



April 2005: de första magneterna i LHC installeras i tunneln. Foto: CERN

ATLAS Data Challenge 2 på Swegrid

På det europeiska partikelfysiklaboratoriet CERN i Genève pågår för närvarande bygget av en ny partikelaccelerator. I denna accelerator, kallad den stora hadronkollideraren (*Large Hadron Collider* – LHC), kommer två motriktade protonstrålar att färdas åt var sitt håll i en ring med en omkrets av 27 kilometer i en tunnel c:a 100 meter under markytan. På ett fåtal platser längs ringen kommer de två strålarna att kollidera med varandra. På dessa platser byggs stora detektorsystem som utformas för att samla in information om de partiklar som bildas ur den energi som frigörs när protonerna krockar med varandra.

En av de frågor som länge har intresserat forskare världen över är uppkomsten av partiklarnas massor. Enligt standardmodellen som beskriver elementarpartiklarna och deras växelverkan är massan hos en partikel ett resultat av att den växelverkar med ett kvantfält kallat Higgsfältet. Förutom att ge upphov till massa hos de partiklar som växelverkar med det kommer närvaron av Higgsfältet enligt standardmodellen också att ge sig till känna genom existensen av en Higgspartikel. Denna partikel har dock ännu inte detekterats experimentellt. En av uppgifterna för de detektorsystem som byggs vid kollisionspunkterna längs LHC är att leta efter denna Higgspartikel.

Ett av de detektorsystem som byggs vid LHC heter ATLAS. I detta experi-

ment arbetar forskare från universiteten i Stockholm, Uppsala och Lund samt från KTH tillsammans med kollegor från Europa och resten av världen. ATLAS-experimentet kommer när LHC startar 2007 att samla in data om de partiklar som bildas i kollisionerna som sker vid kollisionspunkten i detektorns mitt. Kollisionerna kommer att ske 40 miljoner gånger per sekund. Denna mängd reduceras successivt genom tre olika triggernivåer så att cirka 100 händelser per sekund sparas av datainsamlingssystemet. Mängden data per händelse är i medeltal omkring 1,5 MB.

Som en förberedelse för analysen av verkliga data från detektorn har ATLAS-kollaborationen utfört en serie test för att testa den mjukvara som skrivits för att analysera data. Dessa tester går under namnet *Data Challenges*. För varje omgång av dessa tester ökas komplexiteten och omfånget för att så småningom nå den storlek som förväntas behövas då datainsamlingen vid LHC startar. De Monte Carlo-data som produceras används även för förberedande studier av möjliga fysikanalyser och för design av triggeralgoritmer.

Hela den datorberäkningskraft som behövs för att analysera de data som kommer att samlas in från detektorn och för att producera och analysera de Monte Carlo-simulerade data nödvändiga för att

forts sid 3

Nya gridsystem

Dayhoff och Green, de båda kluster som NSC driver åt forskargrupper på IFM (institutionen för fysik och mätteknik), har nyligen blivit anslutna till Nordugrid. Auktoriserade användare kan alltså skicka jobb till systemen genom griden.

Det är framförallt de Dayhoff- och Green-användare som dessutom har SNAC-tilldelning på Swegrid som har nytta av grid-anslutningen. De har nu ett likformigt gränssnitt för att skicka in jobb, och får också automatisk lastbalansering – användarens jobb kommer att exekveras på hemmaklustret eller på något Swegrid-system beroende på var belastningen för ögonblicket är minst.

Grid-anslutningen ger också ökade möjligheter för grupperna att samarbeta och dela på resurserna.

Förutom Dayhoff och Green är även Monolith och Bluesmoke (genom Swegrid) anslutna till Nordugrid. Dessutom har den gamla trotjänaren Ingvar fått ett nytt liv som testsystem inom den produktionsgrid som nu byggs upp genom EU-projektet EGEE.

SNAC-tilldelningen klar!

Vårens ansökningsomgång för beräkningstid till *Swedish National Allocations Committee*, SNAC, behandlades vid kommitténs möte den 26 maj. Samtliga resurser vid NSC, PDC och HPC2N var översökta med en faktor mellan 2 och 5. Undantaget var Swegrid med något mindre sökt än tillgänglig tid.

