



# Aufbau effizienter Staging Area Lösung mittels Oracle Data Guard



Karol Hajdu  
Senior Consultant  
Mathias Zarick  
Senior Consultant

München, 26.05.2011

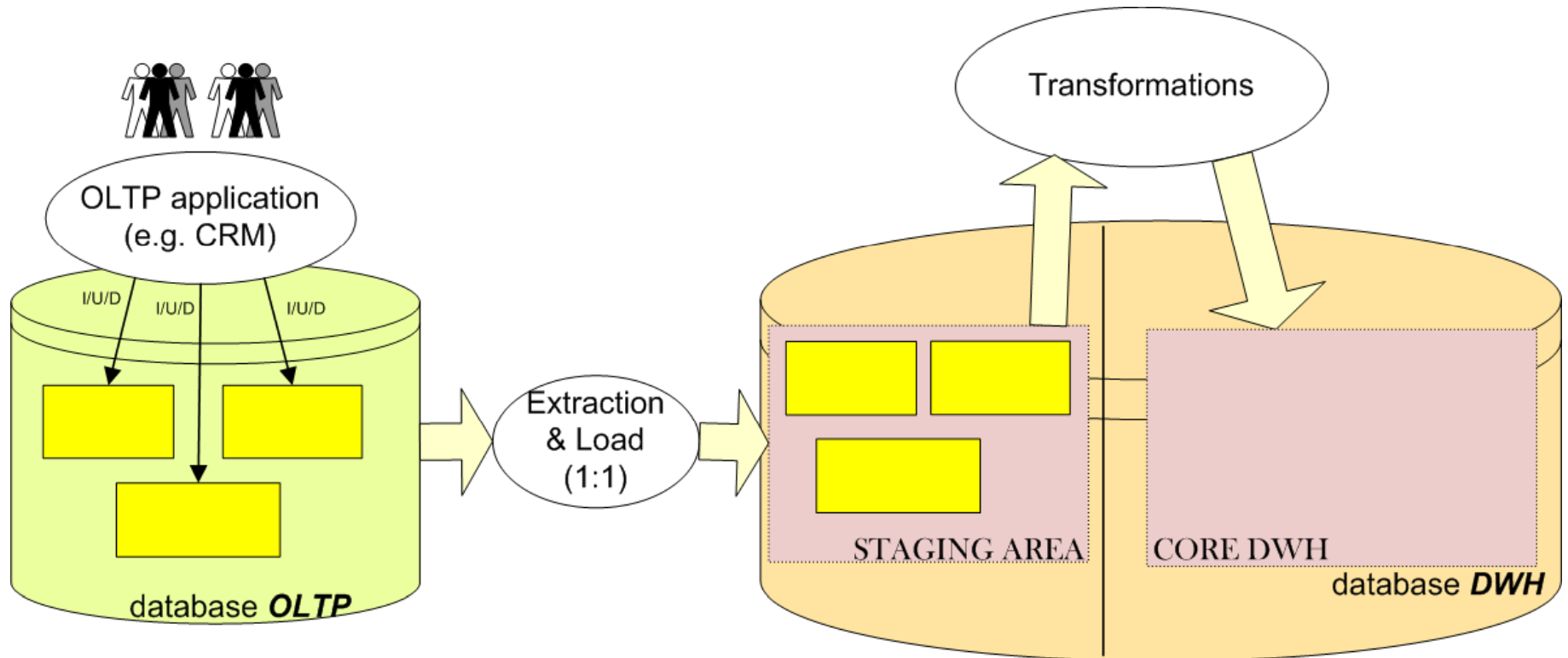
**trivadis**  
makes IT easier. 

# Agenda



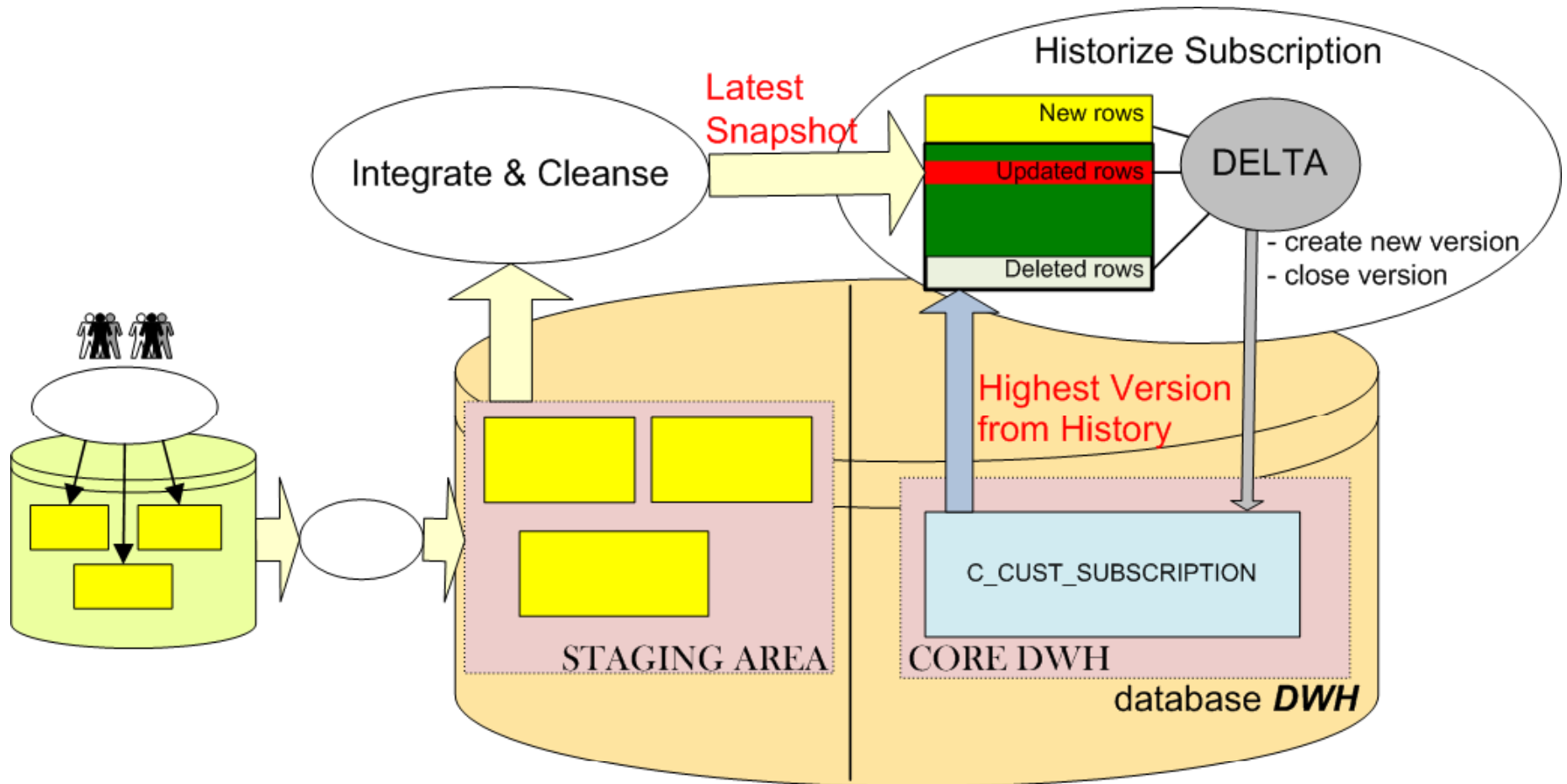
- **Die Herausforderung “Extraktion & Laden“**
- Lösung mittels Oracle Data Guard
  - Konzept
  - Demo
- Zusammenfassung

