Survey of Environmental Data Portals and Reliability Issues

Cyber Infrastructure (UNLV)

Shahram Latifi
David Walker

February 2, 2010

What is a Data Portal?

- Presents information from diverse sources in a unified way
- Enables instant, reliable and secure exchange of information over the Web
- The "portal" concept is to offer a single web page that aggregates content from several systems or servers.

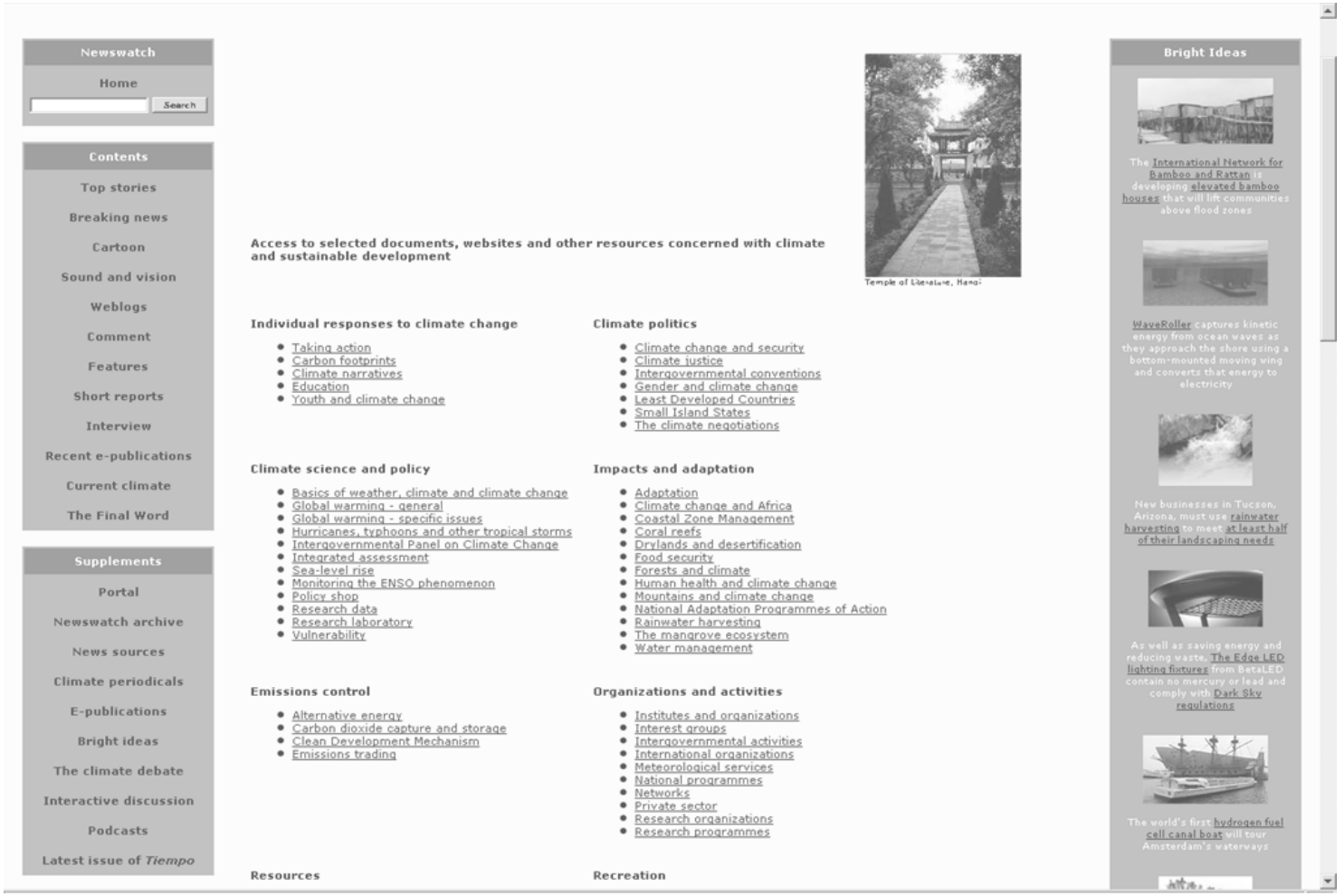
Types of Data Portals

- Commercial– contains basic information for a general audience, including current events information
- Academic – typically includes scientific data (ex. research articles, etc.)
- Also Community and Enterprise Portals

