



2010
DOAG
Konferenz + Ausstellung

Die Änderungen entscheiden

ORACLE Change Data Capture

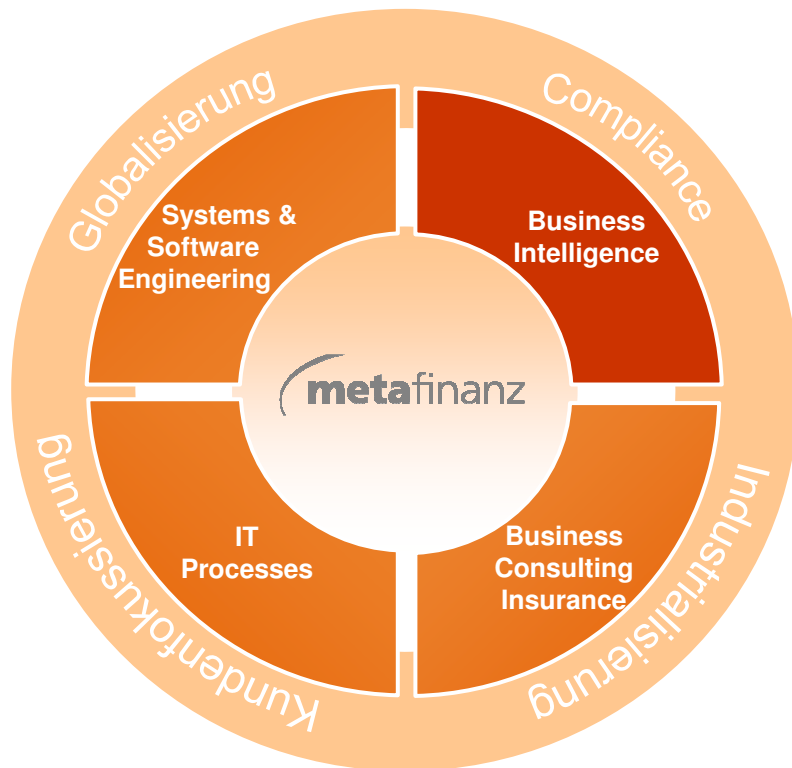
Günther Herzog

Stand: November 2010

metafinanz bietet Fach-, Prozess- und Technologiekompetenz aus einer Hand

20 Jahre dynamic X²cellence

metafinanz GmbH ist ein Software- und Beratungshaus mit Schwerpunkt in der Finanzdienstleistungsbranche. Als Generator effizienter Business-Prozesse übernehmen wir End-to-End-Verantwortung und unterstützen unsere Kunden in ihren Projektvorhaben.



Ihr Partner für:

- **Datawarehousing**
(Oracle, OWB, Informatica)
- **Business Intelligence**
(SAS, Cognos, Oracle, Microsoft)
- Schwerpunkte:
Datenqualität, Sicherheit und Compliance
- Offene **Schulungen** und maßgeschneiderte **Coachings**

Partnerships:



Was ist Change Data Capture (CDC)

- Identifizierung von neuen, geänderten oder gelöschten Daten

Gliederung

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

Anforderung an DWH-Architektur

- Effizient
- Ressourcenschonend
- Reduktion der zu verarbeitenden Datenmenge

Befüllung DWH

- Für Aktualisierung DWH nur Datenänderung notwendig
- Zeit für periodische Aktualisierung meist sehr knapp
- Daher Bereitstellung der Daten oft sehr spät

ORACLE Change Data Capture (CDC)

- Methodik zur Erkennung von Datenänderungen
- Framework zur Bereitstellung der erkannten Änderungen
- zu stellen und dabei so wenig wie möglich Ressourcen zu verbrauchen.

Grafische Frontends

- Oracle Warehouse Builder
- Oracle Data Integrator

Tabellenvergleich

- Vergleich Quell- Zieltabelle
- Hoher Zeitaufwand für Übertragung/Kopiervorgang
- Hoher Ressourcen-Aufwand für Deltaermittlung
- Evtl. Datenkonsistenz muss eigens implementiert werden

Attributsabhängige Selektion

- Fachliche/technische Attribute in der Quelltablelle (z.B. Insert-Date, Verarbeitet-Flag, Version)
- Evtl. Änderungen an der Quelldatenbank notwendig
- Evtl. Datenkonsistenz muss eigens implementiert werden

Trigger

- Jede Änderung in einer Tabelle wird sofort per Trigger in eine Änderungstabelle geschrieben
- Quelltransaktion wird erweitert
- Änderungen an der Quelldatenbank notwendig
- Hoher Verwaltungsaufwand
- Nicht immer ressourcenschonend
- Evtl. Datenkonsistenz muss eigens implementiert werden

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

Change Data Capture von ORACLE

- CDC Framework
- Bei Datenbankinstallation enthalten
- Weniger Hardwareressourcenverbrauch
- Weniger Datenübertragung
- Leichtes Wiederaufsetzen nach Connection-Abbrüchen
- Beinhaltet Auditing
- Beinhaltet Exception Handling
- Berücksichtigung nur der notwendigen Spalten möglich
- Transaktionelle Konsistenz für Änderungen über mehrere Quelltabellen über Change-Sets