Vid pressläggningen återstod bara de sista finjusteringarna. Den slutliga tilldelningen kommer att nå användarna ungefär samtidigt som detta nummer av Nytt från NSC.

NSC och omvärlden

Sven Stafström
föreståndare NSC
stafstrom@nsc.liu.se



En vår av intensivt klusterbyggande på NSC går nu mot sitt slut. SMHI:s nya resurs Blixt börjar leverera prognoser från den 1 juni. Ytterligare två kluster, Tornado och Dunder, som ni kan läsa om i detta nummer av Nytt från NSC, är så gott som klara att tas i bruk. Dessa nya HPD-resurser är tydliga tecken på att kluster är på fortsatt stark frammarsch. Kanske att tänka på för ett helt annat men även det aktuellt ämne, nämligen statlig inköpsamordning. Statskontorets ramavtal har absolut sitt berättigande, som helhet underlättar de våra inköp och gör dem i många fall också billigare. För enskilda mer specialiserade verksamheter som HPD blir det dock inte alltid till vår fördel. Med den utveckling som exemplifieras ovan så kan man önska sig att klusterresurser kommer att finnas med mer explicit inom något anbudsområde vid nästa ramavtalsupphandling.

Det är säkert välkänt för de flesta läsare av Nytt från NSC att Vetenskapsrådet har inrättat Kommittén för infrastruktur (KFI). Det är angeläget att lyfta fram infrastruktur på detta sett och det borde kunna leda till Sverige på ett tydligare sätt nu kan agera inom t.ex. EU, stora infrastrukturella satsningar är med nödvändighet gränsöverskridande. Jag är övertygad om att detta också kommer svensk HPD-verksamhet till godo. Den samverkan som kan komma inom beredningsgruppen "Infrastruktur för e-Science" kan också ges oss möjligheter att expandera och attrahera användare av våra resurser från områden som t.ex. medicin och samhällsvetenskap. Det känns som en spännande utmaning att hitta dessa nya grupper av HPD-användare och kunna erbjuda dem våra tjänster.

Förberedelserna för höstens workshop LCSC och SNIC interaction pågår för fullt.

Vi har redan sett ett mycket stort intresse för mötet och vi hoppas kunna samla stora delar av den svenska "HPD-eliten" med SNIC samt HPD-expertiser från SNIC-centra, HPD-användare från universitet, forskningsinstitut och företag, samt försäljare och representanter för tillverkare. Missa inte detta unika tillfälle till nationell samling inom HPD, vi ses i Linköping den 17–19 oktober. Men innan dess så kommer sommaren, vi hoppas att Blixt-prognoserna kommer att bli goda och att vi fullt ut kan njuta av en skön semester!

Höjd kvalitet på gridkörningar via EU-projekt

När du skickar ditt beräkningsjobb till en grid, specificerar du ett antal krav på de beräkningsservrar som skall köra jobbet, såsom att det finns vissa programvaror och ett visst antal lediga beräkningsservrar, och jobbet körs sedan vid en ospecificerad tidpunkt på några passande beräkningsservrar. Skulle någon av beräkningsservrarna sluta fungera under körningen, får du i efterhand besked om detta och kan köra om ditt jobb från början.

Projektet HPC4U försöker höja kvaliteten i denna hantering på olika sätt, enligt målbilden nedan.

När du skickar in jobbet, skall du kunna tillfoga en SLA (*Service Level Agreement*), som du kräver skall vara uppfylld för ditt jobb. Om ditt jobb accepteras av griden, så fungerar denna SLA som ett slags garanti för hur din körning kommer att köras.

Om du i jobbet SLA specificerar att jobbet måste vara klart vid en viss tidpunkt, så används flera mekanismer för att se till att det verkligen blir så. Ditt jobb kommer att köras på ett beräkningskluster, som klarar att checkpointa ditt jobb regelbundet, d.v.s. att spara undan all aktuell status för ditt jobb. Om någon av jobbet beräknings-servrar slutar att fungera, så tar i stället en annan beräkningsserver över i dess ställe, med hjälp av checkpointing-informationen.