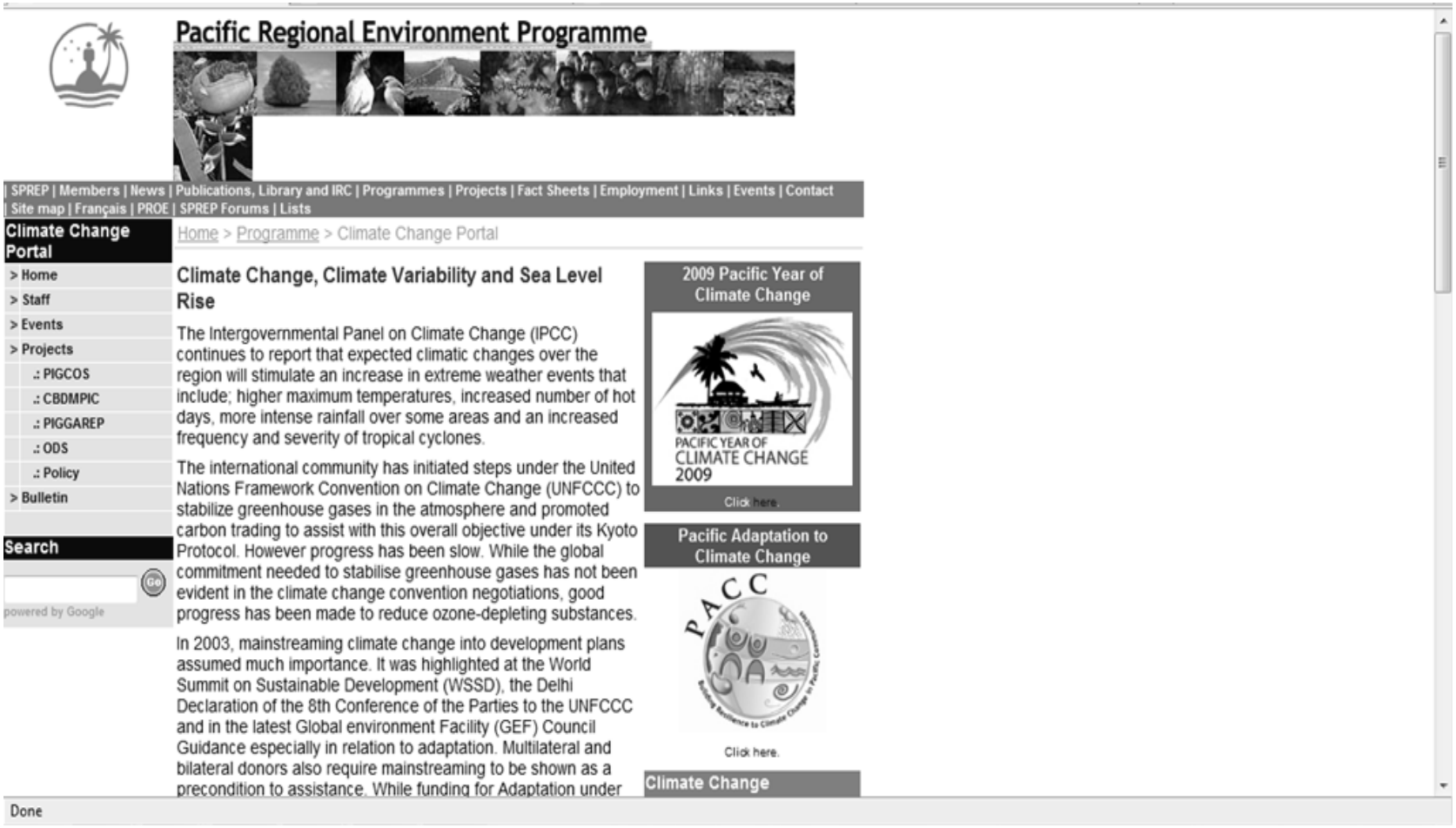
Examples of Data Portals

- California Climate Change Portal (community)
- UN Climate Change Portal (community)
- EPA Climate Change Portal (enterprise)
- Tiempo Climate Change Portal (academic)
- SPREP Climate Change Portal (academic)
- Climate Ark Portal (community)