Versionsgeschichte ORACLE CDC

- DB 9i (2001):
 - Einführung Synchronous Change Data Capture
- DB 10gR2 (2006):
 - Einführung Asynchronous Change Data Capture (CDC)
- DB 11gR1 (2008):
 - CDC ist nach Bedarf zu aktivieren oder zu deaktivieren.
 - Das Füllen (Extend) der Subscriber Views wurde flexibilisiert, so daß man den Zeitpunkt angeben kann, bis zu dem gefüllt werden soll.
 - Das Füllen Löschen (purge) der Subscriber Views wurde flexibilisiert, so daß man den Zeitpunkt angeben kann, bis zu dem logisch gelöscht werden soll.
- DB 11gR2 (2010):
 - Keine Neuerungen

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

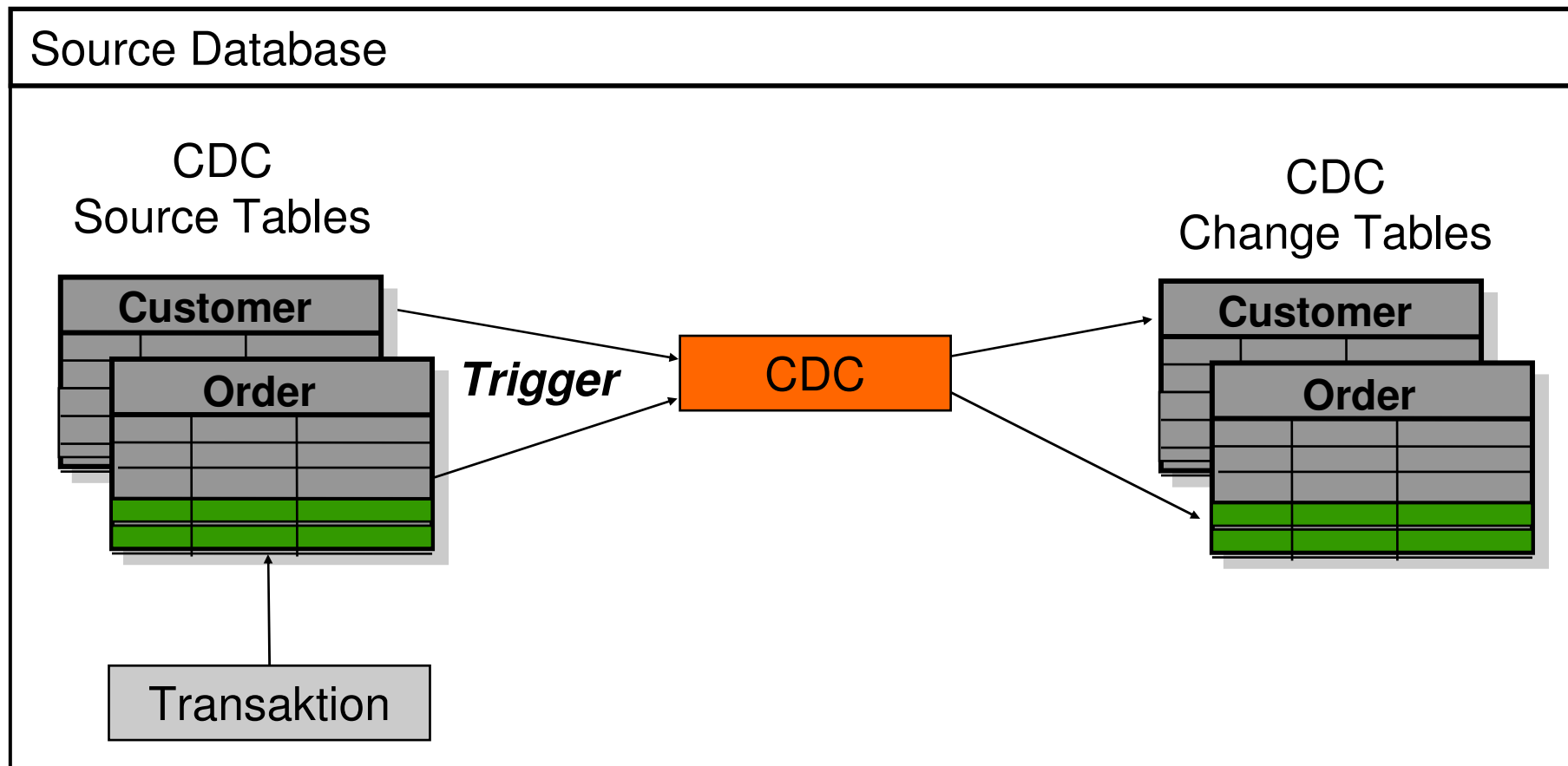
IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

