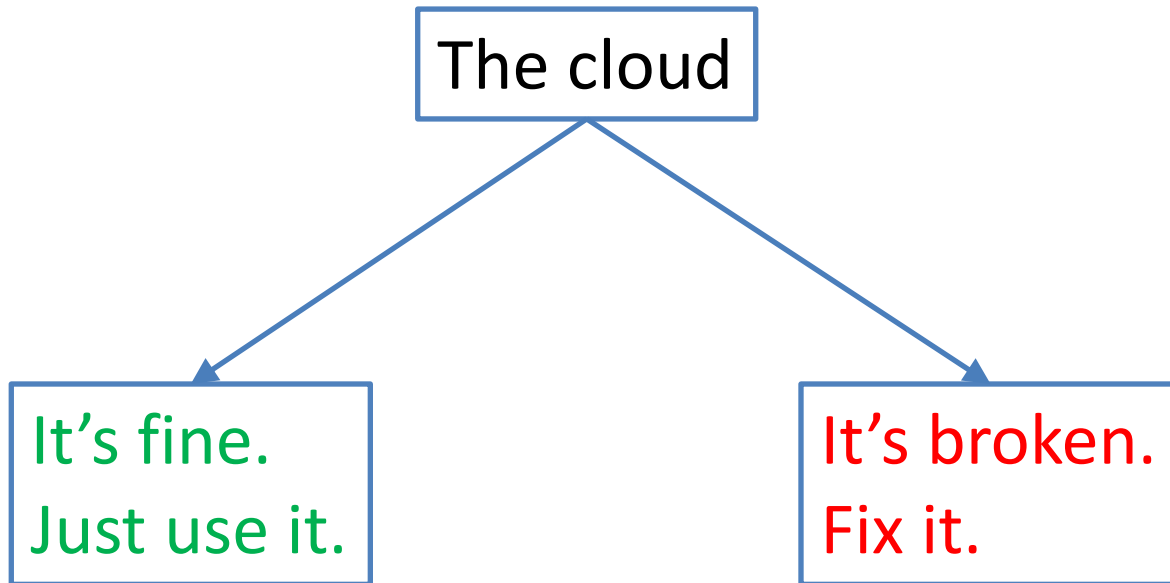# Why is the cloud a good fit for research? (How are researchers using the cloud?)

Marty Humphrey
Associate Professor
Department of Computer Science
University of Virginia
2012 Chicago Marathon: 3:39:08

The obvious (in theory)

- Seemingly infinite data storage
- Seemingly infinite data processing
- Cheap
- No late-night calls to sysadmins
- Best aspects of competition (e.g., more features, less cost, etc.)
- Hadoop
- *Don't build the infrastructure*

How are researchers using the cloud?

```
The cloud
```

It's fine.
Just use it.

It's broken.
Fix it.

National Science Foundation
WHERE DISCOVERIES BEGIN

QUICK LINKS

SEARCH

News

Email   Print

News

News From the Field

For the News Media

Special Reports

Research Overviews

NSF-Wide Investments

Speeches & Lectures

NSF Current **Newsletter**

Multimedia Gallery

News Archive

News by Research Area

Arctic & Antarctic

Press Release 11-082
The Sky Is No Limit: 13 Research Teams Compute in the Clouds

NSF awards nearly $4.5 million to innovative projects to participate in NSF / Microsoft cloud computing collaboration

1. **Cornell University (Kenneth Birman) -** Building Scalable Trust in Cloud Computing
2. **J. Craig Venter Institute, Inc. (Andrey Tovchigrechko) -** Bettering Interactive Protein-Protein Docking
3. **State University of New York (SUNY) at Buffalo (Tevfik Kosar) -** Enhancing Stork Data Scheduler for Azure
4. **University of California, San Diego (Kenneth Yocum) -** Utilizing Continuous Bulk Processing
5. **University of Colorado Boulder (Richard Han) -** Enabling Mobile Cloud Computing
6. **University of Michigan, Ann Arbor (Qiaozhu Mei) -** Refining Language Models using Web-scale Language Networks
7. **University of North Carolina at Charlotte (Zhengchang Su) -** Predicting Transcription Factor Binding Sites for Genes
8. **University of South Carolina Research Fund (Jonathan Goodall) and the University of Virginia (Marty A. Humphrey) -** Managing Large Watershed Systems
9. **University of Southern California (Viktor Prasanna) -** Tackling Large Scale Graph Problems
10. **University of Texas at Austin (Michael Walfish) -** Storing Data with Minimal Trust
11. **University of Washington (Magdalena Balazinska) -** Understanding Relational Data Markets
12. **Virginia Tech (Wuchun Feng) -** Conducting Intensive Biocomputing
13. **Virginia Tech (Kwa-Sur Tam) -** Effectively and Widely Using Renewable Energy Sources

It's broken.

(HotCloud'12, ACM Symp. on Cloud Computing'12, etc.)

- Resource management / scheduling
- I/O
- Big Data processing
- Workload characterization
- Real-time processing / predictable HPC
- Data consistency

Why is the cloud a good fit for research?

It's fine.

It's broken.

- Cheap, persistent data repository

- Cheap, instantaneous large data processing

- Costs will only decrease

- Companies will continue to innovate to survive

- Academic computer science will continue with cloud as focus