Tiempo Climate Change Portal



SPREP Climate Change Portal



Climate Ark Portal

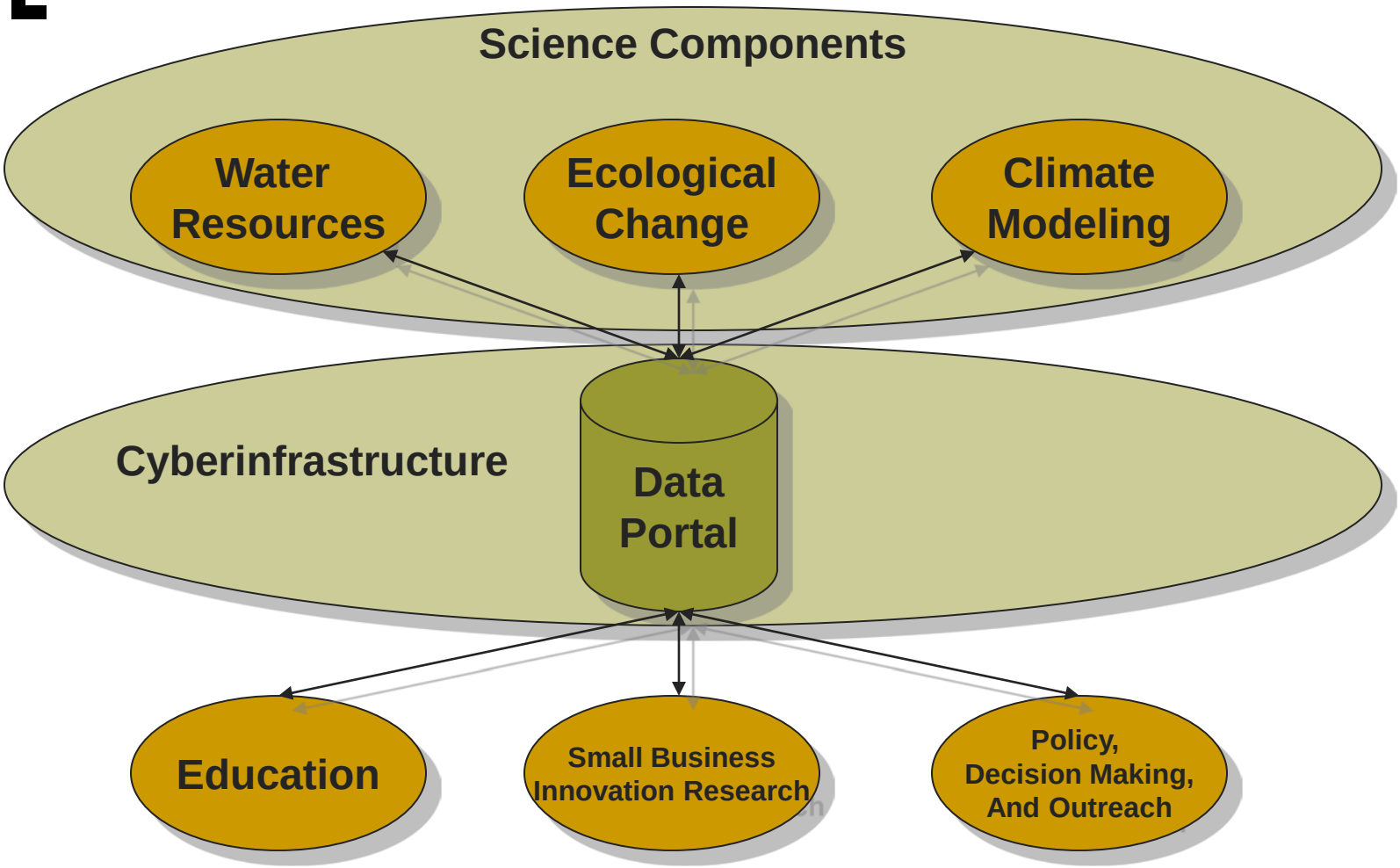
Reliability Issues

Why?

