Matt

# C3.ca Association Inc.

## Technical Analyst Support Program (TASP)

Attn: C3 CCO <cco@c3.ca>

***Project Title***: **Geophysical Disaster Computational Fluid Dynamics Centre - Part 5: A Maturing HPC Complex - Infrastructure Upgrade Planning**

***Principal Investigator***: **Roland Stull**, Professor, Atmospheric Science Program
Dept. of Earth & Ocean Sciences, University of British Columbia
6339 Stores Rd., Vancouver, BC V6T 1Z4
(604) 822-5901, fax: (604) 822-6088, email: rstull@eos.ubc.ca

***Dates***: 1 July 2005 - 30 June 2006 (employ George Hicks II)

***Budget***: $ 29,560.

***TASP Person***: Contact and relevant information listed below. CV is attached.

| *Continue During 1 July 2005 - 30 June 2006* |
|---|
| **George Hicks II**<br>(same mailing address as PI)<br>604-822-4760<br>Fax 604-822-6088<br>ghicks@eos.ubc.ca |
| *Comments*: Primary HPC systems administrator since January 2003. Also manager of the Emergency Weather Network, programming project leader, and technical support for cluster users. |
| *Special circumstances*: (none) |
| *Status*: Hired and in place now. Agrees to support C3 goals, and to serve on TASP-TC. |

***Eligibility***: Stull is PI of a team of 15 scientists who won a $1.3M grant from CFI during July 2000. This infrastructure funding allowed the purchase of a HPC at UBC as the foundation for a new Geophysical Disaster Computational Fluid Dynamics Centre (GDCFDC). At this Centre, the internal users solve fluid-dynamical codes to study weather-related disasters (cyclones, forest-firestorms, avalanches, floods, etc). As per the CFI agreement, we are a Resource Provider (RP) within C3.

***RP Hardware***:

a) Number Cruncher: 256-processor IBM Linux cluster became operational in January 2002.

- 128 dual-processor PIII, 1 GHz nodes, 256kB cache, 1 GB RAM, 18.2 GB disk
- 2 head nodes, each dual PIII, 1 GHz, 256kB cache, 2 GB RAM, 144 GB RAID
- 2 management nodes, dual PIII 1 GHz, 256 kB cache, 512 MB RAM, 72 GB RAID
- 2 storage nodes, dual PIII, 1 GHz, 256 cache, 2 GB RAM, 72 GB RAID, LTO tape
- fast storage: 1.1 TB RAID with Netfinity Fibre Channel

- Networks: Myrinet 2000 & 100 MB ethernet & 10 MB ethernet & serial
- Misc: 5 racks of 1U processors, UPS APC Silicon, PGI & GNU compilers

b) Input & Output Computer Systems:

- 2 quad IA64 Itanium nodes, 733 MHz, 6 GB RAM, 2 MB L3 cache.
- 1 SGI Origin 300 with 4 processors.

***Commitment Guarantee***: We agree to make at least 20% of our computer time available to outside users, as specified by CFI. This outside time will be allotted/managed by C3, as specified by CFI. Technical and training support for these outside users will be provided by the TASP person who will be partially supported via this C3-TASP grant.

***Budget Proposal***

1. **Salary**: George's proposed full time salary = $4,034/month + 18% non-discretionary fringe benefits. Will be employed 1 July 2005 - 30 June 2006 = 12 mo x 50% TASP support x ($4,760/month) = **$28,560 requested from C3.**

2. **Travel Costs**: To cover the cost of the TASP-Tech. Committee meeting scheduled for 2006, I request **$1k** from C3 to help support half of travel costs for George.

**NET Budget**: Item 1: **$28,560**
Item 2: **$ 1,000**
**TOTAL $29,560 requested of C3.**

***Deliverables***: Deliverables include monthly status reports on external user usage. An annual report will include benefits accrued, theses resulting, new projects enabled, new personnel trained, and other info as specified by C3. Also, George will participate in the TASP Technical Committee (TC), via internet, phone, database, and at face-to-face meetings as scheduled by TASP.

