Linux új méretarányban: az SGI Altix 3000 rendszer

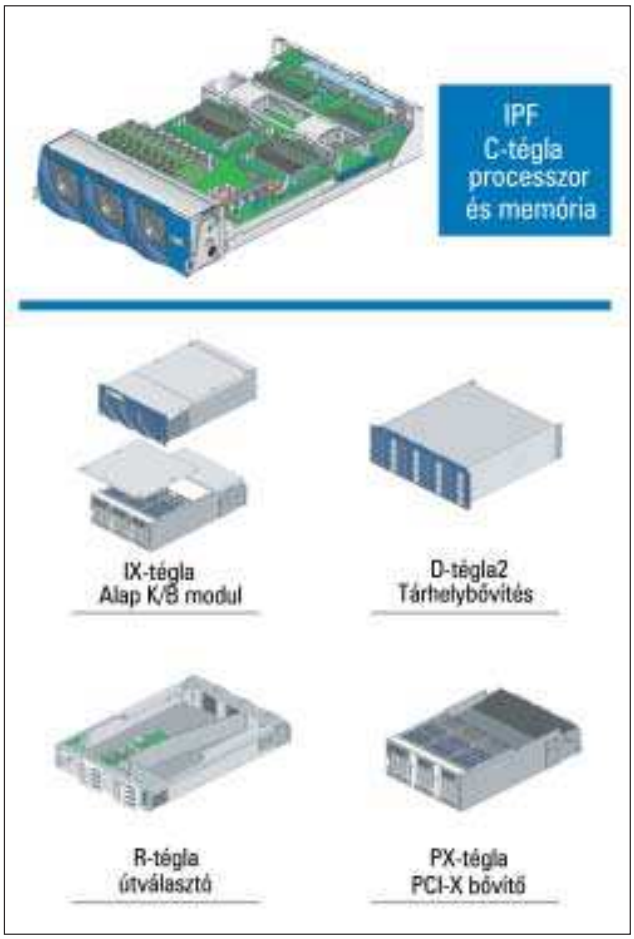
64 processzorával és az 512 GB memóriájával az SGI igényt tart a világ legerősebb Linux-rendszerének a címére.

Az SGI nemrégiben lépett színre új 64-bites, 64 processzoros, Intel Itanium 2 lapkákon alapuló Linux rendszerével – ami jelentős lépés mind a cég, mind a Linux számára. Ez a rendszer új távlatokat nyit, hiszen az összetett és igényes magas teljesítményigényű számításokon (High-Performance Computing, azaz HPC) dolgozó tudósok olyan körülmények között használhatnak és telepíthetnek Linuxot, amelyre ez idáig még nem volt példa. A HPC környezetek mindig az operációs rendszerek végső határait feszegetik, folyton több központi egységet, magasabb K/B sávszélességet és gyorsabb, hatékonyabb párhuzamos programozási támogatást követelve. A rendszer fejlesztésének korai szakaszában az SGI úgy döntött, hogy a Linuxot választja új felületének kizárólagos operációs rendszeréül, mivel bizonyítottan erős és megfelelő operációs rendszer az SGI által megcélzott számítási környezetekhez. A SGI NUMAflex globálisosztottmemória-rendszerével, az Intel Itanium 2 központi egységekkel és a Linux használatával már jóval a rendszer tényleges bemutatása előtt megdöntött minden rekordot.

Az új rendszer, amely az SGI Altix 3000 nevet kapta, legfeljebb 64 processzorral és 512 GB memóriával rendelkezik. A következő változatok azonban már 512 processzort és 4 TB-ot kínálnak majd. Ebben a cikkben az új SGI-rendszer mögött rejtőző alkatrésztervezést fedezzük fel, leírjuk, hogy milyen programfejlesztéseket kellett végrehajtani, hogy az új rendszert kivihessük a piacra, illetve megmutatjuk, milyen készségesen méretezhető és alkalmazható a Linux a legigényesebb HPC-környezetekben is.