- Portal needs to be operational.
- Portal needs to be resilient to changes.
- Data must be authentic.
- New applications are integrated into the portal.

Note. Reliability vs. Availability

Priorities, Sensitive Data



How to Achieve Reliability?

Data Portal Components

- A data base
- Communication links
- A computing cluster
- Web-based application software

Redundancy (H/W, S/W, Info.)

- Redundancy at main failure points

Typically, these failure points are:

- Network Switching Devices
- Application Web Servers
- Database
- Executable Code Repository

Network Switching Device

- First entry point to the portal architecture
- Hardware Switches

  - Active/Standby Capability

Application/Web Servers

- The core processing of the portal
- Can be configured in a cluster
- Servers configured identically and run application software

Database

- Use a standby database

- All the transactions that occur on the master database are run on the standby in real-time.

Executable Code Repository

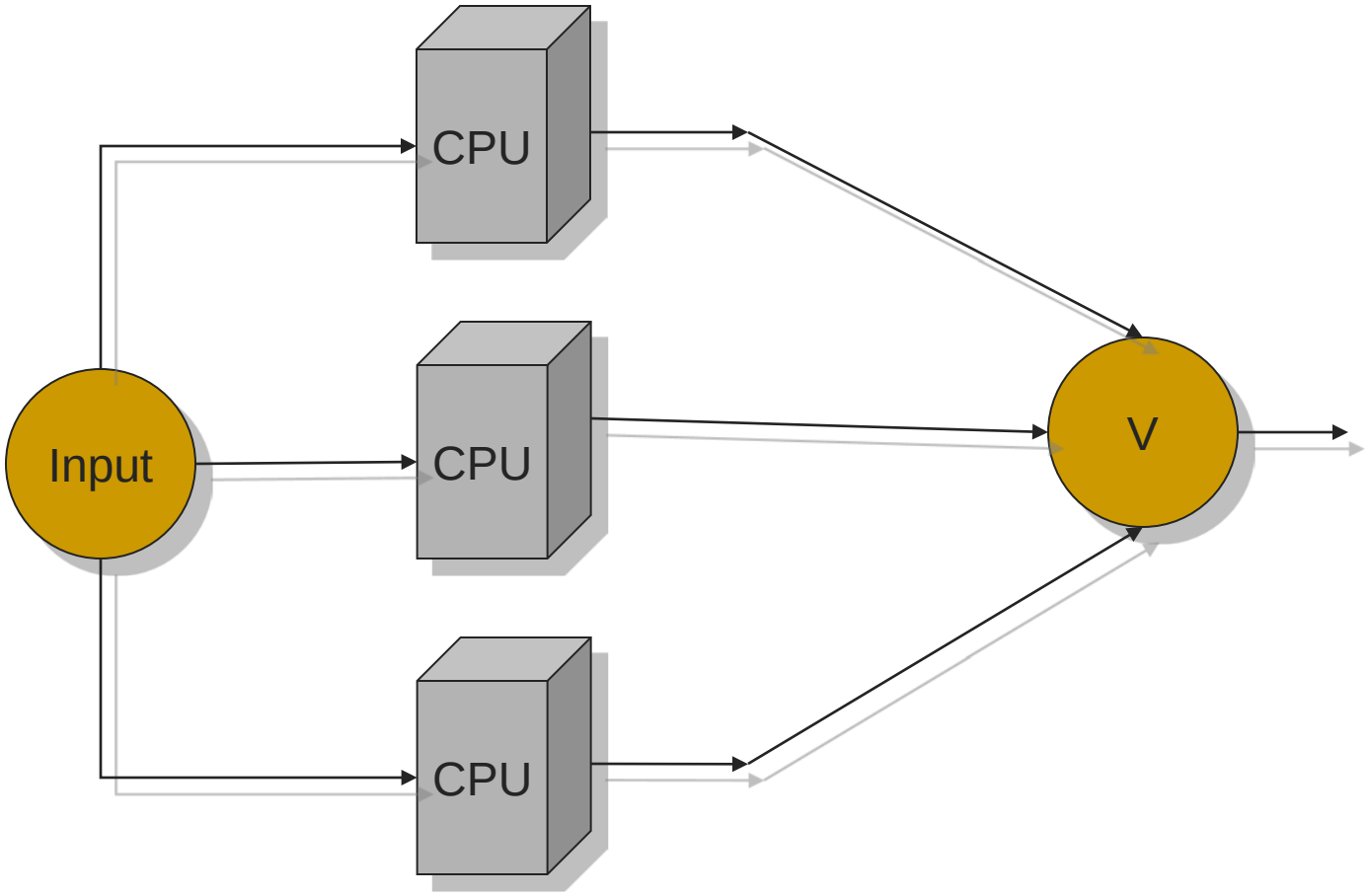
- Codes can be run on servers that are physically separated.

