A Cray World Wide Web Document Server

Gregory Matthews, Cray Research, Inc., Eagan, MN, USA

ABSTRACT: *Cray Research software publications have been available online, through the use of the CrayDoc viewer, since UNICOS 8.0. With the customer demand for information on the World Wide Web, we are moving our documentation to the Internet for viewing with industry-standard browsers such as Netscape and Mosaic. The Cray Software Publications group document server dynamically converts existing SGML source into HTML. This talk will discuss the strategy, design, and implementation details of a Cray document server.*

Introduction

Cray customers continue to demand that our documentation be available online. Customers have indicated their preference for workstation-based viewers such as the World Wide Web (WWW) browsers and the CrayDoc browser. However, these customers overwhelmingly prefer a single interface for accessing local and Cray information. HTML is the format most requested by these customers. Given our online strategy, we are implementing a WWW document server for all of our software publications.

Customer Requirements

The requirements for moving our documentation to the WWW are clear. In 1995, the Publications department conducted an online documentation access survey of 50 Cray Research customers. The results have been used to help set both our online and printed documentation strategies. These results indicated the following trends:

- The WWW is by far the preferred method of accessing information. Web clients (browsers) were preferred more than 3 times over the CrayDoc online reader (or any other company's reader) for accessing information.
- Nearly everyone agrees that all online documentation should be accessible from the same interface.
- Hypertext Markup Language (HTML) is the format requested at most sites.
- Netscape was the most popular browser followed by Mosaic
- Documentation storage and access from workstations and terminals is preferred over storage on a Cray system.
- Line-mode access to information is still important to users and administrators. For example, UNICOS man pages, currently stored on the Cray system, are very important.

A glance through the proceedings of the Cray User Group meetings also indicates solid movement toward making information available on the WWW. Use of the WWW has been shown to reduce the number of phone/email contacts at user help desks and front-line support groups. Web information servers have been used for many things, including documentation, database applications, software updates and upgrades, and distributed applications. The use of the WWW as a powerful information tool is self-evident.

In addition to these trends, security and performance are also important to our customers. Accessing Web servers through security firewalls and outside of local area networks is a problem for some. Performance issues in accessing a centralized, public Web server prevent timely use of servers internationally. This is also true in designing a method for updating a Web server on a regular basis. That is, how can we provide a simple method of upgrading our online documentation in light of geographic and security constraints?

Design Objectives

Using these requirements and looking at our current publishing practices and documentation needs for Cray software, an initial set of design objectives were designed.

The document server design objectives are as follows:

- We must maintain our documentation strategy of publishing primarily for online and secondarily for hard copy.
- The online document system must be a WWW-based server that allows customers to use the browser of their choice, as dictated by their site requirements.

- We must ensure that the server will accommodate all Cray Research documentation. Our information database contains nearly 50,000 formatted pages of documentation, spanning operating system releases and programming environment releases.
- We must maintain the most useful features of our current online system, including a point-and-click interface, hypertext linking, an efficient search and retrieval mechanism, and graphics support.
- Availability on customer local area networks and accessibility of the server on a wide area network are both desirable. We must allow the flexibility required to integrate the server into existing customer information systems.
- A line-mode interface is still highly desirable.
- For productivity reasons, the document server needs to integrate into our current writing, publishing, and software release processes. This includes supporting online updates and just-in-time publishing.
- We would like to adhere to industry standards as much as possible. SGML is our documentation standard, and HTML is evolving into a defacto presentation standard. We don't want to invent standards.

Implementation

Upon taking these objectives and requirements into account, we have decided to base our document server upon the DynaWeb product produced by Electronic Book Technologies, Inc. (EBT). DynaWeb is commercial-grade Web server software that serves large SGML documents to Web browsers for rapid navigating and searching. DynaWeb provides important benefits to end users by enabling remote Web browsers to fully exploit CrayDoc-like functionality on the Web.