Om klostret får så mycket problem att det inte finns någon extra beräkningsserver ledig, flyttas i stället hela ditt jobb över till ett annat beräkningskluster och körs färdigt där, för att du skall hinna få körningsresultatet före den garanterade sluttiden.

Det här sättet att köra jobb kan komma att kosta mycket, eftersom det måste finnas reservservrar som kan ta över när ordinarie beräkningsservrar slutar fungera, men för vissa kategorier av beräkningar kan det naturligtvis vara värt det.

En mindre krävande SLA skulle kunna vara att utan krav på en garanterad sluttid få checkpointing gjord med en viss regelbundenhet, så att jobbet startar om från sin senaste checkpoint om det blir något problem med beräkningsservrarna.

NSC deltar i projektet HPC4U, tillsammans med flera företag och organisationer i andra länder. I projektet företräder NSC slutanvändarna och har i projektet även byggt ett testverktyg för HPC4U.

Forskningsprojektet är finansierat av EU inom det sjätte ramprogrammet. Mera information hittar du på <http://www.hpc4u.org>.

LENNART KARLSSON

Ny programvara vid NSC

På Monolith har vi installerat DIRAC04, vilket är ett program för relativistiska molekylära elektronstrukturberäkningar på 4-komponentsnivå. Vidare har versionen av Dalton uppdaterats till 2.0 på Monolith och SweGrid – den enskilt största förändringen i Dalton gentemot tidigare versioner är inkluderandet av DFT.

Mer detaljerad information finns att läsa under NSC:s kemisida:
<http://www.nsc.liu.se/software/chemistry/>