Caveat. A “common mode” software bug can bring the system down.

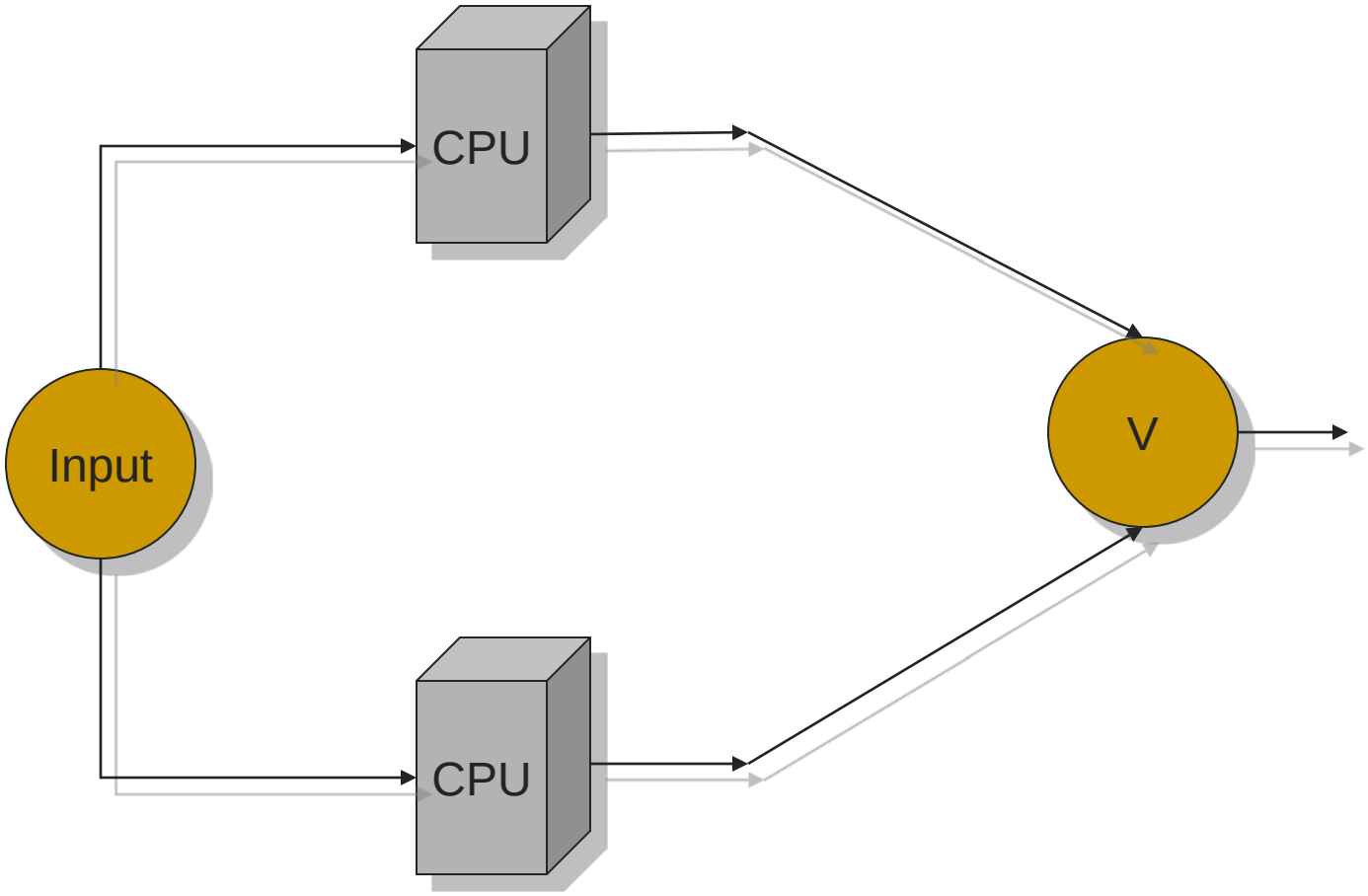
Hardware Vs. Software Redundancy

- Hardware Redundancy
- TMR (Triple Modular Redundancy)
  - System is operational as long as at least two outputs match.