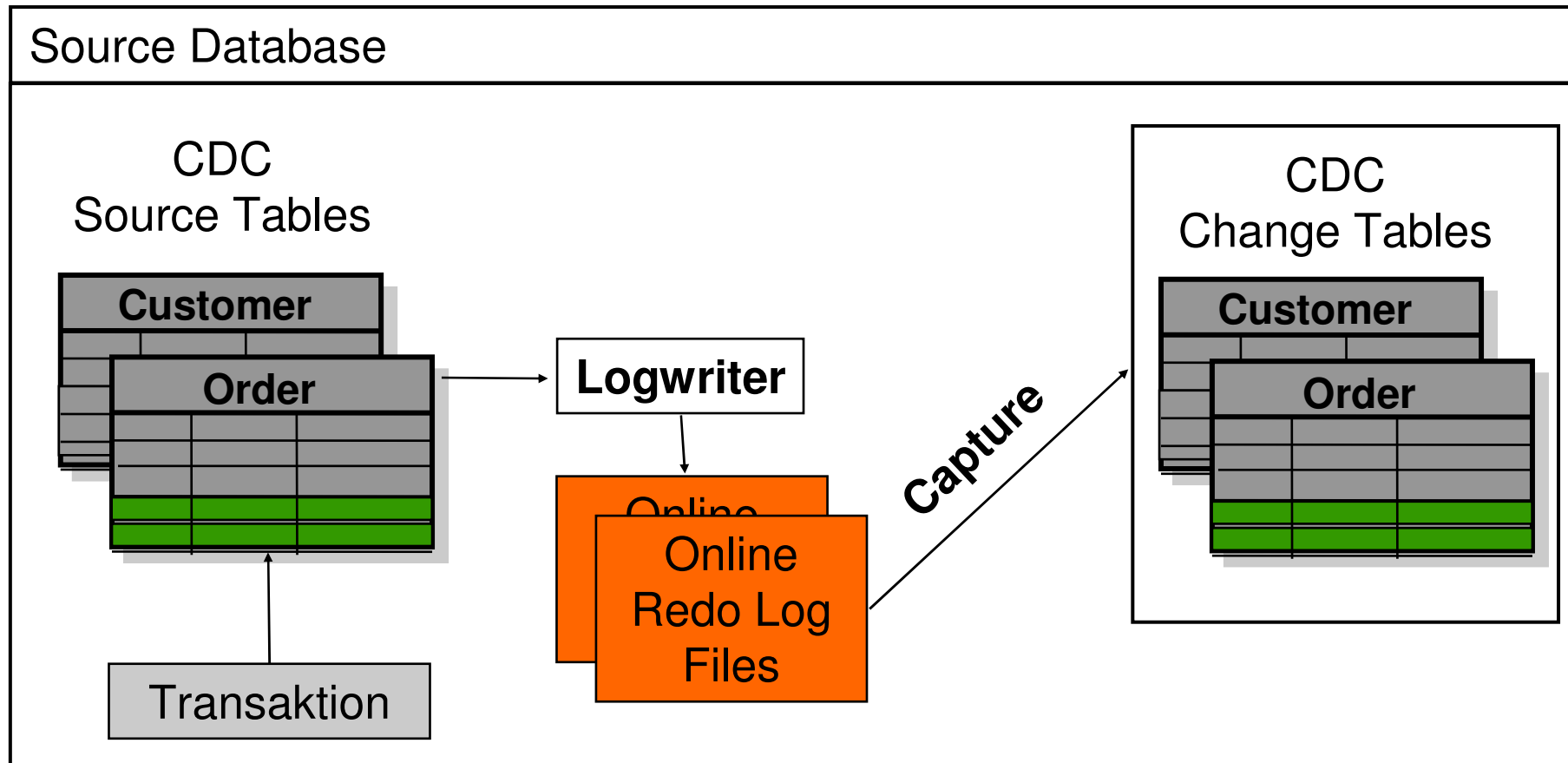
VI. Bewertung & Fazit

Synchronous CDC HotLog

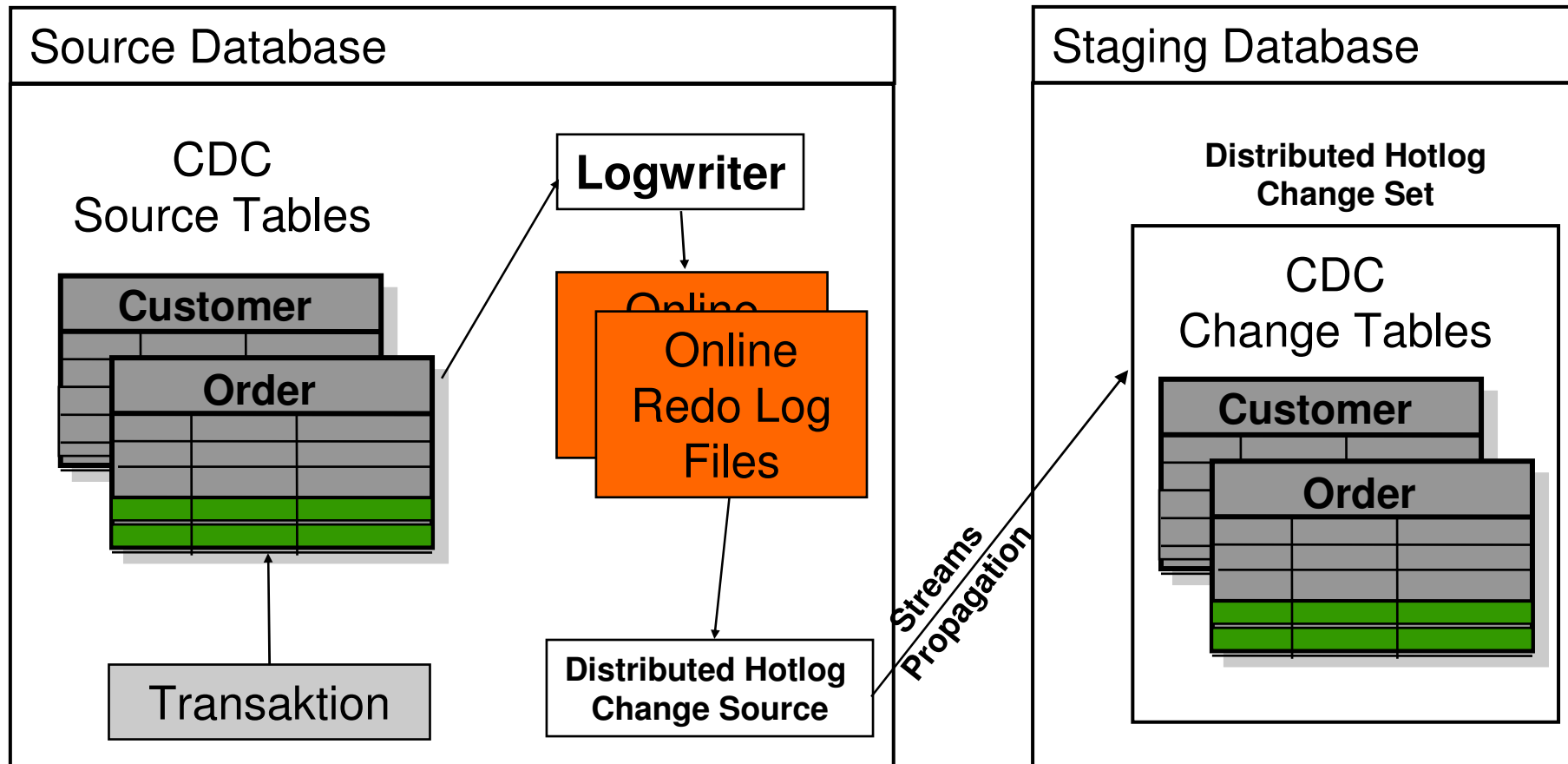
- Systemtrigger
- Echtzeit-Capturing
- Gleiche Performance-Auswirkungen wie User-Trigger



