

Mit dem Datenbankservice des RZ ins Web

Einleitung

Die steigende Informationsflut und der hohe Anspruch der Web-Benutzer an die Aktualität der im WWW verfügbaren Daten zwingen jeden Informationsanbieter, ständig nach einfachen Möglichkeiten der Speicherung und Aktualisierung seiner Informationen zu suchen. Die Lösung – eine Datenbank! Mit dieser Entscheidung entsteht eine Vielzahl von Fragen. Datenbanksysteme gibt es viele, für welches entscheide ich mich? Wie hoch ist der Aufwand, meine Daten in eine Datenbank zu übernehmen? Wie bringe ich meine Daten dann ins Web? Wo bekomme ich Unterstützung für mein Vorhaben? Was kostet das?!

Die Antwort des Rechenzentrums, der Datenbankservice! Hier wird ein leistungsstarker Datenbankserver betrieben. Die Software zur Übernahme der Daten in eine Datenbank und zur Anbindung von Datenbanken an das Web ist vorhanden. Praktische Erfahrungen liegen vor.

Der Datenbankserver des Rechenzentrums

Das Rechenzentrum der Humboldt-Universität zu Berlin betreibt einen Datenbankserver mit dem Namen „Adaptive Server Enterprise (ASE)“ in der Version 11.5. Der ASE ist ein Produkt der Firma Sybase. Er besteht aus einem leistungsstarken „Relationalen Datenbank-Management-System“ (RDBMS), das den Datenbankserver betreibt, sowie einer Reihe von Anwendungen und Bibliotheken, die auf Datenbank-Klienten betrieben werden. Der Datenbankserver versorgt über ein Netzwerk mehrere Klienten. Er bildet die Basis der Client/Server-Architektur von Sybase. Die Datenbanksprache des ASE ist SQL (Structured Query Language). Detailliertere Informationen zum Datenbankserver des RZ und dessen Nutzungsbedingungen sind auf der Homepage des Datenbankservice unter <http://www.hu-berlin.de/rz/dbsv> zu finden.

Zur Sicherung der Datenbanken wird die Software „SQL-BackTrack“ der Firma BMC verwendet.

Aufbau und Pflege einer Datenbank

Prinzipiell gibt es zwei Möglichkeiten, eine Datenbank auf dem Datenbankserver des RZ zu erstellen.

Als erstes können bestehende Datenbanken auf den ASE portiert werden. Dabei werden Datenbanken unterschiedlichster Typen (z. B. Access, dBase etc) in eine Datenbank auf dem ASE konvertiert. Bei Datenbanken, von denen nur ein geringer Teil des Datenbestandes im WWW präsentiert werden soll und deren Daten wenig verändert werden, kann es sinnvoll sein, nicht die gesamte Datenbank zu portieren. In solchen

Fällen ist es günstiger, eine neue weniger komplexe Datenbank auf dem Datenbankserver des RZ einzurichten, nur die ausgewählten Daten zu übernehmen und diese periodisch zu aktualisieren.

Existiert noch keine Datenbank, so ist eine neue zu entwerfen. Im Unterschied zur Portierung einer bestehenden Datenbank muss dann ein komplettes Datenbank-Design durchgeführt werden. Nähere Erläuterungen zum Datenbank-Design sind in [3] beschrieben.

Zur Unterstützung der Portierung und des Designs von Datenbanken wurde vom Rechenzentrum eine Lizenz des „PowerDesigner DataArchitect“ der Firma Sybase erworben. Eine Dokumentation ist unter <http://sybooks.sybase.com/pdbooks.html> zu finden.