Alkatrészháttér és rendszerfelépítés

Az SGI Altix 3000 rendszer Intel Itanium 2 processzorokat használ, és az SGI NUMAflex memóriakezelési szerkezetén alapul, ami a nem egységes memóriaelérés- (Non-Uniform Memory Access – NUMA) szerkezet SGI-féle megvalósítása. A NUMAflex 1996-ban mutatkozott be, és azóta használják a cég megújított SGI Origin kiszolgálócsaládjában, illetve a MIPS központi egységen alapuló szuperszámítógépeiben, valamint az Irix 64-bites operációs rendszerben. A NUMAflex-tervezés lehetővé teszi, hogy a processzort, memóriát, K/B rendszert, a kapcsolatokat, a grafikát és a tárakat moduláris alkotórészekbe, úgynevezett téglákba csomagoljuk. Ezek a téglák hihetetlen rugalmassággal kombinálhatók és állíthatók be, hogy az eredmény a vásárló erőforrás- és munkaterhelési igényeinek mind jobban megfelelhessen. Ezt a harmadik nemzedékbeli tervezést módosítva az SGI ilyen téglák haszlatával képes volt felépíteni az SGI Altix 3000 rendszert a Ki/Bemenet (IX- és PX-téglák), a tárhely (D-téglák) és a kapcsolatok (útválasztó téglák/R-téglák) részekhez. Az új rendszer fő eltérése a processzortégla (C-tégla), amely az Itanium 2 processzorokat tartalmazza. Az SGI Altrix 3000 rendszerében alkalmazott téglatípusokat az *1. ábra* mutatja be. A 2. *ábra* azt írja le, miként lehet ezekből a téglákból két keretet összeállítani, létrehozva egy egységes rendszerlenyomatú 64 processzoros rendszert (single-system-image 64-processor system).

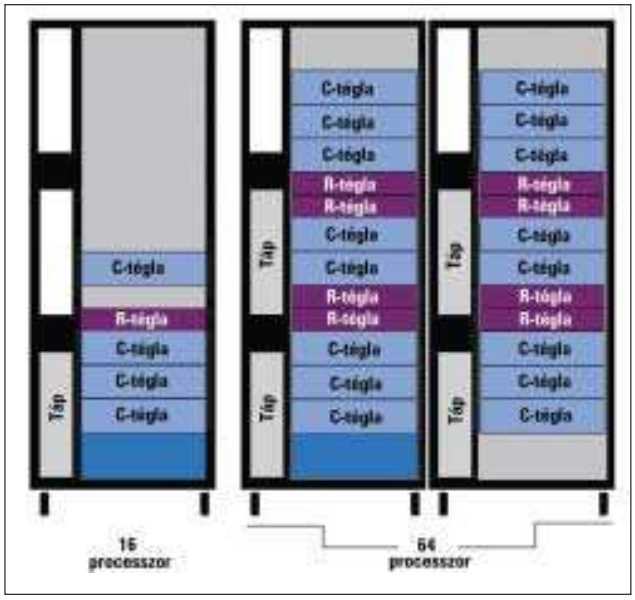


1. ábra NUMAflex-téglatípusok

A Linux felkészítése az új alkatrészkészletre

Egy olyan jól megtervezett és kiegyensúlyozott alkatrészrendszeren, mint a NUMAflex, az operációs rendszernek kell gondoskodnia arról, hogy a felhasználók és az alkalmazások az alkatrészeket teljes mértékben kihasználhassák anélkül, hogy közben a pocsékoló erőforrás-kezelés vagy a valahol egy szűk keresztmetszet hátráltatná őket. Hogy a nagy NUMA-rendszeren kiegyensúlyozott alkatrészerőforrás-kezelő rendszert tudjunk létrehozni, a rendszermagfejlesztést jóval azelőtt meg kellett kezdenünk, hogy az első Itanium 2 lapkák és alkatrészprototípus-rendszerek megérkeztek volna. Jelen esetben felhasználtuk az Itanium lapkák első nemzedékét, hogy a keresett HPC-környezethez szükséges processzorméretezést, a K/B teljesítménynövelést és az egyéb változtatásokat a Linux rendszeren elvégezhessük.

A program előkészítésének első lépése, még mielőtt az alkatrész-prototípusok megérkeztek volna, annak a lehető legpon-



2. ábra Két lehetséges NUMAflex-összeállítás

tosabb megállapítása volt, hogy milyen alacsony változásokat kell a rendszermag alacsony szintű (regiszterek és alkatrészek szintjén) kódjában eszközölni, hogy elinduljon és megbízhatóan fusson. és futtatásához. Azok a rendszerkészítők, akik saját, különösen fejlett rendszerekhez szánt ASIC tervezésébe fognak, általában szimulációs programokat és eszközöket használnak alkatrészterveik kipróbálásához. Mielőtt még a vasat megkaptuk volna, kifejlesztettünk és széles körben használtunk szimulátorokat a rendszerprogramokhoz (firmware) és a rendszermag fejlesztéséhez egyaránt, hogy segítségükkel elkészíthessük a rendszerszintű programokat.