NSC inbjuder till

LCSC

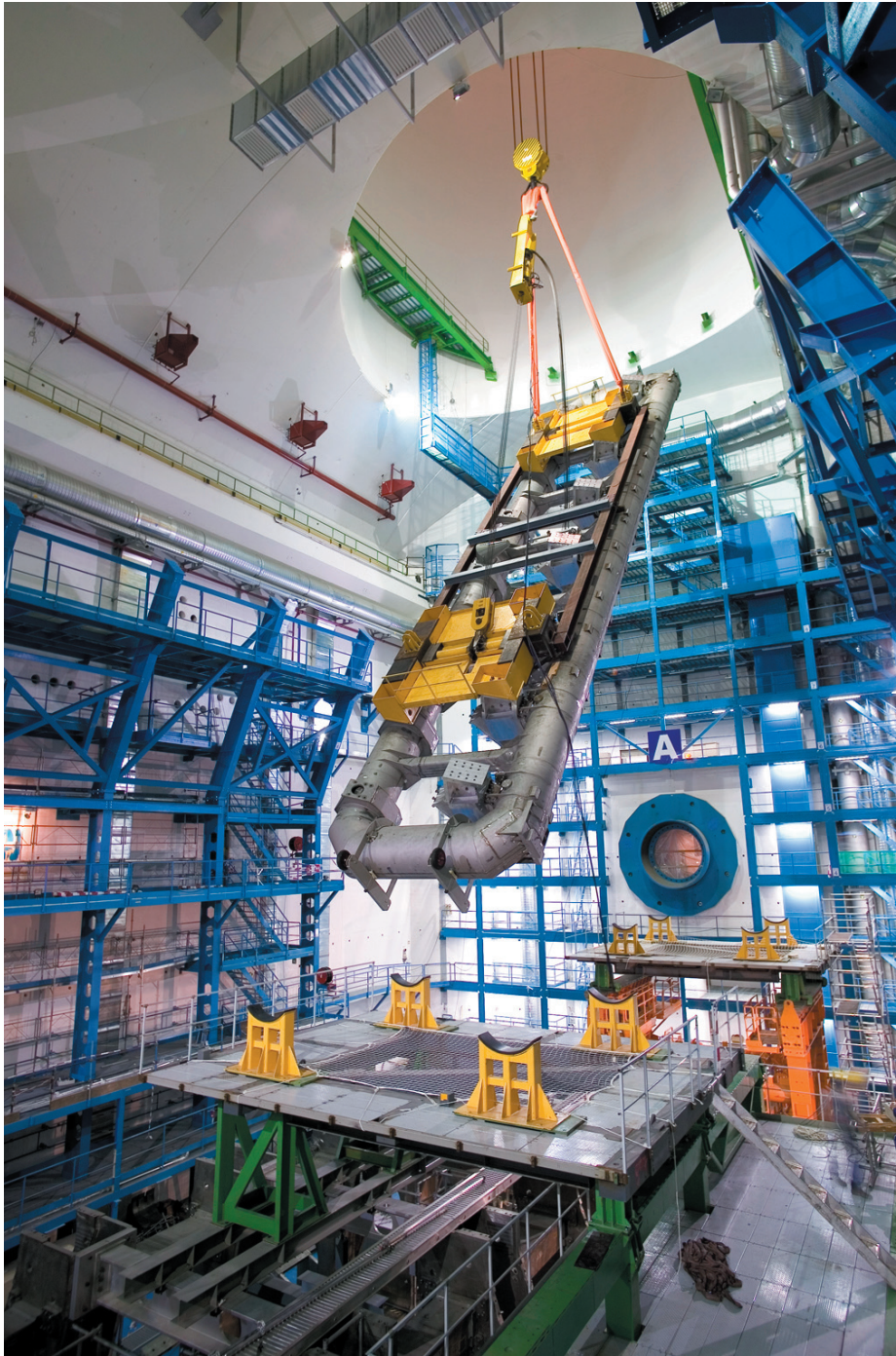
6TH ANNUAL WORKSHOP ON
LINUX CLUSTERS FOR SUPERCOMPUTING



SNIC Interaction

2ND USER MEETING FOR
SWEDISH NATIONAL INFRASTRUCTURE
FOR COMPUTING

Linköping 17–19 oktober 2005
<http://www.nsc.liu.se/lcsc/>



ATLAS-detektorn under uppbyggnad. De första toroidspolarna sänks ner till bergrummet. Foto: CERN

forts fr sid 1

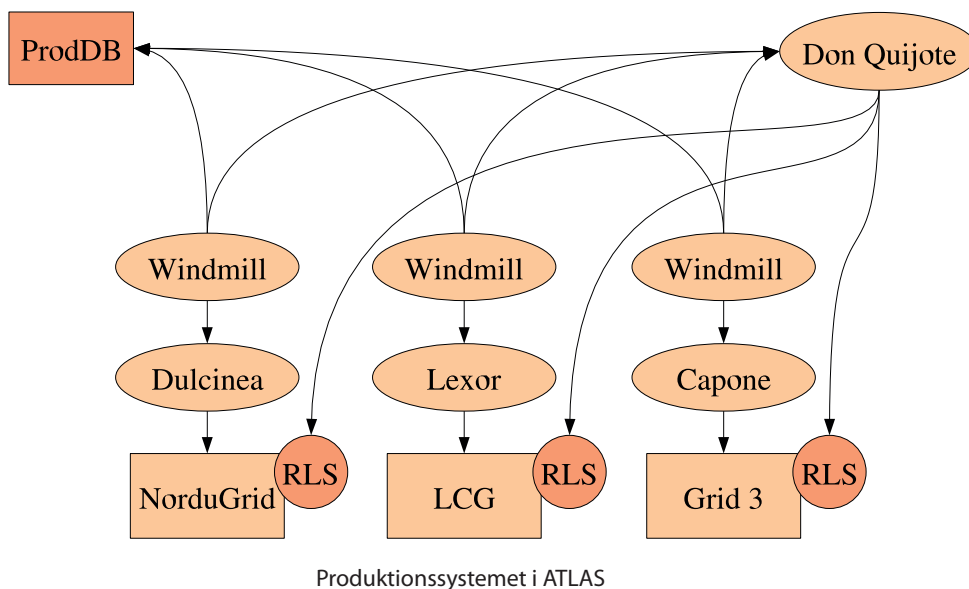
tolka dessa insamlade data kommer inte att finnas att tillgå centralt på CERN:s eget datorcentrum. Av denna anledning kommer det att vara nödvändigt att genomföra stora delar av dessa beräkningar utanför CERN. Därför är ett av målen med *Data Challenge*-testen att visa att betydande datorresurser utanför CERN kan integreras i produktionssystemet.

