

# Networking & Information Technology Research and Development Program

Andrew W. Moore ([awm@cs.cmu.edu](mailto:awm@cs.cmu.edu))  
Carnegie Mellon, School of Computer Science  
Formerly VP Engineering, Google

---

***Good natured provocative remarks: It  
is, basically, a collision***



# A Big Data Machine Learning Problem

le high performance computing

Web News Books Images Videos More Search tools

About 117,000,000 results (0.52 seconds)

**Change the Way You Engage** - ibmmarketingcloud.com

Ad [www.ibmmarketingcloud.com/](http://www.ibmmarketingcloud.com/)

The IBM Marketing Cloud Built on Silverpop Engage. Learn More.

**High Performance Computing** - nsghcp.com

Ad [www.nsghcp.com/](http://www.nsghcp.com/)

Experts in 2-64 node HPC clusters for finance, oil/gas & engineering.  
What Is HPC - Operating Systems - Software Tools - Cluster Platforms

**Scholarly articles for high performance computing**

... transformations for **high-performance computing** - Bacon - Cited by 1074

...-scale study of failures in **high-performance computing** ... - Schroeder - Cited by 701

GPU cluster for **high performance computing** - Fan - Cited by 492

**High Performance Computing** most generally refers to the practice of aggregating **computing** power in a way that delivers much higher **performance** than one could get out of a typical desktop **computer** or workstation in order to solve large problems in science, engineering, or business.



en.wikipedia.org

**What is high performance computing?** - insideHPC

[insidehpc.com/hpc-basic-training/what-is-hpc/](http://insidehpc.com/hpc-basic-training/what-is-hpc/)

More about Supercomputer

Feedback

**What is high performance computing?** - insideHPC

[insidehpc.com/hpc-basic-training/what-is-hpc/](http://insidehpc.com/hpc-basic-training/what-is-hpc/)

High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get

# A Big Data Machine Learning Problem

Serving-time task:

$P(\text{user clicks ad} \mid \text{query} = \text{"hpc"} \text{ and context})$



le high performance computing

Web News Books Images Videos More Search tools

About 117,000,000 results (0.52 seconds)

**Change the Way You Engage - ibmmarketingcloud.com**  
Ad www.ibmmarketingcloud.com/  
The IBM Marketing Cloud Built on Silverpop Engage. Learn More.

High Performance Computing - nsgnpc.com  
www.nsgnpc.com/  
Experts in 2-64 node HPC clusters for finance, oil/gas & engineering.  
What Is HPC - Operating Systems - Software Tools - Cluster Platforms

Scholarly articles for high performance computing  
... transformations for high-performance computing - Bacon - Cited by 1074  
...-scale study of failures in high-performance computing ... - Schroeder - Cited by 701  
GPU cluster for high performance computing - Fan - Cited by 492

**High Performance Computing** most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business.

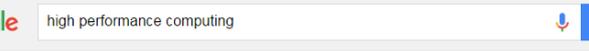
What is high performance computing? - insideHPC  
insidehpc.com/hpc-basic-training/what-is-hpc/

