

МОДЕЛИРАНЕ И АНАЛИЗ НА СОФТУЕР

Павел Кюркчиев

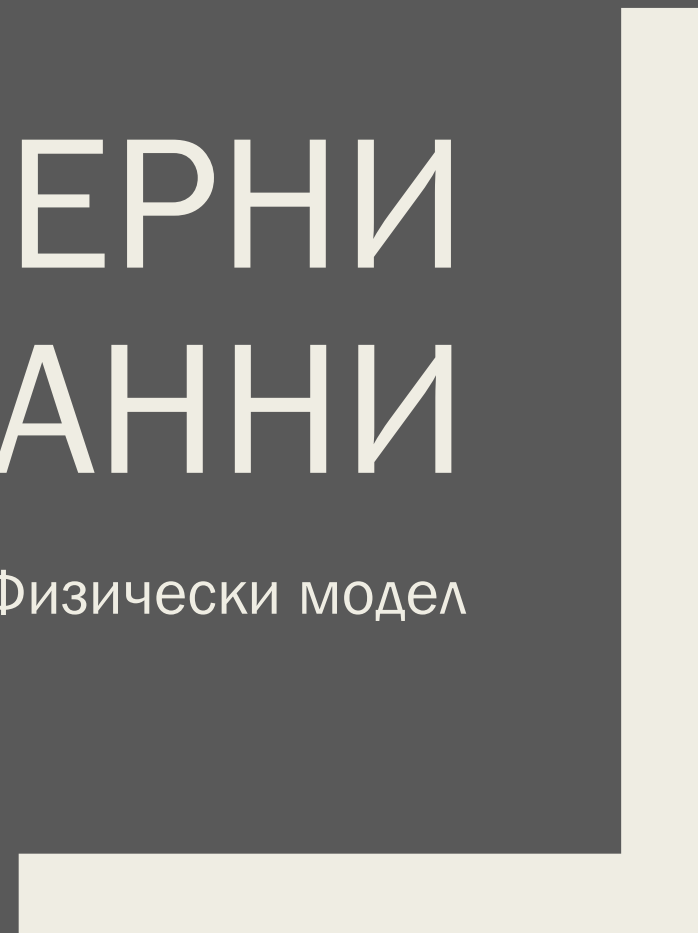
Ас. към ПУ „Паисий Хилендарски“

<https://github.com/pkyurkchiev>

@pkyurkchiev

МНОГОМЕРНИ МОДЕЛИ НА ДАННИ

Физически модел



Многомерни модели на данни (Multidimensional data model (MDDM))

- Многомерните модели на данни са създадени за разработване на складове и матрици от данни. MDDM предоставят както механизъм за съхраняване на данни, така и начини за бизнес анализ.

Елементи на MDDM

- Основните елементи са Факти и Измерения (Величини).
 - *Факти – цифров обем за анализ на бизнеса.*
 - *Измерения (Величини) - атрибути, служещи за анализ на данните.*
- Фактологични таблици (Fact Tables)
- Таблици измерения (Dimension Tables)

Фактологични таблици (Fact tables)

- Фактологичните таблици записват измервания или показатели за конкретно събитие. Фактологичните таблици се състоят от цифрови стойности и чужди ключове на данните за измерване, където се съхранява описателната информация.

- Фактологичните таблици са проектирани за ниска степен на уникално детайлно представяне (наричани по-долу "детайлност" или "зърно"), което означава, че фактите могат да записват събития на много ниско ниво. Това може да доведе до натрупването на голям брой записи във фактологична таблица с увеличаване на времевия период.

Таблицы измерения (Dimension tables)

- Таблиците с измерения обикновено имат сравнително малък брой записи в сравнение с фактологичните таблици, но всеки запис може да има много голям брой атрибути, за да опише фактическите данни.

Най-често срещани атрибути в таблица измерения

- Таблиците с величини за време описват времето на най-ниско ниво;
- Таблиците с географски величини описват данни за местоположение, държава, област или град;
- Таблиците с величини на продукти описват самите продукти;

- Таблиците за величините на служители описват служителите и тяхната специализация, като например продавачи, касиери и други.
- Таблиците с диапазони на обхват описват диапазони на времето, стойности във валута или други измерими количества за опростяване на отчетите.