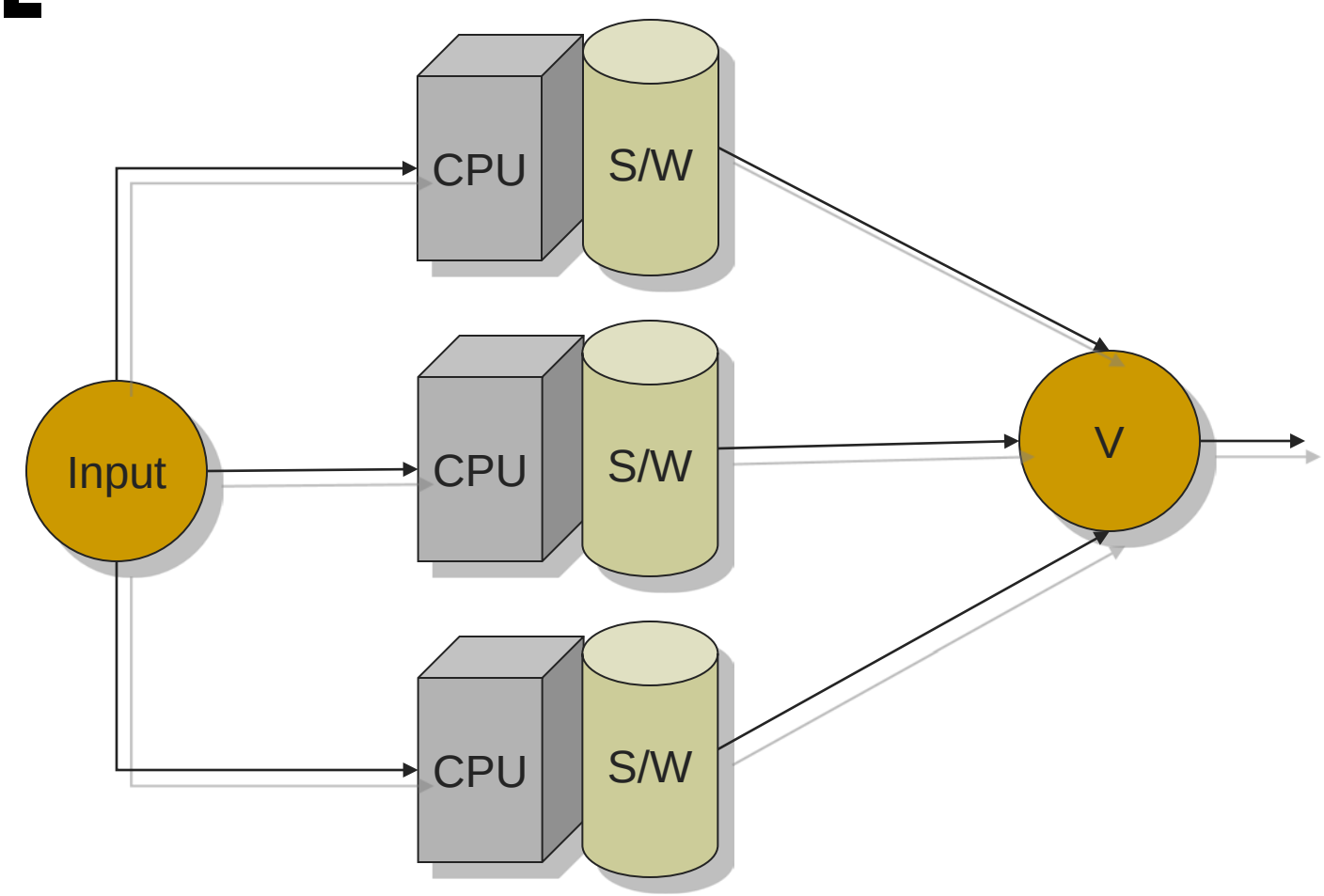
TMR H/W



TMR H/W (after one failure)



A TMR System
Does it really Work?



Software Reliability

- Fallacy 1- The outcome of a