

Historisierung mit Flashback Database Archive (FDA)

Wolfgang Tanzer
metafinanz Informationssysteme GmbH
München

Schlüsselworte

Oracle 11g, Total Recall, Flashback Data Archive, Historisierung, Datawarehouse,

Einleitung

Im ersten Teil des Vortrages wird die Datenbankoption Flashback Database Archive erklärt. Im zweiten Teil wird ein Proof of Concept erläutert, mit dem untersucht werden sollte, ob man FDA in einem vorhandenem Data Warehouse als Historisierungsmöglichkeit einsetzen kann.

In jedem Unternehmen müssen Daten gespeichert und eventuell lange aufbewahrt werden. Dauer und Anforderung hängen teilweise von gesetzlichen, externen (Wirtschaftsprüfern) und internen Anforderungen ab.

Es muss sichergestellt werden, dass Änderungen an Daten jederzeit nachvollzogen werden können. Des Weiteren muss gewährleistet sein, dass historisierte Daten nicht nachträglich manipulierbar sind.

In vielen Fällen übernehmen Anwendungen, welche die Daten produzieren auch die Verwaltung und Historisierung der Änderungen an den Datensätzen.

Die Oracle Datenbankoption Flashback Database Archive, auch unter den Namen Total Recall bekannt, kann in vielen Fällen diese Probleme lösen.

Mit dem Einsatz des Flashback Database Archive (FDA) werden alle Änderungen an einer Tabelle automatisch versioniert und archiviert. Jede Änderung eines Datensatz kann dadurch jederzeit nachvollzogen und abgerufen werden. Diese Versionen können nicht mehr verändert werden und sind daher sogar Revisionsicher. Auch gelöschte oder umbenannte Spalten bleiben im Archiv erhalten. Dies bedeutet, dass selbst Änderungen an der Tabellenstruktur und nicht nur an den Datensätzen archiviert werden.

Die historisierten Daten sind über einen speziellen Parameter im Abfrage Statement abrufbar. Für die Anwendungen, die diese Datenbank nutzen (Reporting, operative Systeme), ändert sich erst einmal nichts. Sie können weiterhin auf vorhandene Daten zugreifen und in die Tabellen schreiben ohne geändert werden zu müssen. Sollen historisierte Daten abgefragt werden, ist dies jederzeit ohne Probleme und große Aufwände möglich, in dem nur kleinere Änderungen an den Abfragestatements gemacht werden müssen.

Im Data Warehouse Umfeld kennt man für einen gewissen Teil solcher Anforderung die Möglichkeit der sogenannten „Slowly Changing Dimension – Typ 2“ (SCD2). Mit dieser Methode wird jede Änderung an einem Dimensionsdatensatz historisiert und es entstehen mehrere Versionen (ein aktueller und mehrere historisierte). Wird dies nachträglich eingeführt, werden hierbei aber auch eine Änderungen an den Abfragen dieser Tabellen notwendig.

In dem Vortrag soll darauf eingegangen werden, ob Flashback Database Archive (FDA) in einem Data Warehouse (DWH) anstatt SCD2 eingesetzt werden kann.

In dem gegebenen Fall, ist das DWH bereits seit Jahren vorhanden, wurde jedoch ohne Historisierung geplant und umgesetzt. Mittlerweile haben sich die Anforderungen (gesetzlich, Wirtschaftsprüfer) verändert, so dass eine Historisierung notwendig ist.

Um zu Prüfen, ob FDA eingesetzt werden kann, wurde ein Proof of Concept aufgesetzt. Dabei wurde das Verhalten von FDA in diesem produktiven DWH untersucht. Es wurden verschiedene Kriterien, wie zum Beispiel ETL Performanz, Speicherbedarf, Abfragegeschwindigkeit von Berichten und der Aufwand für die Umsetzung betrachtet.

Funktionsweise Flashback Data Archive (FDA)

Voraussetzungen:

Um FDA in der Datenbank verwenden zu können, muss die entsprechende Lizenz vorhanden sein. Des Weiteren müssen ein paar Vorbereitungen getroffen werden.

Für FDA werden Archive angelegt, in denen die versionierten Datensätze gespeichert werden. Um diese Archive anlegen und verwenden zu können, muss das Undo Management der Datenbank auf „AUTO“ konfiguriert sein.