# Übliche Data Warehouse Architektur

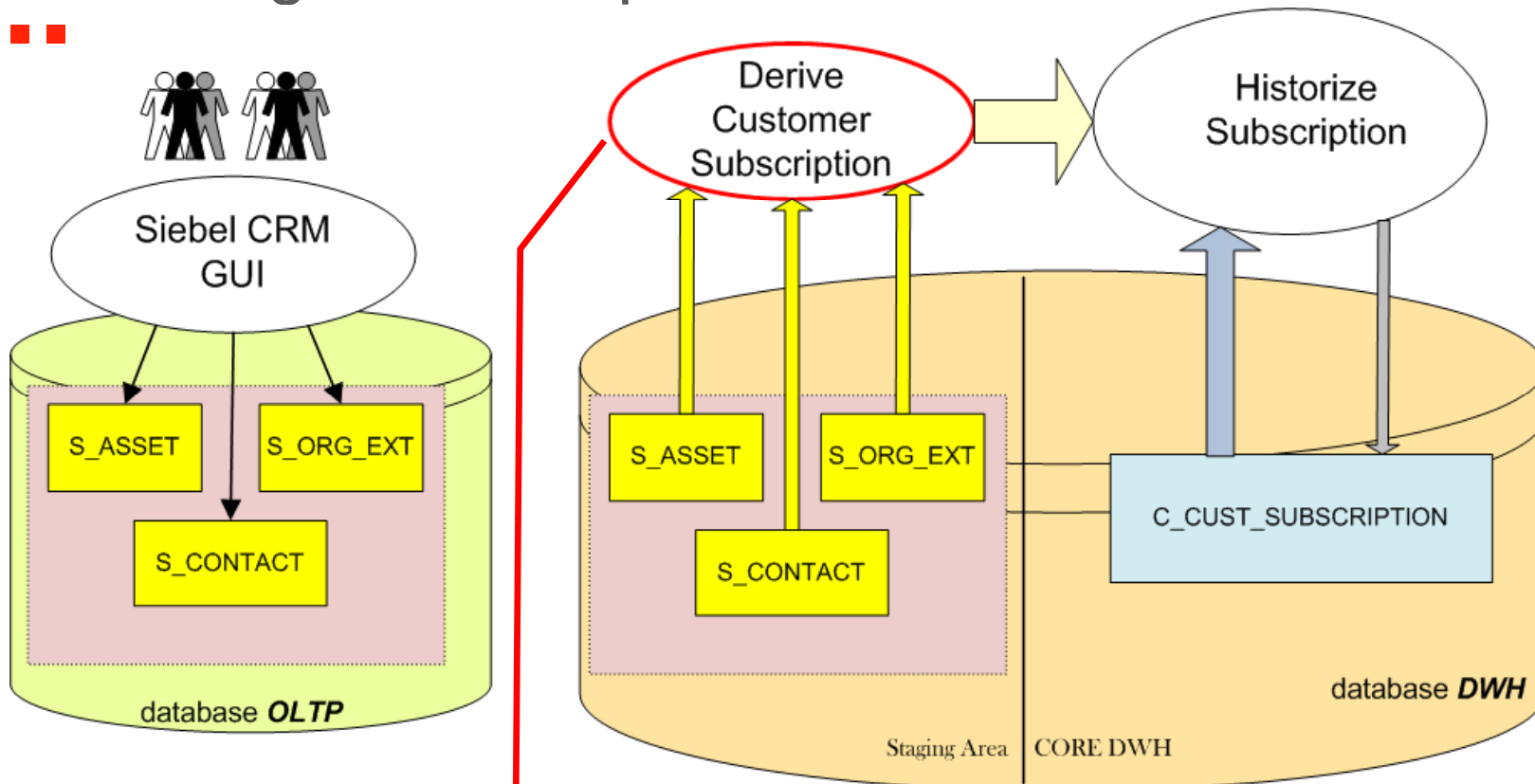


- Extrahierte Daten werden erst in die Staging Area (SA) geladen

# CORE DWH enthält auch historische Daten: Stammdaten werden versioniert (valid\_from, valid\_to)



# Die Integration- & Cleansing-Regeln sind häufig sehr komplex



```
CREATE VIEW i_cust_subscription AS
SELECT      ...
FROM        sa.s_asset ch
INNER JOIN  sa.s_asset p ON  p.row_id = ch.par_asset_id AND ...
INNER JOIN  sa.s_org_ext acc ON  acc.row_id = p.ou_id AND ...
LEFT OUTER JOIN sa.s_contact c ON acc.row_id =c.ou_id
WHERE ...
```

# Stammdaten mit komplexen Beziehungen: Versionierung basiert auf integrierten & bereinigten Daten



## Was bedeutet dies konzeptionell?

- Nicht jede Datenänderung im OLTP System veranlasst eine neue Version im CORE DWH.
- Neue Versionen entstehen aufgrund der Differenzen zwischen zwei integrierten Snapshots
  - neuester Stand in SA vs. zuletzt bekannter Stand im CORE)

## Was bedeutet dies technisch?

- erst integrieren (ableiten und bereinigen), dann versionieren
- Staging Area enthält immer den vollen Datenbestand aus OLTP