Asynchronous CDC HotLog

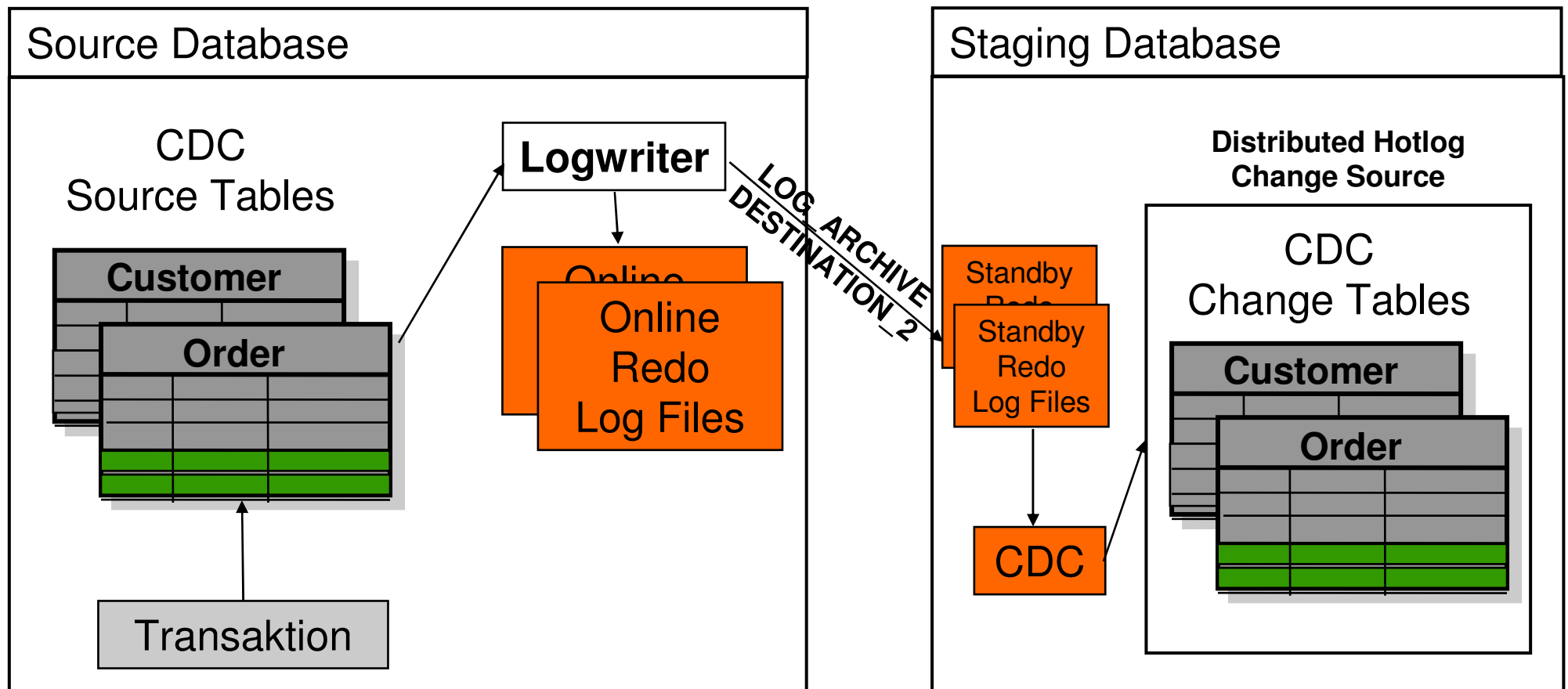


Asynchronous CDC distributed HotLog



Asynchronous Autolog CDC

- **Autolog Online CDC Mode:** Logwriter schreibt Daten für RedoLog local und auf die StagingDB
- **Autolog Archive CDC Mode:** Archiver-Prozess schreibt bei einem Redo Log File Switch die Daten in ein lokales Archive Redo Log und ein Redo Logfile auf die Staging DB



ORACLE CDC-Varianten

	Synchron	Asynchron		
	Synchronous Hotlog CDC	Asynchronous Hotlog CDC	Asynchronous Distributed Hotlog CDC	Asynchronous Autolog CDC
Quelle/ Mechanismus	Systemtrigger	Online Redo Logfile	Online Redo Logfile, Übertragung durch Streams	OnlineVersion: RedoLog ArchiveVersion: ArchiveLogs Transport managed by Redo transport services
Teil der Quell- transaktion	Ja	Nein	nein	Nein
Anzahl Systeme	1	1	2	2
Zeitversatz zur Quelltransaktion	Kein	Nahe Echtzeit. ChangeTabelle wird befüllt, sobald die neue abgeschlossene Transaktion ankommt	Abhängig von der Topologie. ChangeTabelle wird befüllt, sobald die neue abgeschlossene Transaktion ankommt	Abhängig von der Topologie und der LogSwitch-Intervalle
Seit DB Version	9i	10g	10g	10g
Voraussetzung: Quell- und ZielDB gleich?	-	-	<ul style="list-style-type: none"> • Source DB Version größer 9.2 • Target DB Version größer 10.2 • Betriebssystem kann unterschiedlich sein. 	<ul style="list-style-type: none"> • Hardwarearchitektur • Betriebssystem • DB-Version müssen gleich sein

Voraussetzungen

- DB: Archive-Log-Mode
- Supplemental Logging auf Datenbank- oder Tabellenebene
- Forced-Logging-Mode

Lizenzbedingungen

- Synchronous CDC ist ab der Standart Edition enthalten
- Asynchronous CDC ist ab der Enterprise Edition enthalten

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

Change Table

- Tabelle in der die Änderungen eingetragen werden
- Teilmenge der Quelltabellenattribute möglich
- Metadatenattribute für CDC

Change Set

- Logische Sammlung von Änderungsdaten die transaktionell konsistent benötigt werden
- Beinhaltet ein oder mehrere Change Table
- Transaktionelle Konsistenz innerhalb der Change-Tabellen eines Change Sets.

Changetable

	OPERATIONS\$	CSCN\$	RSID\$	COMMIT_TIME STAMP\$	employee_id	first_name	last_name	salary
Before Update	UO	1624263	10046	31.10.10 22:29:08	118	Guy	Himuro	54933
After Update	UN	1624263	10046	31.10.10 22:29:08	118	Guy	Himuro	7433
Before Update	UO	1624263	10047	31.10.10 22:29:08	119	Karen	Colmenares	5833
After Update	UN	1624263	10047	31.10.10 22:29:08	119	Karen	Colmenares	7333
Insert	I	1624271	10048	31.10.10 22:29:08	210	Paul	Jeep	6500
Delete	D	1626536	10049	31.10.10 22:41:17	210	Paul	Jeep	8000

Meta-Attribute

OPERATION\$ (CHAR(2))

- 'I' : Insert Operation
- 'UO': Version vor Update
- 'UN': Version nach Update
- 'D' : Delete Operation

CSCN\$ (NUMBER): Commit System Change Number

- Synchronous CDC: CSCN der EXTEND_WINDOW - Operation.
- Asynchronous CDC: CSCN der DML operation

Meta-Attribute

RSID\$ (NUMBER)

- Unique Row Sequenz-ID der Transaktion (nicht transaktionsübergreifend)
- Identifizierungsmöglichkeit der Source-Table-Version
- Die RSID\$ des Updates entspricht der Version vor dem Update.