Amikor az első nemzedékbeli Itanium processzorokkal ellátott eredeti alkatrész-prototípus megérkezett, eljött az üzembe helyezés ideje. Az egyik legfontosabb mérföldkő a rendszer első bekapcsolása, a processzor újraindítása (reset), majd az első utasítások PROM-ból történő kiemelése és végrehajtása volt. Az indítás után az alkatrészfejlesztési laborban hosszú órákon és hétvégeken keresztül tartott az igazi móka. Ez volt az a labor, ahol az alkatrészeket, a kipróbálást végző és a felületet tervező mérnökök szorosan együttműködtek egymással a rendszer hibaellenőrzésében, keresztülsegítve a processzort számos lényeges állomáson: eljutottak az PROM-tól az indítási promptig, a Linux-rendszermag futtatásától a felállásig, továbbá túljutottak a gyökérfájlrendszer olvasásán és befűzésén, az egyfelhasználós mód elérésén, majd a többfelhasználós módba lépésen, végül a hálózati csatlakozáson. Ezt követően ugyanezt több processzorral és több csomóponttal – többnyire párhuzamosan haladva – néhány, más állomásokon dolgozó felhozó csapattal is végigcsináltuk, akik szorosan követték a vezető-csapat fejlődését.

Miután az első nemzedékbeli Itanium processzoros prototípus-rendszereken sikerült a Linuxot futásra bírni, a programmérnökök nekiláthattak a munkának, immár tudva, hogy a Linux fut, és ami talán még fontosabb: jól méretezhető a NUMA-rendszereken. Számos házon belüli, első nemzedékbeli Itanium alapú rendszert építettünk fel és használtunk, így meggyőződhettünk róla, hogy a Linux a nagyrendszereken tényleg jól teljesít. 2001 elején kategóriájában elsőként sikeresen futtattunk egy 32-processzoros Itanium alapú rendszert.

1. kép Az alkatrészeket tervező mérnök, a PROM-ot megalkotó mérnök és a rendszermérnök megvitatnak egy hibát

2. kép A szerző fia egy korai 32 processzoros Itanium alapú rendszer előtt 2001 nyarán

3. kép Az Itanium 2 alapú C-tégla első indulása

Ezek az első nemzedékbeli Itanium alapú rendszerek kulcsfontosságúak voltak, mert a segítségükkel tudtuk a Linuxot az igényes HPC-követelményeknek megfelelőre formálni. Így már jóval azelőtt, hogy az Intelnél az első Itanium 2 processzorok

Feladatmegoldások valós környezetben

Következő példáinkban három tudományos HPC-alkalmazás teljesítményét mutatjuk be Linuxot futtató SGI Altix 3000 rendszeren. Három példarendszerünk a bioinformatikához szánt FASTA, a számítógépes kémiai feladatokra tervezett Gaussian és a folyadékdinamikai modellezésre használt STAR-CD lesz. Az összes próbát az SGI vezette.