# Periodische Aktualisierung der Staging Area – Wo liegt die Herausforderung?

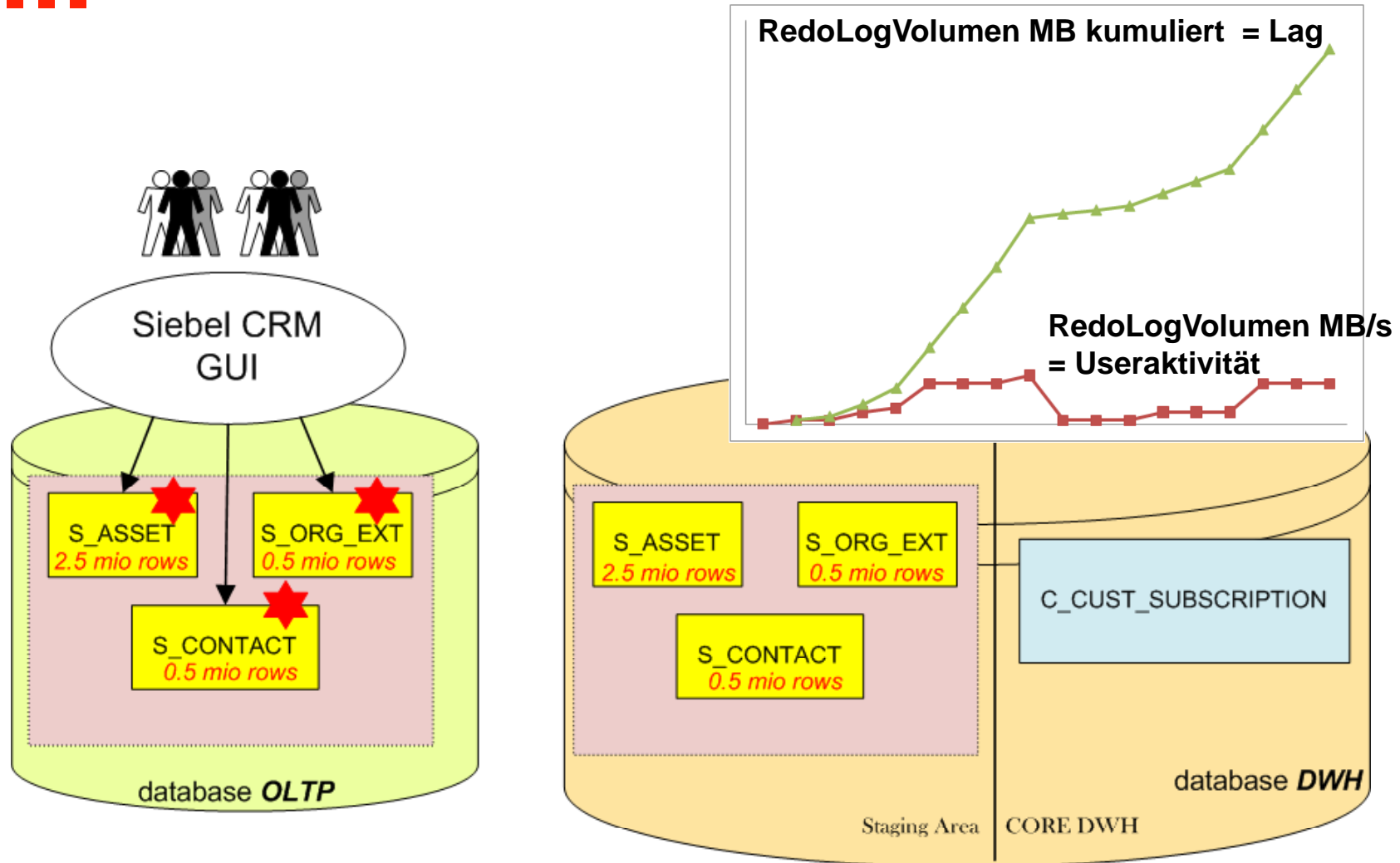


- Staging Area soll den vollen Datenbestand enthalten
- Herausforderung für die Aktualisierung:
  - der volle Datenbestand bedeutet oftmals **großes Datenvolumen**, und
  - die Staging Area soll **sehr oft** aktualisiert werden, oder
  - sie soll **in sehr kurzer Zeit** aktualisiert werden

Wie aktualisiert man die Staging Area auf **effiziente Weise**?



Während die Anwender mit OLTP Applikation arbeiten, divergieren die Daten in OLTP DB und Staging Area → Lag



**Effiziente Aktualisierung = Schneller Abbau vom Lag.**



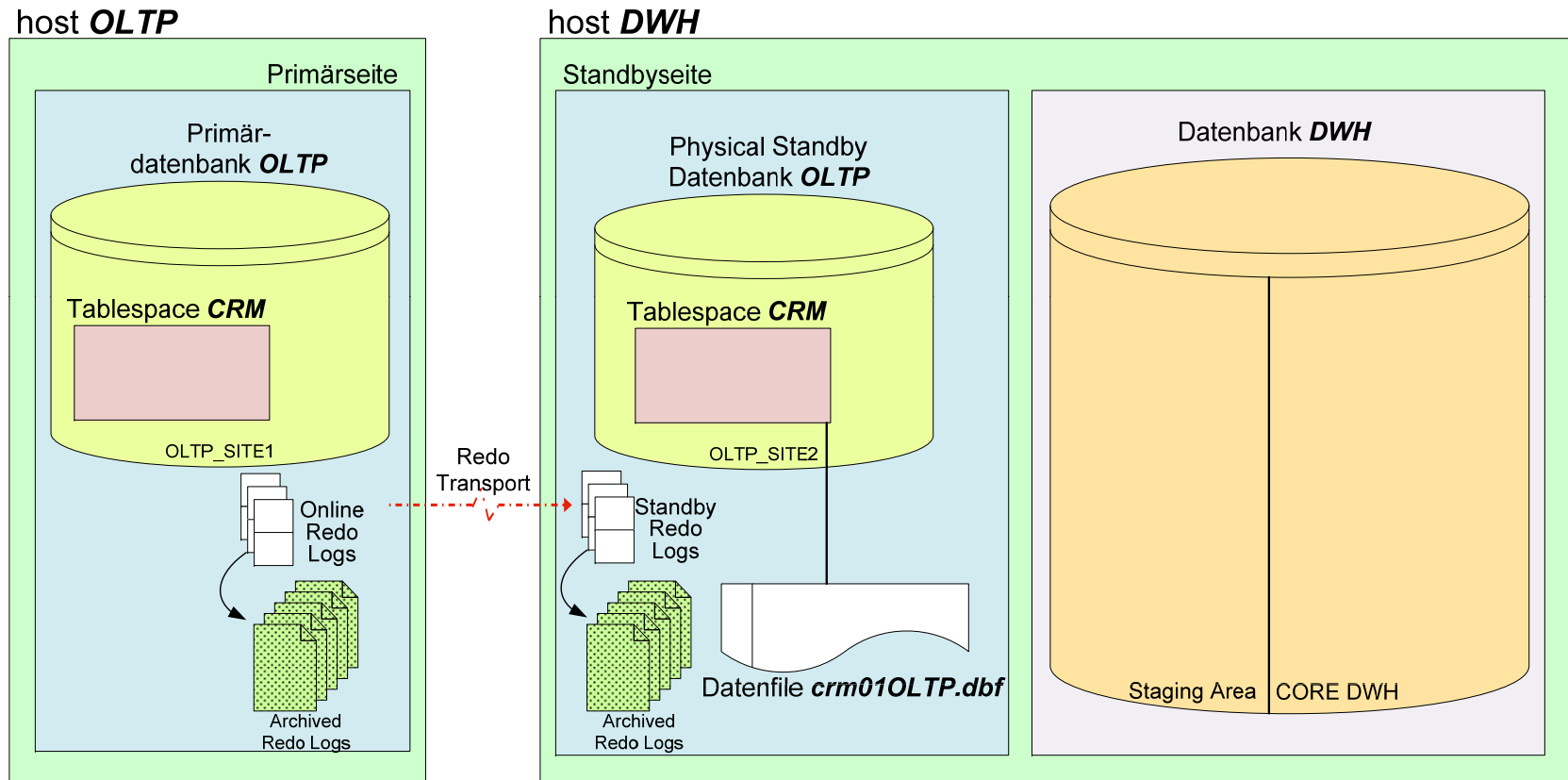
# Agenda



Wissen vermitteln ist der Anfang. Wissen umsetzen das Entscheidende.