COMMIT_TIMESTAMP\$ (DATE):

- Commit-Zeitpunkt der Transaktion

TIMESTAMP\$ (DATE):

- Zeitpunkt der Operation in der Quelltable

USERNAME\$ (VARCHAR2(30)):

- Username des Users, der die Operation angestoßen hat. Vor 10.2 NULL.

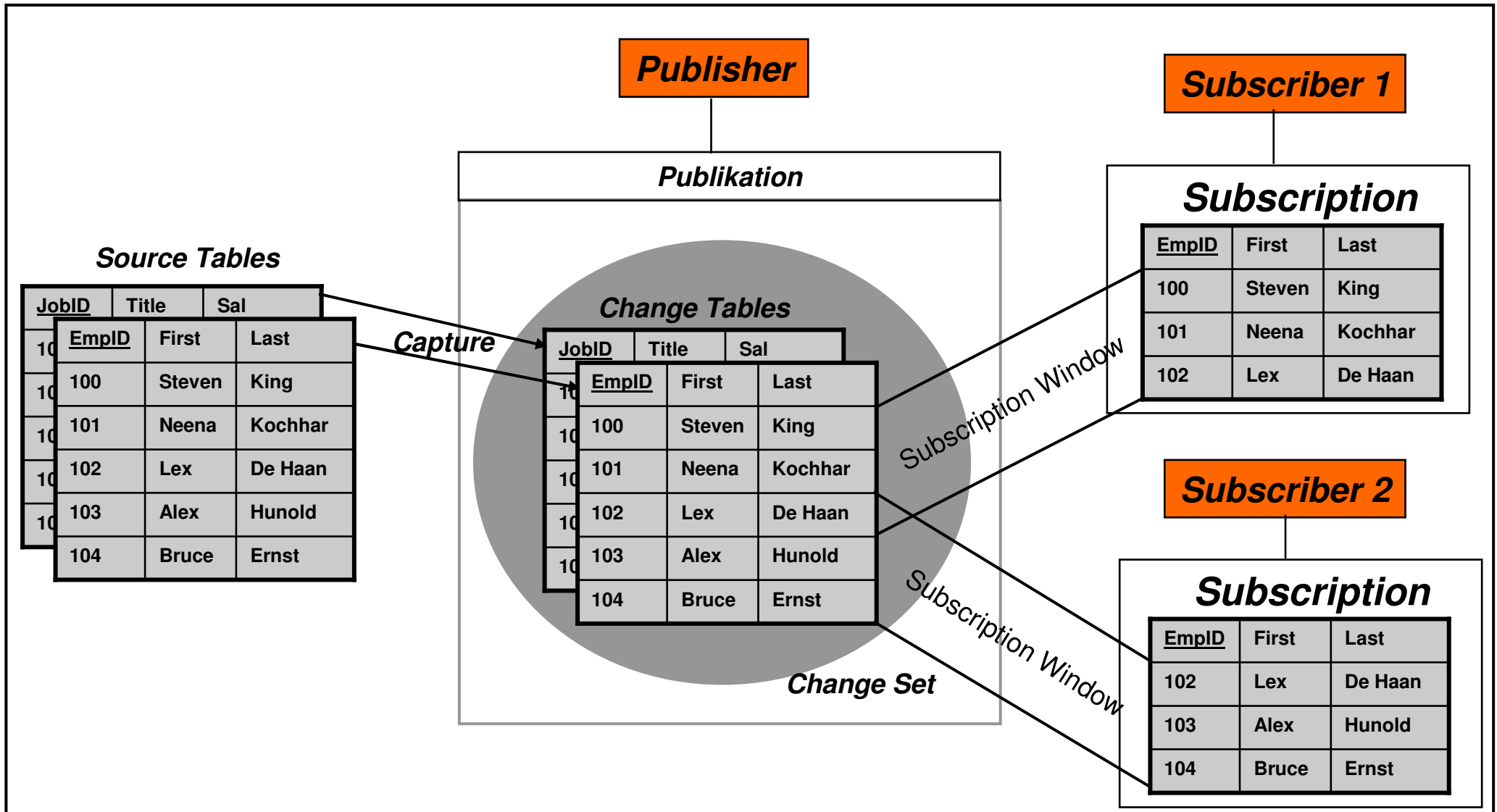
ROW_ID\$ (ROW_ID):

- ROW-ID des betroffenen Datensatzes in der Quelltable.

