

Impact-Analyse und Data Lineage in Siebel CRM und Oracle BIEE mit Hilfe von Oracle BI

Metadaten-Tool von Riverland

**Robert Bitunjac
Riverland Solutions GmbH
München**

Schlüsselworte:

Impact-Analyse, Impact, Siebel, Siebel CRM, Migration, Reporting, Reports, Oracle BI, Oracle, IO, Integration Object, BO, Business Object, Metadaten-Tool, Informatica, Datenmodell, Data Lineage, Datenherkunftsanalyse, Datenbank, Schnittstellen, OLTP, OLAP, DWH, Data Warehouse, RPD

Einleitung

Oft stehen die Siebel-Entwickler vor dem Problem einzuschätzen, welche Auswirkungen eine Feldänderung zur Folge hat. Das gleiche trifft auch im Oracle-BI-Umfeld zu. Das wird oft besonders schwer, wenn das System stark angepasst ist und wird noch schwieriger, wenn das System schlecht dokumentiert ist. Eine Hilfe soll hier die Impact-Analyse sein, die Riverland mit Hilfe von Oracle BI entwickelt hat. Im Vortrag soll mit der Impact-Analyse genau gezeigt werden, welche Auswirkungen Feldänderungen im Siebel-CRM-System und im Oracle-BI-System haben können. Diese Informationen helfen einzuschätzen, welche Aufwände erwartet werden können. Zusätzlich wird eine Data-Lineage-Analyse gezeigt, die Ende-zu-Ende-Beziehungen zwischen Spalten aus dem OLTP und dem OLAP-System auswertet.

Impact-Analyse vs. Data-Lineage-Analyse

Die Impact-Analyse verfolgt ein anderes Ziel als die Data-Lineage-Analyse (Datenherkunftsanalyse). Sie habe auch unterschiedlich Ausmaße. Die Data-Lineage-Analyse soll schlicht die Herkunft eines Datenfeldes als Ergebnis liefern. Während die Impact-Analyse die Auswirkung einer Feldänderung innerhalb von Siebel CRM und Oracle BIEE analysiert. Ein Beispiel: Für die Befüllung eines Datenfeldes, W_ORG_D.Col_A, im DWH (Data Warehouse) sind zwei Spalten, Tbl_x.Col_A und Tbl_x.Col_B, verantwortlich. Zusätzlich ist eine weitere Spalte (Tbl_x.Col_C) innerhalb eines Filters dafür verantwortlich, dass die Ergebnismenge eingeschränkt wird. Für die Datenherkunftsanalyse ist die Information über den Filter nicht relevant, aber für die Impact-Analyse kann diese Information sehr wohl relevant sein. Darüber hinaus kann es für die Datenherkunftsanalyse schwieriger herauszufinden sein, wo der Ursprung eines Feldes liegt. Denn ein Lookup innerhalb eines Informatica-Mappings kann genauso die Quelle für ein Datenfeld sein, wie eine eigene Feldzusammensetzung innerhalb einer Transformation, die auf bestimmte Inhalte eines Datenfeldes reagiert und eigene Werte einsetzt.

Hier nochmal eine Übersicht der Unterschiedlichen Zielsetzungen beider Analysen.

| Data Lineage | Impact-Analyse |
|---|---|
| <ul style="list-style-type: none">• Die Data Lineage zeigt eine „Ende-zu-Ende“-Beziehung der Daten. Sie ist auf die fachliche Information fixiert, die hinter einem Feld steckt. Der technische Hintergrund spielt hier keine Rolle.• Typischerweise zeigt sie den Fluss der Informationen von der operativen Anwendungs-GUI (hier Siebel CRM) hin zu den Berichten (hier Oracle-BI-Berichte), wo die Informationen gebraucht werden.• Der Startpunkt und Endpunkt der fachlichen Information kann dynamisch sein. Er muss nicht statisch sein. | <ul style="list-style-type: none">• Die Impact-Analyse ist auch eine Data-Lineage-Analyse. Sie ist aber auf einer viel detaillierteren und technischeren Ebene.• Sie hat das Ziel die Auswirkungen von Informationsänderungen nachzuvollziehen. Sie versucht eine Indikation für die möglichen Aufwände zu geben.• Sie soll möglichst alle betroffenen Elemente innerhalb des Siebel-CRM- und dem Oracle-BI-Systems aufzeigen.• Es gibt mehrere Endpunkte für einen Startpunkt. Die Endpunkte enden nicht zwangsläufig in einem Oracle BI Bericht. |