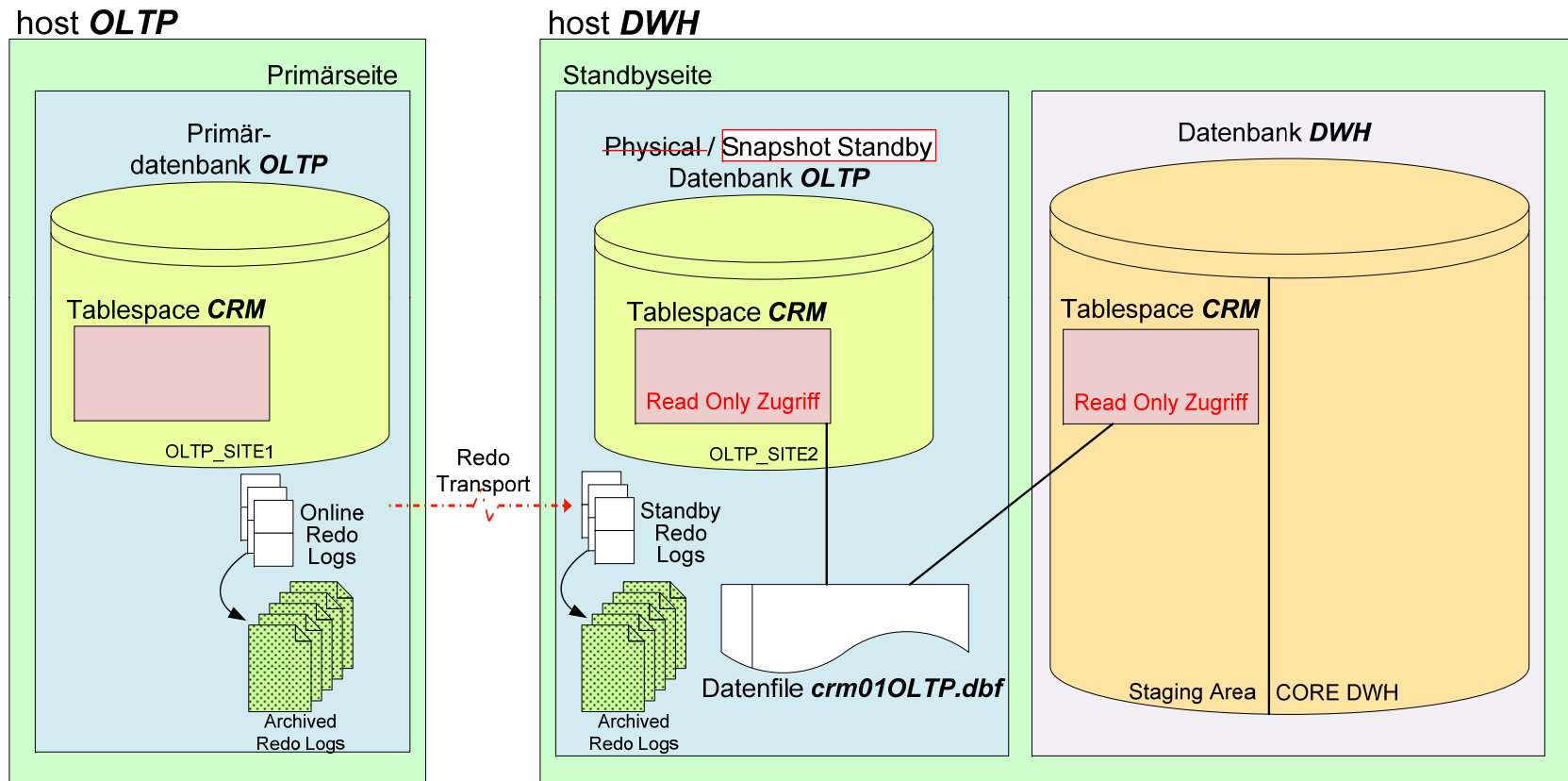
- Die Herausforderung “Extraktion & Laden“
- **Lösung mittels Oracle Data Guard**
  - Konzept
  - Demo
- Zusammenfassung

