Warenkorbanalyse

Dokumentation zum Hauptprojekt

Inhalt

[1 Einleitung 2](#_Toc359186021)

[2 Systemlandschaft 3](#_Toc359186022)

[3 Datenarchitektur 5](#_Toc359186023)

[3.1 Quelltabellen 5](#_Toc359186024)

[3.2 Datenaufbereitung in DI Jobs 7](#_Toc359186025)

[3.2.1 Jobs 7](#_Toc359186026)

[3.2.2 Datentransfer 10](#_Toc359186027)

[3.2.3 Job-Management 10](#_Toc359186028)

[3.2.4 Historisierte Tabellen in DWH (SAS\_Analyse\_Mart) 11](#_Toc359186029)

[3.2.5 Tabellen in WKA (SAS) 11](#_Toc359186030)

[3.2.6 Tabellen in DB (WebShop) 11](#_Toc359186031)

[4 Scoring-Modul 11](#_Toc359186032)

# Einleitung

Die Warenkorbanalyse soll für die Identifikation häufig gemeinsam konsumierter Warengruppen/Produkte genutzt werden. Ziel ist es, jedem Kunden im Globus Drive Webshop eine möglichst individuell abgestimmte Leistung, in Form von Produktempfehlung, anbieten zu können und dem Kunden einen Mehrwert zu schaffen.

ToDo: In der Einleitung sollte noch kurz erwähnt werden, was dieses Dokument beschreibt. Systemlandschaft, Architektur, usw.

Teilnehmer des Hauptprojektes (04.06. – 11.06.2013) sind:

|  |  |  |  |
| --- | --- | --- | --- |
| **Abteilung / Firma** | **Verantwortlichkeit** | **Mitarbeiter** | **E-Mail** |
| Firma SAS | Coaching | Christoph Sieb | christoph.sieb@sas.com |
| Controlling | Datenhaltung, SAS | Desislava Fehringer | d.fehringer@globus.net |
| Marketing | Projektleitung | Andrea Linsler | a.linsler@globus.net |

Weitere Beteiligte:

|  |  |  |  |
| --- | --- | --- | --- |
| **Abteilung / Firma** | **Verantwortlichkeit** | **Mitarbeiter** | **E-Mail** |
| Fachabteilung Drive | Project Owner | Anja Weirich | a.weirich@globus.net |
| Firma UNIC | Implementierungspartner | Jens Saade | jens.saade@unic.com |

