

PACE Advisory Committee Activity Summary

August, 2021

Introduction

The PACE Advisory Committee is a joint effort between the Office of the Executive Vice President for Research and the Office of Information Technology. The goal of the committee is to ensure that shared research computing services are meeting faculty needs and resourced in a sustainable way. The committee consists of a representative group of faculty who depend on the advanced computational and data capabilities provided by OIT's research cyberinfrastructure. An important goal of the committee is to provide essential feedback, which will help continuously improve this critical service.

The formal charge of the committee, jointly given at the kickoff meeting by Dr. Chaouki Abdallah, EVPR (Executive Vice President of Research), and Daren Hubbard, VP-IT and CIO is to

- Function as a communication channel between the broader research computing community and PACE.
- Serve as a sounding board for major changes to the PACE infrastructure.
- Maintain an Institute-level view of the shared resource; and
- Help craft strategies that balance the value and benefits provided by the resources with a sustainable cost structure in the face of ever-increasing demand.

The committee currently consists of:

Srinivas Aluru, IDEaS Director

Omar Asensio, Public Policy

Dhruv Batra, Interactive Computing/ML@GT

Mehmet Belgin, PACE

Annalisa Bracco, Earth and Atmospheric Sciences

Neil Bright, PACE

Laura Cadonati, Physics

Sudheer Chava, Scheller College of Business

Yongtao Hu, Civil and Environmental Engineering

Lew Lefton, EVPR/Math (ex-officio)

Steven Liang, Mechanical Engineering/GTMI

AJ Medford, Chemical and Biomolecular Engineering

Joe Oefelein, Aerospace Engineering

Annalise Paaby, Biological Sciences

Tony Pan, IDEaS

David Sherrill, Chemistry and Biochemistry, Director of [Center for High Performance Computing](#)

Huan Tran, Materials Science and Engineering

The committee began its regular meeting cadence in March 2021 and has met four times to date. Below is a summary of the meetings and a review of progress made.

Meeting date 3/30/2021

Kickoff meeting. Introductions. Committee formally charged (see above) by Dr. Chaouki Abdallah and Daren Hubbard.

Neil Bright reviewed the new cost accounting model. Further details are at <https://pace.gatech.edu/update-gts-research-cyberinfrastructure-cost-model>. Key points: PACE charges based on actual computing and storage resources used. This consumption model is based on an approved rate study and cost per CPU-hour is based on corresponding hardware cost and standard benchmarks. Sponsored funds used for PACE services and commercial cloud services are granted an F&A Waiver (through at least FY22). The new model has more transparent accounting, can support rapid provisioning (vs old model where 6-8 month delays were the norm), hardware flexibility, a free tier for all PI-eligible faculty, and operational capacity planning and forecasting. Solicitations that explicitly target hardware/equipment in the RFP (e.g., DURIP, MRI, CC*) can still be handled, and such equipment would typically not be included in the model.

Low Lefton reviewed feedback and concerns already voiced. Key points: The decision to change the model was made with insufficient faculty input, concerns about expiration of credits (both existing credits from converted Rich Building equipment and startup/retention), concerns about wait times in the new model as compared with dedicated resources, request to pool credits (library model) across larger groups, concern that F&A will be start being charged on PACE after initial 2 year evaluation period, cost share uncertainties (including using existing credits as cost share), concerns about negative sponsor reaction, concerns that new model will increase cost on grants, concerns that unfunded computational explorations (e.g., graduate student research and thesis work) will not be supported, concerns that this will create additional workload on faculty who will have to submit secondary proposals for compute time on national resources, concerns that national resources cannot provide the necessary level of computing support and control that local resources provide, e-routing checkbox confusion.

Additional concerns discussed, including queue system, how to manage credits on multiple projects and grants, need for HPC researchers to have limited but dedicated access to a large cluster for benchmark runs, software version churn

Action items:

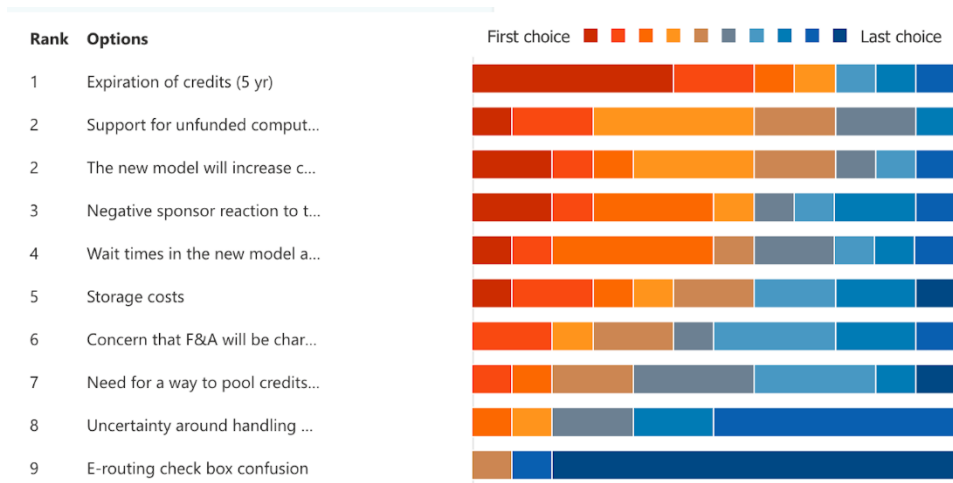
We will poll the Committee to rank/prioritize the concerns and begin working on the highest priority ones.

Meeting date 4/30/2021

Review survey results and come to consensus around top three priorities. These were the results:

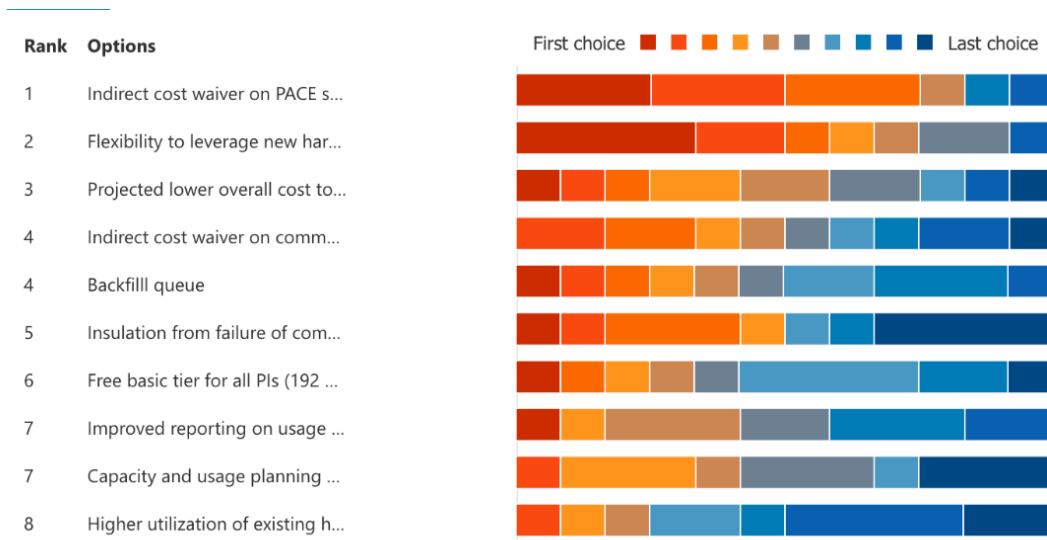
PAC ranking of concerns (highest to lowest)

1. Expiration of credits (5 yrs.)
2. Support for unfunded computational explorations (e.g., graduate student research and thesis work)
3. The new model will increase costs on grants, which is a competitive disadvantage
4. Negative sponsor reaction to the new model
5. Wait times in the new model as compared with dedicated resources
6. Storage costs
7. Concern that F&A will be charged on PACE after initial 2-year waiver
8. Need for a way to pool credits (library model) across larger groups
9. Uncertainty around handling of cost share
10. E-routing check box confusion



PAC ranking of speculative benefits (highest to lowest)

1. Indirect cost waiver on PACE services
2. Flexibility to leverage new hardware releases instead of being restricted to hardware purchased at a specific point in time
3. Projected lower overall cost to PIs versus the equipment model
4. Indirect cost waiver on commercial cloud services
5. Backfill queue
6. Insulation from failure of compute nodes
7. Free basic tier for all PIs (10000 CPU hours and 1 Tb project storage)
8. Improved reporting on usage per project, user, job, etc. for data-informed resource management
9. Capacity and usage planning allows for pre-purchase of resources, drastically reducing procurement wait times for new PIs
10. Higher utilization of existing hardware improves environmental impact



Extensive and robust discussion of top concern – expiration of credits from Rich Building Refresh. Various options considered. Settled on the following proposal:

Credits from Rich Building Refresh will last 7.5 years (instead of 5 years)

PI may request 1 year no-cost extension

Neil announces his departure to NZ.

Homework is to review CoS memo.

Meeting date 5/27/2021

Proposal regarding credit expiration reviewed by PACE team, and the following has been approved:

Existing compute credits from the Rich Building migration, which were originally set to expire in 5 years (April 2026), will remain available for 7.5 years (October 2028). If there are still credits available at that time, PIs may request a one-time no-cost 1 year extension.

