

Modeling of fact tables in Data Warehouse

Zlatinka Kovacheva*, Ina Naydenova**

**Middle East College, Knowledge Oasis Muscat, P.B. No. 79, Al Rusayl, PC: 124, Sultanate of Oman*

***University of Sofia St. Kliment Ohridski, 5, James Bourchier boul., Sofia, Bulgaria*

Abstract

Nowadays the warehouses store data that is not always suitable for presentation in a dimensional manner. This requires the use of advanced approaches to modeling fact tables especially when using a bottom-up data warehouse architecture. In this paper we provide a review of both basic and more complex techniques for fact tables modeling. We also discuss how some fact modeling techniques are suitable for a given query, but are not suitable for others. The given examples address the need of a methodology which combines the presentation flexibility of the top-down Inmon's approach and the quick results given by Kimball's bottom-up modelling approach.

Though the BI solutions are famous for their ad-hoc functionality, there are no universal models of the data. Moreover, some instruments (such as data mining tools) require more specific representations. For Kimball's approach, there is no central atomary model which should be the source of all representations needed for the goals of the reports. To overcome this shortcoming, advanced techniques of fact tables modeling appear. Also, the necessity arises to keep the same data in different structures, i.e. a denormalization and duplication of the data in the facts appear. Dependences between the metrics are created, which the multidimensional model does not reflect.

We present the architecture approach of several successful data warehouse projects as well as the assessment of a data warehouse supporting team about the risk of data inconsistency during the life time of the system. Some indicators were assessed in 5 degree scale (0 – none, 1 – very low, 2 – low, 3 – medium, 4 – high, 5 – very high). The provided information confirms the conclusions.

Despite all efforts of designer research community to develop the best data warehouse architecture, for the moment there is no technique described to effectively combine the flexibility of Inmon's approach with the low cost and quick results given by Kimball's approach. Our experience in the development and maintenance of Data Warehouse systems shows that an important place in such new architecture must be taken by a system following and maintaining the dependences between the data and its duplication.

Моделиране на факт таблици в склада на данни

Златинка Ковачева *, Ина Найденова **

**Близкоизточен колеж, Knowledge Oasis Muscat, P.B. № 79, Al Rusayl, PC: 124, Султанат Оман*

*** Софийски университет "Св. Климент Охридски", бул. Джеймс Баучер 5, София, България*

РЕЗЮМЕ

В наши дни складовете от данни (data warehouses) съхраняват данни, които не винаги са подходящи за многомерно представяне. Това изисква използването на нетривиални подходи за моделиране на факт таблици, особено когато използваната архитектурата на склада от данни е от тип „отдолу нагоре“ (bottom-up architecture). В тази статия предлагаме преглед както на основни, така и на по-сложни техники за моделиране на факт таблици. Също така дискутираме ситуациите, при които определени техники за моделиране на факти са подходящи за даден тип заявка, но не са подходящи за други. Приведените примери са насочени към необходимостта от методология, която съчетава гъвкавостта на представянето на подхода на Inmon “отгоре надолу” и бързите резултати, които дава подходът на Кимбъл „отдолу нагоре“.

Въпреки че BI решенията са известни със своята гъвкавост и динамична ad-hoc функционалност, универсални модели на данните няма. Освен това, някои средства (като средствата за извличане на знания) изискват по-специфични представяния на данните. Всеки подход има свои плюсове и минуси, които трябва да се прецизират за всеки конкретен случай, спрямо поставените цели. Например при подхода за моделиране на Кимбъл липсва централизиран атомарен модел, който да е източник на всички представяния, необходими за целите на изготвяне на справки и анализи. За да се преодолее този недостатък, се появяват съвременни техники за моделиране на факт таблици. Също така възниква необходимостта да се запазват едни и същи данни в различни структури, т.е. появява се денормализация и дублиране на данните във факт таблиците. Създават се зависимости между показателите, които многомерният модел не отразява.

В тази статия представяме архитектурния подход на няколко успешни проекта за съхранение на данни, както и оценката на екипа по поддръжка на склада на данни относно риска от неконсистентност на данни по време на функционирането на системата. Използвана е 5-степенна скала за оценка (0 - няма, 1 - много ниска, 2 - ниска, 3 - средна, 4 - висока, 5 - много висока). Представената информация потвърждава изводите.

Въпреки всички усилия на дизайнерската изследователска общност за разработване на най-добрата архитектура за съхранение на данни, за момента няма описана техника, която ефективно да комбинира гъвкавостта на подхода на Inmon с ниската цена и бързите резултати, дадени от подхода на Kimball. Нашият опит в разработването и поддръжката на складове на данни показва, че важно място в такава нова архитектура трябва да заеме система, която следи и поддържа зависимостите между данните и тяхното дублиране.