Helping With Science From Home: Volunteer Computing

Yawa Adonsou, Jennifer Griffiths,
Darius McDaniel, Florencia Tolentino
Mentor: Mary Kathryn Cowles, Ph.D.

The University of Iowa
Iowa Summer Institute in Biostatistics (ISIB)
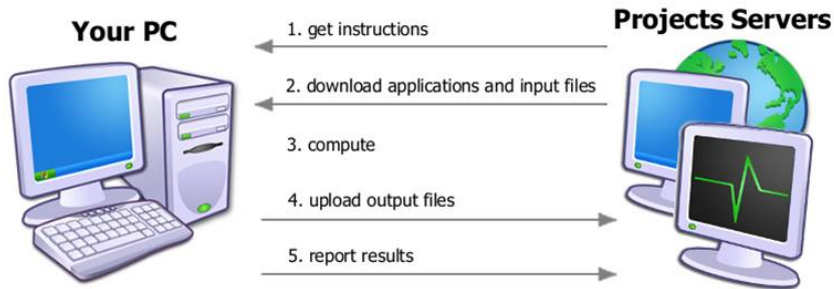
Summer 2010

Volunteer Computing

- Volunteer computing is a type of distributed computing in which computer owners donate their computing resources (such as processing power and storage) to one or more "projects".
- It uses one's computer idle time to help advance science.
- Projects are typically academic (university-based) consisting of various scientific research areas.

Volunteer Computing

- Suitable computing problems for volunteer computing:
  - Complex computing problems that can be divided into parts that can be executed independently at the same time at different volunteers' computers.
  - When the complex problem is separated amongst thousands of volunteers' computers, the time it takes to complete the thousands of computations is greatly decreased when compared to the amount of time it would take a super computer to analyze the data.

BOINC

- Berkeley Open Infrastructure for Network Computing
- Free, open source software: a "'Distributed Computing Framework"', developed by a team in Berkeley University California, led by Dr. David Anderson.
- System Requirement Link: http://boinc.berkeley.edu/wiki/System_requirement.
- BOINC software allows scientists to use the power of people's home computers to carry out scientific research.
- With BOINC, researchers get all the benefits of a virtual supercomputer for free.

How BOINC Works

Your PC → **Projects Servers**

1. get instructions
2. download applications and input files
3. compute
4. upload output files
5. report results

↪ Graphic from http://boinc.berkeley.edu/wiki/How_BOINC_works

BOINC Rules and Privacy

- Run BOINC only on computers that you own, or for which you have obtained the owner's permission.
- To participate in a BOINC-based project, you must give an address where you receive email.
- BOINC uses part of the computer's CPU power, disk space, and network bandwidth. You can control how much of your resources are used by BOINC, and when it uses them.
- On the project website, your computer's location (e.g. its domain name or network address) is not revealed and is not visible to anyone but you.

BOINC Projects

↪ Graphic from http://www.irelandboinc.com/what-is-boinc
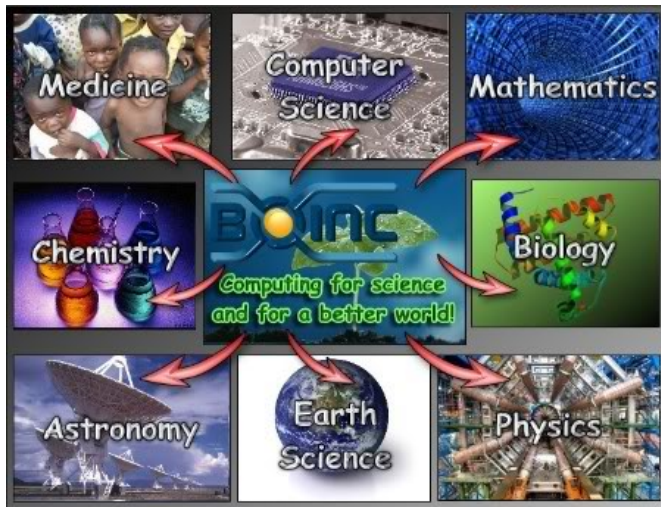
World Community Grid (WCG)

- Mission: to create the world's largest public computing grid to tackle projects that benefit humanity.
- IBM Corporation has donated the hardware, software, technical services and expertise to build the infrastructure for WCG and provides free hosting, maintenance and support.