computer program is a deterministic rather than a probabilistic event.

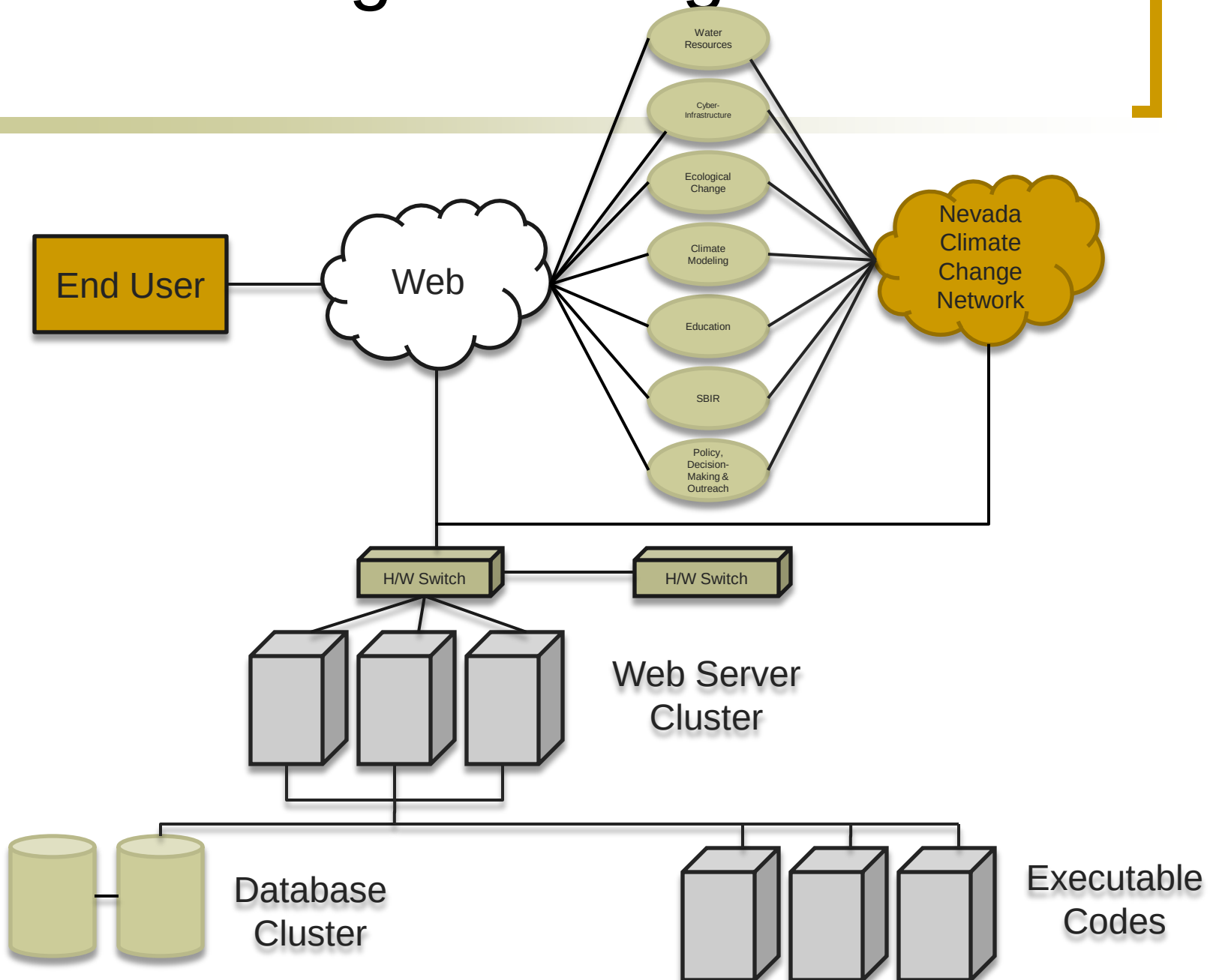
- Fallacy 2. Instructions do not wear out; software cannot fail.