FASTA bioinformatikai példa

Bár a biokémia és a számítógépes biológia már a 1980-as évek közepétől kezdve létezik, a bioinformatika viszonylag fiatal tudományága, amelyet az adatigényes élettudományok és a laboratóriumautomatizálási technológiák közeledése, illetve a hatalmas adatmennyiséget gyorsan szervező, feldolgozó és szétosztó számítógépes adatbázisok és algoritmusok megjelenése tett megvalósíthatóvá. A bioinformatika segítségével az új gyógyszerek hamarabb kerülhetnek piacra, megakadályozhatjuk a genetikai fertőzéseket, fertőzés- vagy szárazságtűrő kukoricát hozhatunk létre, meghosszabbíthatjuk az ételek szavatosságát, választási lehetőségünk lesz az olajkérdésre vonatkozóan, és létrehozhatjuk a jövő ételeit, amelyek segítenek majd a koleszterinszint szabályozásában és megelőzik a rákot. A gyorsan növekvő nyilvánosan elérhető biológiai adatok mennyisége miatt a szekvencia-adatbáziskeresés a bioinformatika egyik legkényesebb területe. A Smith–Waterman-féle (*T. F. Smith* és *M. S. Waterman*, 1981, Journal of Molecular Biology 147: 195–197) klasszikus keresési módszerek nyújtják a biológiai adatbázisok leghatékonyabb módszerét a szekvenciahasonlóságok kereséséhez. Csakhogy a Smith–Waterman–módszer elég számításigényes, így az eljárás hatékony felhasználásához különösen fontos a párhuzamosítás. Az egyik ilyen párhuzamosított Smith–Waterman-megoldás a FASTA bioinformatikai csomag (W. R. Pearson, 1991, Genomics 11: 635–650). A Smith–Waterman-algoritmus FASTA 3.4 változatát (ssearch34_t) – ami egyébként a Virginiai Egyetem lapján (alpha10.bioch.virginia.edu/fasta) fellelhető – használtuk 64 processzoros SGI Altix 3000 rendszerünk párhuzamos teljesítményének a mérésére.
A Smith–Watermant P-szálakkal (Pthread) párhuzamosítottuk, az SGI ChemBio Applications csapat pedig a magalgoritmust tovább finomhangolta, hogy az még jobban kihasználhassa az SGI Altix 3000 rendszer képességeit. Végeredményül a Smith–Waterman-algoritmus közel eszményi méretezhetőséget mutatott (a tökéletes méretezhetőséget a pontozott vonal jelzi), 64 processzoron futtatva közelítőleg 59×-es sebességnövekedést értünk el.

I. ábra
A FASTA teljesítménye csaknem lineáris

Gaussian számítógépes kémiai példa

Nemzeti kutatólaboratóriumok, egyetemek, gyógyszeripari és biotechnológiai cégek és vegyészeti vállalatok is használnak a Gaussian Inc. (➲ http://www.gaussian.com) Gaussian 98 rendszeréhez hasonló számítógépes kémiai alkalmazásokat a molekuláris energiák, tulajdonságok és reakciók kutatásában, elektronikus szerkezetekkel igen nagy molekularendszereket modellezve.

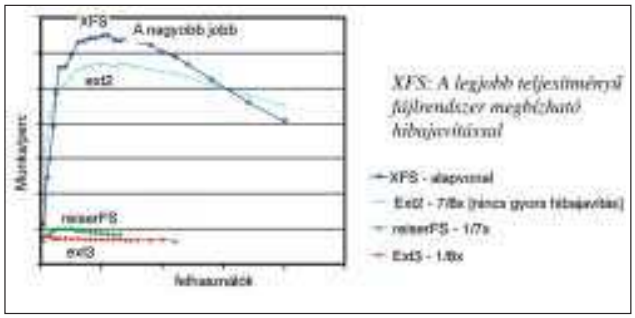
II. ábra
Gaussian 98 eredmények

A *II. ábrán* a Gaussian 98 méretezhetőségére vonatkozó eredményeket figyelhetjük meg, amelyeket a Gaussian QA-suite nevű, széles körben alkalmazott teszttel készítettünk, egy korai SGI Altix 3000 prototípusrendszeren az inkább Itanium, semmint Itanium 2 processzorokhoz szánt Intel-fordító korai változatát használva. Az itt látható eset a Valynomycin molekulának (C54H90N6O18) a sűrűségalapú elmélet (Density Functional Theory) szerinti erőszámításait mutatja be. A grafikonon megfigyelhetjük a tesztben kialakult párhuzamos sebességnövekedést. Az eltelt időt másodpercben adtuk meg. Az eszközök és a rendszer korai változatai ellenére a számításhoz szükséges idő az SGI Altix 3000 rendszer húsz processzorának felhasználásával körülbelül 4,5 óráról mindössze 25 percre csökkent.