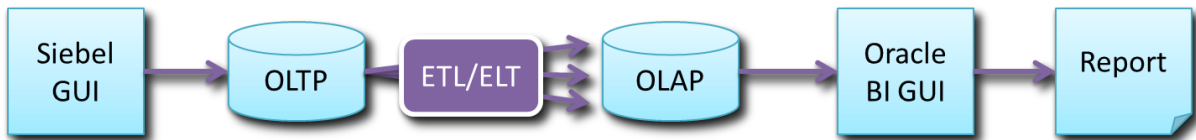


Abbildung 1 Der Datenfluss aus der Sicht des Oracle-BI-Systems

Komplexität der Analysen

Die ETL-Strecken, die mit Informatica realisiert werden, stellen die größte Herausforderung bei der Analyse dar. Dabei ist es nicht wichtig, ob sich dabei um die Data-Lineage-Analyse oder um die Impact-Analyse handelt. Bei beiden Analysen befindet sich die größte Komplexität innerhalb der Beladungen des DWHs. Das liegt vor allem daran, dass mit Hilfe der Informatica-Mappings komplexe Transformationen möglich sind.

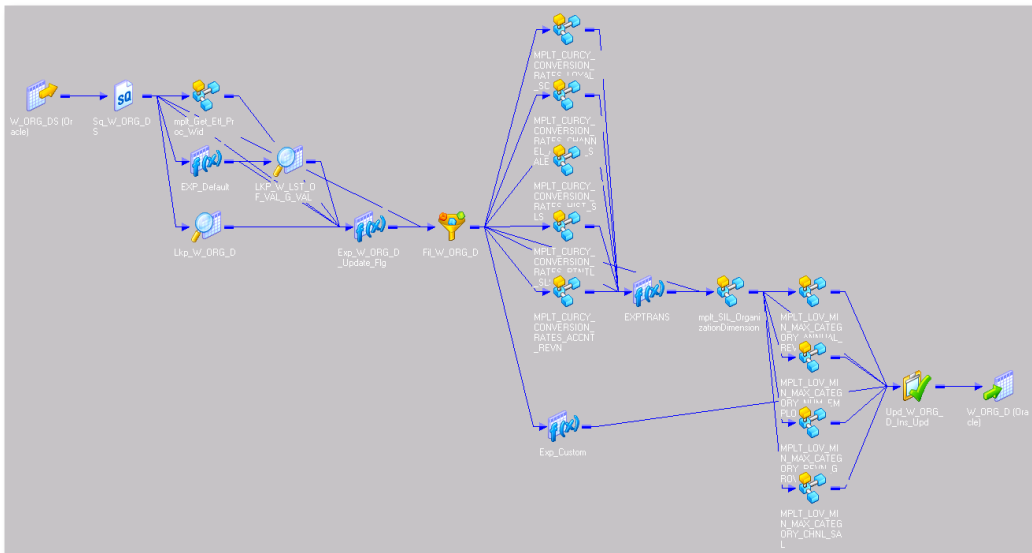


Abbildung 2 Beispiel eines Informatica-Mappings

Nicht alle Felder im OLTP landen in einem Oracle BI Bericht. Manche Spalten werden nur aus technischen Gründen in Mappings verwendet (das trifft zum Beispiel bei Auxiliary-Mappings zu). Sie haben also keinen Endpunkt im RPD. Das muss berücksichtigt und als Hintergrundwissen vorausgesetzt werden, will man die Analysen korrekt auswerten.

