

Übungsblatt 3 Stand: 02.05.2021

Bearbeiten Sie alle Aufgaben aller Übungsblätter in festen **Zweiergruppen**. Beide Gruppenmitglieder müssen die Aufgabe gemeinsam bearbeiten und die Lösung alleine erläutern können. **Die Lösungen dürfen nicht mit anderen Gruppen gemeinsam erarbeitet werden.**

Aufgabe 4 (ROLAP)

Ihr **Mondrian-/Saiku-Prototyp** ist ein voller Erfolg. Nicht nur Robert, sondern auch seine Manager wollen gar nicht mehr ohne die neuen Analysemöglichkeiten arbeiten. HyperCube@DirtBikes soll also weitergehen. Daher beauftragt Robert Sie, für die Analysen eine performante Produktivumgebung für den Regelbetrieb aufzusetzen. Erstellen Sie daher auf Basis Ihres ADAPT-Modells (Übung 2) ein **Star-Schema** in der Produktivumgebung von Dirt Bikes (aka ihr Bereich auf der HAW Oracle-Datenbank).

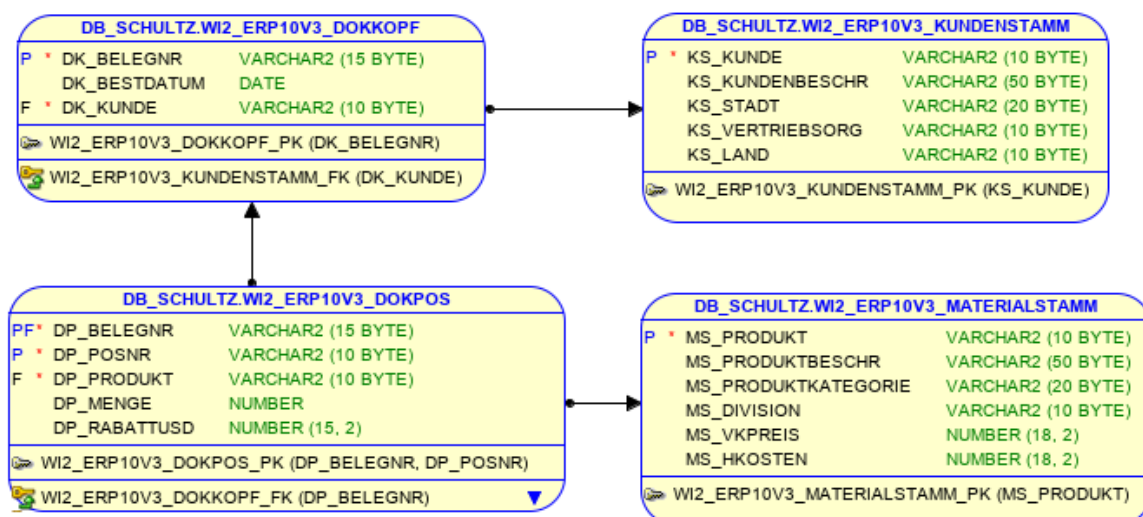
Sie machen sich also an die Arbeit und sammeln nach und nach folgende Informationen ein (Datenquelle, Berichtsformen, Änderungen und Performance). Auf Basis der Informationen legen Sie entsprechende Tabellen in Ihrem Bereich in der (HAW) Oracle-DB an. Am Ende erstellen Sie sogar noch eine schöne Darstellung Ihres DB-Schemas.

Datenquelle

Robert fällt aktuell nur das DirtBikes ERP-System (ERP10V3) als mögliche Datenquelle ein. Das System wurde zum 01.01.2011 in Version 3 eingeführt und basiert ebenfalls auf einer Oracle-DB für die Datenhaltung. IT-Mitarbeiter db_schultz stellt Ihnen das nachfolgende Datenmodell für die relevanten Tabellen des ERP-Systems zur Verfügung. Aus Sicherheitsgründen werden die Daten also wieder durch das Schema von db_schultz geleitet. Sie können also in Oracle mit Ihrer **A-Kennung** z.B. über

```
SELECT * FROM DB_SCHULTZ.WI2_ERP10V3_KUNDENSTAMM
```

auf die Daten zu den Kunden zugreifen.