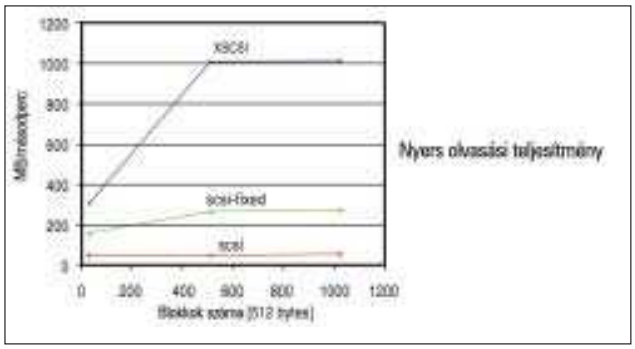
STAR-CD számítógépes folyadékdinamikai példa

A számítógépes folyadékdinamikát (Computational Fluid Dynamics, azaz CFD) számos hagyományos ipari ágazat is használja, többek közt az autóipar, az űrkutatás és az energiatermelési ágazat. A Computational Dynamics Limited (➲ http://www.cd-adapco.com) STAR-CD programja a folyadékáramlás témakörében a CFD-módszer egyik vezéralakja. A CFD-felhasználót végigsegíti a kezdeti alapvető tervezéstől kezdve a valós értékekkel bíró tanulmányokon át a hatékonnyá tételig, felajánlva fejlett fizikai modelljét, illetve azt a képességét, hogy strukturálatlan hálókkal összetett geometriai formákat is kezelni képes. A STAR-CD minden vezető Unix- és NT-felületen üzemel. Párhuzamosított változata, a STAR-HPC osztott memóriájú kiszolgálókon, erősen párhuzamosított rendszereken és munkaállomások telepein fut. A STAR-HPC az MPI könyvtárat használja a magas szintű méretezhetőség eléréséhez. SGI Altix 3000 rendszeren a STAR-CD előzetes kiadását és az autó áramlástani modelljéhez szánt „Model A Class Dataset" felhasználását futtatva az előzetes teljesítménypróbák alapján a Linux ismét bizonyította kitűnő processzorméretezhetőségét – egészen 64 processzorig. A 2002 novemberében a ➲ http://www.cd-adapco.com/support/bench/aclass.htm lapon közölt adatok szerint a Linux jobban méretezhető, mint két másik kereskedelmi rendszer: a Hewlett Packard HP-UX és az IBM AIX rendszere.
A *III. ábrán* a STAR-CD összehasonlító eredményeit láthatjuk a Linux, a HP-UX és az AIX rendszerek között.

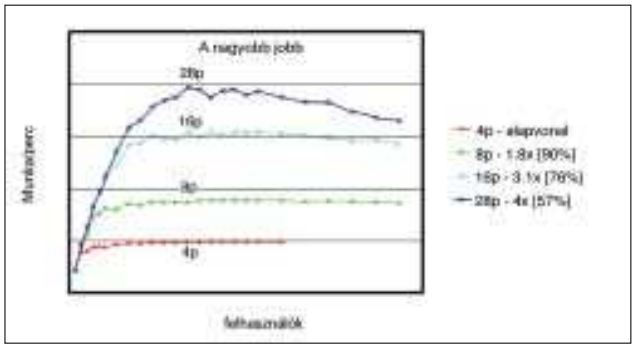
III. ábra
STAR-CD eredmények Linux,
HP-UX és AIX összehasonlításban



3. ábra Fájlrendszerteljesítmény-összehasonlítás: AIM7 többfelhasználós rendszermag terhelése, 2.4.18-as rendszermag 28 P Itanium mintapéldánnyal, 14GB, 120 lemez; különböző fájlrendszerek, az SGI álltal kiegészített és finomhangolt rendszermaggal



4. ábra Linux XSCSI teljesítménypélda: a 2.4.16-os rendszermaggal; 120 folyamat 120 merevlemezről olvas



5. ábra Processzorskálázási példa az AIM7-el: AIM7 többfelhasználós rendszermag terhelése, 2.4.16-os rendszermag; SGI-kiegészítések és finomhangolás a rendszermagban

