alapjául. Bár az előre szerelt házakkal és tápegységekkel sok időt megtakaríthatunk, a tavalyi gép

meglehetősen hangosra sikeredett. Idén – a szokásos megoldáshoz visszatérve – minden alkatrészt külön-külön válogattunk össze.

A *K8QS Pro* két *PCI-X* busszal rendelkezik, az A és a B jelűvel. A B busz két 133 MHz-es *PCI-X* foglalat számára van elkülönítve, az A pedig kettő darab 66 MHz-es *PCI-X* és egy darab hagyományos *PCI* foglalat kiszolgálásáért felelős. A hálózati kapcsolatok létrehozását kettő darab *Broadcom BCM5704C Gigabit Ethernet* csatoló segíti, ezek szintén az A buszra csatlakoznak.

A megszokott kapuk is a rendelkezésünkre állnak, ezek közül mi csak az *USB* kapukat vettük igénybe. A *SCSI* és az *ATA* vezérlő kiegészítő jelleggel kérhető, erre nem árt odafigyelni, ha linuxos erőműünk építése közben az alaplapot hagyományos kiszolgáló szerepet játszó gépbe akarjuk beépíteni.

Ebbe a kiváló alaplapba az elérhető *Opteron* processzorok közül a legjobbakat (*846 HE* típusjelzés, 2 GHz-es órajel, 1 MB másodsztígyorsítótár) helyeztük be, azokból is mindjárt négyet. A rendszer tesztelése közben elérhetővé vált újdonságokat a széljegyzet taglalja. A rendszerbe a lehető legtöbb memóriát, 32 GB-os építettünk.

A házrajongók szerencsétlenségére az alaplap *SSI MEB* formátumú, vagyis mérete 13" x 16", azaz 330,2 x 406,4 mm. Számunkra ez nem okozott gondot, ugyanis idén egyedi házat használtunk, de a méret mindenképpen korlátozza azon házak körét, amelyek közül választhatunk.

1. kódrészlet: A /etc/fstab fájlban szereplő lemezrészek

```

LABEL=/nstor-OS      /          ext3    defaults    1 1
LABEL=/cfboot        /boot      ext3    defaults    1 2
LABEL=/nstor-DATA    /u1        ext2    defaults    1 2
none                 /dev/pts   devpts  gid=5,mode=620 0 0
none                 /dev/shm   tmpfs   defaults    0 0
none                 /proc      proc    defaults    0 0
none                 /sys       sysfs   defaults    0 0

```



■ 1. ábra Mi van az irodádban, Justin? A hűtést egy nulláról induló rendszerrel, az lm_sensors segítségével teszteltük.

Amikor kiválasztjuk egy egyedileg épített rendszer házat, legyen szó akár a folyó év erőművéről, akár más gépről, mindig valamivel nagyobb választunk, mint amit a neves gyártók egy hasonló géphez kínálnának.

A kisebb házakhoz kevesebb anyag kell, és szállítani is olcsóbb őket, mi viszont egyediségre törekszünk, ezért több helyre van szükségünk, egyrészt az eszközök hozzáadásához, másrészt a gépben végzett munkához.

Adattárolás

Ha teljesen csendes rendszert akarunk építeni, akkor az adattárolást a gép házában kívülre kell száműznünk. Régen ezt NFS-sel vagy három méteres kábelekkel csatlakozó külső SCSI-házak segítségével lehetett megoldani, ám azóta bővültek a lehetőségek.

Ha a meghajtók távol tartására USB, FireWire, SCSI (természetesen ez nem maradhat el), Fibre Channel vagy éppen ATA over Ethernet megoldásokat használhatunk, utóbbiról 2005. júniusában mi is írtunk. A különálló meghajtóház többé nem a vállalati adatközpontok kiváltsága.