Zum Aktualisieren und Abfragen der Datenbanken kann auf jede Software zurück gegriffen werden, die durch ein Interface zum ASE unterstützt wird. Im Rechenzentrum sind 3 Arten von Interfaces verfügbar:

- **ODBC** – (Open DataBase Connectivity) ist eine von Microsoft entwickelte Anwendungsschnittstelle (API). Sie ist standardisiert und erlaubt einer Software, auf eine Vielzahl von Datenbanken, für die ODBC-Treiber existieren, zuzugreifen. Für den ASE ist ein entsprechender ODBC-Treiber vorhanden. Er ermöglicht es zum Beispiel, eine Tabelle des ASE in eine Access-Datenbank einzubinden und mit ihr zu arbeiten. Der Vorteil – eine anspruchsvolle graphische Oberfläche zur Pflege der Datenbank.
- **JDBC** – ist eine Java-API. Sie besteht aus einer Menge von Klassen, die in der Programmiersprache Java geschrieben sind. Der JDBC-Treiber für den Zugriff auf den ASE befindet sich unter <http://www.sybase.com/products/internet/jconnect/>. Eine Java-Applikation zum Erstellen und Ausführen von SQL-Anweisungen auf dem ASE ist unter <http://www.sybase.com/products/internet/jconnect/utilities/> verfügbar. Die Software mit dem Namen „jsql“ ist ein einfach zu verwendender SQL-Editor mit einer graphischen Oberfläche.
- **Open Client/C** ist eine von Sybase entwickelte offene API. Sie ermöglicht es, von Client-Anwendungen, Werkzeugen von Drittanbietern und 3GL-Programmen auf den ASE zuzugreifen. Der Open Client/C unterstützt offene Standard APIs wie Embedded SQL/C, ODBC und IDAPI sowie die Netzwerkprotokolle DEC/net, Named Pipes, NetWare SPX/IPX, TCP/IP.

Das Sichern und Wiederherstellen von Datenbanken ist Teil des Datenbankservice des RZ. Die dazu notwendigen Arbeiten werden vom Datenbankadministrator (DBA) ausgeführt. Art und Umfang der Sicherung einer Datenbank, werden durch den Eigentümer der Datenbank (DBO) festgelegt.

Die Datenbank im Web

Die Anbindung einer Datenbank an das Web und der weltweite Zugriff auf die in ihr gespeicherten Informationen erfordern einen zuverlässigen Schutz der Daten vor unberechtigtem Zugriff. Daraus ergeben sich zwei wichtige Anforderungen an die Web-Schnittstelle. Als erstes sind die Zugriffsrechte der Benutzer in der Datenbank zu definieren. Diese Funktion wird durch das jeweilige DBMS realisiert. Der ASE verfügt über folgende Sicherheitsfunktionen:

- **Identifizierung und Authentifizierung** von Benutzern. Alle Benutzer des ASE erhalten ein eindeutiges Login-Kennzeichen. Jede Aktivität eines Benutzers kann diesem Kennzeichen zugeordnet und protokolliert (auditing) werden.
- **Aufteilung von Rollen.** Auf dem ASE können den Benutzern die Rechte zur Ausführung von operativen und administrativen Aufgaben übertragen werden. Dies geschieht durch die Definition von „Rollen“. Der ASE stellt Systemrollen wie „System Administrator“ und „Security Officer“ zur Verfügung und gestattet benutzerdefinierte Rollen, die von einem „Security Officer“ erstellt werden.
- **Discretionary Access Controls (DAC).** Sie werden vom Objekteigentümer festgelegt und sind beliebig bestimmbar. Mit dem SQL-Befehl „grant“ kann ein Objekteigentümer Benutzern, Gruppen und Rollen Berechtigungen erteilen, mit „revoke“ kann er sie entziehen.
- **Auditing.** Beim Auditing werden sicherheitsbezogene Systemaktivitäten in einem Audit-Trail aufgezeichnet, anhand dessen Systemeindringlinge und der Missbrauch von Ressourcen aufgespürt werden können. Durch Auswertung des Audit-Trails kann ein „Security Officer“ die Muster des Zugriffs auf Objekte in Datenbanken untersuchen und die Aktivität bestimmter Benutzer überwachen.