Zudem müssen die Archive in separaten Tablespaces abgelegt werden. Ein Archiv kann sich allerdings über mehrere Tablespaces ausdehnen.

Ein Archiv kann für mehrere Tabellen oder auch nur für eine Tabelle verwendet werden. Zur Besseren Wartbarkeit empfiehlt sich ein Archiv für zusammengehörige Tabellen (z.B. alle Tabellen eines Datamart in ein Archiv) zu verwenden.

Die Tablespaces für diese Archive müssen über das Automatic Segment Space Management (ASSM) verwaltet werden. Es muss außerdem natürlich genügend Speicher für die versionierten Datensätze bereitgestellt werden. Je nach Architektur der ETL Prozesse in der Datenbank kann dieser Speicher sehr schnell wachsen.

Die mit FDA erzeugten Archive sind neue Objekte in der 11g Datenbank. Um mit diesen Objekten arbeiten zu können, muss der User das Privileg FLASHBACK ARCHIVE ADMINISTER zugeordnet bekommen. Die DBA Rolle hat standardmäßig dieses Recht. Nur mit diesem Recht können Archive angelegt, geändert oder gelöscht werden.

Architektur:

Wenn DML Operationen (INSERT, UPDATE, DELETE, TRUNCATE) ausgeführt werden, schreibt die Oracle Datenbank diese Daten in einen UNDO Tablespace, welcher für Rollbacks und Datenkonsistenz genutzt wird.

Nachdem für eine Tabelle FDA aktiviert wurde, werden alle Datenänderungen an dieser Tabelle und die zugehörigen UNDO Datensätze für die Archivierung markiert. Um zu gewährleisten, dass alle Änderungen historisiert werden, werden die UNDO Datensätze solange nicht gelöscht, bis sie im Archiv gespeichert sind.

Ändert sich die Tabellenstruktur der archivierten Tabelle, werden zum Beispiel neue Spalten hinzugefügt, Spalten gelöscht oder umbenannt, bleiben diese trotzdem im Archiv erhalten und können auch wieder abgefragt werden.

Beim Anlegen des Archives wird auch eine Zeitspanne mit angegeben. Diese Zeitspanne beschreibt, wie lange Daten im Archiv vorgehalten werden sollen. Angegeben wird die Zeitspanne mit dem RETENTION Parameter. Die historisierten Daten bleiben nur solange im Archiv erhalten, wie beim Anlegen des Archives angegeben.

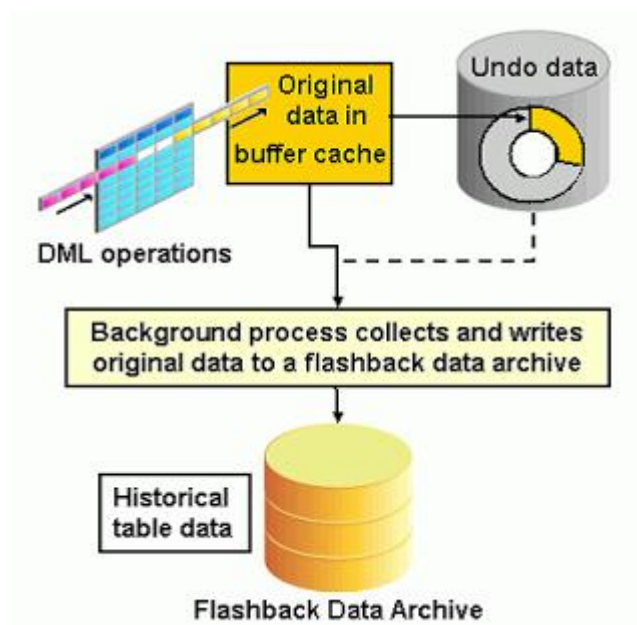


Abb. 1: Flashback Data Archive Architektur

Archiv neu anlegen:

```
CREATE FLASHBACK ARCHIVE fdba1  
TABLESPACE fda_tbs  
RETENTION 15 YEAR;
```

Mit diesem Befehl wurde ein neues Datenbankobjekt `fdba1` angelegt. Verwendet wird der Tablespace `fda_tbs` für dieses Archiv. In dem Archiv werden die historisierten Daten für 15 Jahre archiviert. Nach den 15 Jahren werden sie automatisch gelöscht. Ein DBA mit dem entsprechenden Rechten kann aber jeder Zeit in dem Archiv eingreifen und dieses aufräumen.