További lehetőség a hálózatról végzett rendszerindítás, majd a tárolóhely NFS-en keresztüli befűzése. Mivel a Penguin vállalati kiszolgálószobákban látott eszközökkel dolgozik, és a Fibre Channel lenyűgöző eredményeket hozott a teljesítményteszt során, végül mellette döntöttünk: egy nStor 4320F Fibre Channel RAID-házat választottunk, amelyben 18 GB-os Hitachi meghajtók tárolták az operációs rendszert, további, nagyobb méretű Seagate meghajtók pedig további tárhelyet biztosítottak. Mivel önálló, a rendszerindítás tekintetében más kiszolgálótól nem függő rendszert akartunk összeállítani, a rendszerindítás céljára beszereltünk egy 256 MB-os Sandisk CompactFlash kártyát. Ez a gép számára pontosan úgy látszik, mint bármely ATA meghajtó, vagyis bármilyen személyi számítógépes alaplap képes róla betölteni a rendszert.

USB-kulcs használatára is gondoltunk, ám ahhoz bele kellett volna nyúlni az initrd és a GRUB beállításába. Természetesen annak is vannak előnyei, ha a rendszerindításra használt eszközt ki lehet húzni a gépből, illetve el lehet különíteni, de nem számoltunk azzal, hogy repülőtereken keresztül fogjuk utaztatni a gépet, titkosított, bizalmas adatokkal teli meghajtókkal.

Ha csendesnek készülő linuxos gépünket mindig rajta akarjuk hagyni a hálózaton, akkor rugalmasabban dönthetünk a rendszerindításról, és például PXE-t is használhatunk. Ha viszont alkalmanként valamelyik barátunkhoz is el akarjuk vinni a gépet, és ott szeretnénk zenét lejátszani vele, akkor szükségünk lesz a független rendszerindítás lehetőségére.

A 2005-ös év linuxos erőművének alkatrészei

Alaplap: Tyan Thunder K8QS Pro (S4882)

Processzor: 4 db AMD 846HE Opteron

Memória: 8 db 4 GB-os Registered ECC Samsung DDR PC2700 CL 2,5 DIMM

Tápegység: 510 W-os, átépített PC Power and Cooling Turbo-Cool 510 ATX

Ház: egyedi, Matt Fulvio tervei alapján Trevor Sherard készítette

Fibre Channel: Qlogic 2342 2 Gb Fibre Channel csatoló, két kapuval, 133 MHz, PCI-X

Rendszerindító eszköz: 256 MB-os Sandisk CompactFlash kártya, DCFB-256-A10

Adattárolás: nStor 4320F Fibre Channel RAID-ház

Merevlemezek: 2 db 18 GB-os, 10000 fordulat/perces Hitachi DK32DJ-18FC Fibre Channel meghajtó, RAID 1 tömbben (operációs rendszer) és 6 db 73 GB -os, 10000 fordulat/perces Seagate ST373405FC Cheetah 73LP FC Fibre Channel meghajtó, RAID 10 tömbben