More about Supercomputer

Feedback



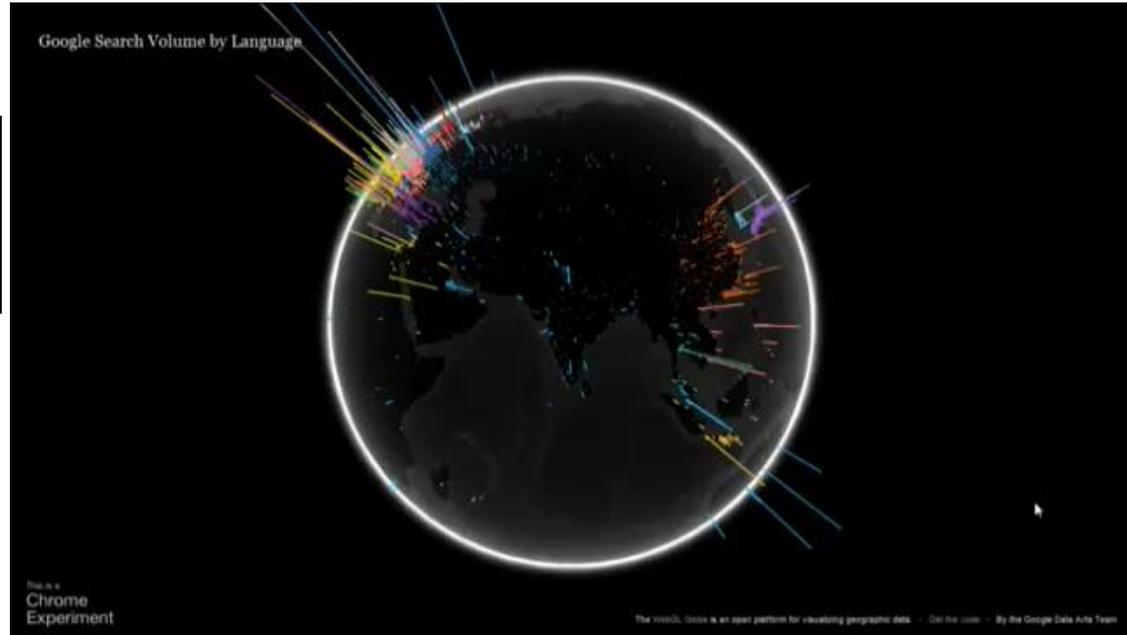
# A Big Data Machine Learning Problem



Serving-time task:

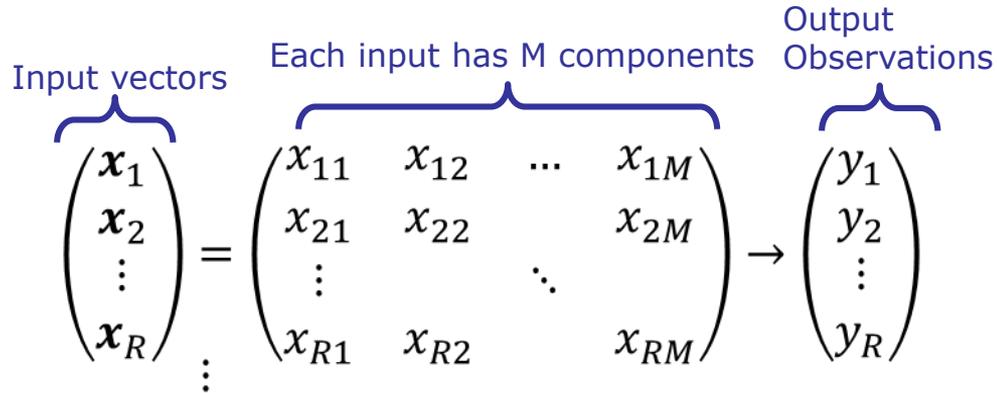
$$P(\text{user clicks ad} \mid \text{query} = \text{"hpc"} \text{ and context})$$

Batch task:  
Learn model  
from data



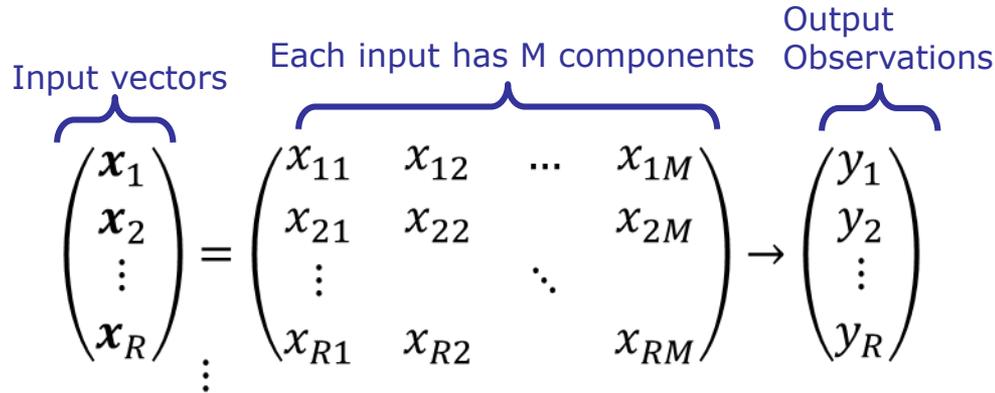
# This is Machine Learning

Training Data consists of  $R$  records:



# This is Machine Learning

Training Data consists of  $R$  records:



## Learning Task:

Find a model represented by a set of  $N$  weights

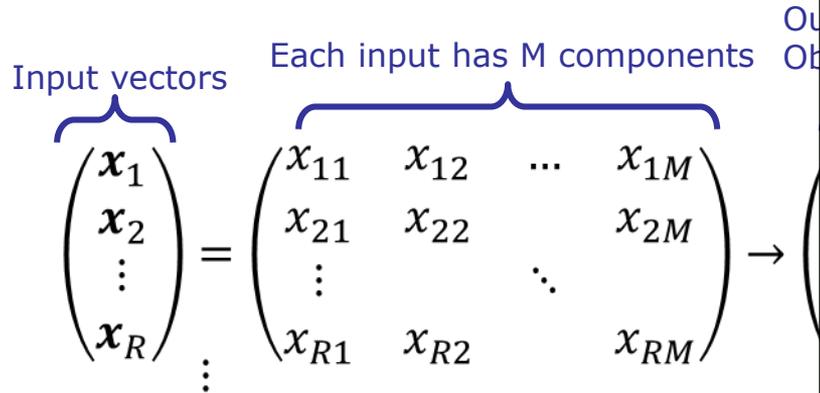
$$\mathbf{w} = w_1, w_2 \dots w_N$$

which accurately predicts

$$P(y_{new} \mid \mathbf{x}_{new}, \mathbf{w})$$

# This is Machine Learning

Training Data consists of  $R$  records:



$$\operatorname{argmax}_w P(\mathbf{w} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R, y_1, y_2, \dots, y_R)$$

## Learning Task:

Find a model represented by a set of  $N$  weights

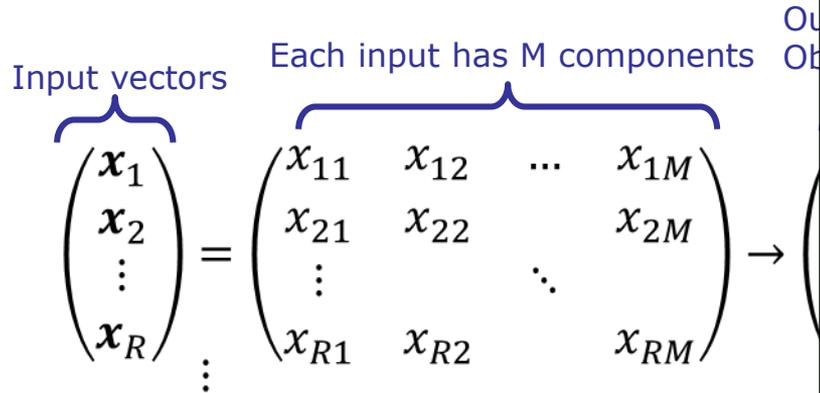
$$\mathbf{w} = w_1, w_2 \dots w_N$$

which accurately predicts

$$P(y_{new} | \mathbf{x}_{new}, \mathbf{w})$$

# This is Machine Learning

Training Data consists of  $R$  records:



## Learning Task:

Find a model represented by a set of  $N$  weights

$$\mathbf{w} = w_1, w_2 \dots w_N$$

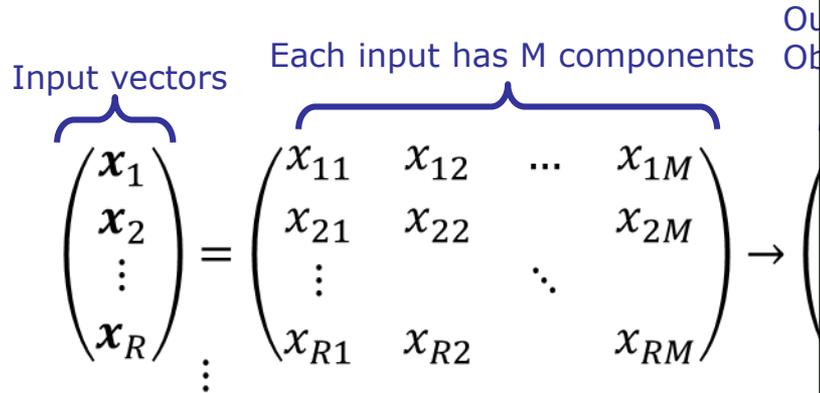
which accurately predicts

$$P(y_{new} | \mathbf{x}_{new}, \mathbf{w})$$

$$\begin{aligned} \operatorname{argmax}_{\mathbf{w}} P(\mathbf{w} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R, y_1, y_2, \dots, y_R) \\ = \\ \operatorname{argmax}_{\mathbf{w}} \log P(\mathbf{w}) + \sum_{i=1}^R \log P(y_i | \mathbf{x}_i, \mathbf{w}) \end{aligned}$$

# This is Machine Learning

Training Data consists of  $R$  records:



## Learning Task:

Find a model represented by a set of  $N$  weights

$$\mathbf{w} = w_1, w_2 \dots w_N$$

which accurately predicts

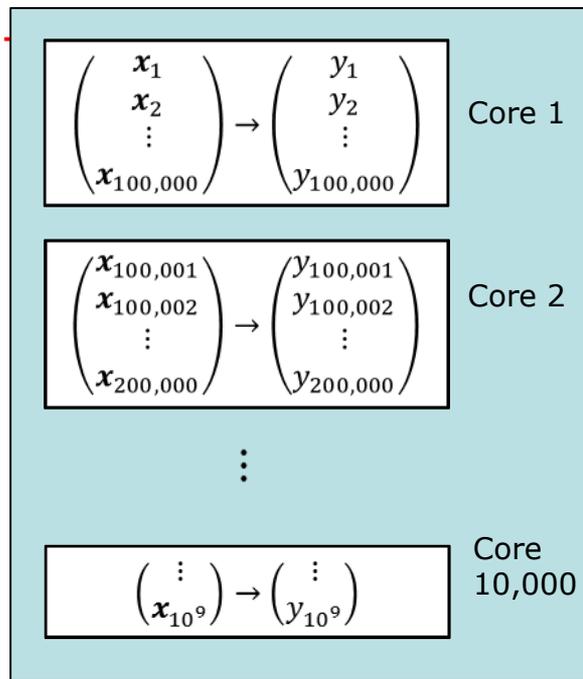
$$P(y_{new} | \mathbf{x}_{new}, \mathbf{w})$$

$$\begin{aligned} \operatorname{argmax}_{\mathbf{w}} P(\mathbf{w} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R, y_1, y_2, \dots, y_R) \\ = \\ \operatorname{argmax}_{\mathbf{w}} \log P(\mathbf{w}) + \sum_{i=1}^R \log P(y_i | \mathbf{x}_i, \mathbf{w}) \end{aligned}$$

If  $w_k$  is one of the current weights  $\mathbf{w}$  then a better guess for  $w_k$  is

$$w_k \leftarrow w_k + h \frac{\frac{\partial}{\partial w_k} P(\mathbf{w})}{P(\mathbf{w})} + h \sum_{i=1}^R \frac{\frac{\partial}{\partial w_k} P(y_i | \mathbf{x}_i, \mathbf{w})}{P(y_i | \mathbf{x}_i, \mathbf{w})}$$

# This is Machine Learning



of  $R$  records:

components Ob

$$\begin{pmatrix} x_{1M} \\ x_{2M} \\ \vdots \\ x_{RM} \end{pmatrix} \rightarrow$$