Neu bei diesem CREATE Befehl ist der `RETENTION` Parameter. Dieser legt fest, wie lange die Daten im Archiv vorgehalten werden sollen. Die Zeitspanne kann in Jahren (`YEAR`), aber auch in Tagen (`DAY`) und Monaten (`MONTH`) angegeben werden. Das Löschen der Daten erfolgt automatisch über ein internes DROP der jeweiligen Partition. Daher nimmt das Löschen keine große Zeit in Anspruch und hat keine Auswirkungen auf die Performance. Der Anwender muss hier nicht eingreifen.

Folgende Befehlen dienen dazu um ein Archiv bearbeiten:

Archiv als Default setzen. Dieses Archiv wird verwendet, wenn kein anderes Archiv angegeben wird.

```
ALTER FLASHBACK ARCHIVE fdba1 SET DEFAULT;
```

Dauer der Archivierung ändern

```
ALTER FLASHBACK ARCHIVE fdba1 MODIFY RETENTION 2 YEAR;
```

Archiv komplett leeren

```
ALTER FLASHBACK ARCHIVE fdba1 PURGE ALL;
```

Archiv teilweise leeren (alle Historisierungen, die älter als 1 Jahr sind)

```
ALTER FLASHBACK ARCHIVE fdba1  
PURGE BEFORE TIMESTAMP (SYSTIMESTAMP - INTERVAL '1' YEAR);
```

Folgendermaßen kann die Archivierung aktiviert werden:

Archivierung für bestehende Tabelle aktivieren:

```
ALTER TABLE tab1 FLASHBACK ARCHIVE fdba1;
```

```
ALTER TABLE tab2 FLASHBACK ARCHIVE; <= hier wird das Default Archiv verwendet
```

Neue Tabellen mit Archivierung anlagen:

```
CREATE TABLE tabname (column1 NUMBER(10) NOT NULL, ...)  
FLASHBACK ARCHIVE fdba1;
```

Nachdem die Archivierung für eine Tabelle aktiviert wurde ist kein Löschen der Tabelle mehr möglich.

Ab sofort werden alle Änderungen an einem Datensatz in dem Archiv für die festgelegte Dauer gespeichert.

Archivierte Daten abfragen:

Nach dem die Archivierung für eine Tabelle aktiviert wurde, stellt sich die Frage, wie die historisierten Daten aus dem Archiv gelesen und verwendet werden können.

AS OF:

```
SELECT * FROM tabname AS OF TIMESTAMP TO_TIMESTAMP ('2013-09-01  
16:00:32, 'YYYY-MM-DD HH24:MI:SS');
```

```
SELECT * FROM tabname AS OF TIMESTAMP AS OF TIMESTAMP (SYSTIMESTAMP -  
INTERVAL '60' MINUTE);
```

Mit „AS OF TIMESTAMP“ wird die Version der Datensätze zurück geliefert, die zu diesem gewählten Zeitpunkt gültig waren. Hierfür muss man allerdings wissen, welchen Zeitstempel man verwenden will.

VERSION OF:

```
SELECT versions_startscn, versions_starttime, versions_endscn,  
versions_endtime, versions_xid, versions_operation, column1, column2  
FROM tabname VERSIONS BETWEEN TIMESTAMP TO_TIMESTAMP ('2013-09-01  
16:00:32, 'YYYY-MM-DD HH24:MI:SS');AND TO_TIMESTAMP('2013-09-04  
16:00:32, 'YYYY-MM-DD HH24:MI:SS');
```

Mit "VERSION OF" können die verschiedenen Versionen der Datensätze und die dazugehörigen Metadaten abgerufen werden. Hiermit kann nachverfolgt werden, welche Operationen die Datensätzen zwischen bestimmten Zeiträumen geändert haben.

