DataTurbine at SDSC

Paul Hubbard
Cyberinfrastructure Lab for Environmental Observing Systems
Science R&D
SDSC/UCSD
Outline

• What is DataTurbine?
• Activities and deployments
• Current status and plans
• Open-source DataTurbine initiative
What is DataTurbine?

• DataTurbine is an open source, Java based network ring buffer for all sorts of data. You can use memory + disk for the ring, and it runs on almost any JVM.
• Started life as a NASA telemetry project
• The basic division of work looks like this:
**DataTurbine tech details**

- Ring buffers are per-source configurable with amounts of memory and disk. By using roughly 10% ratio, very large sets of data can be rapidly accessed. (Empirical result)
- Parent/child/*child routing topologies as well as simple mirrors
- Primary interface is Java API, but can also use
  - ActiveX on Win32
  - TCP/UDP proxy interface
  - WebDAV (most operating systems) via Tomcat app
  - Java proxy with arbitrary interface/protocol
- The killer app is probably the ability to navigate data TiVO-style (scan through, replay fast/slow, play backwards, etc)
- Time synchronization - server/client, NTP necessary
A more complex example

[Diagram of a complex network system with various components connected through TCP/IP and API interfaces.]
Marketing image
More about DataTurbine

- Sources can have multiple channels with varied types - numeric (e.g. sensors), video, audio, text, binary blobs.
- We have a variety of sources and sinks: In-house, from the original vendor Creare and also community contributed.
- Can also use plugins for tightly-coupled computations such as image processing.
- Runs on J2ME, J2EE and 64-bit JVM as well. Extremely scalable.
NSF, SDCI and CLEOS

- In summer 2007, the CLEOS group at SDSC won a two-year NSF award under the SDCI (Software Development for Cyberinfrastructure Improvement) to work on DataTurbine
  - Move from closed-source to community-based open source (Apache 2.0 license)
  - Create and record metrics for performance, scalability
  - Port to 64-bit Java
  - Work with various communities to encourage use and dissemination
  - User workshop
Viewing, browsing and analyzing data

• Getting data into DataTurbine is often the easy part. Once there, you need a good viewer that lets users interact with the data in ways that they find useful.

• There are many clients (sinks) as well as DataTurbine->SQL code, file writers, etc so you can use existing tools
  • Simple interfaces, import/export both lower the difficulty of using DataTurbine. We don’t want to be a one-stop-shop.
RDV is the Real-time Data Viewer, written by Jason Hanley at SUNY Buffalo for NEES. It’s plugin-based Java, handles time series, X vs Y, FFTs, audio, video, TiVO-style navigation, per-channel metadata, events and more.
More RDV
You can also view data via the web

Data, converted to graphs by the PNG plugin
Deployments and usage

• The following is an incomplete selection of various projects using DataTurbine. It’s incomplete and here to give a flavor of the broad communities who are using it right now.
CLEOS/HPWREN deployment at Santa Margerita Ecological Reserve
NEES experiments using DataTurbine

Pictures courtesy of UIUC and SUNY Buffalo

OPEN SOURCE DATA TURBINE INITIATIVE
Empowering the Scientific Community with Streaming Data Middleware

SAN DIEGO SUPERCOMPUTER CENTER, UCSD
One more slide of NEES

- Early software showing the lab at Argonne, from a viewer in Michigan
NCHC (Taiwan)

- Kenting National Park and Yuan-Yang Lake, pictures from Fang-Pang Lin and Ebbe Strandell
Insight Racing

- DARPA autonomous vehicle competition
- Insight is using DataTurbine for their vehicle video in their Lotus
- North Carolina State University, using multiple Axis 206 cameras, 30fps each
- http://www.insightracing.org/
GLEON (Northern Temperate Lakes)

- Picture is taken at Trout Lake Station, which is part of the North Temperate Lakes (NTL) Long-Term Ecological Research program (LTER) in northern Wisconsin, USA. NTL is one of the first GLEON sites. At NTL, scientists have deployed instrumented buoys in lakes to monitor key limnological variables. As seen in photo, each buoy is solar-powered and hosts a datalogger. [Photo source: Dr. Tim Kratz.]
Lake Sunapee

• New Hampshire
Lake Erken in Sweden
NASA Dryden Flight Center

- Intelligent Network Data Server (INDS)
- Fusion of DataTurbine, Google Earth and live telemetry
- Instruments flown on ER-2 (U2) and DC-8
Another NASA slide

Great plot: Google Earth allows you to pan and zoom around images, such as this slice of atmospheric data.
One more NASA
Future plans

• We have NSF (SDCI) funds to improve, extend and enhance DataTurbine over the next two years, and other funds to support a variety of deployments.

• Plan to
  • Add triggered video data (iQeye, Axis)
  • Web display
  • Collaborate as much as possible, with an eye towards building our community
Where to learn more

• Code, documentation, screenshots, developer mailing list, FAQ, Wiki and more are all available at http://dataturbine.org/
• We are very interested in developers, collaborators, users and in generally pushing the technology to new areas and capabilities.
• Thank you!