Eine zweite wichtige Aufgabe ist die Gewährleistung einer sicheren Datenübertragung zwischen dem Web-Client und dem Datenbankserver. Datenübertragungen im Internet sind nicht sonderlich sicher, praktisch kann alles mitgelesen werden. Werden vertrauliche Daten über das Internet versendet, ergeben sich erhöhte Anforderungen an die Sicherheit der Datenübertragung. Zur sicheren Kommunikation unterstützt der Web-Server des RZ www.hu-berlin.de „Secure Socket Layer“-Verbindungen (SSL). Der Unterschied zwischen normalen Seiten und solchen, die über SSL angefordert werden, liegt darin, dass zum Transport nicht das normale HTTP-Protokoll, sondern ein sicheres Übertragungsverfahren angewendet wird.

Zur Gestaltung der Web-Oberfläche steht dem Benutzer des Datenbankservice folgende Software zur Verfügung:

- **PHP** – (PHP Hypertext Preprozessor) ist eine Scriptsprache zur Gestaltung von dynamischen Webseiten. Die Syntax ist ähnlich zu C, Perl und Java erweitert durch PHP-eigene Kommandos. Eine Stärke von PHP ist seine Unterstützung für eine Vielzahl von Datenbanken. Die Arbeit mit dem ASE wird ebenfalls unterstützt. PHP wurde als Modul in den WWW-Server des RZ integriert, so dass in HTML-Text eingebettete PHP-Anweisungen interpretiert und abgearbeitet werden.
- **Java** – ist eine objektorientierte Programmiersprache. Mit Hilfe von Java und dem JDBC-API ist es möglich, Applets und Servlets zu erstellen, die mit Informationen aus einer Datenbank auf dem ASE arbeiten. Die Abarbeitung von Servlets wird durch den WWW-Server des RZ unterstützt.
- **CGI-Scripts** – Das Common Gateway Interface (CGI) ist ein Standard Web-Interface. Es unterstützt den Aufruf von externen Programmen (z. B. Perl-Scripts, C-Programme) und die zur Parameterübergabe notwendigen Mechanismen. Wenn ein Web-Server eine URL enthält, die auf ein CGI-Script verweist, wird es gestartet und das vom Programm erzeugte HTML-Dokument an den Web-Client zurückgegeben. Zur Kommunikation mit dem Datenbankserver des RZ benötigt ein CGI-Script ein eigenes Interface. Das derzeit vorhandene Interface unterstützt den Zugriff aus C-Programmen und PHP.

Gegenwärtig sind fast alle Web-Frontends, die auf den Datenbankserver des RZ zugreifen, mit PHP erstellt. Die Programmierung von Applets und Servlets wurde im Rechenzentrum bisher kaum angewendet. Eine Ursache ist die mitunter mangelhafte Performance von Applets auf Web-Clients mit schwacher Hardware. Die auf der Programmiersprache C basierenden CGI-Scripts werden gegenwärtig durch PHP-Scripts ersetzt.

Literatur

1. GRAHAM HAMILTON, RICK CATTTEL, MAYDENE FISHER: *JDBC™: Datenbankzugriff mit Java™*. Bonn, Addison-Wesley-Longman, 1998.
2. RAINER KLUTE: *Das World Wide Web: Web-Server und -Clients, HTML 2.0/3.0, HTTP*. Bonn, Addison-Wesley, 1996.
3. HERMANN SAUER: *Relationale Datenbanken: Theorie und Praxis – mit einem Beitrag zu SQL-3* von Klaus Grieger, Bonn, Addison-Wesley-Longman, 1998.
4. EGON SCHMID, CHRISTIAN CARTUS, RICHARD BLUME: *PHP – dynamische webauftritte professionell realisieren*. München, Markt und Technik, Buch und Software-Verlag, 1999.

Michael Rybczak
rybczak@rz.hu-berlin.de