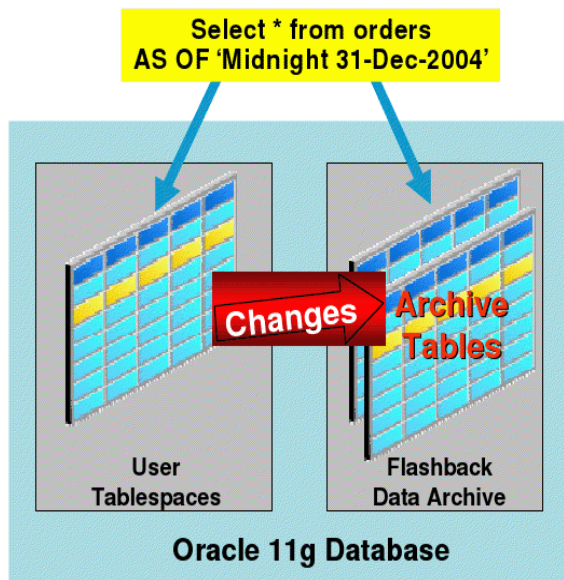


Abb. 2: Abfrage der Daten

Archivierung deaktivieren:

Komplettes Archiv löschen:

```
DROP FLASHBACK ARCHIVE fda1;
```

Dieser Befehl löscht unwiderruflich das Archiv und alle dort gespeicherten Daten.

Archivierung für einzelne Tabellen deaktivieren:

```
ALTER TABLE tab1 NO FLASHBACK ARCHIVE;
```

Dieser Befehl deaktiviert die Archivierung für die Tabelle tab1. Dabei werden aber auch alle archivierten und historisierten Daten dieser Tabelle gelöscht.

Proof of Concept

Anforderungen:

Der Proof of Concept wurde bei einem Kunden der Versicherungsbranche durchgeführt.

Bei diesem Kunden ist ein Data Warehouse vorhanden, welches alle Transaktionen und Bewegungen auf Konten inkl. Verträgen und Schäden enthält. Des Weiteren sind in diesem DWH auch Informationen zu den einzelnen Kunden und verschiedene Mappings für ein SAP System enthalten. Das DWH dient als Schnittstelle zwischen den Systemen zur Verwaltung der Versicherungen und dem SAP System für die Finanzverwaltung.

Bisher wurde keine Historisierung eingesetzt. Eine neue Anforderung der Wirtschaftsprüfer verlangt aber, dass alle Transaktionen und auch die meisten Dimensionen historisiert werden, damit auch in Zukunft Veränderungen der Dimensionen (vor allem der SAP Dimension) nachvollzogen werden können. Aktuell wurden Daten, welche älter als 2 Jahre aus den Faktentabellen gelöscht. Ebenfalls gab es keine Historisierung von Dimensionsänderungen. Die Anforderung der Wirtschaftsprüfer war, dass die Daten mind. 15 Jahr vorgehalten werden und dass Änderungen an den Dimensionen einmalig zum Jahresende gespeichert und daher jährlich nachvollzogen werden können. Dimensionen können sich zwar monatlich ändern, aber für das Auditing reicht wenn die letzte Änderung des Jahres historisiert wird.

Vorhandenes Data Warehouse und Reporting:

Für die gegebenen Anforderungen ist das Konzept der Slowly Changing Dimension – Typ 2 die am meist gebräuchlichste technische Umsetzung. Daher war es in diesem Fall auch der erste Gedanke SCD2 zu verwenden.

Nach einer tiefen Analyse des ETL Prozesses stellte sich allerdings heraus, dass die Verwendung von SCD2 auf den vorhandenen Dimensionstabellen einen kompletten Umbau der ETL Prozesse bedurfte.

In diesem DWH wird kein ETL Tool verwendet. Die kompletten Prozesse sind in PL/SQL Packages und Prozeduren abgebildet. Daher wäre der Aufwand hier sehr hoch und sehr fehleranfällig gewesen, da gleichzeitig noch weitere Projekte diese Prozeduren und Packages bearbeitet haben.

In den bisherigen ETL Prozessen werden als erstes die Faktentabellen befüllt und daraufhin werden die Dimensionen erzeugt. Referenziert werden die Dimension dabei nur über fachliche Schlüssel. Dabei werden die Tabellen der Dimensionen in einem vorherigen Schritt gelöscht (nicht nur geleert) und neu angelegt. Nach dem die Fakten erzeugt wurden, werden die Dimension gefüllt. Um hier SCD2 einzubauen, hätte es einen kompletten Umbau der ETL Prozesse bedurft. Da es sich um über 200 Dimensionstabellen handelt, war dies in der vorgegebenen Zeit nicht realisierbar.