# Lösungsansatz – die Übersicht



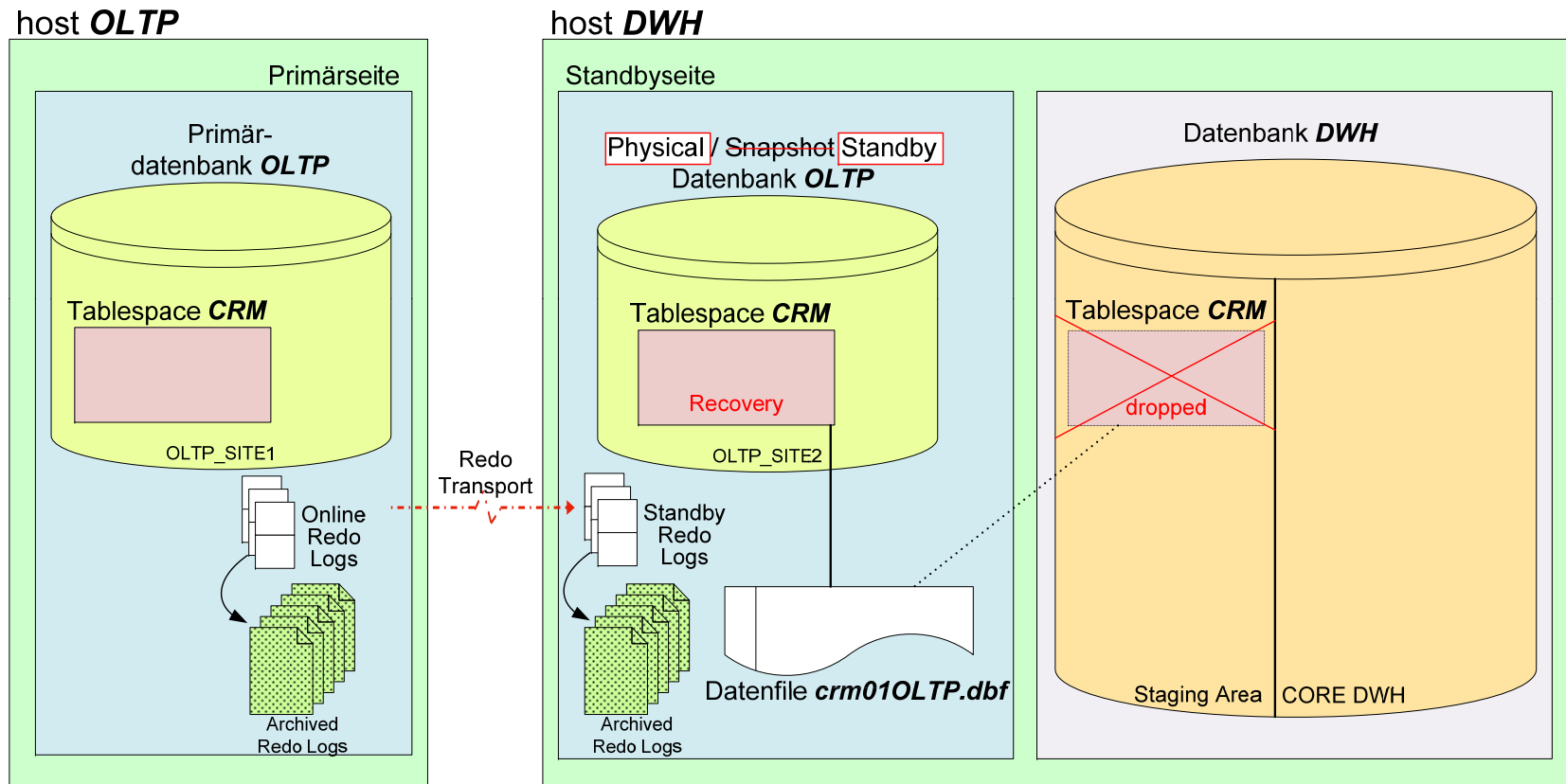
- Physical standby Datenbank von OLTP auf dem DWH host

# DWH-Datenbewirtschaftung will aus Staging Area lesen



- CRM Tablespace ist in beiden Datenbanken angeschlossen

# Staging Area soll aktualisiert werden



- Physikalisches Media Recovery läuft sehr effizient, um den Tablespace zu aktualisieren

# Die wichtigsten Eigenschaften der Lösung



- Vorteile für Architektur des DWHs
  - Staging area (SA) enthält den vollen Datenbestand
  - Anpassen der Statistiken möglich (z.B. Erstellen der Histogramme)
  - Erstellen von zusätzlichen Indexen möglich
  
- Effizient
  - Keine zusätzliche Last auf dem OLTP host
  - Datenfiles (Vollbestand) werden weder transportiert noch kopiert
  - transferiertes Datenvolumen (OLTP → DWH) hängt ausschließlich von dem Volumen der Änderungen ab

# Die wichtigsten Eigenschaften der Lösung (Fortsetz.)



- Performance = kurze Dauer der Aktualisierung
  - Dauer hängt nicht von der Größe der Tablespaces ab, sondern nur von dem Volumen der Datenänderungen seit letzter Aktualisierung
  - Dauer kann noch optimiert werden: solche Tablespaces, welche bei der DWH-Datenbewirtschaftung nicht benötigt werden, können aus dem Recovery process ausgeschlossen werden
  - Kopieren der archivierten Redo Logs vom OLTP zum DWH passiert bereits vor der Aktualisierung, daher beeinflusst es die Dauer nicht
    - Standby Seite empfängt Logs von der Primärseite in beiden Modi: Physical Standby und Snapshot Standby.
    - Im Snapshot Standby Modus werden die Logs gequeued und nicht appliziert
    - Dadurch, dass der Log Transport permanent läuft, sind die ausstehenden Logs bereits verfügbar und registriert, wenn der Recovery Process anstartet

# Vergleich mit Oracle Streams / Oracle GoldenGate



- Aufsetzen & Betrieb von Physical Standby Datenbanken ist wesentlich einfacher
  
- In Streams / GoldenGate Umgebungen muss man im Betrieb einiges beachten:
  - DDL Replikation „richtig“ zum Laufen bringen
  - einige Bugs
  - größere Gaps können schnell unauflösbar werden, da die Performance nicht ausreichend ist
  - Trigger müssen „Replikationsfreundlich“ aufgesetzt werden
  - „on delete cascade“ Constraints stellen ein Problem für die Replikation dar
  - Datendivergenzen müssen erkannt und ausgeglichen werden
  - Troubleshooting für Gecrashte Apply bzw. Replicat Prozesse
  - Alle replizierten Tabellen brauchen einen Primary Key
  - Man braucht Supplemental Logging auf Primary Keys

# Aktualisierung der SA – Alternativansätze im Überblick



Konzept / Implementierung	Transferiertes Datenvolumen	Vollständigkeit	Einfluss auf Performance des Quellsystems	Einfluss auf Verfügbarkeit / Stabilität des Quellsystems	Komplexität des Betriebs
Daten-Vollextraktion					
Marker-basierte Extraktion					
Journal-basierte Extraktion					
Oracle Streams			/		/
Oracle Golden Gate			/		/
Neue Lösung mit Data Guard und Transportable Tablespace					



# Voraussetzungen



- **Identischer Datenbank Zeichensatz**
- **Self-contained Tablespace Sets**
- **Benötigte Oracle Datenbank Releases**
  - *OLTP*: Release 10g oder höher
    - 11g empfohlen: Snapshot Standby Feature ist hier verfügbar
    - 10g möglich
      - Snapshot Standby Feature kann mit einfacher Implementierung emuliert werden
      - Kein Redo log Transfer möglich, während des emulierten „snapshot“ standby Modus
  - *DWH*: Gleiches oder höheres Release als *OLTP*
- **Benötigte Oracle Lizenzen**
  - Oracle Enterprise Edition für sowohl *OLTP* als auch *DWH*
  - Keine zusätzlichen Optionen erforderlich
    - Weder Partitioning noch Active Data Guard