***Continuous Activities***: (Interactions with *external users are highlighted with italics.*)

*Continue to activate new accounts and trains new users (as necessary). Maintain a user-friendly front end showing daily outside-user window for job scheduling, size of queue, max processors requested, and remaining time and processors available. Post usage statistics on the web, to encourage more outside users. Respond to user questions, offer help and user support, and update the FAQ web page.* Maintain and refine the daily operation to maximize usage. Continue to monitor system and file usage, provide security, and recommend changes/enhancements (e.g., replace failed parts, purchase more disks; add firewalls). Share experiences with TASP-TC colleagues, and write monthly reports for C3-TASP RC on routine operations, accessibility of user support, and success in attracting and serving external users.

***Milestones***: (Interactions with *external users are highlighted with italics.*)

Quarters 1 & 2 (July - Dec 2005): Search for potential grant sources to expand/replace existing cluster resources. If alternatives to CFI can be found, proposals will be written. Assistance, where possible, will be given to Dr. Stull and other members of the GDCFDC team on research that may yield additional grant opportunities. This may include assistance in configuring/running forecast models and the production of high resolution forecast products.

*Continue offering individual support and training to external and internal users, particularly in the area of porting and adapting parallel code to the existing cluster and itanium nodes.*

Quarters 3 & 4 (Jan - June 2006): As required, continue researching and applying for grants. Begin/continue short term infrastructure grants. *Continue offering individual support and training to external and internal users, particularly in the area of porting and adapting parallel code to the existing cluster and itanium nodes.*

***Past Performance***: The primary focus of the past year has been on maintaining the aging cluster. Numerous hardware failures occurred during the past year, including some major failures such as the loss of the Myrinet Switch backbone. As the cluster is now in its fourth year, hardware problems will become more prevalent, and this will be a growing concern as maintenance contracts shrink to save money for future systems. Despite this, external and internal user numbers remained steady and cluster utilization continues to grow with consistent resource utilization around 66% and external utilization over 13%.

Our greatest concern now is meeting the needs of our growing forecasting research and external users. Our prediction for usage levels last year was met, and on some occasions exceeded. We already see that there is a demand for our cluster resources beyond what we are able to accommodate (e.g., one external user requesting 80 processors around the clock for three months for continuous simulations). To meet the future needs of the GDCFDC, we will need to replace the aging cluster with a newer, faster, and larger cluster. As we have seen across Canada, this demand is not a passing fancy but a requirement to take the research being done at the GDCFDC to the next level.

Since September 2002 to present, George Hicks II has been in charge of the day-to-day operation and maintenance of the IBM cluster. Registering new users and creating new user accounts has remained the responsibility of Henryk Modzelewski. George's knowledge of the cluster and its management has grown to meet our expectations. Trina Cannon has been working on contracts with the GDCFDC since 2000, and in the past year has been handling user support for the researchers within the GDCFDC and has begun training in systems administration. Trina has become an integral part of the support team for the GDCFDC. Funds for helping to support Trina and Henryk are not requested in this proposal.

***Summary***: Cluster utilization is already often at or near saturation. We will be working to expand our available resources for both internal and external users through whatever grants can be found in 2006 to allow the center to grow and remain a leader in meteorological research and disaster prevention. At the same time, we will maintain the availability of resources and support for our external users.

**NOTE**: The GeoDisaster CFD computer was awarded before CFI allowed operations costs to be included in their grants. Therefore, the C3-TASP grant is critically important to enable the continuing operations of our cluster and services to outside users.

***GEORGE D. HICKS II***
*3948 West 23rd Ave*
*Vancouver, BC V6S 1L2*
*(604) 734-4794 (H)*
*(778) 388-4794 (Cell)*
*ghicks@eos.ubc.ca*

## EMPLOYMENT INTEREST

Seeking challenging full-time programmer/systems administrator position.

