

## Beschreibung und Einsatz einer objektorientierten Datenbank

# Multidimensionale Objekte

Caché ist eine postrelationale, multidimensionale Datenbank, die Objekte samt Eigenschaften speichert und neben den Objekten sogar die Geschäftsanwendung selbst erfasst. **Wolfram E. Mewes**

## Auf einen Blick

### Inhalt

Der Artikel bietet eine detaillierte Vorstellung der multidimensionalen, objektorientierten Datenbank Caché. Kurze Beispiele erläutern das Anlegen einer Objektklasse mit fünf Eigenschaften, die Abfrage von Daten sowie den Zugriff mit Microsoft Excel.

### Technik/Anwendung

Datenbank Caché von InterSystems

### Autor

Wolfram E. Mewes ist Geschäftsführer, Berater, Dozent und freier Journalist. Nach einer technisch-kaufmännischen Ausbildung studierte er Marketing. Seit 1993 hat Mewes mehrere Fachbücher bei verschiedenen Verlagen veröffentlicht. Seit 1988 schult er hauptberuflich und hat bisher circa 2.250 Seminartage gehalten.

Relationale Datenbanken sind Ihnen sicherlich ein Begriff: Es handelt sich dabei um Datenbanken, deren Tabellen eine Beziehung zueinander haben, eben eine Relation.

Die Datenbank namens Caché gehört zu den postrelationalen Datenbanken, obwohl der Hersteller InterSystems [1] diesen Begriff in jüngerer Zeit nicht mehr verwendet. Diese tragen ihren Namen deshalb, weil sie nach den relationalen Datenbanken auf den Markt kamen.

Der bekannte deutsche Informatiker Professor Dr. Gottfried Vossen verwendet den Begriff postrelational für alle Datenbanken, die nicht mehr auf sogenannten „flachen“, normalisierten Tabellen basieren.

Erforderlich wurden solche Datenbanken, um den aktuellen Anforderungen zu genügen. Sie verfügen über eine Technik jenseits der gängigen relationalen oder objektrelationalen Datenbanken. Basierend auf einer multidimensionalen Speicherung, ist Caché die einzige Datenbank, die diese Technik auch für die Transaktionsverarbeitung einsetzt. Allerdings darf man diese Technik nicht mit dem aus dem OLAP- und Data-Warehouse-Bereich bekannten multidimensionalen Datenmodell verwechseln.

Architecture (UDA) bezeichnet. Diese bietet eine einheitliche Beschreibungsschicht für Tabellen und Objekte. Der Zugang von außen, also aus Anwendungen heraus oder aus dem Web, kann über drei unterschiedliche Technologien erfolgen, wie dies auch **Bild 1** zu entnehmen ist.

Eine dieser drei Möglichkeiten stellt der objektorientierte Zugriff über XML, ActiveX, Java/EJB und SOAP dar. Die zweite Möglichkeit, der SQL-Zugriff, kann über so populäre Schnittstellen wie ODBC oder JDBC erfolgen.

Die dritte Möglichkeit ist hausgemacht: Der Zugriff erfolgt über Caché Server Pages.

Damit stehen alle Schnittstellen zur Verfügung, mit der sowohl vorhandene als auch neue Applikationen auf die Daten zugreifen können. Es spielt dabei primär keine Rolle, ob diese Anwendungen nach dem relationalen oder objektorientierten Prinzip funktionieren.

## Die Technik

Beim Entwurf einer neuen Datenbankumgebung wird der Entwickler bestrebt sein, die realen Bedingungen und Beziehungen der Daten zueinander in einem Datenbanksystem abzubilden. Für die Realisierung solcher Modelle werden Datenbanken erforderlich, die nicht nur die Voraussetzungen für eine derartige Entwicklung bieten, sondern es auch gestatten, diese Entwicklung in möglichst kurzer Zeit durchzuführen.