Für die aktuelle Projektphase benötigen Sie das Datenmodell aber ausschließlich für die Wahl der richtigen **Datentypen** für die einzelnen Attribute in Ihrem Star-Schema. Sie werden die Daten in der nächsten Übung dann für das Befüllen ihres Star-Schemas verwenden.

Berichtsformen

Nach Rücksprache mit seinen Managern wünscht sich Robert, dass die Produktivumgebung so eingerichtet wird, dass prinzipiell die Berichtsformen **As-Is** und **As-Of** abgebildet werden können. Dabei fällt ihm ein, dass er im Vergleich zur ersten Anforderungserhebung (Übung 2) noch eine zusätzliche Kennzahl haben möchte: **AnzahlVerkaufsaufträge (NoOfSalesOrders)**. Zudem möchte er bei den Kunden noch das **Level City** haben (zwischen Customer und SalesOrg).

Änderungen

Als DWH-Experte fragen Sie Robert natürlich gleich, wie es mit Änderungen bei den Stammdaten in den Dimensionen aussieht. Robert erklärt Ihnen, dass bzgl. der Produkte keine Änderungen an den Stammdaten zu erwarten sind. Hier wünscht er, dass immer der aktuelle Stand der **Produktstammdaten** für die Analysen verwendet wird. Dabei fällt Ihm noch ein, dass die Produkte neben einen SalesPrice (**MS_VKPREIS**) auch noch **Herstellkosten (MS_HKOSTEN)** hat.

Bei den **Kundenstammdaten** gab es seit 2011 schon einige Änderungen, die irgendwo im System protokolliert sind. Er will noch in Erfahrung bringen, wie er Ihnen diese Protokolleinträge zur Verfügung stellen kann. Zudem erwartet er auch weiterhin hier und da mal Änderungen (einzelne Attribute oder auch die Zuordnung von Kunden zu Verkaufsorganisationen). Hier möchte er eine vollständige Historie zur Verfügung haben, um die oben genannten Berichtsformen für den gesamten Datenbestand ab 2011 abbilden zu können.

Performance

Um auf Nummer Sicher zu gehen, fragen Sie Robert noch mal, was er unter „performant“ versteht. Er versteht ihre Frage nicht wirklich, erläutert aber, dass der Dirt Bikes-Vorstand wohl plant, das Produktportfolio massiv auszubauen. Robert wünscht sich daher, dass HyperCube@DirtBikes auch nach dieser Erweiterung gute Antwortzeiten liefert, insbesondere für Anfragen auf **Produktkategorie-Ebene (ProductCategory)**, da Robert davon ausgeht, dass bei einem massiv erweiterten Produktportfolio sich die Vertriebssteuerung eher auf die überschaubare Anzahl an Produktkategorien konzentrieren wird. Zudem geht er davon aus, dass bzgl. der Zeitdimension fast alle Anfragen auf **Monats- und Jahresbasis** erfolgen werden, und nicht auf Tagesbasis.

Ihnen ist natürlich sofort klar, wie Sie ihr Star-Schema erweitern, um Ihre Lösung zukunftssicher zu machen. Nach dem eingehenden Studium der Anforderungen des Mondrian-Servers an ein **Fact-Constellation-Schema** (siehe Aufgabe 5) berücksichtigen Sie diese beim Anlegen der Tabellen. Robert kennt den Begriff zwar nicht, aber er findet, dass „**Fact Constellation-Schema**“ ziemlich fancy klingt.

Sonstiges

IT-Mitarbeiter db_schultz beäugt das Projekt HyperCube@DirtBikes natürlich sehr kritisch. In seiner Rolle als Head Of Data Management fordert er bei Robert ein, dass er zumindest lesenden Zugriff auf die HyperCube@DirtBikes-Tabellen bekommt. Auf Anweisung von Robert müssen Sie sich fügen und räumen (**GRANT**) dem User **DB_SCHULTZ** Leserechte (**SELECT**) für alle von Ihnen erstellten Tabellen ein.

Zudem möchte db_schultz in seiner Datenbank nachvollziehen, wer welche Tabelle angelegt hat. Daher benennen Sie, alle anzulegenden Tabellen nach dem folgenden Schema.