Grafikus kártya: PNY NVIDIA Quadro NVS 280 PCI

Kijelzők: 2 db ViewSonic VX2000 20"-os LCD monitor, 1600 x 1200-as felbontással

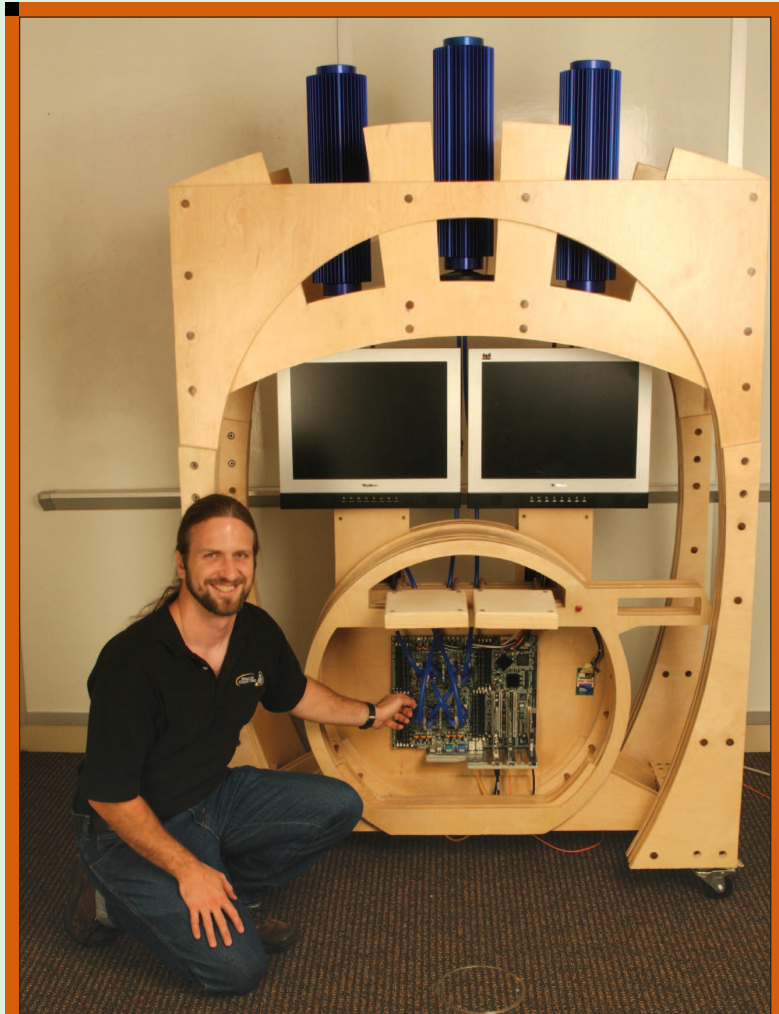
Hangkártya: RME HDSP9652 PCI hangkártya

Hang be- és kiviteli: 36 csatornás, 24 bites, 96 kHz-es RME Multiface be- és kiviteli egység

Hűtőrendszer: 3 db Zalman Reserator 1s

Vízhűtőblokkok a processzorokon: 4 db Zalman ZM-WB2 Gold blokk

Tápegység hűtése: egyedi, tervezte és készítette a Global Precision



■ **2. ábra** Semmi sem szívárog! Mindenki látja?

A *Penguin* csapat úgy tervezte, hogy elviszi a gépet a *LinuxWorld* rendezvényre, márpedig, ha bemutatóra kell cipelni egy gépet, akkor csak jó, ha egyfel kevesebb dolgot kell üzembe helyezni a helyszínen. Aki csendes *Linuxos* gépet épít és telepít, az akár kettős megoldást is kidolgozhat: *NFS*-t alkalmaz a felhasználók kezdőkönyvtárainak, a vállalati `/usr/local/bin/` könyvtár és az egyéb olyan anyagok tárolására, amelyeknek ugyan naprakésznek kell lenniük, de vészhelyzetben nélkülözhetők; míg a gép saját fájlrendszerét a nagyméretű munkafájlok – például a rendszer csúcscategóriájú hangkártyájától érkező adatok – tárolására tartja fenn. Végül, hogy még az egérek kattintások se törjék meg a csendet, a *Penguin* alapí-

tója, *Sam Ockman* egy érintőpadként működő *TouchStream LP* billentyűzetet javasolt, melynek nincsenek mozgó részei. Egyben mutatóeszközként is szolgál, és a műveletekhez egérgesztyűsok hozzárendelését is lehetővé teszi.

Hangrendszer

Ez az első alkalom, hogy a linuxos erőműbe profi hangkártyát szereltünk. Egy csendes gép hol is található jobb helyre, mint egy hangstúdióban?

Az általunk kiszemelt *RME Hammerfall HDSP9652* kártya akár 52 csatorna kezelésére is képes; beszereztünk hozzá egy *Multiface* nevű külső dobozkát is, amelyen 8 darab \square -os aljzat, valamint optikai, koaxiális és *MIDI* csatlakozó található.

A kártya gyakorlatilag kisebb stúdió egyetlen egységben, ugyanis belső keverővel rendelkezik, ami lehetővé teszi a jelek a kártyán belüli irányítását, minimális késleltetésekkel és a processzorra nézve elenyésző mértékű terheléssel; továbbá például a hagyományos, szalagos magnókat idéző szolgáltatásokkal is rendelkezik.