Bei dem ETL Prozess handelt es sich um ein sogenanntes „Monthly Closing“. Bei diesem werden einmal im Monat die Transaktionen in den Quellsystemen abgeschlossen. Dabei werden die Daten von den Quellsystemen über das DWH nach SAP geschickt. Im Data Warehouse werden die Daten entsprechend aufbereitet und verdichtet, damit sie im SAP System weiterverarbeitet werden können.

Als Reportingtool wird Cognos verwendet. Hier musste gewährleistet werden, dass mit den vorhandenen Berichten die historisierten Daten abgefragt werden können.

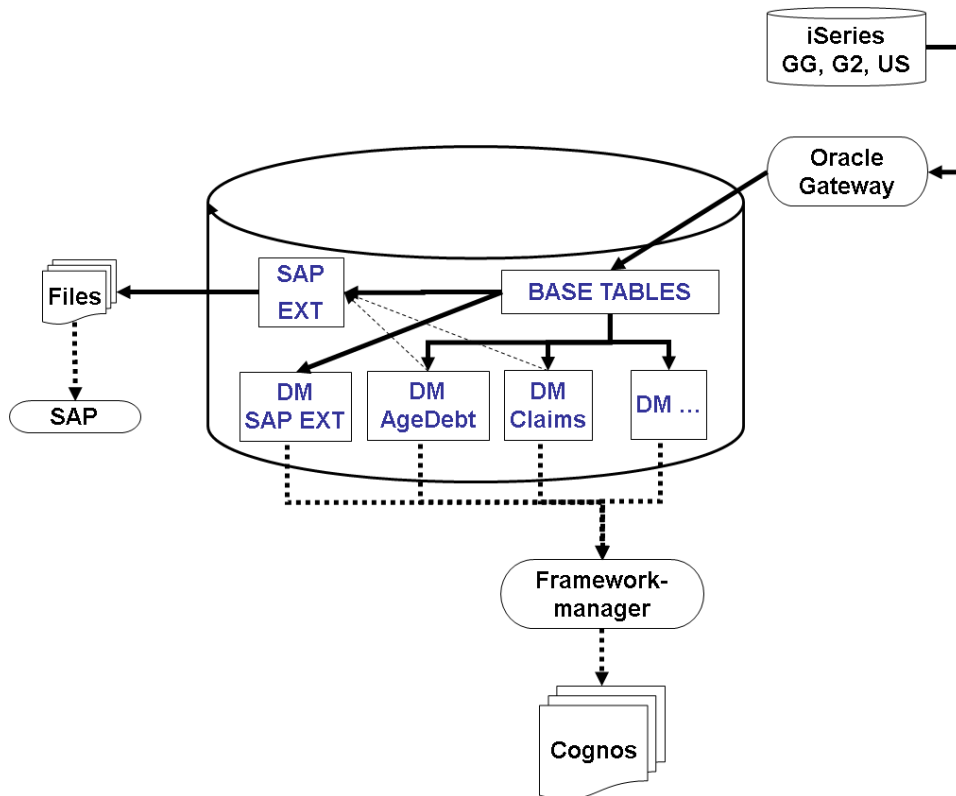


Abb. 3: DWH / ETL Prozess

Um trotzdem in kurzer Zeit die gegebenen Anforderungen umsetzen zu können, wurde beschlossen sich die Option Flashback Database Archiv näher anzusehen. Es wurde festgelegt, dass man die Funktionsweise innerhalb eines Proof of Concepts im DWH testen muss um feststellen zu können, ob mit FDA alle Anforderungen erfüllt werden können.

Umsetzung:

Für den Proof of Concept (PoC) wurde eine Kopie des DWHs angelegt. Ebenfalls wurde für den PoC nur ein gewisser Teil der Tabellen für die FDA aktiviert werden soll ausgewählt. Dabei handelte es sich um die Tabellen, bei denen die meisten Änderungen zu erwarten waren.

Es wurde geplant, mehrere der sogenannten „Monthly Closings“ durchzuführen um die Auswirkungen von FDA auf die Datenbank genauer untersuchen zu können.

Bei den ersten Testläufen stellten sich dann sogleich auch die ersten Unwegsamkeit heraus, die aber durch kleinere Anpassungen der Prozeduren und Packages behoben werden konnten. Da im bisherigen ETL Prozess die „Base“ Tabellen nicht nur geleert sondern gelöscht werden, mussten hier die ersten Anpassungen vorgenommen. Das DROP der Tabellen wurde entfernt und die Daten wurden nur noch gelöscht (per TRUNCATE).