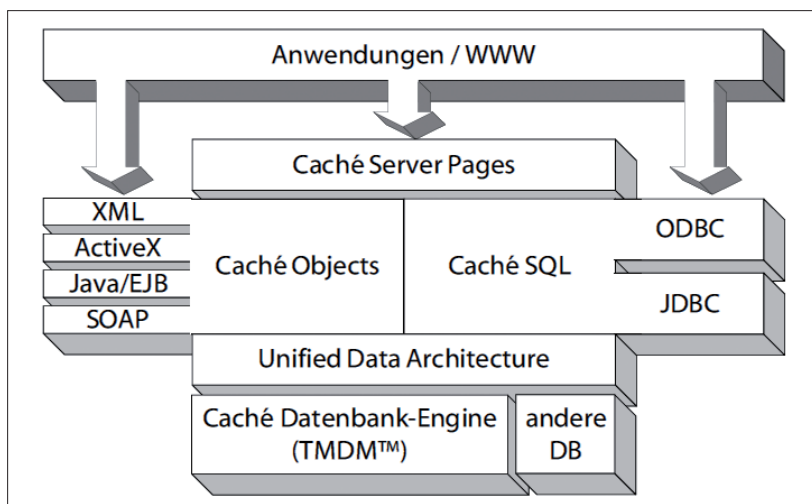
Die Datenbank Caché ist nach dem objektorientierten Prinzip aufgebaut und dadurch in der Lage, die Daten multidimensional zu speichern. Die Objektorientierung wurde entwickelt, um die zu verarbeitenden Daten anhand ihrer Eigenschaften und möglichen Aufgaben zu klassifizieren. Dadurch wird es möglich, Prozesse und Daten aus der realen Welt zu übernehmen und in der virtuellen Welt als Objekte abzubilden. Diese Objekte verfügen über Eigenschaften (Properties) und Methoden (Methods). Sie sind in der Lage, über die Schnittstellen (Eigenschaften und Methoden) mit anderen Objekten zu kommunizieren und zu interagieren. Wie sich ein Objekt verhält, wird durch die Methoden des Objekts bestimmt. Die Eigenschaften dagegen bestimmen die Inhalte, also die Daten.

Der große Vorteil hierbei ist, dass diese Schnittstellen von jedem Objekt genutzt werden

## Der Zugang

Beim Einsatz eines Datenbanksystems, sei es auch noch so gut, muss Rücksicht auf bereits bestehende Strukturen genommen werden. Im Fall von Caché hat der Hersteller dazu einen Weg eingeschlagen, den er als Unified Data

Die Systemarchitektur von Caché (**Bild 1**)



können. Eine Kenntnis der inneren Struktur des angesprochenen Objekts ist dabei nicht erforderlich. Ein weiterer Vorteil besteht in der Wiederverwendbarkeit der Objekte in unterschiedlichen Anwendungen.

Der wesentliche Unterschied zwischen der objektorientierten und der prozeduralen Vorgehensweise besteht darin, dass Strukturen, also Informationen, und nicht Abläufe beschrieben werden.

Relationale Datenbanken arbeiten tabellenorientiert. Dabei werden große Mengen von strukturierten Daten in Tabellen abgelegt. Diese Form der Ablage erlaubt es, auch riesige Datenmengen nachvollziehbar zu verwalten. Zum Auswerten der Daten kommt meist die Abfragesprache SQL (Structured Query Language) zum Einsatz. Weitere Möglichkeiten der Auswertung ergeben sich durch den Einsatz von Excel als Frontend, um mittels einer Pivot-Tabelle große Datenbestände transparent zu machen.

Der Hersteller Intersystems verfolgt mit Caché einen vollkommen anderen Ansatz. Statt die Daten in Tabellen zu speichern, werden multidimensionale Würfel- oder Baumstrukturen verwendet. Sie liefern die Basis für die Abbildung komplexer Inhalte als Objekte.

Durch die Objekte mit ihren Methoden und Eigenschaften wird es möglich, nicht nur die Daten in einer Datenbank zu speichern, wie dies bei rein relationalen Datenbanken der Fall ist, sondern auch die Geschäftsanwendung selbst abzubilden. Dies ermöglichen die Methoden, die das Objektverhalten bestimmen. Damit reduziert sich eine erforderliche 3-Schicht-Anwendung (Three Tier) auf zwei Schichten, da der Datenbankserver nun auch ohne zeitliche Verluste die Aufgaben eines Applikationsservers übernehmen kann.

### Speichern und verarbeiten

Einen herausragenden Vorteil bietet die UDA beim Umgang mit den Daten. Während die Speicherung in multidimensionalen Strukturen erfolgt, die vom Hersteller als „Sparse Arrays“ oder auch „Globals“ bezeichnet werden, kann die Betrachtung und Auswertung der Daten bei Bedarf wie bei normalen relationalen Tabellen erfolgen. In Globals lassen sich Daten mit einer beliebigen Zahl von Subscripts speichern. Die Subscripts sind nicht an einen bestimmten Datentyp gebunden, es können neben Strings auch alle anderen Datentypen auf derselben Subscript-Ebene verwendet werden.