# Systemlandschaft

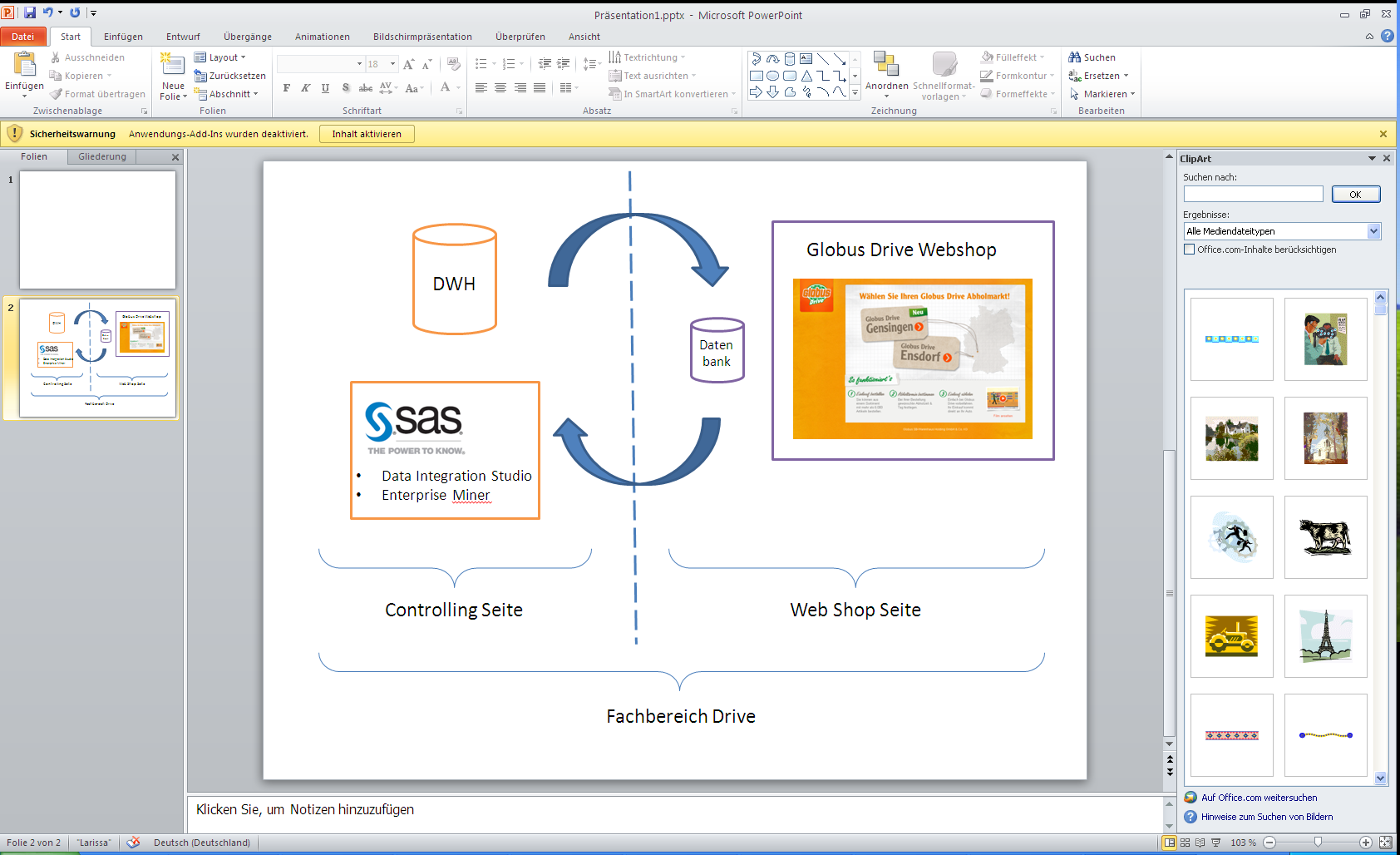


Abbildung 1: Übersicht Systemlandschaft

In dem Unternehmen Globus werden geschäftsrelevante Daten aus verschiedenen Datenquellen in einem konsistenten Datenbestand zusammengeführt und in das Data-Warehouse langfristig gespeichert. Somit ist das Data-Warehouse Ausgangbasis für zahlreiche Datenanalysen. Das Data-Warehause basiert auf Oracle-Technologie (Oracle 11.1).

Für die Regelextraktion kommen zwei SAS Produkte zum Einsatz – das Data Integration Studio 4.2 und der Enterprise Miner 6.1. DI Studio ermöglicht das graphische Design von ETL Stecken, sowie den Zugriff auf eine Vielzahl von Datenquellen und –zielen von Dateien und relationalen Datenbanksystemen. Die Quelldaten werden im DI Studio so aufbereitet, dass sie im dem SAS Enterprise Miner für die Extraktion von Produktempfehlungsregeln eingesetzt werden können.

Die extrahierten Regeln werden anschließend in einer Datenbank in der Nähe des Globus Drive WebShops abgelegt. Bei Produktempfehlungen bedient sich der WebShop aus dieser Regelbasis. Der WebShop ist eine externe Anwendung basierend auf der Hybris-Technologie und wird seit Mai 2013 von der Firma UNIC technisch betreut.

Bei der Interaktion zwischen den Systemen auf Controlling und auf WebShop Seite findet ein Datenaustausch in beiden Richtungen statt. Allerdings ist der Rückfluss ins DWH (Tracking der Produktvorschläge) nicht Teil dieser Projektphase.