After checking with Grants and Contracts Accounting and EVPR office, the following has been approved:

PACE credits can be used for cost share arrangements. See <https://pace.gatech.edu/cost-share-within-pace-cost-accounting-model> for further details on how this should be handled.

Discussion of other expiration concerns, in particular startup and retention. Current practice is that College and EVPR split PACE startup and retention commitments. CoS proposes to have no expiration on their half of the contribution. EVPR commitment would remain at 3 years and ideally spent first. Follow up with other Deans to see if they want to consider something similar. PAC may suggest an updated model for EVPR half after the current Provost review of startup practices is completed.

For future discussion: Propose a consistent model for the Institute where all startup and retention funds would be available for 7.5 years with 1 year no cost extension (like the current Rich building equipment credits). This would create a consistent “floor” which allows startup and retention funds to more closely compare with an equipment purchase. CoS may still choose to keep their half unexpired.

Discussion of PAC communication to the larger community. Use PACE website area.

Consider changes to PACE advisory committee makeup. Decided to leave as is, but make sure PAC activity is communicated more broadly. Also, if unit IT Directors are finding that the new model is increasing the number of ad hoc clusters, then PAC should be engaged to understand the situation and address any gaps in PACE services.

The F&A waiver on PACE charges will include consulting charges (this wasn't clear). Will be reviewed in Spring 2023, when we have more data and at the same time the full F&A waiver is reviewed. PIs who request sponsored funds for consulting charges should handle them in the same section in their budgets (M&S) as other PACE compute & storage charges. In particular, they should NOT be included as personnel lines.

Discussion of CoS memo. General support for the relevant well-articulated concerns. Most are already captured in our working list.

Some recent CoS faculty startup was promised to purchase equipment (presumably in Rich) but ended up being converted fully into credits. These faculty (list to be provided) will be given the better of two options: 7.5 years of credits (with 1 year no cost extension option) or, since they are CoS, 50% of their credits never expire. The latter option is default.

Announcements and discussion topics

- PIs can now optionally put guardrails, soft and hard limits on usage.
- Dashboard for tracking usage. Still in development but available soon. It would be nice to include the cost of nodes in the dashboard.
- Increase Flexibility (reducing one-size-fits-all approach): sharing of credits among small to medium sized groups now widely available through hierarchical credits model. Needs to be put on website.
- Increase Flexibility (reducing one-size-fits-all approach): Use reservation system to guarantee certain capacity for benchmark runs, etc. Note reservation implies you pay 100% of cost, whether used or not.
- If wait times exceed certain thresholds, (e.g., In x weeks y % of jobs have more than z hours wait) then time to get more nodes of appropriate type. Option to also fence off dedicated resources temporarily.
- Scheduling downtimes before a big conference is a problem. Maybe use reservations to see when to schedule downtime.

Other topics:

Seek balance between HPC and HTC and optimize it.

Traffic on Phoenix (I/O can be swamped by HTC).

Discussed future meeting schedule. June is already scheduled. Skip July, and continue monthly through December. Revisit in November.

Meeting date 6/23/2021

External communication.

- Meetings will be recorded, and recordings stored in the private Teams channel. Recordings will not be published (to maintain some privacy and allow committee members to feel comfortable speaking openly).
- Minutes will be taken, and a summary published on PACE website after committee review.
- An update of Advisory Committee activity (approximately quarterly) will be made available on the web and for committee members to share to their units and broadly.

Update on startup and retention funds:

- No expiration date for College of Sciences startup and retention funding is **approved** and will be implemented.
- EVPR funds will be drawn down first, but EVPR will track this to see if this becomes an issue and adjust if necessary.
- Needs coordinated communication with unit finance staff to handle PACE invoices correctly.

Discussion of prepurchase credits for internal funds (startup and retention). Invite Jonathan Jeffries to PAC meeting for conversation on this.

Support for unfunded computational explorations (e.g., exploratory research, thesis work, fellowship recipients w no PI).

- Keep 10000 CPU hours/mo. free tier. Very simple and good way for new researchers to get started.
- Backfill queues are free and are in place for current users with a PI.
- Seed grants for higher tier (e.g., 100,000 CPU hours or more). Good skill for grad students to develop. Run seed grant program for 6-12 months and see how it is used and re-evaluate it if needed. Leverage Tech Fees for free tier student cluster? (Different from ICE since not connected to courses, but still primarily student use).
- Undergrad research projects self-directed (no PI) could be accommodated via seed grants.
- Communicate all tiers/options for free use in one place?