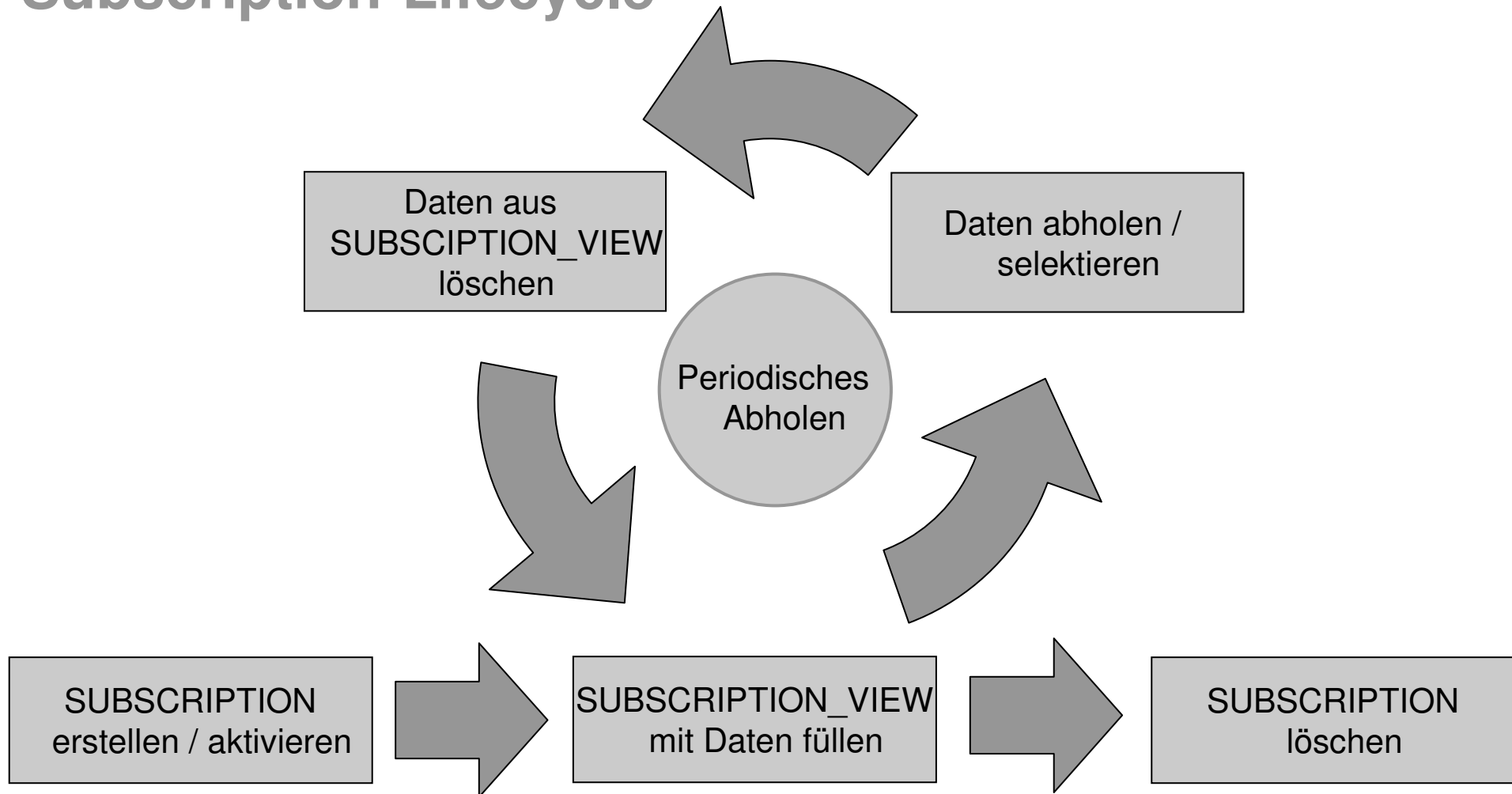
Publisher/Subscriber-Modell

- Datenbereitstellung durch den Publisher
- Datenabholung durch den oder die Subscriber
- Erlaubt unabhängigen Zugriff auf Änderungsdaten
- Kontrolliert Beibehaltung oder logische/physische Löschung der Änderungsdaten

Daten abholen (Publisher/Subscriber)



Subscription-Lifecycle



Periodisches Abholen

Übliches Vorgehen eines Subscribers für den Zugriff auf CDC-Daten:

- Daten in `SUBSCRIPTION_WINDOW` laden

```
DBMS_CDC_SUBSCRIBE.EXTEND_WINDOW(subscription_name, upper_bound);
```

Subscription-Window enthält alle Änderungen seit dem letzten Aufruf von `EXTEND_WINDOW` oder `PURGE_WINDOW` aufgelaufen sind.

- Daten selektieren und verarbeiten.

- Daten logisch aus dem `SUBSCRIPTION_WINDOW` löschen:

```
DBMS_CDC_SUBSCRIBE.PURGE_WINDOW(subscription_name, lower_bound);
```

- Die Änderungsdaten sind in den ChangeTabellen physikalisch noch vorhanden.

Der Publisher

- **Erstellt die ChangeSets:**

```
DBMS_CDC_PUBLISH.CREATE_CHANGE_SET
```

- **Erstellt die ChangeTables:**

```
DBMS_CDC_PUBLISH.CREATE_CHANGE_TABLE
```

- **Aktiviert das ChangeSet:**

```
DBMS_CDC_PUBLISH.ALTER_CHANGE_SET
```

- **Vergibt Select-Rechte für Subscriber:**

```
GRANT SELECT ON <ChangeTable> to <Subscriber>
```

Der Subscriber

- Erstellt für sich eine Subscription
Subscription stellt eine bewegliche View auf die Änderungsdaten zur Verfügung

`DBMS_CDC_SUBSCRIBE.CREATE_SUBSCRIPTION`

- Abbonniert die Daten:

`DBMS_CDC_SUBSCRIBE.SUBSCRIBE`

- Aktiviert die Subscription:

`DBMS_CDC_SUBSCRIBE.ACTIVATE_SUBSCRIPTION`

- Wenn alle Subscriber die Änderungsdaten in ihren Subscription-Windows logisch gelöscht haben, werden diese Daten auch physikalisch aus den ChangeTables gelöscht (default).