The Cray document server will have the following customer benefits:

- Serves Cray electronic books to Web clients (Mosaic, Netscape, lynx)
- Is relocatable, for customer access within LANs
- Accommodates searching across all Cray book collections
- Maintains hypertext links within and between manuals
- Collects access statistics
- Integrates with other WWW servers
- Allows restricting access to manuals, based upon permissions
- Incorporates graphics

The document server utilizes SGML-based electronic books, transforming them on-the-fly to HTML-browsable volumes. This process frees us from mapping URLs, and provides for expansion and inclusion of new manuals and updates as needed. To enable Web browsing of large books, DynaWeb automatically chunks documents and generates a table of contents (TOC). Users can navigate through the TOC hierarchy, selecting only information manageable by the Web browser. References and hypertext links currently found in our CrayDoc electronic books are converted to HTML links within the document server.

The document server allows Web browsers to exploit the full functionality provided in CrayDoc. Using capabilities such as fulltext, wildcard, proximity, and context searches, end users can quickly locate relevant information spanning multiple electronic book collections. The number of search hits are displayed alongside an interactive TOC. Only TOC entries that contain search hits are displayed to end users, greatly improving efficiency when users are searching large collections of electronic books.

Deliverables

The document server software and Cray electronic manuals will be made available via CD-ROM. No separate licensing is required with the server. The server will accompany each major release in support of the software product and, therefore, is available at no charge. If additional servers are necessary, they will be available for a fee through Cray Research.

Web clients (Mosaic, Netscape, lynx) are not supplied with the product. It is the responsibility of each customer site to ensure that a browser is available for end users as needed.

Since the server operates by converting SGML files to HTML on-the-fly, there are no HTML files supplied with the server software. Note, however, that HTML can be viewed using options in both Mosaic and Netscape.

Platform support for the document server is dictated by EBT's platform support for DynaWeb. Currently the following platforms are supported:

Vendor	Hardware	OS levels
Digital Equipment Corp. (DEC)	Alpha series	OSF/1 3.x
Hewlett Packard	HP 9000 Series	HP/UX 9.x
International Business Machines (IBM)	RS600	AIX 3.2+
Novell	—	UnixWare 2.x
Silicon Graphics, Inc.	—	IRIX 5.x
Sun Microsystems	SPARC Series	Solaris 2.x

In addition, document servers will be established at various locations. Initially, a document server supporting all of our products will be available under the CRInform system. We are also looking at having a document server that will be accessible from the Cray Research WWW home page (http://www.cray.com/) and possibly at other locations, as network performance dictates.

Schedules

The document server is currently running as a prototype in the software development environment at Cray Park in Eagan, MN. The timeline for development and delivery follows:

Design and testing	thru 5/96
Field Tests	57/96
Document server ships	3Q96

Upon the initial ship of the document server, we will stop shipping CrayDoc. The document server will supersede CrayDoc as a product. We will, however, continue to support the CrayDoc product as necessary. Again, electronic manuals will not be shipped via CrayDoc beyond the introduction of the Cray document server.

A mechanism for updating the manual set has not been determined at this time. We are looking at providing an online update method similar to our release software updates.

Other Considerations

As we continue to develop the document server design, we need your help on several key design considerations. Your input is valuable to us and we take it seriously.

- Accessibility and ease of use are important factors in making sure our online documents are useful. A cumbersome or non-intuitive interface discourages users and can compound frustrations in solving problems.
- Summaries of each manual will aid in reducing the search for a particular topic. Incorporating our publications catalog with our document server and linking these summaries will provide easy access to specific books. A simple form for ordering hard-copy manuals would be useful.
- Formatted print capabilities (Postscript output) would add functionality beyond our current capabilities and fulfill an often requested format requirement.
- The update process is unknown at this time. Your input on this would be of value to our design.
- Incorporating man pages so that all of our documentation is accessible would make for complete document access. Is this something that you would like to see? Please contact me via email or telephone as listed below

For More Information

As we progress toward the first releases of the document server we will keep you posted via the Cray Research Service Bulletin and by postings to the Publications directory on the Cray Research home page, accessible at the following URL:

http://www.cray.com/

For further information about Cray Research online documentation strategy and plans, please contact:

Greg Matthews - Software Publications
Cray Research, Inc.
655 Lone Oak Drive
Eagan, MN 55121
Email: gregor@cray.com
Phone: (1-612) 683-3814