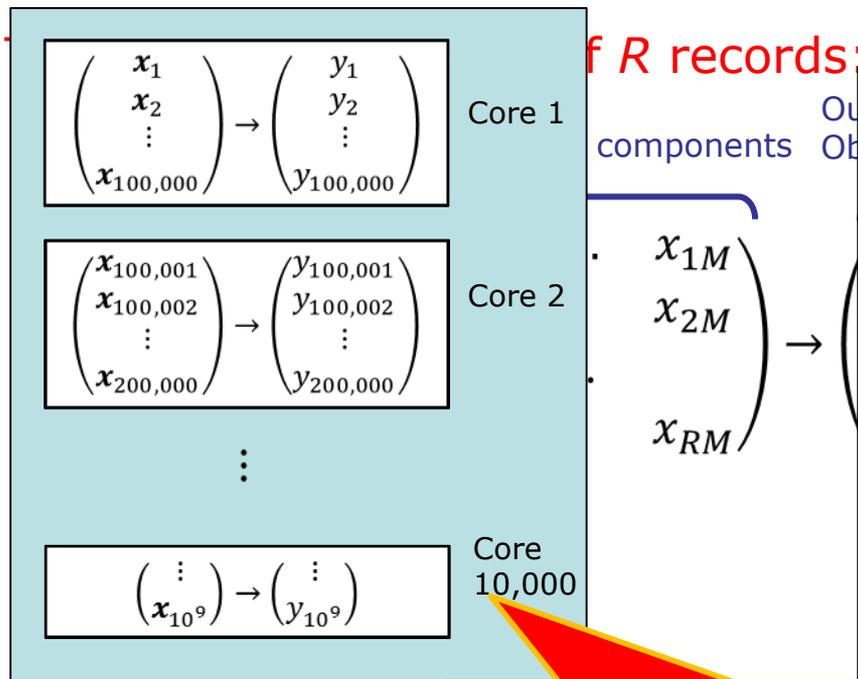
$$\begin{aligned} \operatorname{argmax}_{\mathbf{w}} P(\mathbf{w} | x_1, x_2, \dots, x_R, y_1, y_2, \dots, y_R) \\ = \\ \operatorname{argmax}_{\mathbf{w}} \log P(\mathbf{w}) + \sum_{i=1}^R \log P(y_i | x_i, \mathbf{w}) \end{aligned}$$

If  $w_k$  is one of the current weights  $\mathbf{w}$  then a better guess for  $w_k$  is

$$w_k \leftarrow w_k + h \frac{\frac{\partial}{\partial w_k} P(\mathbf{w})}{P(\mathbf{w})} + h \sum_{i=1}^R \frac{\frac{\partial}{\partial w_k} P(y_i | x_i, \mathbf{w})}{P(y_i | x_i, \mathbf{w})}$$

$\mathbf{w} = w_1, w_2 \dots w_N$   
which accurately predicts  
 $P(y_{new} | \mathbf{x}_{new}, \mathbf{w})$

# This is Machine Learning



$\mathbf{w} = w_1, w_2, \dots$   
which accurately predict  
 $P(y_{new} | \mathbf{x}_{new}, \mathbf{w})$

Very well suited to racks and racks of boring commodity servers with decent network, GPUs and flash

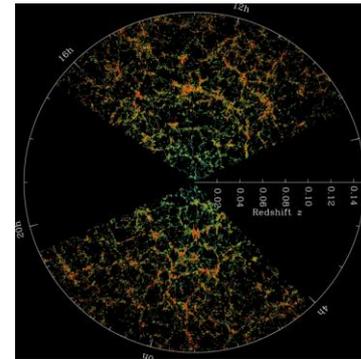
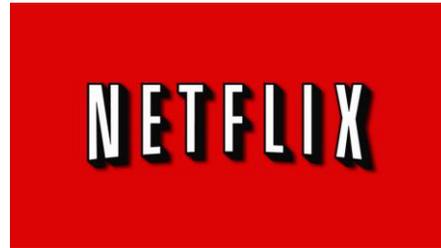
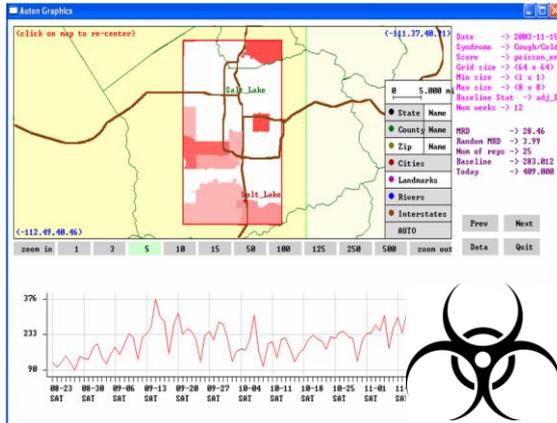
$$\begin{aligned} \operatorname{argmax}_{\mathbf{w}} P(\mathbf{w} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R, y_1, y_2, \dots, y_R) \\ = \\ \operatorname{argmax}_{\mathbf{w}} \log P(\mathbf{w}) + \sum_{i=1}^R \log P(y_i | \mathbf{x}_i, \mathbf{w}) \end{aligned}$$