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

Tabellenänderungen

Um eine Tabelle zu einer Subscription hinzuzufügen oder eine vorhandene Tabelle per DDL zu ändern sind prinzipiell die Subscriptions und die Changetable zu löschen, die Änderungen vorzunehmen und dann die Changetable sowie die Subscription neu zu erstellen.

Einzelsschritte:

1. Gewährleisten, daß auf den publizierten Quelltabellen keine Daten geändert werden.
2. Das Subscription-Window um alle CDC-Daten erweitern
`DBMS_CDC_SUBSCRIBE.EXTEND_WINDOW`
3. Eventuelle neue/verbliebene CDC-Daten abrufen/verarbeiten.

Tabellenänderungen

4. Subscription-Window löschen

```
DBMS_CDC_SUBSCRIBE.PURGE_WINDOW
```

5. Subscription löschen

```
DBMS_CDC_SUBSCRIBE.DROP_SUBSCRIPTION
```

6.a) Um eine neue Tabelle einzufügen:

Tabelle publizieren

```
DBMS_CDC_PUBLISH.CREATE_CHANGE_TABLE
```

6.b) DDL-Änderungen einer Tabelle:

Change-Table löschen

```
DBMS_CDC_PUBLISH.DROP_CHANGE_TABLE
```

Änderungen vornehmen

Change-Table erstellen

```
DBMS_CDC_PUBLISH.CREATE_CHANGE_TABLE
```

Tabellenänderungen

7. Die Subscription neu erstellen

```
DBMS_CDC_SUBSCRIBE.CREATE_SUBSCRIPTION
```

8. Zu allen Tabellen subscriben

```
DBMS_CDC_SUBSCRIBE.SUBSCRIBE
```

9. Subscription aktivieren

```
DBMS_CDC_SUBSCRIBE.ACTIVATE_SUBSCRIPTION
```

I. Einführung & Überblick

II. Change Data Capture

III. Varianten

IV. Daten abholen (Publisher/Subscriber)

V. Betrieb

VI. Bewertung & Fazit

Absolute Performance

- System mit und ohne CDC schlecht zu vergleichen
- Bei CDC verteilt sich die Last über die ganze Zeit
- Ohne CDC entsteht punktuelle Belastung beim Ermitteln des Deltas

Nachteile

- Einrichtungsaufwand
- Überwachungsaufwand
- Overhead bei DDL-Änderungen

Vorteile

- Flexible Datenbereitstellung für mehrere Abnehmer
- Geringerer Zeitaufwand als herkömmliche Datenänderungserkennung
- Bessere Ausnutzung der vorhandenen Zeit für die Periodische Befüllung
- Frühzeitigeres Bereitstellen der DWH-Daten
- Komfortables Framework zum Verwalten der CDC-Funktionalität

Metafinanz auf der DOAG 2010

17.11.2010 / 10:00 Uhr / Raum Helsinki
Einführung OWB Java API

17.11.2010 / 12:00 Uhr / Raum Helsinki
Vergleich von OWB Repositories mittels Skripting

17.11.2010 / 16:00 Uhr / Raum Helsinki
Arbeiten mit Template Mappings

18.11.2010 / 10:00 Uhr / Raum Stockholm
The beast - Umstieg von einer E20K nach Exadata Version V2

▪ 18.11.2010 / 12:00 Uhr / Raum Stockholm
ORACLE Change Data Capture (CDC) - Die Änderungen entscheiden

▪ 18.11.2010 / 15:00 Uhr / Raum Stockholm
MS-SQL als DWH-Quelle, angebunden per JDBC-Framework ORAJDBCLINK

Interesse? Fragen? Austausch?

Besuchen Sie unseren Stand...



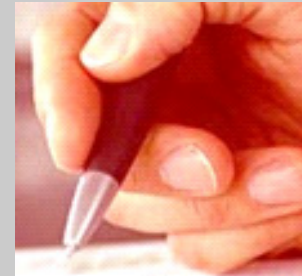
Aktuelle Schulungen:
OWB11gR2 New Features
&
Skripting mit OMB*Plus
Mehr Info an unserem Stand!

Sie finden uns auf Ebene 1, Stand 112 (neben der Mercator Lounge)
Mehr Informationen und Downloads unter: <http://owb.metafinanz.de>



Handelsblatt

Herzlichen Dank!



metafinanz
Informationssysteme GmbH
Leopoldstr. 146
80804 München
phone: +49 89 360531-0
fax: +49 89 360531-15
kontakt@metafinanz.de
www.metafinanz.de