elérhetővé váltak volna, már fejleszteni lehetett és ki lehetett próbálni a méretezést, a K/B teljesítményt és az egyéb változtatásokat.
Miközben az SGI-programmérnökök a teljesítményen, a méretezésen és más feladatokon dolgoztak az első nemzedékbeli Itanium processzorokkal felszerelt prototípusokon, az alkatrészeket tervező mérnökökből és felületet késztő mérnökökből álló másik csapat felkészítette az indításra a következő nemzedékbeli Itanium 2 processzoros SGI C-téglát, elölről megismételve a teljes fejlesztési folyamatot.
2002 közepére a fejlesztő csapat nagyszerű fejlődést mutatott: egyetlen processzor indításától eljutottak a 64 processzoros rendszer működéséig. Az Itanium 2 lapkákkal felszerelt 64 processzoros rendszer ismét úttörőnek számított a maga nemében. Mindezek természetesen egyetlen rendszerlenyomaton lévő (single system image) Linuxon futottak.
Az egész folyamat alatt minden változtatást és hibát, amit csak a Linuxban találtunk, visszaküldtünk a rendszermagfejlesztőknek, hogy a ja vításokat későbbi Linux-terjesztésekbe már beletehessék.

Közelkép a Nagy Vasról

Más Linux-fejlesztők gyakran kérdezik: „Miféle változtatásokat kellett alkalmaznotok a Linuxon, hogy egy ilyen méretű rendszeren fusson?” vagy „A Linux-processzor méretezhetősége nincs nyolc vagy hasonló számú processzorra korlátozva?”. Ahhoz, hogy válaszolhassunk ezekre a kérdésekre, előbb meg kell vizsgálnunk, mit használ az SGI programalapnak, milyen kitűnő változtatásokat végzett el a közösség, és hogy milyen más HPC-vel kapcsolatos fejlesztéseket és eszközöket adott az SGI, hogy a Linux messze túlszárnyalja az ismert nyolcprocesszoros határt.
Az SGI Altix 3000 rendszerek rendszerprogramja az Itanium processzorokhoz szánt szabvány Linux-terjesztést és a Linuxot további képességekkel felruházó SGI ProPack bővítményt tartalmazza. Az SGI ProPack termék egy újabb 2.4 alapú Linux-rendszermagot, HPC könyvtárakat, amelyek az SGI alkatrészeinek legjobb kihasználására vannak kiélezve, valamint NUMA-eszközöket és meghajtókat tartalmaz.
Az SGI Altix 3000 rendszeren használt 2.4 alapú Linux-rendszermag az Itanium processzorokhoz szánt szabványos 2.4.19-es rendszermagot (kernel.org), illetve néhány továbbfejlesztést tartalmaz. Ezek a továbbfejlesztések három osztályba sorolhatók: általános hibajavítások és felülettámogatás, fejlesztések a Linux-közösség más munkáiból, végül SGI-változtatások.
A rendszermag-változtatások első csoportjába a kipróbálás alatt talált hibák javításai tartoznak, illetve az alapot képező felület továbbfejlesztései, valamint a NUMA-támogatás. Ezeket a változtatásokat az SGI a rendszermagot fejlesztő csapat megfelelő karbantartójával együttműködve végezte, hogy e módosítások visszakerülhessenek a rendszermag főáramába.
A rendszermagfejlesztések második csoportjába azok a kitűnő munkák és teljesítményfoltok kerültek, amelyeket a közösségből mások fejlesztettek ki, de hivatalosan még nem fogadtak el, vagy átütemeztek a 2.5 fejlesztői vonalra. Ezek a fejlesztések a következő VA Software SourceForge lapokon találhatók meg: „Linux on Large Systems Foundry” (large.foundries.sourceforge.net) és a „Linux Scalability Effort Project” (sourceforge.net/projects/lse). Mi a projektekből a következő foltokat használtuk: a processzorütemezőt, a nagy rendszermagzár-felhasználást (Big Kernel Lock) csökkentő javítást, a Read–Copy–Update (olvass-másolj-frissíts) spinlock elven alapuló `dcache_lock-usage` csökkentésjavítást, illetve az FRlock zárolási elven működő `xtime_lock` (gettimeofday) felhasználáscsökkentő fejlesztést.
A Linux eszközkezelő fájlrendszerét (`devfs` ➲ http://www.atnf.csiro.au/people/rgooch/linux/docs/devfs.html) is beállítottuk és használjuk, hogy a rendszerünkön elérhető rengeteg lemezt és K/B sínt kezelhessünk. A `devfs` biztosítja számunkra, hogy az eszközelérési utak újraindítás után is megmaradjanak, mégha időközben lemezeket vagy vezérlőket helyeztünk be vagy távolítottunk is el. Az egyik legkellemetlenebb dolog, amivel a nagy rendszerek gazdái találkozhatnak, ha az egyik rendszer tönkremegy, és hirtelen ötven, esetleg még több lemez átszámozódik és átneveződik. A `devfs` megbízhatónak bizonyult olyan különlegesen igénybe vett rend-