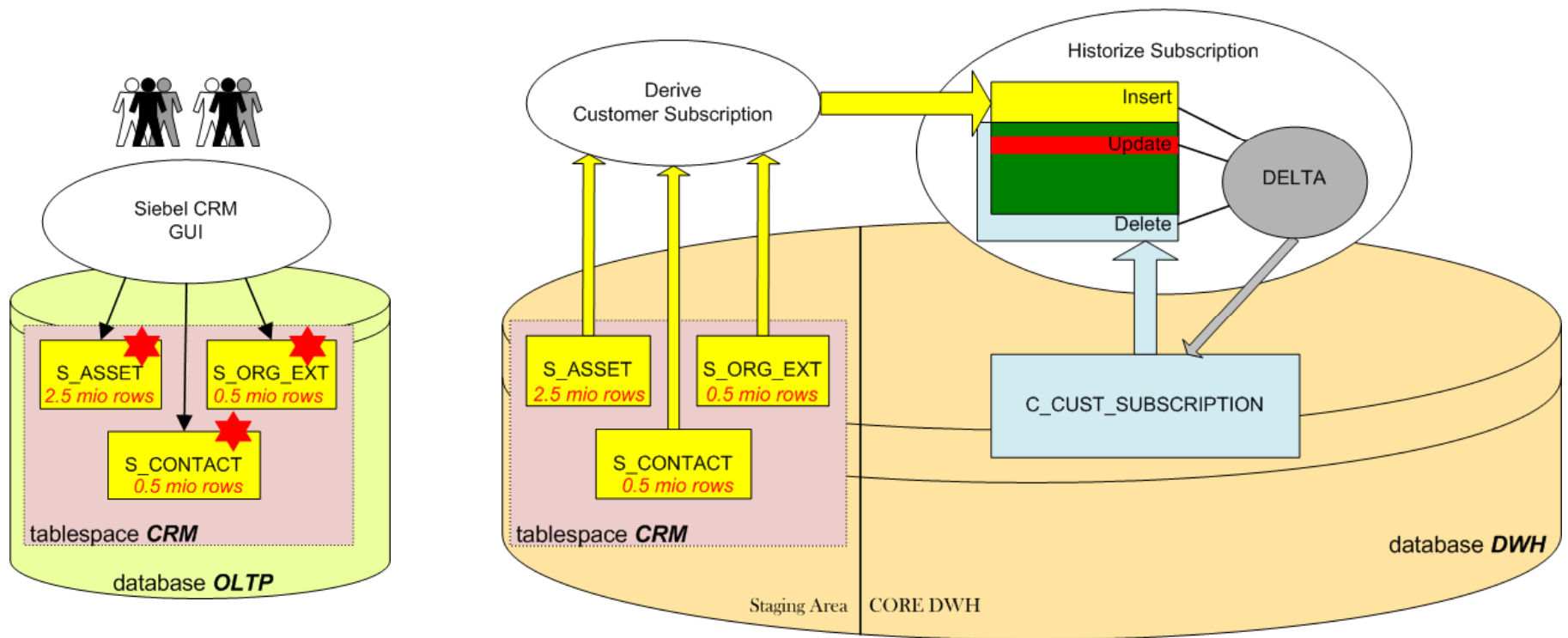
# Agenda



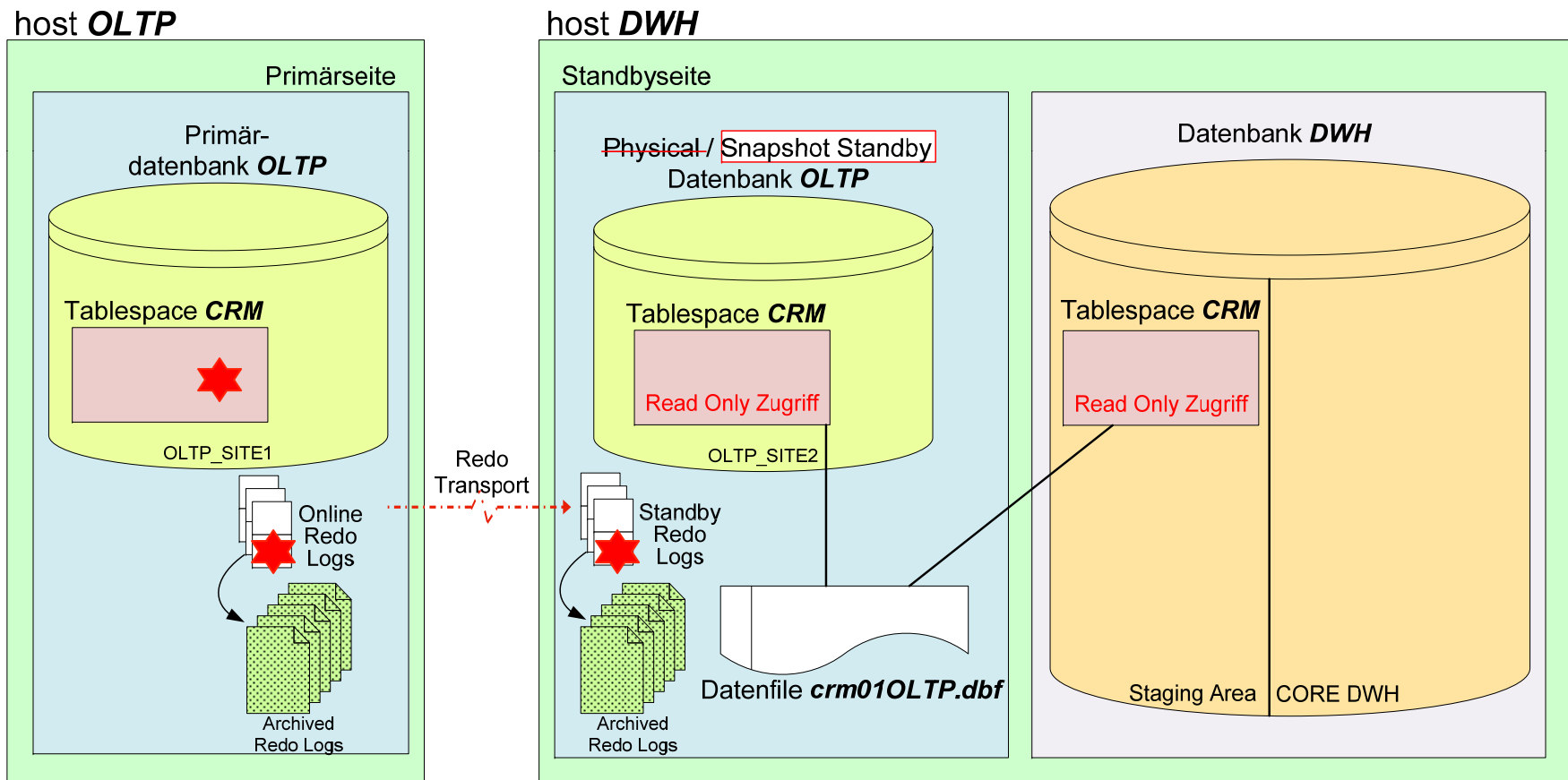
Wissen vermitteln ist der Anfang. Wissen umsetzen das Entscheidende.