Nachdem die Anpassungen an dem Code durchgeführt wurden, konnten die ersten Testläufe stattfinden. Dabei wurden keine weiteren technischen Probleme mehr festgestellt und der ETL Prozess lief durch.

Um festzustellen, ob alle fachlichen und technischen Anforderungen durch FDA erfüllt werden können wurden von verschiedenen Personen, die am Data Warehouse beteiligt sind (DBAs, DataMart Owner, Entwickler) Testkriterien festgelegt. Nur wenn alle dieser Kriterien erfüllt werden, kann FDA als Historisierungstool für das Data Warehouse verwendet werden.

Folgende Kriterien wurden festgelegt (Auszug):

- Datenwachstum für das Neuaufbauen der Tabellen
- Performance der Cognos Berichte
- Einfluss eines TRUNCATE mit aktiviertem FDA auf die Performanz
- Partitionierung der archivierten Tabellen
- Korrekte Wahl des "as-of-timestamp"
- Import & Export / Backup & Recovery
- Einfluss auf den gesamten ETL Prozess
- FDA nur einmal im Jahr aktivierbar
- ...

In dem Proof of Concept wurde keine Zufriedenstellende Möglichkeit gefunden, FDA in diesem Data Warehouse so einzusetzen, dass alle fachlichen sowie technischen Anforderungen erfüllt werden konnten.

Ein großes technisches Problem für die Abfrage der richtigen historisierten Version der Daten war es den richtigen TIMESTAMP für den AS OF Befehl zu finden. Hierfür wäre zum Beispiel eine weitere Mapping Tabelle von Nöten gewesen, mit der der TIMESTAMP zum zugehörigen „Monthly Closing“ gejoint werden kann. Für die Fachanwender musste es aber so einfach wie möglich gehalten werden. Dies war aber in der gegebenen Zeit nicht möglich so umzusetzen.

Des Weiteren wären durch das komplette Löschen der meisten Tabelle und das Neubefüllen die Anforderungen an den Speicherplatz zu hoch gewesen. Um die fachliche Anforderung zu erfüllen, hätte es gereicht, FDA nur beim Jahresabschluss zu aktivieren. Damit hätte man die Änderung zum Vorjahr historisieren können. Allerdings verliert man alle bisherigen archivierten Datensätze, sobald FDA deaktiviert wird. Als Alternative müsste der ETL Prozess so angepasst werden, dass nur noch eine Delta Verarbeitung statt findet. Auch dies war in der gegebenen Zeit nicht möglich.

Auf die Performance des ETL Prozesses hatte FDA kaum spürbare Auswirkungen, zumindest konnten keine festgestellt werden. Dieses Kriterium wurde also erfüllt.

Leider konnte die Administratoren der Datenbank innerhalb des PoC keine Aussagen über das Backup und Restore Verhalten mit aktivierten Archiven treffen, so dass dieses Kriterium auch nicht als erfüllt angesehen werden konnte.

Am Ende des Proof of Concepts waren zu viele Kriterien un- oder nicht ausreichend beantwortet. Daher wurde Flashback Database Archiv nicht für die Historisierung eingesetzt und es wurde nach einer alternativen Lösung gesucht.

Fazit

Man kann sich Flashback Database Archiv genauer ansehen, wenn man eine neue Historisierung plant. Für ein operatives System gibt es momentan keine einfachere Art eine Historisierung und Archivierung, die auch noch Revisionssicher ist für alle Datenänderungen einzuführen.

Flashback Database Archive hat sich jedoch in diesem konkreten Fall nicht bewährt, da während des Proof of Concepts nicht alle erforderlichen Kriterien erfüllt werden konnten und auch weitere Fragen aufgeworfen wurden.

Ob es auch für Data Warehouses geeignet ist müsste in jedem einzelnen Fall überprüft werden. Die Voraussetzungen dafür hätte Flashback Database Archiv.

Kontaktadresse:

Wolfgang Tanzer
metafinanz Informationssysteme GmbH
Leopoldstraße 146
D-80804 München

Telefon: +49 (0) 89 3605315232
Fax: +49 (0) 89 36053115
E-Mail: Wolfgang.Tanzer@metafinanz.de
Internet: <http://www.metafinanz.de>