A legjobb benne mégis az, hogy az *RME 2000* óta támogatja az *Advanced Linux Sound Architecture (ALSA) Projectet*, vagyis esetében a *Linux* felhasználók nem csupán másodrangú polgároknak számítanak. Az *RME* webhelye szerint „a *Hammerfall ALSA-támogatása végre megtöri a mindannyiunk számára unalmassá vált, a tyúk vagy a tojás kérdéséhez hasonló nincs professzionális hardver vagy illesztőprogram – nincs professzionális hardver kört*”.

2003. októberi számunkban *Peter Todd* részletesen is tárgyalta a *Hammerfall HDSP* kártyák használatához szükséges eszközöket.

A megjelenésre egy viszonylag egyszerű kártyát alkalmaztunk (lásd az internetes forrásokat). Hiába szerettük volna érdekes és újszerű megjelenítési képességekkel felruházni a gépet, az illesztőprogramok terén továbbra is problémákat tapasztaltunk (lásd a keretes részt).

Hőmérsékletkezelés

Hogyan tartatjuk megfelelő hőmérsékleten az eszközöket? A legfontosabb, hogy ne kezdjünk el játszani a különféle hardverkombinációkkal, amíg nem tudjuk, hogyan mérhetjük a módosítások által a rendszer hőmérsékletére gyakorolt hatást. Sőt, semmit ne változtassunk meg, amíg nem tudjuk, hogyan fogjuk mérni a változás hatását. A jó hír az, hogy a processzorok és az alaplapok gyártói

A teljesítménytesztek eredményei

dbench 100 szimulált ügyféllel:

```
%dbench 100
Throughput 1234.57 MB/sec (NB=1543.21 MB/sec
=12345.7 MBit/sec)
```

Bonnie++ 1.03 – egy pontosabb, a merevlemezek sebességét mérő program:

- Soros kivétel, karakterenként: 58577 Kb/s, 98 % CPU
- Soros kivétel, blokkonként: 281032 Kb/s, 50 % CPU
- Soros kivétel, újírás: 52603 Kb/s, 18 % CPU
- Soros bevétel, karakterenként: 34717 Kb/s, 58 % CPU
- Soros bevétel, blokkonként: 90097 Kb/s, 11 % CPU
- Véletlenszerű léptetés: 257,5/másodperc
- Soros létrehozás: 5924 fájl/másodperc
- Véletlenszerű létrehozás: 6056 fájl/másodperc

Postmark teljesítményteszt

A *Postmark* egy leterhelt levelezőszolgáltató működését szimulálja. 20000 alapfájllal és 100000 tranzakcióval az alábbi eredményeket kaptuk.

Idő:

- 46 másodperc (összesen), ebből 40 másodperc tranzakciókezelés (2500 db/másodperc).

Fájlok:

- 70128 létrehozás (1524 db/másodperc); Létrehozás önmagában: 20000 fájl (5000 db/másodperc);

Tranzakciókkal vegyesen:

- 50128 fájl (1253 db/másodperc)
- 49656 olvasás (1241 db/másodperc)
- 50199 hozzáfűzés (1254 db/másodperc)
- 70128 törlés (1524 db/másodperc)

Törlés önmagában:

- 20256 fájl (10128 db/másodperc);

Tranzakciókkal vegyesen:

- 49872 fájl (1246 db/másodperc)

Adatok:

- 303,46 MB beolvasva (6,6 MB/s)
- 436,18 MB kiírva (9,48 MB/s)

A rendszermag lefordítása:

- 50 másodperc

a legfontosabb alkatrészeket hőmérsékletérzékelőkkel is el látják, ezek jeleit egy alapvető eszközzel, az *lm_sensors* segítségével tudjuk figyelni. A meghajtók hőmérsékletét nem mértük, hiszen ezek külön házba kerültek, de a *smartmontools* (lásd a forrásokat) segítségével ezt is könnyen megtehetjük. Rendeltünk néhány tételt a kiváló vízűtőkezeseteket kínáló *Zalmantól*. A leglátványosabb darab a *Reserator 1*, mely egy fél méter magas, 2,5 liter vizet tartalmazó, kombinált víztartály és hőleadó. A *Reserator* mellett a processzorokhoz is rendeltünk egy-egy vízblokkot, illetve a szükséges csöveket is beszereztük.