Voraussetzungen für Data Leneage und die Impact-Analyse

Die Impact-Analyse von Riverland geht von zwei Betrachtungswinkeln aus. Erstens, die Auswirkungen einer Änderung im Siebel-CRM-System auf das Siebel CRM selbst und das angeschlossene Oracle BI. Und zweitens, die Auswirkungen einer Änderungen im Oracle-BI-System

auf das Oracle BI und Siebel CRM. Von diesen zwei Betrachtungswinkeln ausgehend, stehen unterschiedliche Dashboards und Berichte zur Verfügung, die eine Analyse ermöglichen. Folgende Punkte sind also Voraussetzungen für eine erfolgreiche Analyse:

- Für beide Analysen, Impact-Analyse und Data Lineage, muss eine Datenflussrichtung vorgegeben werden. Aufgrund der Datenflussrichtung kann dann begonnen werden eine Auswertung zu erstellen.
- Für die Data-Lineage-Analyse gelten als Startpunkte entweder die Tabellen in der physikalischen Schicht des RPD oder die OLTP-Tabellen. Und sie enden dann jeweils im OLTP oder RPD.
- Für die Impact-Analyse können die Startpunkte entweder die Tabellen in der physikalischen Schicht der RPDs oder die OLTP-Tabellen sein. Die Endpunkte sind nicht fest definiert und können überall zwischen dem OLTP und dem OLAP-System liegen. Es werden alle Elemente berücksichtigt und aufgelistet, für die ein Zusammenhang zum Startpunkt besteht.

Um klarzustellen, welche Metadaten-Elemente bei der Analyse berücksichtigt werden, gibt es hier eine Auflistung.

Innerhalb des Siebel-CRM-System werden folgende Elemente berücksichtigt:

- Datenbank (Spalten, Tabellen)
- IO (Integration Objects)
- BO (Business Objects)
- View (Felder)
- Applets

Innerhalb des Oracle-BI-Systems werden folgende Elemente berücksichtigt:

- DWH (Spalten, Tabellen)
- ETL (Informatica Mappings, Workflows)
- RPD (Physical Layer, Logical Layer, Presentation Layer)
- Web Catalog (Berichte, Dashboard)
- BIP (BI Publisher)

Was den Oracle BI Publisher angeht, so wird im Falle der Data-Lineage-Analyse folgender Informationsfluss hauptsächlich von Interesse sein.

BIP-Report(-Feld) → IO → IC → IC Field → BC Field → Database Column

Siebel Elemente: IO = Integration Object, IC = Integration Component, IC Field = Integration Component Field, BC = Business Component

Grenzen und Probleme der Impact-Analyse / Data Lineage

Der Vortrag wird auch die Probleme und Grenzen der Impact-Analyse bzw. Data-Lineage-Analyse aufzeigen. Diese ergeben sich zum einen aufgrund der Eigenarten der ETL-Software (Informatica). Als Beispiel seien hier die Brüche genannt, die innerhalb des Datenflusses der ETL-Strecken entstehen können (Verwendung von Mapplets in Informatica-Mappings). Zusätzlich erschwert wird das Ganze durch die nicht vorhandene Dokumentation darüber, wie Informatica seine Metadaten speichert.

Dashboards und Berichte für die Impact- und Datenherkunftsanalyse

Im Vortrag werden vorgefertigte Dashboards und Berichte präsentiert, die Analysen hinsichtlich der Auswirkungen von Feldänderungen und der Datenherkunft ermöglichen. Sie sollen einen Überblick über die Möglichkeiten geben.