## EDUCATION

**University of British Columbia**
Vancouver, BC

B.Sc. (conferred in 2003) majoring in computer science and minoring in philosophy
B.Ed (six months completed) specializing in high school drama and math
B.A. (conferred in 1995) majoring in theatre (acting, directing, history,lighting and set design), unrecognized minor in mathematics (linear analysis and optimization)
B.Eng (two years completed) specializing in computer electronics

## WORK EXPERIENCE

**Geophysical Disaster Computational Fluid Dynamics Centre** **May. 1, 2001 – Present**
**Department of Earth and Ocean Sciences**
**University of British Columbia**
**Vancouver, BC**

• Web Development: Designed and implemented the online laboratories for the course EOSC 114. Remodeled/maintain the online interface to the Emergency Weather Network. Maintain the site for the GDCFD's weather forecasting products and cluster resources.
• Database Administrator/Programmer: Maintained and extended the Emergency Weather Network database system. Tripled the number of recorded observation stations. Designed and implemented a complementary database system for storing forecast results as well as function libraries for accessing both the observation and forecast databases.
• Systems Administrator: Assisted in administration of a 128 node Linux supercluster, a 12 node beowulf cluster, six SGI multiprocessor systems, and several stand alone workstations. This included hardware maintenance and replacement, software troubleshooting and installation, and assisting users with inquiries.

**Association of British Columbia Professional Foresters** **April 1, 1996 – Aug. 27, 1999**
Vancouver, BC **Summer 1994, Summer 1995**

• Deputy Registrar. Handled all aspects of registration (maintained both hardcopy and electronic

rolls, processed change of status applications, enrolment applications, and Registration Exam applications, performed accreditation analysis of new enrolments).
• Administrator and organizer for the Board of Examiners, Standing Investigations Committee, Discipline Committee "B" Panel, and 26 Academic Standards subcommittees. Internal and external resource and information officer dealing with inquiries and complaints regarding ABCPF policies, procedures, bylaws, the *Foresters Act*, FOI requests, Code of Ethics, and discipline. Junior systems administrator for small (10 station) computer network. Included troubleshooting minor network problems, hardware maintenance, and maintenance of the association database.
• Office clerk/computer consultant. Main focus on database development (GENCAT), document production, and staff training.

## SKILLS

- Computer languages: Perl, C/C++, Fortran, HTML, PHP.
- Operating systems: DOS, Windows 3.11/95/98/NT4.0/2000/XP, Unix, Linux, IRIX.
- Computer maintenance experience (cleaning, replacement, and installation of hardware).
- Experienced in dealing with the public, both in presentations and in handling inquiries.
- Extensive time management and event organization experience.
- Extensive experience with office procedures and business writing.
- Type 60 wpm.

## EXTRACURRICULAR ACTIVITIES

Acting, photography, chess, cycling, basketball, squash, computer programming, strategy games of all kinds, reading, writing

## REFERENCES

***References are available upon request.***

Geophysical Disaster Computational Fluid Dynamics Centre 2004-2005 Annual Report

The Geophysical Disaster Computational Fluid Dynamics Centre (GDCFD) was created to save lives and minimize economic losses caused by natural disasters. The primary focus of the centre has been on weather and disasters particular to the Pacific Northwest.

The GDCFD utilizes a high performance supercluster (purchased in September 2001) for research in numerical weather forecasting and producing daily operational forecasts. The cluster is a 256 processor, 128 node linux cluster controlled by four management nodes. The computational nodes have dual Pentium III 1GHz processors and 1GB of RAM. User storage is located on a 1TB RAID array. Additional data can be offloaded to LTO tape through a single tape drive on one of the management nodes. The cluster's interprocess communication is handled through a Myrinet switch. In 2004, a Max Xserver was added to the cluster. The Xserver is functioning as an additional fileserver with an additional 750GB of storage space.

Although capable of running small single processor jobs, the cluster is particularly well suited to large scale parallel numerical models. It supports the GNU and Portland Group compilers with additional support for MPI. Resource allocation is handled through Maui 3.0.8 layered on top of Torque 1.1.0p6.

Presently, five different numerical models are run daily to provide operational weather and air-quality forecasts over western Canada and parts of the Yukon, Northwest Territories, and Washington State. These models produce 21 different forecasts at varying resolution and length. The weather forecasts are then utilized to produced an Kalman Filtered ensemble forecast with very reliable results. Work continues towards producing ensembles of the air quality output along with output from another air quality model being run by Environment Canada. At present, four air-quality forecasts are made daily. This is being expanded, and it is hoped that a 6-13 member ensemble will be available by mid-2006. This research has never before been attempted by another research group anywhere in the world. The GDCFD continues to study the effectiveness of these ensembles and recommend possible improvements.