If  $w_k$  is one of the current weights  $\mathbf{w}$  then a better guess for  $w_k$  is

$$w_k + h \frac{\frac{\partial}{\partial w_k} P(\mathbf{w})}{P(\mathbf{w})} + h \sum_{i=1}^R \frac{\frac{\partial}{\partial w_k} P(y_i | \mathbf{x}_i, \mathbf{w})}{P(y_i | \mathbf{x}_i, \mathbf{w})}$$

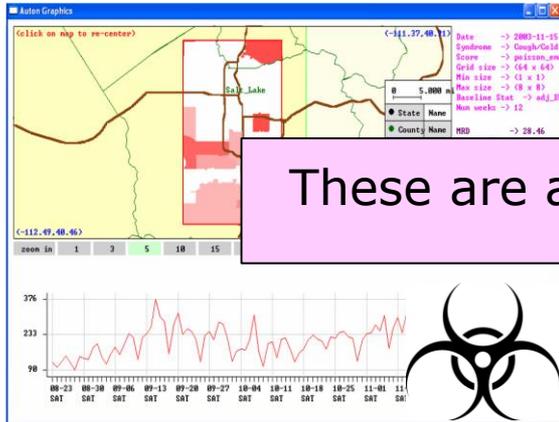
# Use Cases for $R=10^{12}$ records, $M=10^{10}$ input dimensions

## Google AdWords

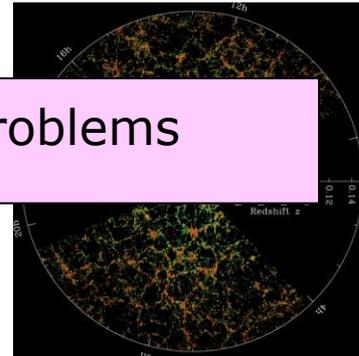


# Use Cases for $R=10^{12}$ records, $M=10^{10}$ input dimensions

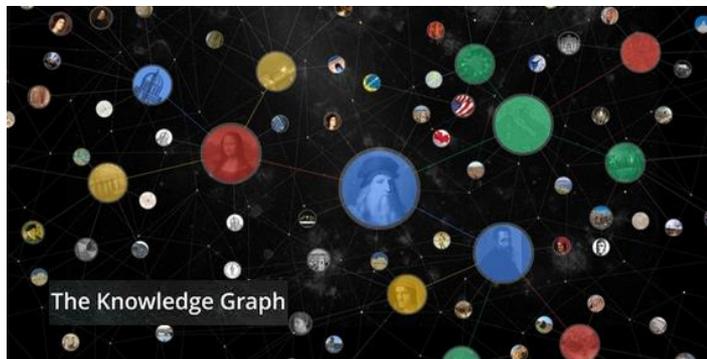
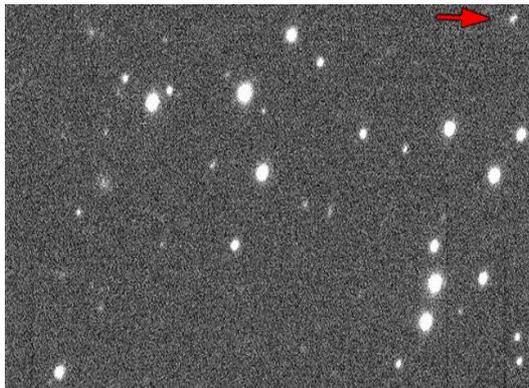
## Google AdWords



These are all  $\operatorname{argmax}_w P(w|\text{data})$  problems



# But there are some with more of a multipole flavor



# My Opinion

My gut feel on HPC and big data:

- Classic HPC is far removed from what is normally needed for big data
- But big data can use a lot of help with
  - Vector compute at nodes
  - Fast RAM cache over Flash
  - Does NOT need accurate RAM
  - Does NOT need reliable compute nodes
- Classic HPC will be much more important for the big-AI that will be built on the big-data



Questions,  
Comments, Ideas?  
[awm@cs.cmu.edu](mailto:awm@cs.cmu.edu)