A hőmérsékleti becslések azt mutatták, hogy nem lesz szükségünk processzoroként egy-egy *Reseratorra*, vagyis két-két processzorhoz egy-egy *Reserator*ot illesztettünk, illetve a tápegység is kapott egyet. A *Reserator* egy 5 wattos szivattyúval rendelkezik, ami sajnos nem zajtalan, ezért a *Reseratorokat* át kellett alakítanunk hőáramlásos működésűre. Alapkiépítésben a *Reserator* bemenete és kimenete közel vannak egymáshoz, ezért mindegyik *Reseratorba* beszereltünk egy a forró vizes bemenetűl a tetejéig futó csövet. Működött. A processzorhőmérséklet körülbelül 50° C-ig kúszott fel, és a processzorokat és a *Reseratorokat* összekötő csövek eléggé felmelegedtek ahhoz, hogy beinduljon a hőáramlás. Normál használat mellett a hőmérséklet 47-48° C körül maradt, teljes terhelés mellett pedig nem ment 50° C fölé.

A tápegység hűtése már keményebb feladat volt. A *Zalman* legerősebb ventilátor nélküli tápegysége 400 W-os, ám a négyutas alaplapnak ez kevés volt. Így a *PC Power and Cooling Turbo-Cool 510 ATX* egysége mellett döntöttünk.

Felmerült a saját tápegység építésének ötlete is, de ezt elvetettük, mert fontos, hogy az összetevők megfelelő sorrendben kapják meg a tápfeszültséget, és a *PC Power and Cooling* már megoldotta ezt a kérdést helyettünk. A hűtés gondja viszont megmaradt. Itt lépett be a képbe a lakatosmesterség. *Phil* elment a *Global Precision* nevű fémmegmunkáló műhelybe, ahonnan három munkafázist rendelünk meg. Először levágták a tápegység hűtőbordáinak eredeti lamelláit, ezzel sima, a vízhűtés blokkjainak felszerelésére alkalmas felületek jöttek létre. Ez után elkészítették magukat a vízhűtés blokkjait; itt kék színű, eloxált alumíniumot választottunk, ez ugyanis illeszkedett a *Zalman* alkatrészekhez. Végül két Y-csatlakozót is kértünk, ezek feladata a vízáram a két hűtőblokk közötti elosztása lett. A tápegységből eltávolítottuk a ventilátorvezérlést – többé úgysem volt rá szükségünk.

Ház

Egy ilyen gép befogadására, igényeinek kielégítésére kevés ház képes. Idén egyetlen választásunk volt: egyedi megoldást kidolgozni. Az ideai ház akrilablakokat kapott, melyeken keresztül látható a hűtőrendszer, valamint beépített támasztékokkal rendelkezik a *Reseratorok* és az *RME Multiface* számára.

Összefoglalás

Mindezek után, ha nehéz is elfogadni, de a való életben a legtöbb számítógépben sem 52 csatornás hangrendszerre, sem *Fibre Channel* alapú adattároló alrendszerre nincs szükség. Azonban a szokatlan összeállítások azok, amelyek képesek segíteni az igazán kreatív tevékenységeket, és örömminkre szolgál, hogy a Linux semmiben nem lesz gátunkra, segítségével bármibe belefoghatunk.

Aki azzal indul, hogy mi a lehetséges, majd kiveszi a számára szükségtelen elemeket, az bízhat benne, a gépe meg fog felelni az igényeinek. Reméljük, hogy olvasóink bármilyen számítógép építése mellett döntenek is, sikerült néhány ötletet meríteniük az ideai év linuxos erőművéből.