Work is also continuing in areas of data assimilation and the effects of the lack of observations over the Pacific ocean on model initialization. Eight graduate students work at the GDCFD maintaining these models and advancing this research.

In order to study the effectiveness of the models and perform verification, the GDCFD operates in conjunction with the Emergency Weather Network (EMWXNet). The EMWXNet collects, stores, and shares weather observation data collected by numerous private and government agencies across western Canada. In addition, the EMWXNet stores forecast data produced by the GDCFD after it has been interpolated to observation stations. It has continued storing verification statistics produced through comparison of the interpolated data and actual observations. These statistics have been crucial for error and bias correction, particularly in the ongoing development of a Kalman Filter. In February 2005, the EMWXNet underwent a significant recoding and is now capable of supporting a more expansive verification and Kalman Filter system. Several members of the GDCFD have been examining ways in which to improve the Kalman Filter system as a result.

In addition to the real-time forecast research, the GDCFD has also supported

the following projects and studies:

- Fire-weather forecasting;
- Study of the heavy rain event, 17-22 January 2005, BC Southern Coast and Interior;
- Wind speed in mountainous terrain;
- Multimode fine-resolution ensembles for short range forecasts in mountainous terrain;
- Parametrization frameworks for Boundary Layer cumuli;
- Numerical simulation of canopy flow and carbon dioxide flux at the Western Canada flux stations;
- Toronto Greenroof Project - high resolution sensitivity studies of vegetation effects in cities;
- Genecomber Project - Ab initio gene prediction in the human genome;
- Image reconstruction for positron emission tomography;
- Finite-difference time-domain studies of photonic crystal defect states;
- Episodic mixing and buoyancy sorting representation of non-precipitating cumulus clouds;
- Assimilation of surface weather station and radar station data in complex terrains; and
- 3D cloud and radiative modelling studies.

In 2004, the primary duty of the administration personnel at the GDCFD remained the maintenance of the GDCFD's high performance cluster as well as six SGI multiprocessor machines, a twelve node beowulf cluster, three intel servers, and a dozen macintosh desktop machines. In addition, personnel offered continued on-hand support to over a dozen researchers at the GDCFD. This included assisting with coding, debugging, and regular operation of four numerical forecasting models. The TASP analyst also offers support to the EMWXNet in the form of coding, debugging, user support, and hardware maintenance.

Problems resulting from the software and firmware upgrades applied to the cluster in August 2003 have been mostly worked out. There is still a lingering problem with the network filesystem running on one of the fileservers forcing us to reboot the machine weekly. This problem only affects internal users as external users access disk through the other fileserver.

Hardware maintenance has been more of a pressing concern during the past year. As the cluster is now in its fourth year, hardware failures are becoming more prevalent and will likely occur more frequently in future. We suffered the following hardware problems over the past year:

- 1 computational node hard drive failed, April 2004 – replaced and node re-initialized
- Cache battery failed on terabyte RAID, May 2004 – replaced
- Myrinet cable failed on a computational node, May 2004 – replaced
- 2 SDRAM memory sticks failed on a computational node, May 2004 – replaced
- 1 KVM (keyboard-video-mouse) switch failed, September 2004 – replaced (this is the third time this switch has been replaced)
- Myrinet backbone and power supply failed, December 2004 – replaced

and system re-initialized

- 1 computational node hard drive failed, December 2004 – replaced and node re-initialized
- 1 computational node motherboard failed, January 2005 – replaced and node re-initialized
- 2 hard drives failed on the terabyte RAID, March 2005 – replaced
- 1 computational node showing recurring failure of its administrative onboard controller, March 2005 – requires motherboard replacement, but does not affect operation so being bypassed

In the fall of 2004, the maintenance contracts with IBM, Myrinet, and APC concluded. Dialogues were opened with these groups to examine the cost of extending the maintenance contracts. The cost of renewing the existing contracts for the cluster was determined to exceed $100,000. The operating funds collected during the previous three years were insufficient to cover these costs, and it was clear that such costs might be better directed towards the purchase of newer hardware to replace existing infrastructure.