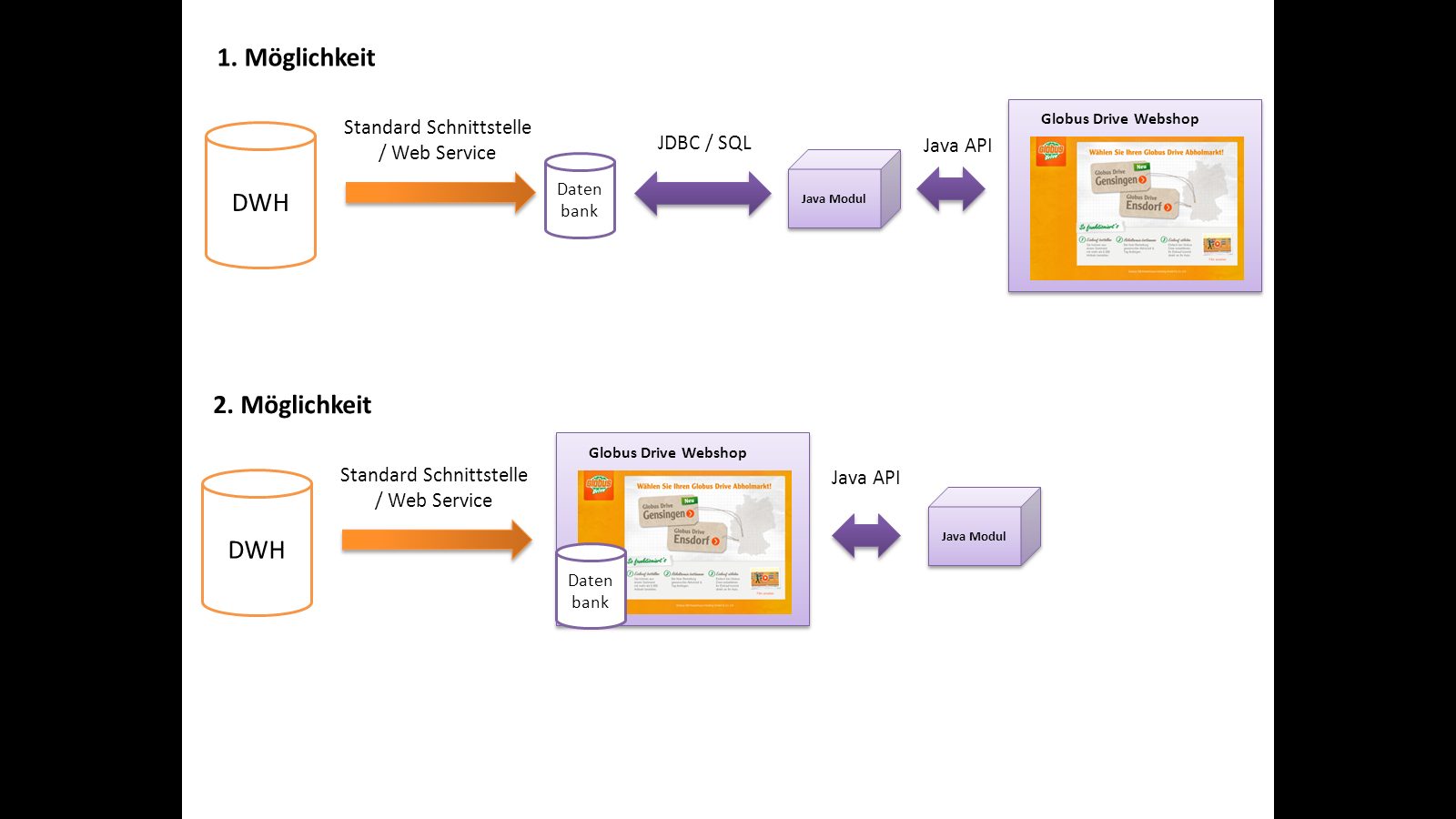


Abbildung 2: Integration der Regelbasis

Um die Integration der Regelbasis zu realisieren gibt es nach momentaner Recherchen 2 Möglichkeiten, die in Abbildung 2 dargestellt worden sind:

1. Die Regelbasis und die dazu gehörigen manuellen Tabellen werden auf separater Datenbank in der Nähe des WebShops abgelegt. Notwendige Schnittstellen und deren Bereitstellung werden noch mit dem Implementierungspartener abgestimmt.
2. Die Regelbasis und die dazu gehörigen manuellen Tabellen werden direkt in dem WebShop abgelegt.

Bevor die Entscheidung für eine der Möglichkeiten getroffen wird, muss eruiert werden wie hoch der Impact auf das produktiv System ist und welche Kosten bzw. Aufwände die Integration mit sich bringt.