Eine daraus resultierende Datenstruktur könnte beispielsweise so aussehen:

^Lager(Marke, Typ, Farbe, Baujahr)

Die zugehörigen Daten beziehungsweise die Abfrage auf den Datenbestand könnte dann so aussehen:

^Lager("Mercedes", "Cabrio", "Rot", 2009)=2

So lässt sich einfach feststellen, ob ein Auto, in unserem Beispiel also ein Cabrio in der gewünschten Farbe und mit dem angegebenen Baujahr, auf Lager ist. Spielen Farbe und Baujahr eine untergeordnete Rolle, so ließe sich die Suche unterhalb des entsprechenden Datenknotens erweitern:

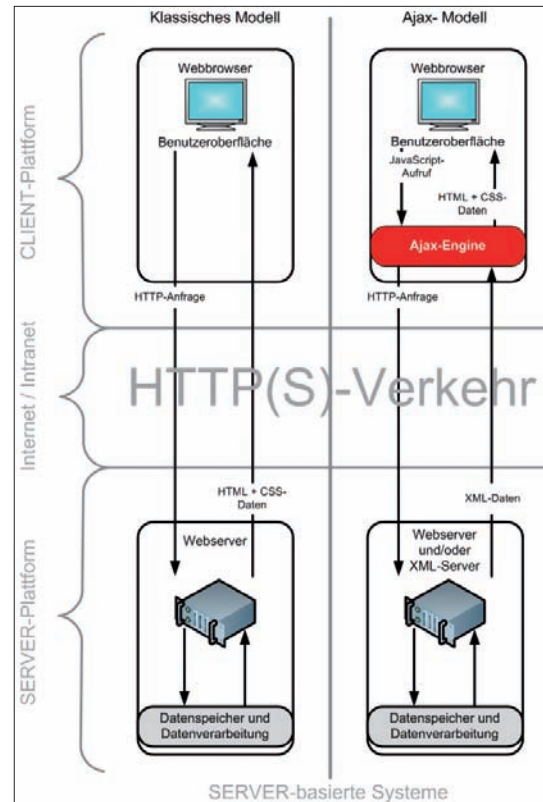
^Lager("Mercedes", "Cabrio")=5

Die Technik, Daten multidimensional zu speichern, spart einen großen Teil des Verarbeitungsaufwands, der dann anfällt, wenn Verbindungen (Joins) zwischen Tabellen hergestellt werden müssen.

Um es noch einmal deutlich zu machen: Die Daten, die in der Datenbank als Objekte gespeichert werden, verfügen über Kapselung, Mehrfachvererbung, Polymorphismus, eingebettete Objekte, Objektreferenzen, Collections, Relationships und Streams/BLOBs.

Zur Auswertung der so gespeicherten Daten können bereits vorhandene Werkzeuge und Berichtsgeneratoren herangezogen werden. Dies ist deshalb möglich, weil Caché bereits beim Anlegen einer Datenbank-Objektklasse dafür sorgt, dass im Hintergrund automatisch die SQL-Zugriffsmethoden eingerichtet werden, die zur Manipulation der Daten erforderlich sind.

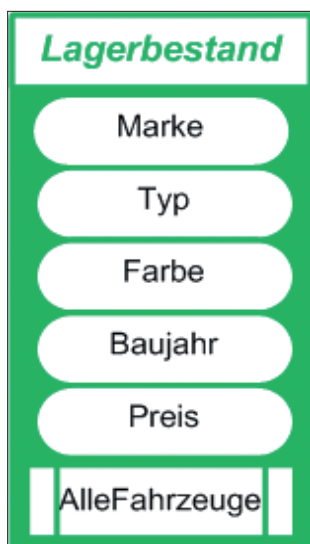
Aber dies ist keine Einbahnstraße in Hinblick auf die Auswertung. Neben der Möglichkeit, mit SQL-Werkzeugen die Daten auszuwerten zu können, lassen sich mithilfe von DDL, der Data Definition Language, die ebenfalls zu SQL ▶



Das Funktionsprinzip von AJAX (Bild 2)

Tabelle 1: Ein Beispiel für die Bitmap-Indizes von Caché

Farbe	Fahrzeug 1	Fahrzeug 2	Fahrzeug 3	Fahrzeug 4	Fahrzeug 5
Rot	1	0	0	0	0
Weiß	0	1	0	1	0
Schwarz	0	0	1	0	0
Silber	0	0	0	0	1
Blau	0	0	0	0	0



Das Schema der Objektklasse Lagerbestand (Bild 3)

gehört, bereits existierende relationale Datenbanken in Caché zu übernehmen. Dort wird dann die Objektdefinition erzeugt. Vorhandene Daten lassen sich nach Wunsch objektorientiert oder relational bearbeiten.