- Die Herausforderung “Extraktion & Laden“
- **Lösung mittels Oracle Data Guard**
  - Konzept
  - **Demo**
- Zusammenfassung

# Während CRM-Anwender die Daten in OLTP DB ändern, Staging Area hält den letzten Snapshot



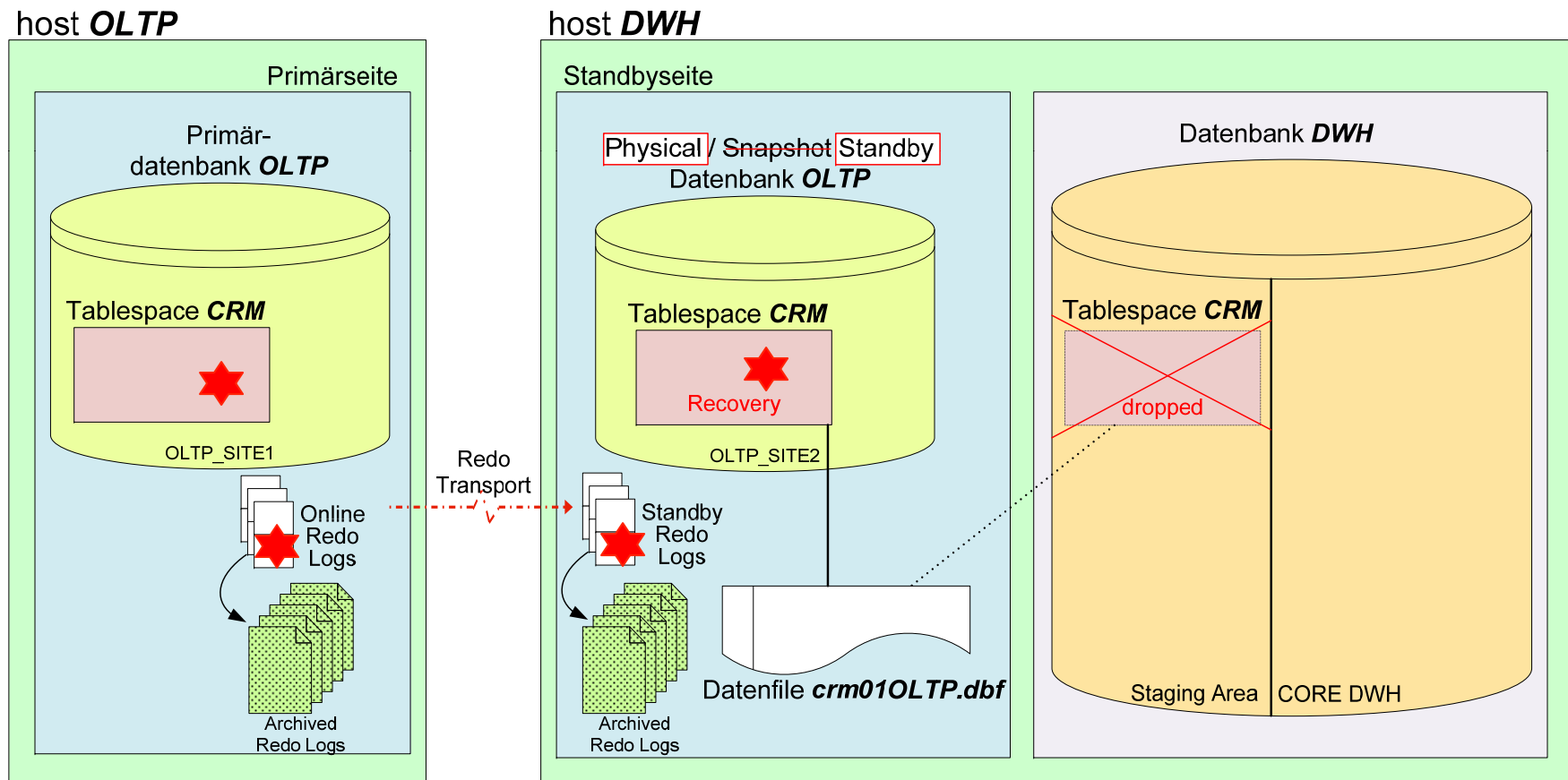
# Staging Area soll aktualisiert werden: Snapshot Standby Datenbank → Physical Standby Datenbank



# Physical Standby Datenbank wird synchronisiert



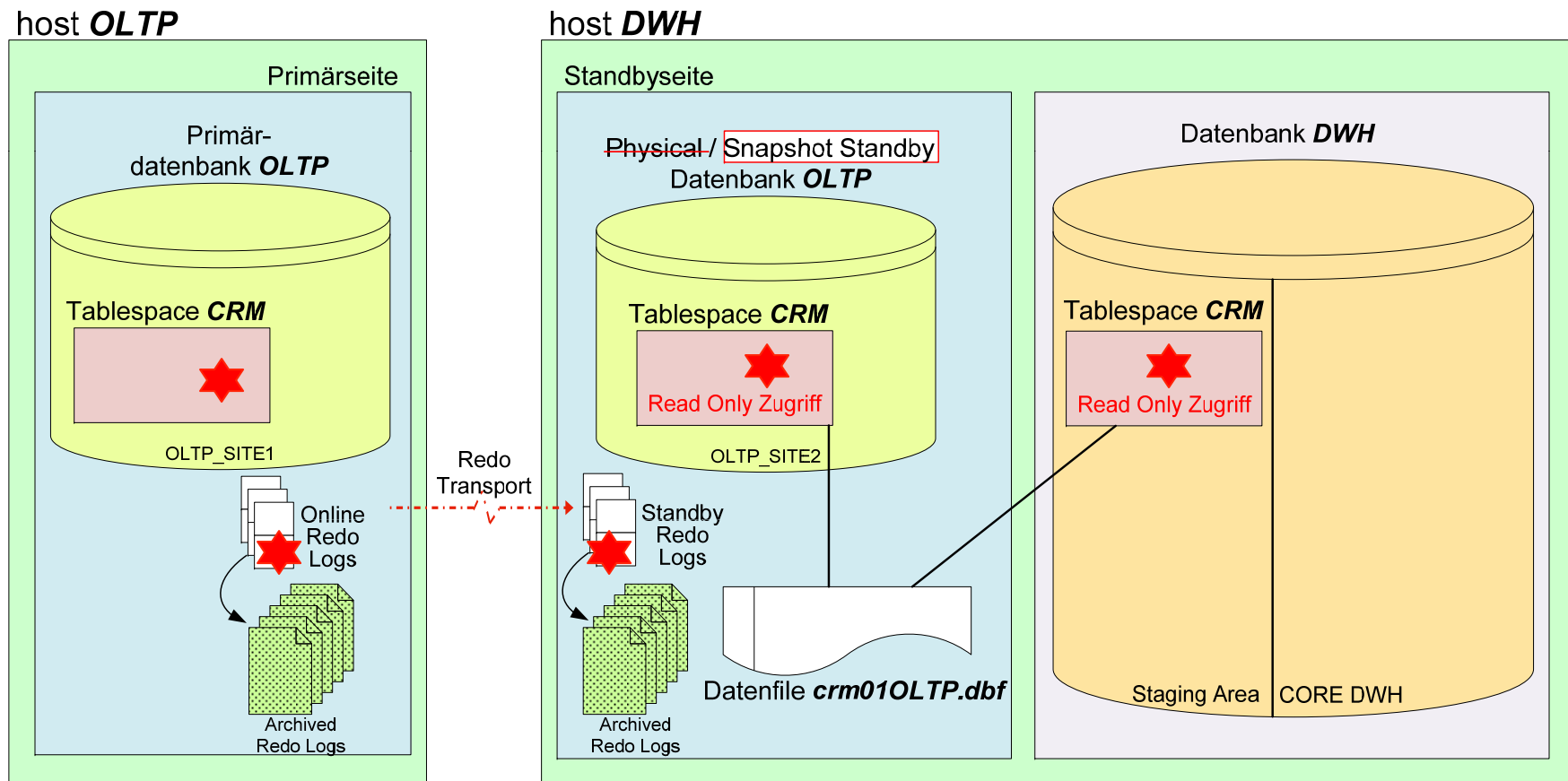
- „kein“ Lag zwischen Primary and Standby (in-sync)
- Staging Area ist für DWH Load nicht verfügbar