szerkörnyezetekben is, amelyekben akár 64 processzorral és rostcsatornák (Fibre Channel) tucatjaival rendelkező összeállítások működnek együtt százszámra befűzött lemezekkel.
A `devfs` kiegészítő része a 2.4-es Linux-rendszermagnak, így további foltra nem volt szükség.
A rendszermag-változtatások harmadik csoportjába kerültek az SGI által végzett módosítások, amelyeknek a Linux fővonalába történő illesztése jelenleg is folyik, s a 2.4-es változatkövetőkben kapnak helyet, vagy a folt különleges felhasználási területe vagy természete miatt elkülönítve maradnak. Ezeket a nyílt forrású fejlesztéseket az „Open Source at SGI" honlapon találjuk (➲ http://oss.sgi.com). Az általunk elvégzett módosítások a következők voltak: az XFS fájlrendszer-program, folyamat-aggregátumok (Process AGGregates, PAGG), CpuMemSets (CMS), rendszermag-nyomkövető (kdb) és a Linux rendszermag összeomláslistázó (Linux kernel crash dump, azaz lkcd). Ezen felül az SGI beépítette az Irix alól áthozott SCSI alrendszerét és vezérlőit. A Linux 2.4 SCSI K/B alrendszerrel végzett első kipróbálások megmutatták, hogy a terület komolyabb fejlesztése nélkül nem tudjuk vásárlóink jelentős tárigényeit kielégíteni. Bár a fővonalbeli rendszermagfejlesztők a későbbi kiadásokhoz már dolgoznak ezen a feladaton, az SGI-nak azonnali megoldásra volt szüksége a 2.4 alapú rendszermagokhoz, így az Irixről áthozott SGI XSCSI hátteret és meghajtókat használtuk ideiglenes megoldásként.
A *3–5. ábra* azt mutatja be, milyen kezdeti javulást értünk el az SGI Altix 3000 rendszeren futó Linux alatt a fent említett változtatásokat követően. A *3. ábra* az XFS fájlrendszert hasonlítja a többi Linux-fájlrendszerhez. (Megjegyzés: ha a Linux-fájlrendszerek teljesítményét összehasonlító részletesebb írást keresünk, nézzük meg a 2002-es Usenix Annual Technical Conference „Filesystem Performance and Scalability in Linux 2.4.17" című cikket, amely az ➲ http://oss.sgi.com lapról szintén elérhető). A *4. ábra* az XSCSI-t hasonlítja 2.4-es Linux SCSI rendszeréhez, végül a *5. ábra* az AIM7-el elérhető processzor méretezhetőségét szemlélteti.
Bár az SGI inkább a nagyteljesítményű, illetve műszaki számítási környezetekre összpontosít – ahol a processzorciklusok többsége általában a felhasználói szintű kódra és alkalmazásokra fordítódik, nem a rendszermagra – az AIM7 teljesítménypróba megmutatta, hogy a Linux a főként az üzleti környezetekre jellemző, más típusú terhelés alatt is jól méretezhető. A HPC-alkalmazások linuxos teljesítmény- és méretezhetőségi példáit a széljegyzetben „Feladatmeghatározások valós környezetben" címmel olvashatjuk.
A Stream Triad teljesítménypróba segítségével az SGI megmutatta, hogy kettőtől 64 processzorig csaknem lineáris méretezhetőség érhető el, és eljutott a másodpercenkénti 120 GB-os sebességig. Ez az eredmény jelentős mérföldkőnek számít az iparban, hiszen új világrekordot állít fel a mikroprocesszor vezérelte rendszerek világában, és ezt az eredményt egy egyetlen rendszerlenyomatot futtató Linuxon értük el! Ez a lenyűgöző eredmény egyben azt is megmutatja, hogy a Linux a megszokott nyolcprocesszoros korlátozáson felül is ténylegesen jól használható. A Stream Triadról további tájékoztatást a ➲ http://www.cs.virginia.edu/stream oldalon olvashatunk.
Ha megnézzük az SGI ProPackben felsorolt rendszermagfrissítés-listáját, a felsorolást meglepően rövidnek találjuk majd, ami ékesen bizonyítja a Linux eredeti kitűnő tervezését. Ami talán még lenyűgözőbb, hogy e javítások nagy része már benne van a 2.5-ös fejlesztői rendszermagban. Azt mondhatjuk, hogy a Linux gyorsan HPC operációs rendszerré növi ki magát.