Under hösten 2004 genomförde ATLAS-kollaborationen *Data Challenge 2*. Genom att använda olika gridtekniker byggdes ett produktionssystem upp som kopplar ihop datorresurser från många olika datorcentra och universitetsinstitutioner.

På de svenska datorcentra används de olika Swegrid-klustren som är sammankopplade med gridsystemet ARC som utvecklats av NorduGrid. För att integrera resurser från olika gridsystem skapade ATLAS-kollaborationen ett gemensamt produktionssystem. Detta system kretsar kring en databas där de beräkningsjobb som skall utföras definieras – vilket program skall köras, vilken version, vilka in- och utdata. Från denna databas hämtas jobb-beskrivningarna av ett övervakarprogram kallat Windmill. Windmill skickar jobbdefinitionerna vidare till ett utförarprogram som översätter jobbdefinitionen till ett format som kan förstås av det gridsystem som utförarprogrammet stöder. ATLAS-

kollaborationen använde under *Data Challenge 2* beräkningsresurser från tre olika gridsystem. Förutom ARC användes även resurser sammankopplade med LCG och Grid3. För vart och ett av dessa system skrevs ett särskilt utförarprogram. Det utförarprogram som skrevs för att kunna skicka jobb till ARC kallas Dulcinea. Eftersom ARC:s grid-manager sköter nedladdning av indata och uppladdning av utdata automatiskt behöver Dulcinea bara se till att in- och utdata beskrivs korrekt i den jobb-beskrivning som skickas till klustren för att initiera den nödvändiga datahanteringen.

För att hålla reda på alla filer som skapas används en *Replica Location Server* (RLS). Denna håller reda på på vilka lagrings-



servrar en viss fil finns lagrad. ARC:s grid-manager kan automatiskt ladda upp utdata-filer till en lagringsserver med ledigt utrymme och registrera dessa i RLS. Ett datahanteringssystem kallat Don Quijote skapades också. Med detta system, som kopplar samman informationen från RLS-servrarna i de olika gridsystemen, kan man söka efter data i och kopiera data mellan de olika gridsystemen.

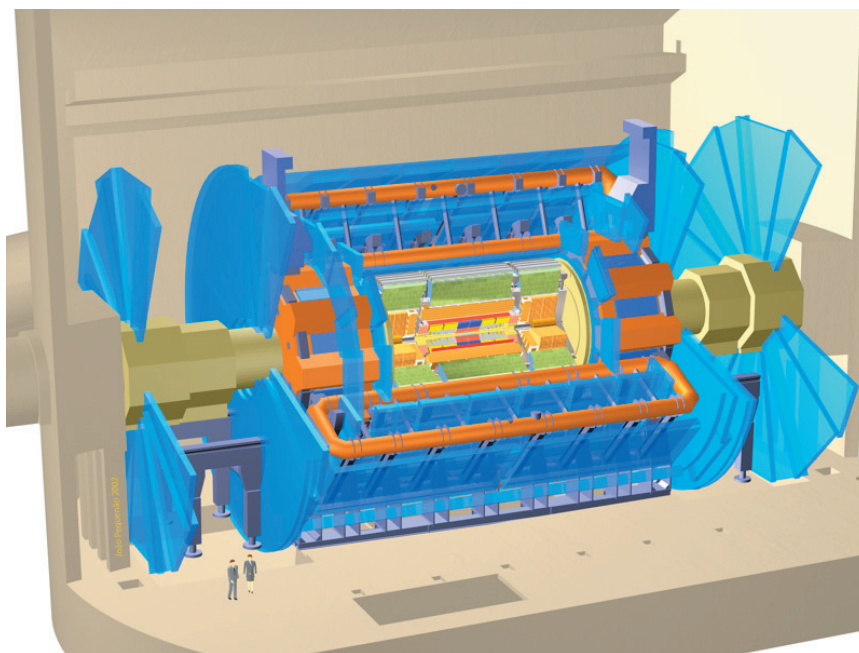
Programvarupaketet som behövs för Monte Carlo-genereringen av kollisionshändelser, simuleringen av ATLAS-detektorns respons och rekonstruktion av händelserna med hjälp av mönsterigenkänning installerades som en programvarumiljö (*runtime environment*) på de kluster som stod till vårt förfogande. I jobbeskrivningen i databasen definieras vilken version av programvaran som skall användas. Denna information använder Dulcinea för att skapa ett programvarumiljövillkor i den jobbeskrivning ges till ARC. På detta sätt skickas ett jobb till något av de tillgängliga klustren som har rätt version av programvaran installerad.