Ein kleines Praxisbeispiel am Ende dieses Artikels soll demonstrieren, wie man mithilfe von Excel auf eine Caché-Datenbank zugreifen kann.

### Die Sprachen

Zur Umsetzung wird hier beispielsweise die Sprache Caché ObjectScript verwendet. Weiterhin stehen für die Programmierung Caché Basic, Java, C++ und C#, die moderne .NET-Sprache von Microsoft, zur Verfügung.

Um nun mithilfe einer objektorientierten Programmiersprache wie Java oder C# aus einer Applikation auf die Informationen, also die Datenbankobjekte zugreifen zu können, bedarf es keiner Umformung mehr. Ein weiteres positives Feature ist die Tatsache, dass bei einer multidimensionalen Speicherung eine höhere Leistung bei der Verarbeitung der komplexen Datenstrukturen erreicht wird.

### Webzugriffe über Caché Server Pages

Active Server Pages (ASP) haben durch die .NET-Technik und die Verfügbarkeit schnellerer Verbindungen im Web eine Renaissance erlebt. Davon profitieren auch die Caché Server Pages (CSP), die sich mithilfe eines vorhandenen Wizards schnell und komfortabel erstellen lassen.

Das diesen Seiten zugrunde liegende Konzept sind HTML-Seiten mit speziellen Tags. Diese werden bei jedem Aufruf ausgeführt und liefern einen individuellen Inhalt zurück, eben die CSP. Integrieren lassen sich in diesen Seiten aber auch XML-Dateien, ebenso wie Bilder und andere Binärdateien.

Der Vorgang der Erstellung ist nicht schwierig, denn der Wizard verwendet eine vorhandene Objektklasse als Basis und generiert daraus die dann bereits lauffähige CSP. Diese muss aber nicht das Endprodukt darstellen, mittels entsprechender Werkzeuge lässt sich diese Seite den Kundenwünschen noch weiter anpassen.

Den Ablauf, wie sich mithilfe des Wizards eine einfache CSP erstellen lässt, zeigt der kleine Workshop am Ende des Artikels.

Das ist auch deshalb relativ problemlos möglich, da Intersystems ein Plug-in für das Produkt Dreamweaver der Firma Macromedia anbietet. Damit lassen sich Caché-Klassen in Dreamweaver einbinden.

Die Weiterentwicklung der CSP-Seiten ist allerdings nicht auf Dreamweaver begrenzt, es

lässt sich auch jede andere Web-Entwicklungsumgebung einsetzen.

Eine neue Technologie beim Umgang mit Webseiten ist AJAX. Das Akronym steht für Asynchronous JavaScript and XML. Bezeichnet wird damit ein Konzept der asynchronen Datenübertragung zwischen einem Browser und einem Server. Dieses Konzept ermöglicht es, innerhalb einer HTML-Seite eine Http-Anfrage durchzuführen, ohne dass die Seite komplett neu geladen werden muss. Nachgeladen werden nur bestimmte Teile, meist die Variablen, also die Nutzdaten (Bild 2).

Aufbauend auf CSP wird nun auch das erweiterbare Framework ZEN angeboten, das als Erweiterung der AJAX-Technologie gedacht ist. Damit lassen sich die Komplexitäten datenintensiver Web-Anwendungen reduzieren.

ZEN beinhaltet eine umfangreiche Bibliothek aus vorgefertigten Komponenten, Tabellen und Auswahlbäumen. Dies erlaubt neben einer schnelleren Entwicklung auch ein einfacheres Management komplexer Prozesse.

Tabelle 2: Klasse Lagerbestand

Name	Typ	Merkmale
Marke	%String	Erforderlich Indiziert
Typ	%String	Erforderlich Indiziert
Farbe	%String	Erforderlich
Baujahr	%Integer	Erforderlich
Preis	%Currency	Erforderlich

### Schneller Zugriff – schnelle Ergebnisse

Nachstehend soll eines der Geheimnisse des schnellen Zugriffs betrachtet werden. Bei relationalen Datenbanken ist es teilweise erforderlich, diese mehrfach zu indizieren oder Felder zu kombinieren.