# Datenarchitektur

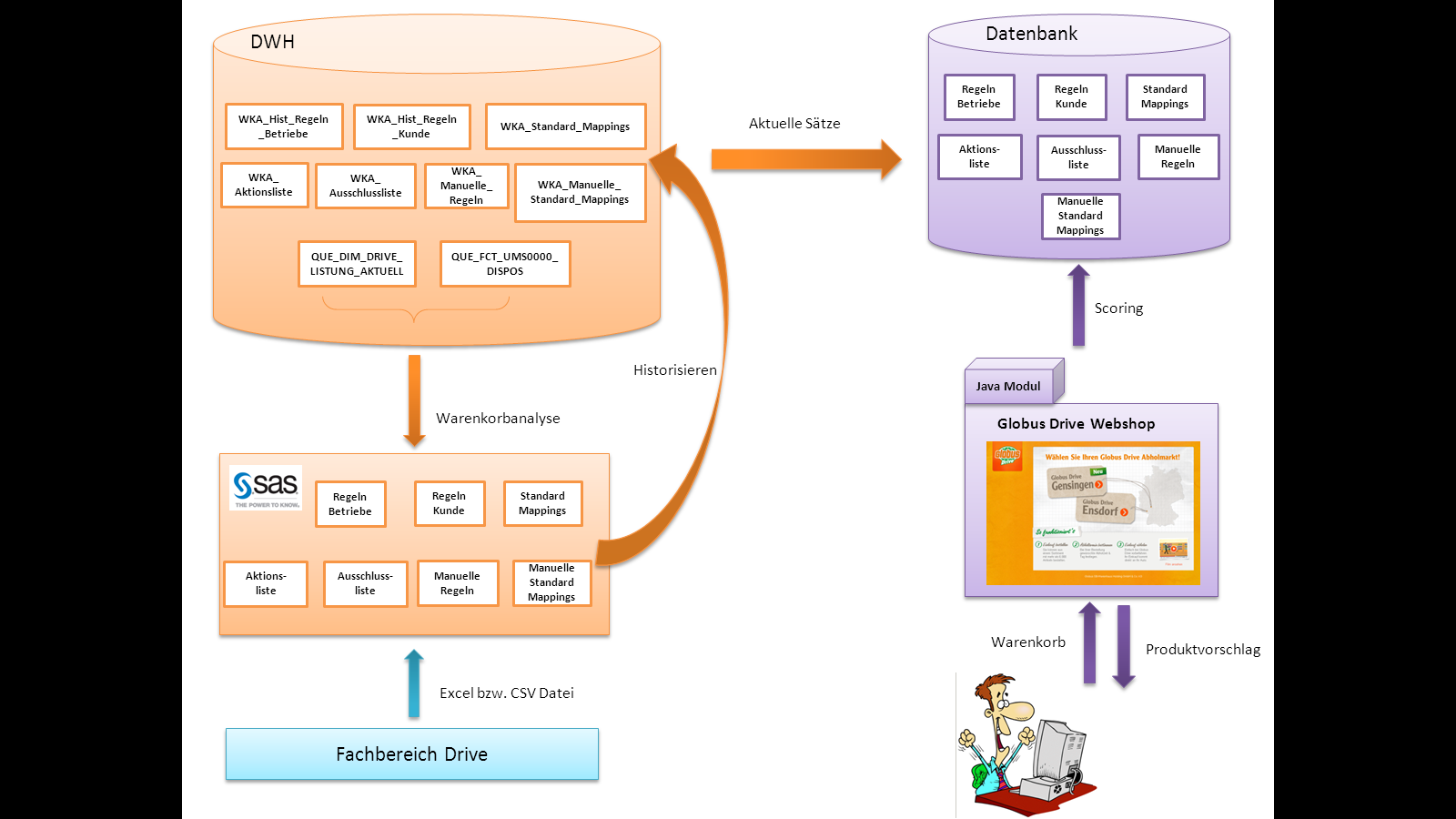


Abbildung 3: Datenaustausch

## Quelltabellen

Als Ausgangsbasis für die Warenkorbanalyse dienen die Views QUE\_DIM\_DRIVE\_LISTUNG\_AKTUELL und QUE\_FCT\_UMS0000\_DISPOS aus der „SAS\_Analyse\_Mart“ Schema in DWH.

**QUE\_FCT\_UMS0000\_DISPOS** – enthält alle Kundentransaktionen (Bonpositionen), die aus DISPOS kommen. Die SAP Bondaten werden zurzeit vernachlässigt, da Drive Märkte noch nicht auf SAP umgestellt sind.

|  |  |  |
| --- | --- | --- |
| **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** |
| BETRIEB\_ID | NUMBER(12) | No |
| DATUM | DATE | No |
| STUNDE | NUMBER | Yes |
| WOCHE\_ID | NUMBER | Yes |
| MONAT\_ID | NUMBER | Yes |
| QUARTAL\_ID | NUMBER | Yes |
| HALBJAHR\_ID | NUMBER | Yes |
| JAHR\_ID | NUMBER | Yes |
| GJ\_JAHR\_ID | NUMBER | Yes |
| BON\_ID | CHAR(40) | Yes |
| KD\_IDX | NUMBER(12) | No |
| KZ\_TANKSTELLE | NUMBER(1) | No |
| MENGE | NUMBER(16,4) | No |
| BRTWA | NUMBER(16,4) | No |
| NETWE | NUMBER(16,4) | No |
| VANR\_ID | NUMBER(12) | No |
| ARTNR | NUMBER | Yes |
| WG1NUM | NUMBER | Yes |
| WG10NUM | NUMBER | Yes |
| WG20NUM | NUMBER | Yes |
| WG30NUM | NUMBER | Yes |
| WG40NUM\_DISPOS | NUMBER(12) | Yes |
| WG40NUM\_SAP | NUMBER(12) | Yes |
| WBID | NUMBER(12) | Yes |
| GWBID | NUMBER(12) | Yes |
| BID | NUMBER(12) | Yes |
| GESAMT\_ID | NUMBER(12) | Yes |

**QUE\_DIM\_DRIVE\_LISTUNG\_AKTUELL** –Übersicht aller gelisteten Drive Artikel, die in dem Drive

Webshop angeboten werden. (Achtung: Die Tabelle enthält nur aktuell gelistete Produkte, d.h., dass Produkte, die ausgelistet wurden, oder Produkte, deren Gültigkeitsdatum in der Zukunft liegt, nicht angezeigt werden). Hier findet das Mapping zwischen Basis\_Variante, SAP-Mini WGR und die SAP B-Ebene statt um eine Produkthierarchie darstellen zu können. Die SAP-Mini WG ist die SAP Hierarchieebene mit der niedrigsten Granularität. Die darüber liegende Ebene ist die B-Ebene. Hierarchie-Ausschnitt:

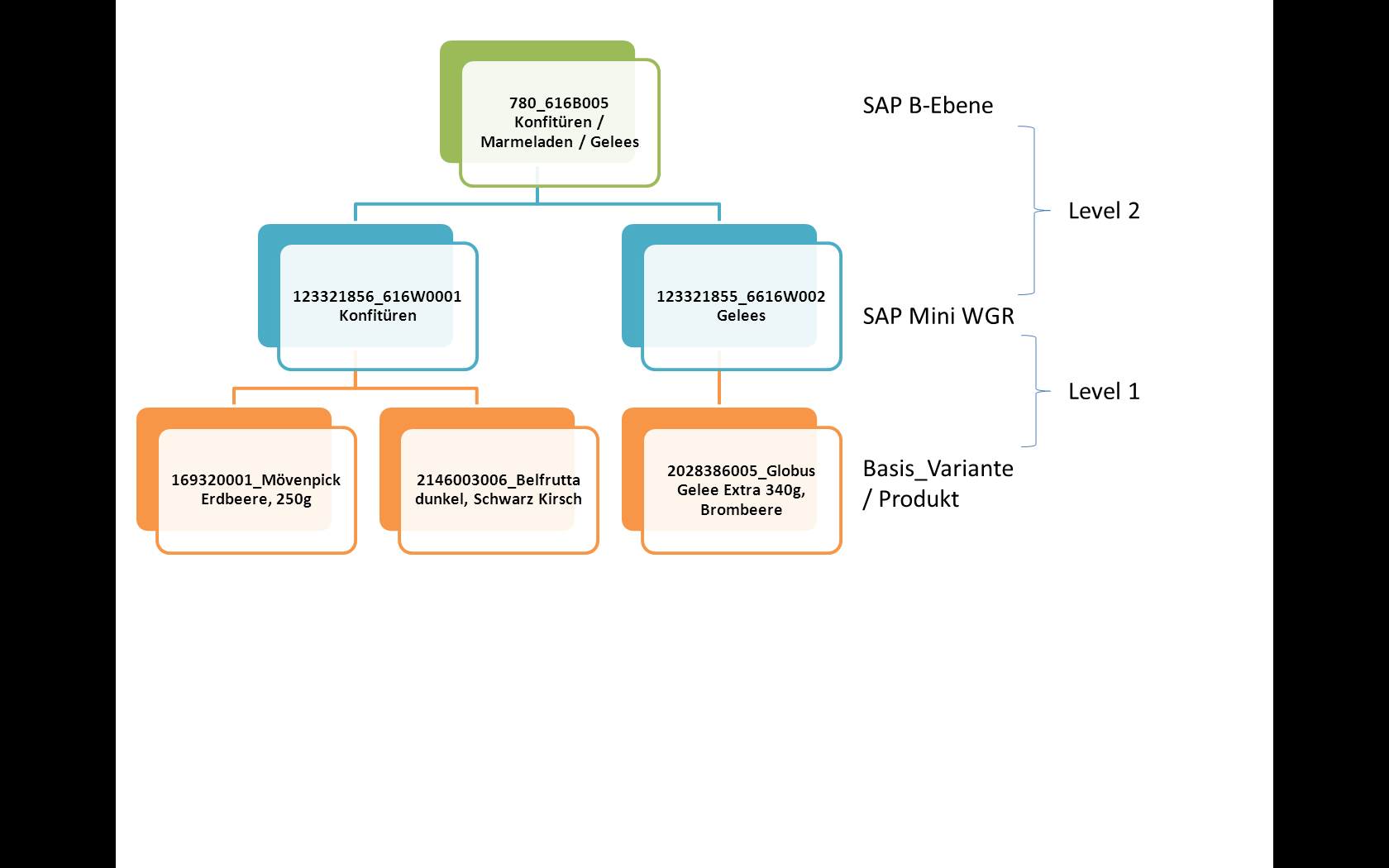


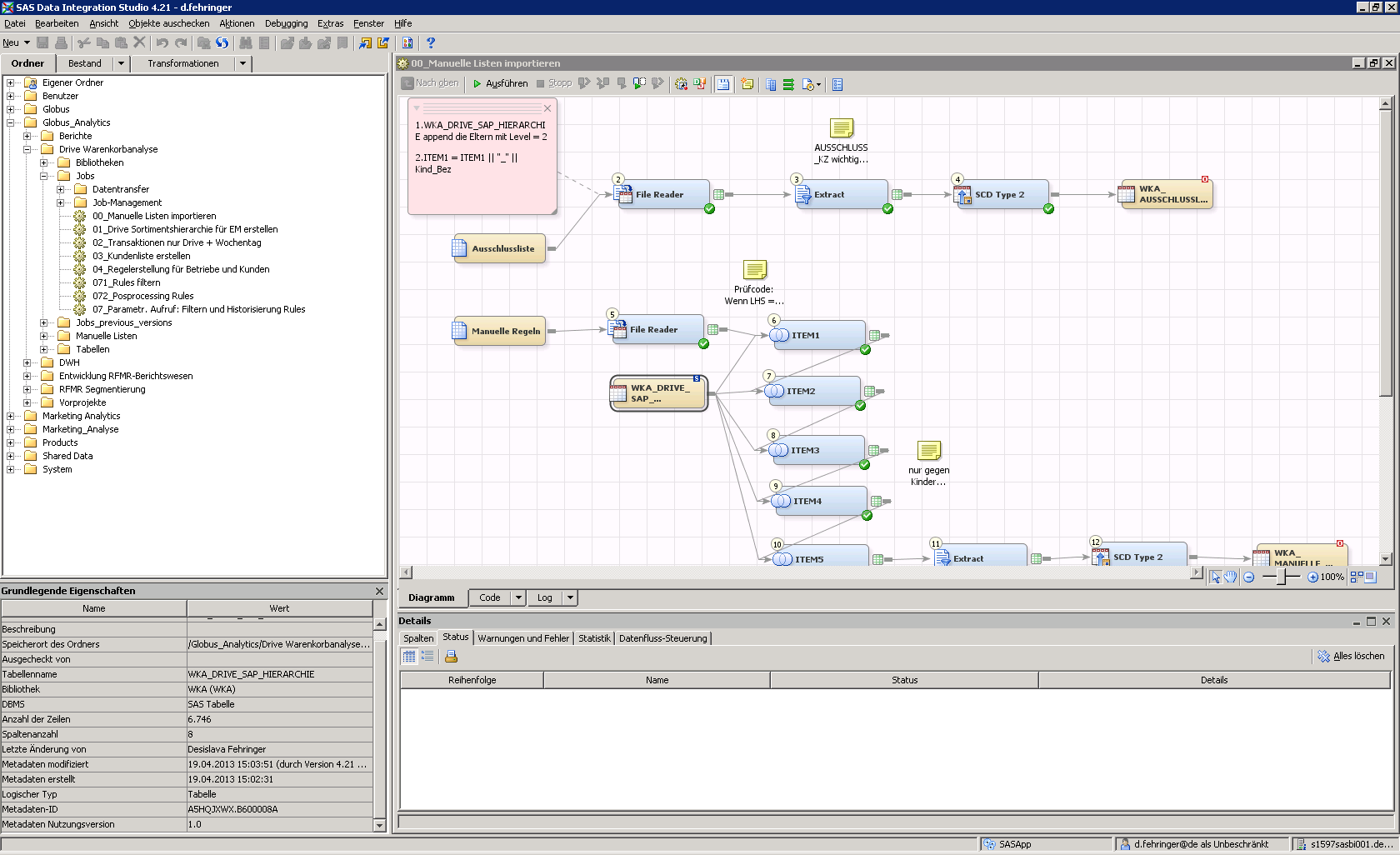
Abbildung 4: Ausschnitt aus der Produkthierarchie

Eindeutiger Schlüssel ist das Attribut „Basis\_Variante“. Die View besteht aus folgenden Feldern:

|  |  |  |
| --- | --- | --- |
| **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** |
| BASIS\_VARIANTE | NUMBER(11) | Yes |
| ARTTXT | VARCHAR2(100 CHAR) | No |
| SAP\_WG\_ID | NUMBER | Yes |
| SAP\_WG\_VALUE | VARCHAR2(721) | Yes |
| B\_EBENE\_ID | NUMBER | Yes |
| B\_EBENE\_VALUE | VARCHAR2(721) | Yes |

## Datenaufbereitung in DI Jobs

Im DI Studio wird für die ETL-Strecken einen eigenen Projektordner angelegt: /Globus\_Analytics/Drive Warenkorbanalyse. In diesem Ordner sind die Bibliotheken, Jobs, Tabellen, und Manuelle Listen enthalten, die in dem Projekt zum Einsatz kommen.