Information Redundancy

- Data Integrity

- Use known Error Detection/Correction Codes

Putting it All Together



References

- Takac, L., Ligus, J., Sarnovsky, J. "Systematic Design of the Web Portals", *EAEEIE Annual Conference*, 2008, pp. 60-65.
- Soreide, N., Sun, C., Kilonsky, B., Denbo, D. Zhu. W., and Osborne, J. "A Climate Data Portal", *OCEANS IEEE Conference and Exhibition*, 2001, vol. 4, pp. 2315-2317.
- Christ, M., Krishnan, R., Nagin, D., Gunther, O., "Measuring Web Portal Utilization", *Proceedings of the 35th International Conference on System Sciences*, 2002.
- Caro, A., Calero, C., Caballero, I., Piattini, M., "Defining a Quality Model for Portal Data", *International Conference on Web Engineering*, 2006, pp. 115-116.
- Aragones, A. and Hart-Davidson, W. "Why, When, and How Do Users Customize Web Portals?", *IPCC'02*, 2002, pp. 375-388.
- Nam, J., "Web Portal Quality", *International Conference on Service Operations, Logistics, and Informatics*, 2009, pp. 163-168.
- Blanshard, L., Tyer, R., Drinkwater, D., Manandhai, A., Sufi, S., van Da, K.K., Dove, M., "Case Study: Using Web Services for the Management of Environmental Data", *Proceedings of the International Conference on Services Computing*, 2004, pp. 132-136.
- DeRose, P., Shen, W., Chen, F., Doan, A., Ramakrishnan, R., "Building Structured Web Community Portals: A Top-Down, Compositional, and Incremental Approach," *ACM Very Large Databases Conference*, 2007, pp. 399-410.

References

- Atluri, V. and Gal, A., “An Authorization Model for Temporal and Derived Data: Securing Information Portals,” *ACM Transactions on Information and System Security*, Vol. 5, No. 1, February 2002, pp. 62-94.
- Atlingovde, I., Ozcan, R., Cetintas, S., Yilmaz, H., Ulusoy, O., “An Automatic Approach to Construct Domain-Specific Web Portals,” *CIKM’07*, November 6-8 2007, pp. 849-852.
- Witten, I.H., Moffat, A., and Bell, T.C., *Managing Gigabytes: Compressing and Indexing Documents and Images*, Van Nostrand Reinhold, New York, 1994.
- Taylor, C.A., Gittens, M.S., Miranskyy, A.V., “A Case Study in Database Reliability: Component Types, Usage Profiles, and Testing”, *DBTest’08*, Vancouver, 2008, pp. 1-6.
- Smith, Michael Alan, “Portals: Toward an Application Framework for Interoperability”, *Communications of the ACM*, Vol. 47, No. 10, October 2004, pp. 93-97.
- Hadzilacos, V., “A Theory of Reliability in Database Systems”, *Journal of the Association for Computing Machinery*, Vol. 35, No. 1, January 1988, pp. 121-145.
- Corcho, O., Cima, A.L., Perez, A.G., “A Platform for the Development of Semantic Web Portals”, *International Conference of Web Engineering*, 2006, pp. 145-152.
- Priebe, T. and Pemul, G., “Towards Integrative Enterprise Knowledge Portals”, *CIKM’03*, 2003, pp. 216-223.
- Hammer, H. and Shipman, D. “Reliability Mechanisms for SDD-1: A System for Distributed Databases”, *ACM Transactions on Database Systems,* Vol. 5, No. 4, December 1980, pp. 431-466.