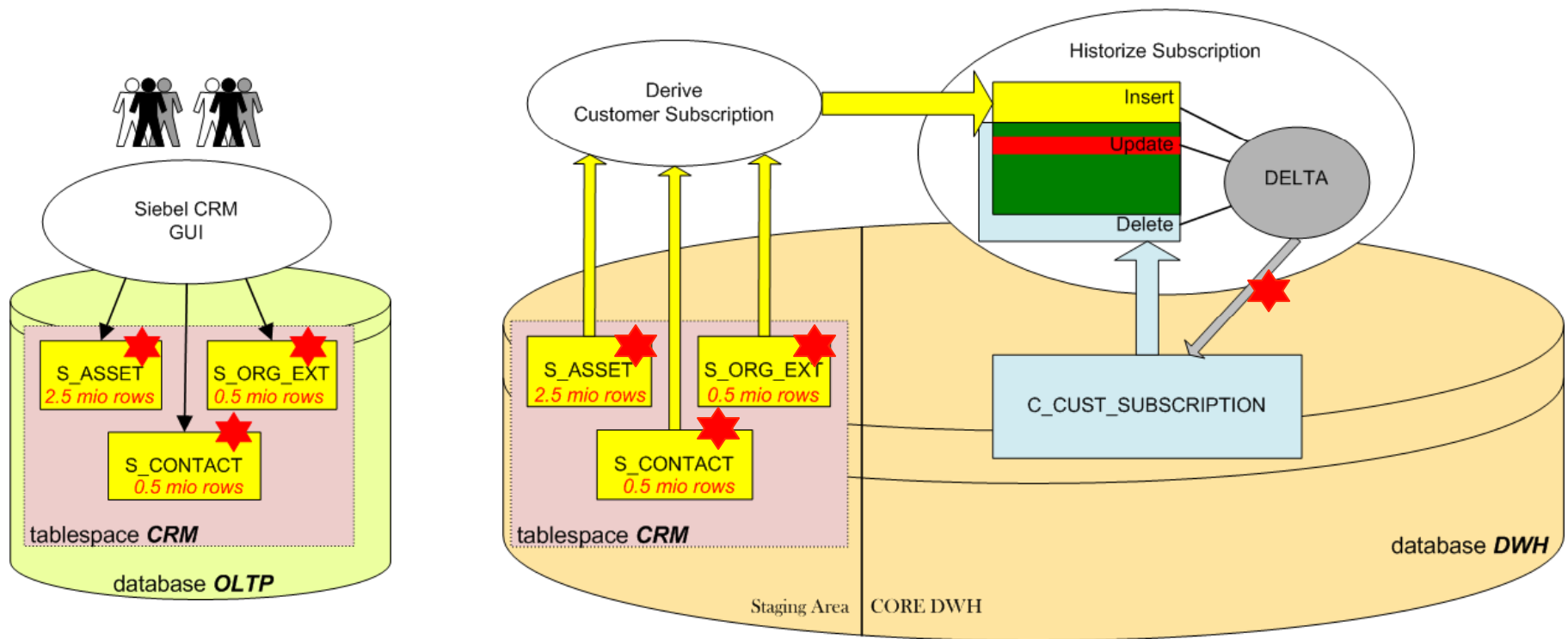
# Snapshot in Staging Area zur Verfügung stellen



- Physical Standby Datenbank → Snapshot Standby Datenbank
- Macht die SA für DWH verfügbar (Tablespace Plug-in)



# Neuer Snapshot in Staging Area wird verwendet, um die Stammdaten im CORE DWH nachzuführen



# Ihre Fragen zur Demo?



Weitere Details unter <http://www.trivadis.com> :

**Solution for Staging Area in Near Real-Time DWH –  
Efficient in Refresh and Easy to Operate**

**Technical White Paper**

Mathias Zarick, Karol Hajdu  
Senior Consultants  
March-2011



# Agenda



Wissen vermitteln ist der Anfang. Wissen umsetzen das Entscheidende.

- Die Herausforderung “Extraktion & Laden“
- Lösung mittels Oracle Data Guard
  - Konzept
  - Demo
- **Zusammenfassung**

# Zusammenfassung – die Vorteile dieser Lösung



- Jede periodische Extraktion & Laden (Ora→Ora) kann profitieren
  
- Meiste Vorteile sind in Fällen:
  - A. Extraktion aus OLTP datenbanken mit **komplexen Datenmodell** und **großem Datenvolumen**
    - OLTP Applikationen, welchen den Lify-Cycle von **Millionen individuellen Subjekten** verwalten (z.B. Kunden, Lieferanten, Verträge, Produkte, Policies usw.)
    - **CRM, SCM o.ä.**
  
  - B. Extraktion aus **Transaktionalen Systemen**
    - **Hohe Insert-Rate**, aber auch einige Änderungen
    - Bookings, Rechnungen, Zahlungen, Transaktionen, rated CDRs,...

■ ■ ■ Vielen Dank für Ihre Aufmerksamkeit!



?

[www.trivadis.com](http://www.trivadis.com)

**trivadis**  
makes IT easier. ■ ■ ■



Basel

Bern

Lausanne

Zurich

Düsseldorf

Frankfurt/M.

Freiburg i. Br.

Hamburg

Munich

Stuttgart

Vienna