### Jobs

*00\_Manuelle Listen importieren*

**Input:** WKA\_DRIVE\_SAP\_HIERARCHIE sowie die Manuellen Listen, die wir als Excel oder CSV-Datei von dem Fachbereich Drive bekommen. Das sind im Einzelnen:

* Aktionsliste: Produkte, die in Aktionen teilnehmen z.B. Osteraktion mit Ostereier-Suche. Diese Produkte tragen nicht zum Erhöhen des Bon-Umsatzes, kommen aber trotzdem auf dem Bon als einzelnen Positionen drauf. So würden sie zur Verfälschung der Empfehlungsregeln führen. Damit das nicht passiert, werden sie im Voraus aus den Bon-Transaktionen herausgefiltert.
* Ausschlussliste: Beinhaltet alle Produkte / Produktgruppen, die für das Vorschlagswesen ausgeschlossen werden sollen (Bspw. Lebensmittelskandal 🡪 Lasagne).
* Manuelle Regeln: Ergibt die Möglichkeit manuell bestimmte Regeln fest zu legen. Jeder Regel bekommt einen Importance-Wert, der hoch genug ist (z.Z. 100.000) um später bei der Sortierung vor den automatisch-erstellten Regeln vorgezogen zu werden.
* Manuelle Standard Mappings: In dem Fall, dass genaue Vorstellungen über Produkte bestehen, die beim Auftreten einer bestimmten WGR vorgeschlagen werden sollen bspw. Produkte mit großer Gewinnspanne.

**Job-Beschreibung:** die Manuellen Listen werden, wenn notwendig, mit Zusatzinformationen angereichert (Beispiel: für bessere Lesbarkeit wird die Artikelbeschreibung an der jeweiligen Basis\_Variante angehängt). Darüber hinaus ist das Ziel des Jobs die Tabellen mit einem Von\_ und Bis\_Datum zu versehen und diese anschließend als historisierten Tabellen ins SAS\_Analyse\_Mart abzulegen.

**Output:** WKA\_AKTIONSLISTE, WKA\_AUSSCHLUSSLISTE, WKA\_MANUELLE\_REGELN, WKA\_MANUELLE\_STANDARD\_MAPPINGS (Oracle, SAS\_Analyse\_Mart)

*01\_Drive Sortimentshierarchie für EM erstellen*

**Input:** QUE\_DIM\_DRIVE\_LISTUNG\_AKTUELL

**Job-Beschreibung:** In diesem Job wird aus der aktuellen Listung für Drive Produkte eine Hierarchie gebaut, die den formalen Anforderung des Enterprise Miners entspricht (Kind-Eltern-Beziehung und der entsprechende Level (s. Abbildung 4). Zusätzlich wird die Pfand SAP WG rausgefiltert.

**Output:** WKA\_DRIVE\_SAP\_HIERARCHIE (SAS Table, WKA Bibliothek)

*02\_Transaktionen nur Drive + Wochentag*

**Input:** CTL\_WKA\_DRIVE\_MAERKTE, QUE\_FCT\_UMS0000\_DISPOS, QUE\_DIM\_SORTIMENT, WKA\_DRIVE\_SAP\_HIERARCHIE

**Job-Beschreibung:** Zunächst werden die Bon-Transaktionen gefiltert: nur Bons aus den Drive Häusern und Bons, die für den Berechnungszeitpunkt (die letzten x Monate) relevant sind. Die Drive Häusern sind in der Tabelle CTL\_WKA\_DRIVE\_MAERKTE aufgelistet. Bei Neueröffnung muss die Betriebe\_ID des neuen Hauses in die Tabelle hinzugefügt werden. Die Länge des relevanten Transaktionszeitraumes wird in Monaten angegeben und als Parameter in der Tabelle SAS\_Analyse\_Mart.CTL\_PARAMETER gespeichert.

Als nächstes wird der Wochentag, an dem die Bons generiert wurden, berechnet und an die restlichen Bon-Transaktionen angehängt.

In den Bons ist leider nicht die Basis\_Variante hinterlegt, sondern die Varianten ID (VANR\_ID). Aus diesem Grund wird aus der Tabelle QUE\_DIM\_SORTIMENT die entsprechende Basis\_Variante dazugejoint. Anschließend filtert man die Varianten raus, die in der Aktionsliste stehen. Dafür werden die Einträge der Aktionsliste in einer Macro Variable (als Liste) aneinander gehängt. In der Job wird geprüft, ob die jeweilige Basis\_Variante in dieser Liste enthalten ist. Im letzten Schritt werden die Artikelbezeichnungen dazugejoint.

**Output:** WKA\_TRANSAKTIONEN\_SUBSET (SAS Table, WKA Bibliothek)

*03\_Kundenliste erstellen*

**Input:** WKA\_TRANSAKTIONEN\_SUBSET

**Job-Beschreibung:** Damit für jeden Drive Kunden Regeln erstellt werden können, wird eine Liste mit allen in den Bon-Transaktionen vorkommenden Kunden\_IDX benötigt. Die Kundenliste mit den distinct KD\_IDX wird in diesem Job angelegt. Vorher werden die Kunden herausgefiltert, die weniger als 2-mal (zwei Bons) eingekauft haben und für die es aus einem Einkauf keine aussagekräftigen Regeln ableiten lassen.