↪ Graphic from http://www.ibm.com/podcasts/howitworks/021307/index.shtml

WCG Projects
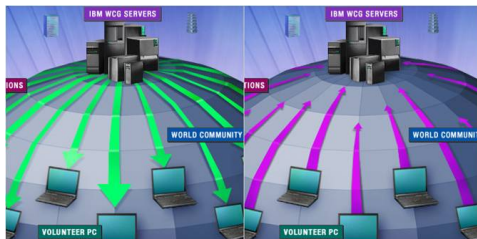
- Active Research:
  - The Clean Energy Project - Phase 2
  - Discovering Dengue Drugs - Together - Phase 2
  - Help Cure Muscular Dystrophy - Phase 2
  - Help Fight Childhood Cancer
  - Help Conquer Cancer
  - Human Proteome Folding - Phase 2
  - FightAIDS@Home
- Completed Research:
  - Influenza Antiviral Drug Search
  - Nutritious Rice for the World
  - AfricanClimate@Home
  - Genome Comparison

World Community Grid (WCG)

- Volunteers are typically members of the general public. Organizations such as schools and businesses may also volunteer the use of their computers.
- More than half a million people in 218 countries have volunteered some 1.5 million laptops and desktops.
- WCG members have volunteered more than 366,584 total years of computing time.
- Communicate and connect with people from all over the world.
  - Forums
 - Teams

World Community Grid (WCG)

- Validation of returned results
  - Redundant Computation: two copies of the work unit are sent to members computers.
  - Single Validation Type I: Only one copy of a work unit will be sent to a computer.
  - Single Validation Type II: Many copies of each work unit will be sent.
- Wingman: People running the same unit for validation.

↪ Graphic from http://advisor401k.info/grid-computing.html

Help Cure Muscular Dystrophy - Phase 2

- Muscular Dystrophy
  - Group of disorders that involve muscle weakness and loss of muscle tissue that get worse over time.
  - Treatment: no known cure for the various muscular dystrophies
    - Goal of current treatment is to control symptoms
  - Causes: many are inherited disorders

↪ Graphic from http://gallery.czechnationalteam.cz/displayimage.php?album=8&pos=27
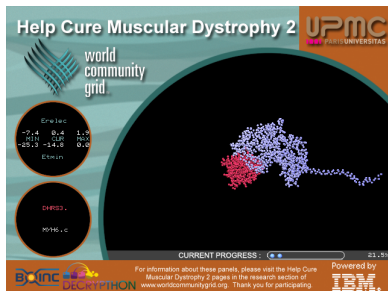
Help Cure Muscular Dystrophy - Phase 2

- Hopes to identify pairs of interacting proteins, or determine binding of protein to DNA.
- Must determine how well the surface shapes of the proteins complement each other and also measure the quality of the atomic interactions between the two molecules.
  - For complex structures like proteins it takes considerable computing time to determine the correct fit of protein-protein interactions.

Help Cure Muscular Dystrophy - Phase 2

- Phase 1: December 2006 - June 2007: looked at 168 proteins
  - Without WCG these 168 protein computations would have taken 8,000 years
- Phase 2: started May 2009: targets 2200 proteins
  - 200 of interest for neuromuscular disease
  - 200 of interest for muscular dystrophy
  - 1800 with unknown involvement in muscular dystrophy
  - Can help biomedical researchers design new strategies to neutralize these protein interactions and develop therapies for neuromuscular disease.

Help Conquer Cancer

- There are around 100 different types of cancer that can affect the human body.
- This project is based on the significant need of improving the understanding of cancer and its treatment.
- Optimize the protein structure related to cancer, using X-ray Crystallography.
- Optimization Beta testing was released to production on Thursday, March 25, 2010.

Help Conquer Cancer

- Mission:
  - Help improve and annotate unknown parts of the human proteome.
  - Enable potential pharmaceutical interventions to treat this deadly disease.

Help Conquer Cancer
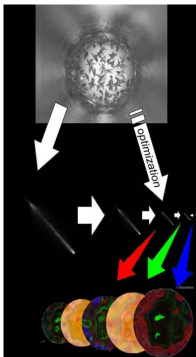
- Grid Computing:
    - X-ray detects resultant protein crystal structure at high resolution.
    - Simulating 12,375 primary features and augment them by a set of 2,533 features, creating a final set of 14,908 features.
    - Without WCG researchers will take 100,000 years to analyze the existing pictures (Computer takes 10 hrs to analyze a single figure).

Conclusion

- "Volunteers aren't paid, not because they are worthless, but because they are priceless." -Anonymous

References

- BOINC. <http://boinc.berkeley.edu/>.
- Volunteer Computer Grids. <http://www.extremetech.com/article2/0,1697,1963295,00.asp>.
- Volunteer Computing. <http://en.wikipedia.org/wiki/Volunteer_computing>.
- World Community Grid. <http://www.worldcommunitygrid.org/>.