Egyéb változtatások a HPC Linuxon

Az SGI ProPack tartalmaz pár olyan eszközt és könyvtárat, amelyekkel a nagy NUMA-rendszerek teljesítményét fokozhatjuk, ha olyan összetett feladatokat akarunk megoldani, amelyek sok processzort és memóriát használnak, vagy amikor több alkalmazás fut egy időben ugyanezen a nagyrendszeren. Linux alatt az SGI a `cpuset` és a `dplace` parancsokat használja, amelyek jól becsülhető és fejlett CPU és memóriafelhasználás-vezérlést biztosítanak a HPC-alkalmazások felett. Ezek az eszközök segítenek nekünk kimetszeni a szükségtelen folyamatokat, segítenek úgy használni a minden feladathoz a szükséges erőforrásokat, hogy ne keresztezzék egymás útjait, illetve megakadályozzák, hogy a kisebb feladatok véletlenül nagyobb mennyiségű erőforrást pazaroljanak el, mint amennyit hatékonyan használni tudnak. Ilyen módon a rendszer erőforrásait hatékonyan használjuk fel, és az eredményeket rendszeres időközönként kapjuk meg – ami két jellemzően kényes pontja a HPC-környezeteknek.
Az SGI ProPackben található SGI Message Passing Toolkit (MPT) az SGI számítógépekre optimalizált, ipari szabványú üzenetküldő könyvárat teszi elérhetővé. Az MPT-ben megtaláljuk az MPI és a SHMEM API-kat, amelyek átlátszóan helyezik üzembe és használják az SGI-alkatrészek alacsony szintű képességeit, például a gyors memóriá belüli másolásokhoz használt blokkátviteli motort (BTE), és a memóriaeszközvezérlő-kiragadó művelet (fetchop) támogatását. A fetchop-támogatás közvetlen kapcsolattartást és összehangolást tesz lehetővé több MPI folyamat között, miközben semlegesíti az operációs rendszer rendszerhívásaival kapcsolatos többletmunkát.
Az SGI ProPack NUMA-eszközök, a HPC könyvtárak és a szabványos Linux-terjesztésre épített egyéb programtámogatás együtt hatékony HPC programozási környezetet jelent a nagy számítás- és adatigényű munkaterhelésekhez. Ahhoz hasonlóan, ahogy az egyedi ASIC „ragasztólogikaként" felhasználhatóvá teszi a processzorokat, a memóriát és a K/B részeket az alkatrészeken, az SGI ProPack program ahhoz biztosítja a „ragasztó logikát", hogy a Linux operációs rendszert a nagy HPC-környezetek általános építőkövévé tegye.

Összefoglalás

Senki sem hitte volna, hogy a Linux ilyen jól és ilyen hamar méretezhetővé válik. A Linux és az SGI NUMAflex rendszerkiépítés ötvözésével, valamint az Itanium 2 processzorokkal az SGI megépítette a világ legerősebb Linux-rendszerét. Az SGI Altix 3000 rendszer piacra viteléhez rengeteg erőfeszítés kellett, és úgy véljük, ez még csak a kezdet. Az SGI által folytatott kemény szabványalapú stratégia, amit az Itanium 2 alapú rendszereken futó Linuxoknál használt, feljebb helyezi a Linux képességeit jelző lécet, miközben vásárlóinak érdekes, megalkuvások nélküli választási lehetőséget kínál a HPC-kiszolgálók és szuperszámítógépek terén. Az SGI mérnökei – így tulajdonképpen a teljes cég – tökéletesen megbízik a Linux képességeiben és folytatni szeretné a megkezdett utat, még több érdekes áttörést hozva a Linux- és a HPC-közösségnek.

Linux Journal 2003. február, 106. szám

Steve Neuner
Az utóbbi 19 évben Unix-rendszermagfejlesztéseken dolgozik. Jelenleg az SGI-nál Linux-mérnök-igazgató, és négy éve, amióta csatlakozott az SGI-hoz, Linuxon és Itanium alapú rendszereken dolgozik.