WI2_[IhreAKennung]_[Tabellename]

Abzugeben:

- Darstellung ihres relationalen Schemas
- Leseberechtigungen auf Ihre Tabellen für DB_SCHULTZ

Aufgabe 5 (OLAP-Server Mondrian)

Da Sie antizipieren können, wie das Projekt weitergeht, erstellen Sie schon mal eine neue Version Ihres Mondrian-Schemas, welches Ihr **Star-Schema** verwendet. Hierbei ist wohl insbesondere auf die Fact Constellation-Schema-Thematik zu achten, die beim OLAP-Server Mondrian über **Aggregate Tables** (https://mondrian.pentaho.com/documentation/aggregate_tables.php) abgebildet werden kann.

Allgemeiner Hinweise:

- Setzen Sie bitte bei jeder Tabellenangabe das Datenbankschema auf Ihre A-Kennung `<Table name="DBTableName" schema="[IhreAKennung]" />`. So kann ich ihr Schema auch direkt bei mir testen.
- Die Oracle-DB erfordert, dass im XML-Schema alle Spalten- und Tabellennamen sowie die Schema-Angabe (i.d.R. ihre A-Kennung) in **Großbuchstaben** eingetragen werden.

Hinweise zu **Aggregate Tables**:

- Der Name Ihrer Tabelle für vorberechnete Aggregate in Oracle soll mit „**AGG_**“ beginnen, da der Mondrian-Server per Regular Expression mögliche Aggregate Tables identifiziert.
- Die Mondrian-Implementierung erfordert, dass in der Aggregate Table ein Eintrag der Form `<AggFactCount column="[ihrSpaltenName]" />` enthalten ist (siehe Beschreibung unter dem oben aufgeführten Link)
- Bei der Abbildung der Klassifikationsstufen in der Aggregate Table haben Sie zwei Gestaltungsoptionen: 1) `Collapsed` und 2) `Non Collapsed`. Entscheiden Sie anhand der SCD-Typen der jeweiligen Dimension, welche Option zielführend ist, z.B. **sollte die Zeitdimension immer collapsed** sein.

Testphase

- Testen Sie, ob bei entsprechenden Anfragen über Saiku ihr Fact-Constellation-Schema (also Ihre Aggregate Table) von Mondrian adäquat verwendet wird.
- Da Ihre Tabellen für das Star-Schema noch leer sind, liefert Saiku natürlich noch keine Ergebnisse. **Füllen sie Ihre Tabellen daher mit ein paar sinnvollen Datensätzen Ihrer Wahl** (min. zwei Produkte, zwei Kunden und zwei Verkaufsvorgänge), so dass im Saiku-Frontend etwas angezeigt wird.

In den Logfiles ist die Verwendung des Fact-Constellation-Schemas nachvollziehbar. Hierzu werden vom Mondrian-Server mehrere Logfiles zur Verfügung gestellt (`mondrian.log`, `mondrian_mdx.log`, `mondrian_sql.log`). Nach der Auswahl Ihres Cubes in Saiku sollte in der `mondrian_sql.log` ein SQL-Statement stehen das auch Ihre Aggregate Table verwendet.

Abzugeben:

- das angepasste Mondrian-Schema (XML)
- textuelle Erläuterungen zu Ihren wesentlichen Design-Entscheidungen
- SQL-Log-File, in dem die Verwendung der Aggregate Table nachvollziehbar ist

Allgemeine Hinweise (gelten für alle Übungsblätter)

Ich akzeptiere nur Lösungen, auf denen **die Namen beider Bearbeiter, die Veranstaltungsbezeichnung, die Nummer des Übungsblattes und die Nummer der Praktikumsgruppe (1/2/3/4) notiert ist**. Außerdem müssen sie eine **Versionsnummer** auf dem Lösungsblatt notieren und die Versionen bitte bei erneuter Abgabe hochzählen.

Je Zweier-Gruppe ist in EMIL nur von einem Gruppenmitglied das Lösungsdokument hochzuladen. Die Abgabe der Lösungen muss spätestens am **Tag der Übung in EMIL** erfolgen (**Deadline: Ende des jeweiligen Praktikumstermins**).

Sind Lösungen von zwei/ mehreren Gruppen zu ähnlich, bekommen alle beteiligten Gruppen mindestens einen Fehlversuch und alle beteiligten Gruppen müssen eine überarbeitete Version einreichen. Ggf. werden alle beteiligten Gruppen von der Übung ausgeschlossen.