Andelen framgångsrikt avslutade jobb ligger runt 65–70%. En av de vanligaste orsakerna till att ett jobb misslyckas är problem med att lokalisera och hämta indata. Lagringsservrarna där datafilerna lagras har haft ett antal planerade och oplanerade servicestopp under perioden *Data Challenge 2* pågick, liksom RLS-servern själv. Denna typ av fel är dock inte så allvarliga eftersom om en indatafil saknas kan inte jobbet starta vilket betyder att ingen beräkningstid går till spillo. Om ett jobb misslyckas registreras detta av Windmill och ett nytt försök görs automatiskt vid ett senare tillfälle.

Förutom Swegrids beräkningssurser användes andra resurser med ARC:s

gridsystem i Norge, Danmark, Tyskland, Schweiz, Slovenien och Australien. Under *Data Challenge 2* producerade de datorresurser som kopplats samman med ARC ca 30% av den totala mängden data. ATLAS *Data Challenge 2* har visat att ARC på ett effektivt sätt har integrerats med ATLAS produktionssystem.

MATTIAS ELLERT
INSTITUTIONEN FÖR
STRÅLNINGSVETENSKAP
UPPSALA UNIVERSITET



Sprängskiss av ATLAS-detektorn i färdigt skick. Illustration: CERN



Det nya klustret Tornado

Nya resurser för väder- och klimatberäkningar

På NSC är det bråda dagar med installation och driftsättande av två nya klusterresurser för väderberäkningar. De nya systemen heter Tornado och Dunder. Tornado är avsedd för klimatstudier på Rossby Center vid SMHI och Meteorologiska Institutionen vid Stockhoms Universitet (MISU) och finansieras via medel från Knut och Alice Wallenbergs stiftelse. Dunder är en SMHI-resurs och avsedd för SMHI:s forskning och utveckling.

Båda systemen är installerade hårdvarumässigt på plats i NSC:s datorhall (som fått en rejäl uppgradering av sin kylkapacitet för att klara de nya systemen). Det som nu tar sin början är programvaruinstallation och konfiguration: operativsystem, Infiniband, övervakning, kompilatorer, kösystem, MPI med mera. Tidig användning kommer att kunna ske under sommaren och full drift är planerad till 1 september.

Tidigare i år har ju, som vi berättade om i förra numret av Nytt från NSC, Blixt

driftsatts. Blixt är avsedd för prognosberäkningar och är nu installerad på SMHI i Norrköping och sedan 1 juni huvudresurs för de dagliga prognoskörningarna vid SMHI.

Med dessa tre system på plats avslutas två aktiviteter som pågått i drygt ett år:

- Den första fasen av klimatberäkningsprojektet SweClics med NSC, Rossby Center vid SMHI och MISU som medintressenter. SweClics blir ett dedicerat system för klimatberäkningar. Invigningen sker den 23 augusti.
- Ett projekt som ger en omfattande uppgradering av de beräkningsresurser som NSC tillhandahåller sin partner SMHI för produktion och forskning.

I projekten har en omfattande förstudie genomförts bestående av tre olika faser:

- Marknadsstudie av hårdvarulösningar, främst processorer, moderkort, nätverk och lagring.

- Marknadsstudie av programvara, främst operativsystem, kösystem, filsystem, kompilatorer och matematikbibliotek.
- Omfattande tester av de olika applikationerna inom meteorologi, klimatologi och oceanografi som skall användas.

Testerna gjordes på befintliga kluster på olika håll i världen för att kunna testa och analysera såväl hårdvaruaspekter (minne, processor, nätverk och lagring) som programvara (främst kompilatorer, matematikbibliotek och MPI).

Testkluster har införskaffats för att testa och utvärdera olika lösningar.