**Output:** KUNDENLISTE (SAS Table, WKA Bibliothek)

*04\_Regelerstellung für Betriebe und Kunden*

**Input:** WKA\_TRANSAKTIONEN\_SUBSET, CTL\_WKA\_DRIVE\_MAERKTE, KUNDENLISTE

**Job-Beschreibung:** Der Code aus dem Warenkorbanalyse-Knoten im EM (E:\SAS\_Eminer\_Projekte\Drive Warenkorbanalyse Dessi) wird exportiert und in einem Macro unter E:\SAS\_Programme\Macros\extract\_WKA\_rules.sas gespeichert. Das Macro „extract\_WKA\_rules“ ist parametrisiert (mit dem Parameter Support) und wird anschließend in einem User Written Code-Knoten sowohl für alle Betriebe, als auch für alle Kunden aufgerufen. Der Parameter Support wird beim Aufruf initialisiert. Als Resultat entstehen 2 Tabellen. Die Tabelle RULES\_BETRIEBE enthält alle Regeln pro Haus (Regeln + Betrieb\_ID) und RULES\_KUNDEN enthält analog alle Regeln pro Kunde (Regeln + KDX\_ID). In dem User Written Code ist auch die Möglichkeit beschrieben den Job in einem Trainingsmodus nur für Kunden auszuführen. Dabei werden keine Regeln erstellt, sondern nur das optimale Support für jeden einzelnen Kunden berechnet und in der Tabelle REGEL\_LERN\_STATISTIK\_KUNDEN dokumentiert, so dass es bei der tatsächliche Berechnung der Regeln diesen optimalen Supportwert benutzt werden kann. Weiteres ist in den Kommentaren im Code nachzulesen.

**Output:** RULES\_BETRIEBE, RULES\_KUNDEN (SAS Table, WKA Bibliothek)

*071\_Rules filtern*

**Input:** PARM\_RULES\_TABLE, WKA\_DRIVE\_SAP\_HIERARCHIE

**Job-Beschreibung:** Der Job ist parametrisiert, je nachdem welche Rules-Tabelle (Betriebe oder Kunden) der Input ist. Es findet das Filtern von Regeln statt, die Items enthalten, die entlang des gleichen Hierarchiepfads laufen.

**Output:** WKA\_GEFILTERTE\_RULES (SAS Table, WKA Bibliothek)

*072\_Posprocessing Rules*

**Input:** WKA\_GEFILTERTE\_RULES

**Job-Beschreibung:** Als erstes wird das Kennzeichen “Importance” = Konfidenz\*Lift berechnet. Im nächsten Schritt werden Regelkonsequenzen mit mehr als 1 Item herausgefiltert. Um die Regeln zu historisieren werden sie mit einem Von\_ und Bis\_Datum versehen und in die Tabelle PARM\_HIST\_RULES\_TABLE historisiert.

**Output:** PARM\_HIST\_RULES\_TABLE 🡪 WKA\_HIST\_REGELN\_BETRIEBE oder WKA\_HIST\_REGELN\_KUNDEN (Oracle, SAS\_Analyse\_Mart)

*07\_Parametr. Aufruf: Filtern und Historisierung Rules*

**Input:** 071\_Rules filtern, 072\_Posprocessing Rules

**Job-Beschreibung:** Parametrisierten Job, der die Jobs 071 und 072 jeweils für Betriebe und Kunden aufruft.

### Datentransfer

Es werden drei Jobs angelegt, die den Transfer der aktuellen Datensätze aus DWH ins DB der WebShop darstellen sollen.

* Aktuelle Regeln
* Drive Hierarchie
* Manuelle Listen

### Job-Management

Hier werden bereitgestellte und geschedulte Jobs gespeichert. Momentan sind das nur die Jobs

* 01\_Automatisierte Regelerstellung
* 02\_Training Job

### Historisierte Tabellen in DWH (SAS\_Analyse\_Mart)

* WKA\_AKTIONSLISTE
* WKA\_AUSSCHLUSSLISTE
* WKA\_MANUELLE\_REGELN
* WKA\_MANUELLE\_REGELN
* WKA\_HIST\_REGELN\_BETRIEBE
* WKA\_HIST\_REGELN\_KUNDEN

### Tabellen in WKA (SAS)

* KUNDENLISTE
* RULES\_BETRIEBE
* RULES\_KUNDEN
* WKA\_DRIVE\_SAP\_HIERARCHIE
* WKA\_TRANSAKTIONEN\_SUBSET
* WKA\_GEFILTERTE\_RULES
* REGEL\_LERN\_STATISTIK\_KUNDEN

### Tabellen in DB (WebShop)

* WEBSHOP\_ AKTIONSLISTE
* WEBSHOP\_AUSSCHLUSSLISTE
* WEBSHOP\_MANUELLE\_REGELN
* WEBSHOP\_MAN\_STANDARD\_MAPPINGS
* WEBSHOP\_REGELN\_BETRIEBE
* WEBSHOP\_REGELN\_KUNDEN
* WEBSHOP\_REGELN\_HIERARCHIE

# Scoring-Modul