Je größer die Vielfalt bei der Auswertung wird, umso stärker erhöht sich die Anzahl der Indizes und desto größer wird auch der Aufwand für die Verwaltung. Unter Indizes versteht man eine interne Liste von Werten, die die physisch geordneten Datensätze (Reihenfolge der Eingabe) in der gewünschten logischen Reihenfolge anordnen, ohne deren physische Position zu verändern.

Caché beschreitet hier einen anderen Weg. Mithilfe von transaktionalen Bitmap-Indizes gibt es für jeden möglichen Wert einen Bitstring, wobei das einzelne Bit die spezielle Ausprägung eines Datensatzmerkmals repräsentiert. Einfacher lässt sich dies anhand eines Beispiels erklären. Tabelle 1 zeigt die Farbe von Fahrzeugen. Besitzt ein Fahrzeug eine bestimmte Farbe, so erhält es für diese Ausprägung den Wert 1 (wahr), sonst 0 (falsch).

Eine neue Eigenschaft anlegen (Bild 4)



Basierend auf diesem Prinzip lassen sich auch komplexe Abfragen in einer Folge binärer AND- und OR-Operationen auf die Indizes ausführen. Und das sogar, ohne dass die entsprechenden Datensätze überhaupt geöffnet werden müssen. Es liegt auf der Hand, dass dadurch die Antwortzeiten sehr kurz werden.

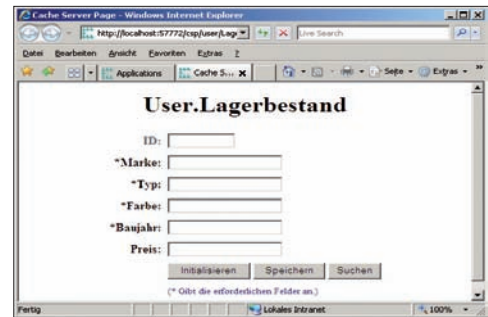
Einen Nachteil, der sich daraus unter Umständen ergeben kann, gleicht Caché auch sofort wieder aus. Da die Bitmaps sehr groß werden können, dauert die Aktualisierung von Datensätzen in anderen Datenbanken oft sehr lange. Aus diesem Grund sind die transaktionalen Bitmap-Indizes von Caché als platzsparende mul-

tidimensionale Datenstrukturen implementiert. Durch verschiedene Optimierungen im Performance-Bereich lässt sich diese Technik sogar bei zeitkritischen transaktionsverarbeitenden Anwendungen einsetzen, bei denen häufig Datenbank-Updates durchgeführt werden.

[bl]

[1] Intersystems, Hersteller von Caché;  
www.intersystems.de

[2] Kostenlose Single-User-Lizenz von Caché-Lizenz;  
www.intersystems.de/cache/downloads/index.html



Die erfolgreich angelegte Webseite (Bild 5)

## Ein einfaches Praxisbeispiel: Anlegen einer Objektklasse und der Zugriff auf Excel

Da das Anlegen einer multidimensionalen Datenbank so gänzlich anders als bei einer relationalen Datenbank funktioniert, sollen die nachstehenden Schritte ein erstes Gefühl dafür vermitteln.

Damit Sie dieses Beispiel auch nachvollziehen können, laden und installieren Sie die kostenlose Single-User-Lizenz von Caché nach einer kurzen Registrierung unter [2].

Erstellt werden soll eine Objektklasse *Lagerbestand*. Mit dieser sollen Lagerdaten von Autos verwaltet werden. Die Eigenschaften dieser Klasse sind *Marke*, *Typ*, *Farbe*, *Baujahr* und *Preis* (Bild 3). Die erforderlichen Methoden zum Speichern und Wiederfinden werden von Caché erstellt.

### Anlegen der Objektklasse

- Starten Sie *Studio*, nachdem Sie auf den Caché-Würfel geklickt haben.
- Erstellen Sie eine Caché-Klassendefinition mithilfe des Menübefehls *Datei | Neu*.
- Der Assistent führt Sie nun durch die weiteren Punkte.
- Vergeben Sie den Klassennamen *Lagerbestand*. Die Beschreibung der Klasse ist optional.
- Nachdem Sie auf *Weiter* geklickt haben, haben Sie die Möglichkeit, den Klassentyp zu bestimmen. Dieser sollte dauerhaft, also *persistent* sein.
- Klicken Sie dann auf *Fertig stellen*, wird die Klasse erstellt.