Det som återstår är en del i den arkiveringsresurs som för tillfället är under införskaffande vid NSC. Kompletterande disklagring är också under införskaffande, såväl för Tornado som för Dunder.

De tre systemen Blixt, Tornado och Dunder är alla Linuxkluster baserade på noder bestyckade med två Intel EM64T Xeon-processorer och 2 GiB minne. Noderna är sammankopplade med Infinibandnätverk från Mellanox via en PCI-Express buss. Intel-kompilatorer, matematikbibliotek och MPI från Scali kompletterar bilden. Kösystem är Torque och operativsystem CentOS 3.4. Systemen har separata diskservrar och inloggningsnoder. Tornado är dessutom bestyckad med en så kallad analysnod med 12 GiB minne, avsett för analys av stora dataset i primärminnet. Tornado och Dunder har levererats av Dell. Blixt har levererats av Advanced Computer Technology i Linköping.

Kluster	Processorer: Intel EM64T					Total toppprestanda (Gflops)
	Klocka (GHz)	Minne per processor (GiB)	Cache (MiB)	Topprestanda (Gflops)	Antal processorer/ Antal noder	
Blixt	3,2	1	1	6,4	120/60	768
Dunder	3,4	1	2	6,8	104/52	707
Tornado	3,4	1	2	6,8	264/132	1795

Prestanda hos de nya klustren

Kalendarium

ICPP 2005, 34th Annual International Conference on Parallel Processing

14–17 June, 2005, Oslo, Norway
<http://www.dnd.no/icpp2005>

The 2005 World Congress in Applied Computing

20–23 June, 2005, Las Vegas, USA
<http://www.world-academy-of-science.org/WCAC2005/ws>

2nd International IEEE Symposium on Mass Storage Systems and Technologies

19–24 June, 2005, Sardinia, Italy
<http://storageconference.org/2005>

Compframe 2005, A workshop devoted to frameworks and components in high performance computing

22–23 June, 2005, Atlanta, Georgia, USA
<http://compframe.org/CFP.html>

The 2005 International Multiconference in Computer Science and Computer Engineering

27–30 June, 2005, Las Vegas, USA
<http://www.world-academy-of-science.org/IMCSE2005/ws>

The 3rd International Workshop on High-level parallel programming and applications, HLPP 2005

4–5 July, 2005, Warwick University, Coventry, United Kingdom
<http://hlpp2005.free.fr>

SAMOS – V: Embedded Computer Systems: Architectures, Modeling and Simulation

18–20 July, 2005, Greece
http://samos.et.tudelft.nl/samos_v/

The 11th International Conference on Parallel and Distributed Systems ICPADS 2005

20–22 July, 2005, Fukuoka, Japan
<http://www.takilab.k.dendai.ac.jp/conf/icpads/2005/>

The 14th IEEE International Symposium on High Performance Distributed Computing HPDC-14

24–27 July, 2005, NC, USA
<http://www.caip.rutgers.edu/hpdc2005>

Challenges of Large Applications in Distributed Environments CLADE 2005

24 July, 2005, NC, USA
<http://www.cs.umd.edu/CLADE2005/>

Euro-Par 2005

30 August–2 September, 2005, Universidade Nova de Lisboa, Portugal
<http://europar05.di.fct.unl.pt/>

PaCT-2005 Parallel Computing Technologies

5–9 September, 2005, Krasnoyarsk, Russia
<http://ssd.sssc.ru/conference/pact2005/>

PPAM 2005, 6th International Conference on Parallel Processing and Applied Mathematics

13–16 September, 2005, Poznan, Poland
<http://ppam.pcz.pl>

HPCC-05, The 2005 International Conference on High Performance Computing and Communications

21–23 September, 2005, Sorrento, Italy
<http://www.hpcc05.unina2.it/>

Linux Clusters for Supercomputing and SNIC Interaction

17–19 October 2005, Linköping, Sweden
<http://www.nsc.liu.se/lcsc/>



Linköpings universitet

Nationellt Superdatorcentrum, Linköpings universitet, 581 83 Linköping

Tel: 013-28 26 18, fax: 013-28 25 35, e-post: info@nsc.liu.se

www.nsc.liu.se