A jövő hardvere és a múltba révedő ügyvédek

Ez mindig bejön. Azok az új termékek, amiket ki szeretnénk próbálni az év linuxos erőművében, mindig pontosan akkor jelennek meg, amikor már a munka közepén járunk. A hőmérsékleteszték elvégzéséhez már túlságosan későn jelentek meg az *AMD* kétmagos *Opteron* processzorai, amelyekkel a meglévő, négy foglalatot tartalmazó alaplapunkat felhasználva is építhetünk nyolcutas rendszert – elég egy *BIOS*-frissítés végrehajtásunk. Ma még tízezer dollár (körülbelül kétfélmillió forint) négy ilyen processzor, ám várakozásaink szerint az árak hamarosan csökkenni fognak. Figyelemmel kísérjük a *LinuxBIOS* tervezet előrehaladását is, és a jövő évre egy általa támogatott alaplapot szeretnénk majd beszerezni. Tudjuk, a türelem fontos erény, ám a néhány másodperces rendszerindítás lehetősége önmagában is vonzó. Az ideai gép hangja annyira megtetszett, hogy jövőre is csendes gépet fogunk építeni. Az adattárolás terén jövőre az *Ed Cashin* által a 2005 júniusi számban tárgyalt *ATA over Ethernet* alkalmazását is számításba fogjuk venni. A megjelenítés továbbra is gyenge pont, de nem a hardver, hanem a gyártók jogászai miatt. Aki 3D-s megjelenítéssel foglalkozik, szinte szükségszerűen megsérti mások szabadalmait, az illesztőprogramok kódját pedig szigorú végfelhasználói szerződések védik, megtiltva a visszafejtést, és lelassítva az egész iparág fejlődését. Ha a rendszermag fejlesztése során egy széles körben használt hardverelem illesztőprogramja eltűnik, akkor vele fog menni a hardverelem is. Grafikus kártyák gyártói, fogjatok össze, kössétek meg a hardverekre vonatkozó keresztlicenclési megállapodásokat, majd készítsetek olyan használati szerződéseket a szoftverekre és a leírásokra, amelyek alapján a fejlesztők el tudják készíteni a grafika iránt érdeklődő felhasználók által igényelt kódokat! Hosszú távon mindenki jól fog járni – például az *NVIDIA* kizárólag egy licenclési döntés miatt párhuzamos szoftver-

terjesztői rendszert tart fenn. Vajon nem lenne jót a költségvetésnek, ha ez megszűnne?

A borulátók azt mondják, ők a realisták, és hajlamosak elfogadni a zárt illesztőprogramokat. A valóság azonban az, hogy az 1990-es évek *UNIX*-gyártói közül egy sem támogatta a *Linuxot*. Ma minden olyan *UNIX*-gyártó, amely egyáltalán létezik még, a *Linux* mögé állt. Aki realistának hiszi magát, gondolkozzon el egy kicsit ezen.

Linux Journal 2005. 136. szám

A cikkhez tartozó források elérhetősége:
 ➔ www.linuxjournal.com/article/8330

Justin Thiessen Linux mérnök a Penguin Computingnél. Az év linuxos erőműve tervezet vezetőjeként ő volt felelős a rendszer tervezéséért, az összeépítésért és a tesztelésért, illetve az alkatrészek kiválasztásába is bekapcsolódott. Ha nem a linuxos erőművel foglalkozik, akkor termékfejlesztési feladatokat végez, az *lm_sensors* tervezet adott hozzájárulásaival pedig a Penguin hardvereinek linuxos támogatását igyekszik javítani.

Matt Fulvio szabadúszó ipari tervező és építészmérnök. A San Franciscoi Építészeti Intézetben matematikát tanít, weboldala a www.mattfulvio.com címen érhető el.

Phillip Pokorny a Penguin Computing mérnökgazdátja. Az ő feladata volt a megfelelő tápegység felkutatása, valamint a vízhűtés beépítéséhez szükséges módosítások egyeztetése. Amikor éppen nem ezzel foglalkozott, akkor csak ácsorgott, és a tüsihajú főnökökre jellemző ostoba kérdésekkel bombázta a munkatársait.

Trevor Sherard a linuxos erőmű házában elkészítésében közreműködő kézműves volt, a san franciscoi öböl környékén szabadúszó szobrászként és asztalosként dolgozik. A www.woodentemple.com címen érhető el.

Don Marti a *Linux Journal* főszerkesztője, a cikk szövegének szerzője.