Jetzt fehlen noch die Eigenschaften: Klicken Sie dazu auf das Symbol *Neue Eigenschaft* (Bild 4) und erstellen Sie die in **Tabelle 2** gezeigten Eigenschaften. Danach sieht die fertige Klasse aus, wie es das folgende Listing zeigt.

```

/// Verwaltung des Lagerbestands
Class User.Lagerbestand Extends %Persistent
{
/// Marke des Pkw
Property Marke As %String [ Required ];
Index MarkeIndex On Marke;

```

```

/// Limousine, Coupé, Cabrio, Kombi,
PickUp
Property Typ As %String [ Required ];
Index TypIndex On Typ;

Property Farbe As %String [ Required ];

Property Baujahr As %Integer [ Required ];

Property Preis As %Currency ;
}

```

Nun soll noch eine Abfrage erstellt werden, die *alle Fahrzeuge* anzeigt. Das Symbol hierzu finden Sie in derselben Symbolleiste, in der sich auch das Eigenschaften-Symbol befindet.

- Vergeben Sie für die Abfrage den Namen *Alle Fahrzeuge*.
- Fügen Sie alle Spalten in der Reihenfolge der Erstellung hinzu.
- Die Abfrage soll die Daten aufsteigend sortiert nach *Marke* und *Typ* ausgeben. Diese Einstellung können Sie im nächsten Fenster vornehmen.
- Klicken Sie anschließend auf *Fertig stellen*, und die Abfrage wird erstellt:

```

/// Alle Fahrzeuge auswählen
Query AlleFahrzeuge() As
%SQLQuery(CONTAINID = 1)
(
SELECT %ID, Marke, Typ, Farbe,
Baujahr, Preis FROM Lagerbestand
ORDER BY Marke, Typ
)

```

Die Klasse ist nun fertig und muss nur noch kompiliert werden. Das klappt mit [Strg]+[F7] oder einem Klick auf die Schaltfläche *Kompilieren*.

Nun folgt der dritte Akt, die Erstellung einer Caché Server Page. Auch hierbei begleitet Sie wieder ein Assistent.

- Markieren Sie im Fenster *Arbeitsbereich* den Knoten *CSP-Dateien* und rufen Sie mithilfe der rechten Maustaste den Befehl *Neue CSP-Datei erstellen* auf.

- Markieren Sie im erscheinenden HTML-Fenster die Passage *My page body*.

- Rufen Sie dann aus der Menüleiste den Befehl *Einfügen | Formular-Assistent* auf.

- Wählen Sie als Klasse *Lagerbestand*, fügen Sie alle Eigenschaften in der Reihenfolge hinzu, in der Sie sie eingegeben haben, und klicken Sie dann auf *Fertig stellen*.

- Nun muss diese Seite noch gespeichert werden. Verwenden Sie als Namen *Lagerbestand.csp*. Damit haben Sie es geschafft. Um die Seite aufzurufen, müssen Sie diese im Arbeitsbereich markieren und den Menübefehl *Ansicht | Webseite aufrufen* (Bild 5).

### Ein Zugriff mit Excel

Wollen Sie nun die Daten mit einer externen Applikation wie Excel 2007 auswerten, so können Sie auf Caché wie auf eine „normale“ relationale Datenbank zugreifen.

- Wählen Sie in *Excel Daten | Externe Daten abrufen | Aus anderen Quellen | Von Microsoft Query*.
- Verwenden Sie *CACHÉ User* und bestätigen Sie dann mit OK.

Nun können Sie die Tabellen hinzufügen, allerdings ist die Auswahl bei nur einer vorhandenen Tabelle mit dem Namen *Lagerbestand* eingeschränkt.

- Fügen Sie nun alle Feldnamen hinzu. Bei Bedarf können Sie diese Abfrage auch speichern. Die Vorgehensweise unterscheidet sich hier nicht von der Verwendung mit Access, SQL Server oder anderen Datenbanken.

- Klicken Sie nun auf *Datei | Daten an Microsoft Office Excel zurückgeben*. Hier haben Sie die Wahl, ob die Daten direkt in eine Excel-Tabelle übernommen werden oder die Basis für eine Auswertung in einer Pivot-Tabelle bilden sollen.