Data Management at Kenting’s Underwater Ecological Observatory

Ebbe Strandell, Sameer Tilak, Hsiu-Mei Chou, Yao-Tsung Wang, Fang-Pang Lin, Peter Arzberger, Tony Fountain, Tung-Yung Fan, Rong-Quen Jan, and Kwang-Tsao Shao

National Center for High-performance Computing, San Diego Supercomputer Center, Academia Sinica, National Museum of Marine Biology and Aquarium

Presented by Ebbe Strandell at ISSNIP07 in Melbourne, Australia. December 6, 2007
Outline

- Introduction
- Background
  - Ecological Observation Networks
  - Open-Source DataTurbine
  - Storage Resource Broker
- Kenting's Underwater Observatory
- Laboratory results
- Results and Conclusions
- Q&A
Introduction

- Real-time data streaming in large-scale networks presents major processing, communication and administrative challenges.
  - Large data collections.
  - Heterogeneous sensors and data streams.

Objective

- Bridge sensors, users and storage pools in scalable and decentralized manners.
- Provide real-time data and legacy data to users.

- Our approach adopts Open-Source DataTurbine and SRB technologies.
Ecological Observation Networks

- **ECO-Grid**
  - Collaboration between NCHC, Taiwan Ecological Research Network (TERN) and Taiwan Forestry Research Institute (TFRI).
  - Observatories include Fu-Shan, Yuan-Yang Lake, Ta Ta Ja, Nan Ren Mountain and Kenting.

- **Coral Reef EON (CREON)**
  - Striving to design and build marine sensor networks to understand the stresses that shape the marine world.
  - French Polynesia, USA, Australia and Taiwan.

- **Global Lake EON (GLEON)**
  - Network of lake ecology observatories, limnologists and computer scientists.
  - Australia, New Zealand, USA, Korea, Finland, Sweden, Israel, Taiwan and China.
Open-Source DataTurbine

- **Open Source (Apache License)**
  - Originally developed by Creare Inc.
  - Released as open source in collaboration with SDSC: Open Source DataTurbine Initiative (http://www.dataturbine.org)
  - Java based, with a very active development team.

- **Streaming data middleware**
  - Supports various data streams: numerical, audio, video.

- **Satisfies Critical infrastructure requirements**
  - Reliable data transport, framework for integration of heterogeneous sensors, Matlab interface.
  - Data management tools include data mirroring, routing, archiving, synchronization, and visualization.
Storage Resource Broker

- SRB is a Data Grid Management System (DGMS) developed at San Diego Supercomputer Center (SDSC).
- SRB supports shared data collections that can be distributed across multiple organizations and heterogeneous storage systems.
- Storage resources can be added dynamically.

SRB setup at NCHC in Taiwan. SRB is installed at two of our main sites. Data inserted into one storage pool is automatically replicated to the other site, thus providing both localization and disaster prevention. Current setup capacity is around 4TB.
Kenting's Underwater Observatory

- Deployed in southern Taiwan in 2004.
  - Features 10 underwater cameras setup to monitor different habitats on the coral reef.
  - Currently used by Academia Sinica and NMMBA in Taiwan for coral reef monitoring and fish behavior studies.

- On-shore video servers are used to convert analog signals to digital MJPEG video streams.
  - Remote observatory, low bandwidth.
  - Video resolution: 320x240px.
  - Effective transfer rate: 2-3 fps, 0.128-0.288mbps (estimated).
- Features (soft) real-time data streaming from the underwater cameras in Kenting.
- RDV provides TIVO-like playback of historical data.
- Data collection in growing with roughly 20GB of data per day.
Real-time Data Viewer, RDV

Available Sources

Child node

Source info

Start/Stop Realtime streaming

Playback rate

Time scale

Timeline

Server metrics
A series of tests to address scalability.

Inserted 70 HD video streams into a single child node.
- Video streams: 1280x1080px, 1fps, ~1900kbps.
- Data rate ~130mbps.

RDV clients were used to subscribe to 140 HD video streams concurrently from a single parent node.
- Outgoing data rate ~110mbps.

Data replication would allow a greater number of clients.
Conclusions

- **RBNB DataTurbine**
  - presents a compelling solution to manage streaming data
  - addresses a core set of cyberinfrastructure requirements common across several environmental observing systems initiatives.

- **The deployment in Kenting**
  - proof-of-concept: RBNB DataTurbine provides a modular and robust solution to manage streaming video data.

- **Laboratory tests**
  - indicate that RBNB DataTurbine is suitable to manage data from observatories orders of magnitude larger than the one in Kenting.

- **Future research**
  - target large scale, world-wide sensor networks of magnitudes larger than the one in Kenting.
  - integrate numerical and video data.
Thank you for listening!