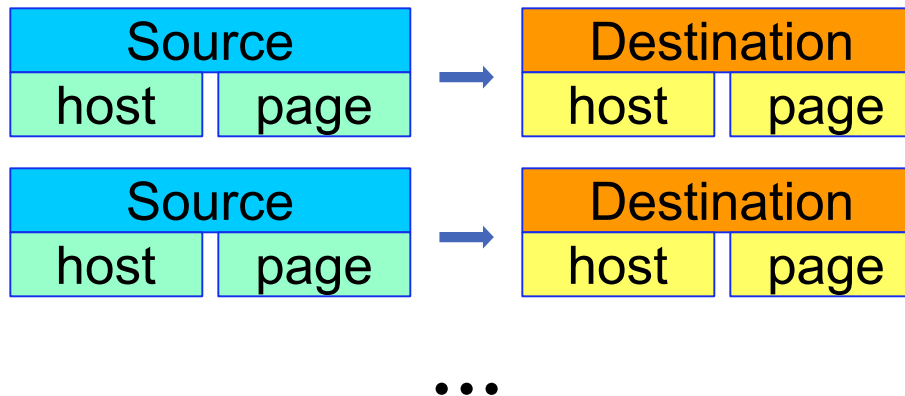# Slides for Netflix Visitor
Sandia Cloud Effort

- **Hadoop, Cassandra, other "big data" frameworks**
  - Graph algorithms
  - Machine learning
- **Commodity cluster ("Nebula")**
  - 280 cores + 280 disk drives
  - 560 TB raw disk capacity
  - 876 GB RAM
  - 2 racks

1

Sandia National Laboratories

Web Refer Graphs

- Directed graph based on HTTP Referer header and URL destination

- Count degree, triangles, SSSP, connected component, page rank, graph spectra

Source		→	Destination	
host	page		host	page

Source		→	Destination	
host	page		host	page

• • •

Sandia National Laboratories

Power Law Distribution of In-Degree

Log-Log Histogram of In-Degree, 26 weeks

Sandia
National
Laboratories