

An interview of Prof. Miron Livny, a pioneer in High Throughput Computing

A conversation about an open computing system that contributed to two Nobel Prizes in Physics



Prof. Miron Livny speaking at TIFR on High Throughput Computing.

[Prof. Miron Livny](#) of University of Wisconsin-Madison visited TIFR on October 24-25, 2019 for the 1st Asian HTCondor workshop. He is a pioneer in High Throughput Computing (HTC), a distributed computing system. Among other things, he leads HTCondor, a distributed resource and job management system for HTC. His work has been crucial to enabling cutting edge science experiments worldwide, the likes of the Large Hadron Collider, the IceCube Neutrino Observatory and more.

Jatan Mehta, a science communicator at TIFR, took the opportunity to have a chat with Prof. Livny on his pioneering work that has grabbed two Nobel Prizes in Physics.

Can you tell us how is High Throughput Computing (HTC) different from say, supercomputers or grid computing that people hear about?

Imprecise language and buzzwords govern the computing ecosystem. I coined the term HTC in the mid-nineties in order to differentiate it from traditional High Performance Computing, known to many as supercomputers. HTC is a distributed system that enables even individual researchers to get effective access to large compute capabilities. Recently, a National Academy of Science report in the US said that high throughput computing is not just important for scientific discovery, but is enabling scientific discovery in a growing number of domains.

How did you start HTC and how did HTCondor develop?

What I'm doing today is anchored in my PhD work in the late 1970s. I always joke that I've been working on the same problem for over 40 years and it's still not done! My thesis was on load balancing and distributed systems. I was always fascinated by the simple problem that you have a quest for work sitting and waiting in one place and a resource capable and willing to serve it is idling in another place. How do you bring them together? It turns out it's an unsolvable problem so I can work for 40 more years!

Then in the early eighties, when I came to Wisconsin, I was introduced to the notion of a workstation where computing power was basically assigned to an individual. And then we added the concept of a distributed ownership where you're not thinking about the system as homogeneous ownership with one master, but that there are many masters because every person now owns a computer capability. How do you bring all of them together while preserving the rights of the individuals? That led to developing Condor and its first deployment in 1985. Twenty years later we had to change the name to HTCondor for legal reasons.

What was the most interesting or challenging problem that HTCondor helped solve?

There are two Nobel Prize discoveries whose computation system was powered by HTCondor - the Higgs Boson in 2012 and then recently detection of gravitational waves by the LIGO collaboration. So I am always joking, I'm looking for the triple crown. But I can't say that they're more important or challenging than other works of science powered by HTCondor.

How do you enable smaller science projects and individual researchers using HTC?

In Wisconsin-Madison in the 90s, we did a lot of simulation of database systems, study algorithm performance and similar high throughput application to make it robust. I always believed that computing can act as a fertilizer for science. That if you give researchers more effective computing, they can do more, better science.

HTC, as opposed to traditional High Performance Computing, is much more of a democratic tool. With HTC, our goal is to allow a younger professor with no army of people in his/her lab to do large scale computing. We have at Wisconsin-Madison an economics professor that can do 200,000 hours a day using the HTC we offer that is powered by HTCondor. So what gives us most satisfaction is when we see the smaller guys being able to bring computing to do their science and then them saying, *"Without you we couldn't do it."*

How has the Open Science Grid project been progressing?

The Open Science Grid, designed as a national shared distributed HTC resource for researchers, has been a wonderful experience in many directions. Obviously one is that we were able to deliver the USA computing commitment to the Large Hadron Collider. That was our first funded mission and we had to deliver.

But the personal angle is that we always saw what we are doing as expanding from the desktop to the world. And that's how we went from the campus to nation-wide and beyond. We now share HTC capabilities across more than 125 institutions. And

that brought with it many complications, not only in terms of volume of users, but also in diversity of science domains, types of institutions and politics. I always listed sociology as the top obstacle to high throughput computing and we have our fair share in the Open Science Grid.

What is a future development in the field that you're looking forward to and where do you see things like quantum computing fit into it?

Powerful forces are driving computing capabilities globally and they are economical, social and political in nature. Our task is that whatever complex dynamics these forces create, it's our responsibility to bring the required computing to the researchers.

We always stayed with distributed systems. Now will distributed systems include quantum computers? Fine. But we have to resist being pulled into the debate on will quantum computing work and what application will it have and so on. We are collaborating with a group from the Large Hadron Collider's ATLAS experiment, who are looking at how to use quantum computing for physics. So we say, if you'll make it work, our job is to get you the resources you need.

Personally, I believe that some problems will run on quantum computers down the road. But will all problems run on that? I don't think so. In the same way that today some of the workload goes to GPUs and some doesn't. Our job is to give researchers something stable. If we expose researchers to the waves, we harm them.

HTCondor is available as open source software on all major operating systems (Windows, Linux, FreeBSD, etc.) What led to making it freely available?

If you believe in high throughput, then you believe in bringing maximum capacity to the researchers. If you make the software that does it subject to lawyers and to purchasing, then you are reducing the amount of capacity that is available to be shared. Therefore, open source is a natural decision. It's also a personal decision. We wanted to stay independent of commercial considerations.

We do help others to build businesses around HTCondor, the most recent example being Cycle Computing that was bought for quite a bit of money by Microsoft. We said we're happy to help you but we don't want any financial dependency. Once we become financially dependent, we lose our freedom.

So even commercial users don't have to pay anything to use HTCondor?

That's correct. HTCondor is open source and freely distributed. It's in our interest that people use it. We also earn trust if someone like Dreamworks says they're using HTCondor. Fun fact, if you've watched a movie that was made by Dreamworks since 2011, its rendering was managed by HTCondor.

Such recommendations make it less hesitant for people to use HTC systems. The other thing is that any user of the system is a resource for understanding its limitations or exposing a requirement that hadn't been considered before. So it's extremely valuable to our quest to advance the state of the art of distributed HTC. We view users as an opportunity to find bugs in the system, to find things that we don't do right. So that again goes back to the open source commitment. We want as many users as possible because they help us make HTC better.