| Newid | Rownumid | Levelh | Direction | Subarea | Mappingname | Sourcetable | Sourcecolumn | Targettable | Targetcolumn | Lkptable | In_lkpcolumn | Out_lkpcolumn | Comments | Datasource | Datarget |
|-------|----------|--------|-----------------|---------|-------------------------------------|-------------|---------------|-------------|---------------|-----------|--------------|---------------|--|------------|----------|
| 50,00 | 50,00 | 1,00 | | SDE | Extracting Organisation Dimensio | S_ORG_EXT | TYPE | W_ORG_DS | PAR_ORG_NAME | S_ORG_EXT | PAR_TYPE | ORG_NAME | Scenario 2 Part 1 - Normal SDE extracting data with join (lkp) with 4 lkp values and getting back 3 returnvalues | SIEBEL | STAGE |
| 52,00 | 52,00 | 1,00 | SIEBEL -> STAGE | SDE | Extracting Organisation Dimensio | S_ORG_EXT | TYPE | W_ORG_DS | PAR_ORG_NAME2 | S_ORG_EXT | PAR_TYPE | ORG_NAME2 | Scenario 2 Part 1 - Normal SDE extracting data with join (lkp) with 4 lkp values and getting back 3 returnvalues | SIEBEL | STAGE |
| 54,00 | 54,00 | 1,00 | | SDE | Extracting Organisation Dimensio | S_ORG_EXT | TYPE | W_ORG_DS | PAR_ORG_NAME3 | S_ORG_EXT | PAR_TYPE | ORG_NAME3 | Scenario 2 Part 1 - Normal SDE extracting data with join (lkp) with 4 lkp values and getting back 3 returnvalues | SIEBEL | STAGE |
| 50,00 | 51,00 | 2,00 | | SIL | Loading into Organisation Dimension | W_ORG_DS | PAR_ORG_NAME | W_ORG_D | PAR_ORG_NAME | | | | Scenario 2 Part 2 - Normal SIL loadin data coming from Siebel | STAGE | DWH |
| 52,00 | 53,00 | 2,00 | STAGE -> DWH | SIL | Loading into Organisation Dimension | W_ORG_DS | PAR_ORG_NAME2 | W_ORG_D | PAR_ORG_NAME2 | | | | Scenario 2 Part 2 - Normal SIL loadin data coming from Siebel | STAGE | DWH |
| 54,00 | 55,00 | 2,00 | | SIL | Loading into Organisation Dimension | W_ORG_DS | PAR_ORG_NAME3 | W_ORG_D | PAR_ORG_NAME3 | | | | Scenario 2 Part 2 - Normal SIL loadin data coming from Siebel | STAGE | DWH |

Abbildung 3 Impact-Analyse-Bericht

Der Einsatz beim Kunden

Es wird kurz darauf eingegangen, was nötig ist, um beim Kunden die Impact-Analyse und Data Lineage einsetzen zu können. Einige kleine Vorbereitungen und Analyse der Ist-Situation sind nötig, bevor die Metadaten-Analyse starten kann. Vor allem das Maß der Abweichung vom Standard des Siebel-CRM-Systems und deren Dokumentation entscheidet über den Aufwand, der betrieben werden muss, damit eine funktionsfähige Impact- und Data-Lineage-Analyse ausgeführt werden kann.

Nutzen

Welchen Nutzen kann durch die automatisierte Impact- und Data-Lineage-Analyse gezogen werden? Ist erst einmal die Metadaten-Analyse beim Kunden etabliert, so sollten sich folgende Vorteile einstellen:

- Eine schnelle Einschätzung der Aufwände bei Feldänderungen und -Umzügen.
- Das wiederum reduziert den Aufwand, der ohne das Metadaten-Tool für die Beantwortung dieser Frage nötig wäre.
- Business Analysten und Tester können einen großen Teil ihrer Analysen und Tests ohne die Unterstützung der Entwickler durchführen.
- Man erhält eine Dokumentation der Datenflüsse und Datenherkunft

Voraussetzungen (Software)

Für die Analyse wird Oracle BIEE zusammen mit Oracle Business Intelligence Applications (Informatica, DAC) benötigt. Riverland hat hierfür ETL-Strecken und Datenmodelle entwickelt, welche die Auswertung einer Impact-Analyse und einer Data-Lineage-Analyse ermöglichen.

- Siebel CRM ab 7.8.x
- Oracle BIEE 10 oder 11g
- Oracle BI Applications 7.9.6.x
- Riverland Metadaten-RPD
- Riverland Metadaten-Mappings (Informatica)

Kontaktadresse:

Robert Bitunjac
Riverland Solutions GmbH
Holbeinstr. 22
D-81679 München

Telefon: +49 89 41 073 860
Fax: +49 89 41 073 862
E-Mail: robert.bitunjac@riverland.com
Internet: www.riverland.com