With this in mind, smaller, although still somewhat expensive, parts only maintenance contracts were established with IBM for the server nodes, and APC for the main UPS system. It was decided that no support would be put forward to the computational nodes, and Myrinet service would be done on a need only basis. We have a small stockpile of parts for the computational nodes, but expect a gradual loss of nodes in future as the cluster moves towards retirement. Useful parts will be cannibalized from dead nodes in future to help maintain other nodes and slow the deteriation of the cluster.

In the meantime, any operational funds that are collected will be pooled to continue the smaller maintenance contracts, and, hopefully, split off to help fund newer hardware. The GDCFD is making, and will continue to make, an active effort to search for grants and other funds to help provide new hardware.

Despite the fact that the cluster is aging, demand for and utilization of the cluster have increased. Usage and user statistics for the past year are:

| Month/Year | External Users | External Usage | Internal Usage | % of Available Resources Used |
|---|---|---|---|---|
| 03/2005 | 6 | 20.40% | 79.60% | 67.30% |
| 02/2005 | 5 | 14.70% | 85.30% | 78.10% |
| 01/2005 | 5 | 12.00% | 88.00% | 72.80% |
| 12/2004 | 5 | 5.90% | 94.10% | 66.10% |
| 11/2004 | 5 | 15.60% | 84.40% | 74.80% |
| 10/2004 | 5 | 10.30% | 89.70% | 66.80% |
| 09/2004 | 6 | 14.00% | 86.00% | 66.50% |
| 08/2004 | 5 | 9.59% | 90.41% | 58.10% |
| 07/2004 | 5 | 4.01% | 95.99% | 55.10% |
| 06/2004 | 5 | 16.37% | 83.63% | 58.20% |
| 05/2004 | 6 | 14.77% | 85.23% | 66.00% |
| 04/2004 | 5 | 16.56% | 83.44% | 62.50% |

The usage levels predicted last year were realized. Our external user base

has remained consistent during the past year at roughly half the number of users of previous peak years. However, external usage has been higher than in previous years, setting the highest usage levels yet. The average external usage was 12.85%. At present, most external users are working with climatological or short range weather forecasting models. Because of this, these users require large numbers of processors for long simulation runs. Even though we charge for accounts, these users have discovered that other computational farms are incapable of supporting the types of simulations they wish to run and have turned to our cluster to support their research. Overall, charging for accounts has actually made our external user base shrink, but the types of jobs run by our external users (typically requiring 30-40 processors for periods of 60-70 hours per job) has increased the utilization and demand.

Coupled with the continued expansion of the real-time forecasting research, external users are pushing the cluster use towards saturation. It is important to note that because of the size of the jobs being run on the cluster, it is rarely possible for all computational nodes to be utilized simultaneously. We rarely have users running short duration jobs (ie 4 hours or less), and we have no users that run single processor, or small parallel jobs. As such, daily gaps between scheduling of larger jobs makes the cluster appear to be less than 100% utilized. However, given the types of jobs currently being run, the cluster is often at or near capacity.

On fees, the GDCFD remains one of the few institutions which charges all users fees for using the cluster. There has long been debate about the fairness or usefulness of charging users for accounts. The year we received CFI funding for the cluster, operational funding was not yet granted by CFI. It was under the CFI guidelines that we were instructed to cover operational and maintenance costs through user account charges. Although we have endeavoured to keep the charges low, it is clear that charging for accounts has hindered the growth of external accounts on the system. Even so, account fees have also proven insufficient for meeting operating costs. Until a granting agency is willing to give funds specifically directed to infrastructure support, institutions such as the GDCFD will have difficulty maintaining computational infrastructure.

Finally, in 2005-2006, we expect high workloads on the cluster and will continue to expanding the computational resources of the GDCFD to allow it to maintain its status as a leading edge research facility in weather and air quality forecasting.