Нормализация на измерения

Доста често разработчиците не нормализират:

- Нормализацията прави структурата по-сложна;
- Бързодействието може да се забави, заради много външни таблици и връзките към тях (joins);
- Денормализацията увеличава заеманото пространство на сървъра;
- Bitmap indexes пространства могат да бъдат използвани;

- Производителността на заявките се намаля. Често 3-та нормална форма води до загуба на производителност при извършване на агрегатни функции върху многомерни данни.

Проектиране на модел

- Изберете бизнес процес (Choose the business process)
- Декларирайте детайлността (Declare the grain)
- Определете измеренията (Identify the dimensions)
- Определете факта (Identify the fact)

Типове MDDM

Типове Многомерни модели на данни

- Модел на куба с данни (Data Cube Model)
- Модел на схема звезда (Star Schema Model)
- Модел на схема снежинка (Snowflake Schema Model)
- Модел на схема фактически съзвездия (Fact constellations)

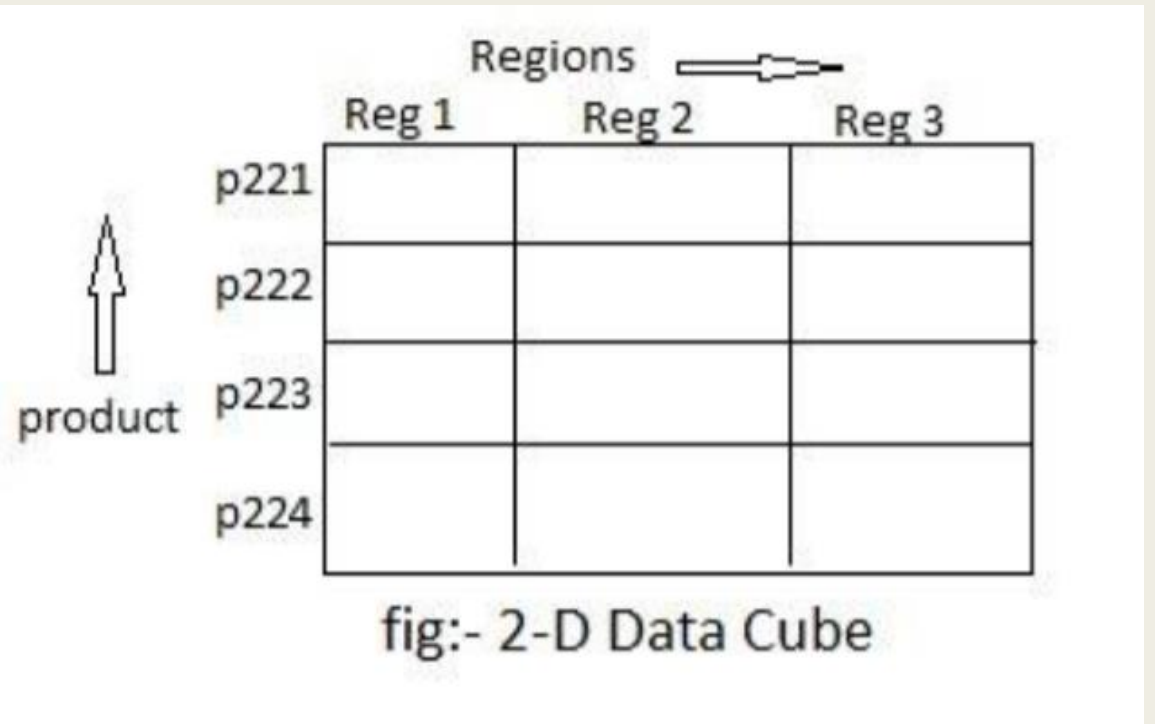
Модел на куб с данни (Data Cube Model)

- Когато данните са групирани или комбинирани в многомерни матрици, наречени Кубове от данни. Кубовете от данни могат да бъдат съставени от две, три или повече измерения.

Две измерения (величины)

■ Пример:

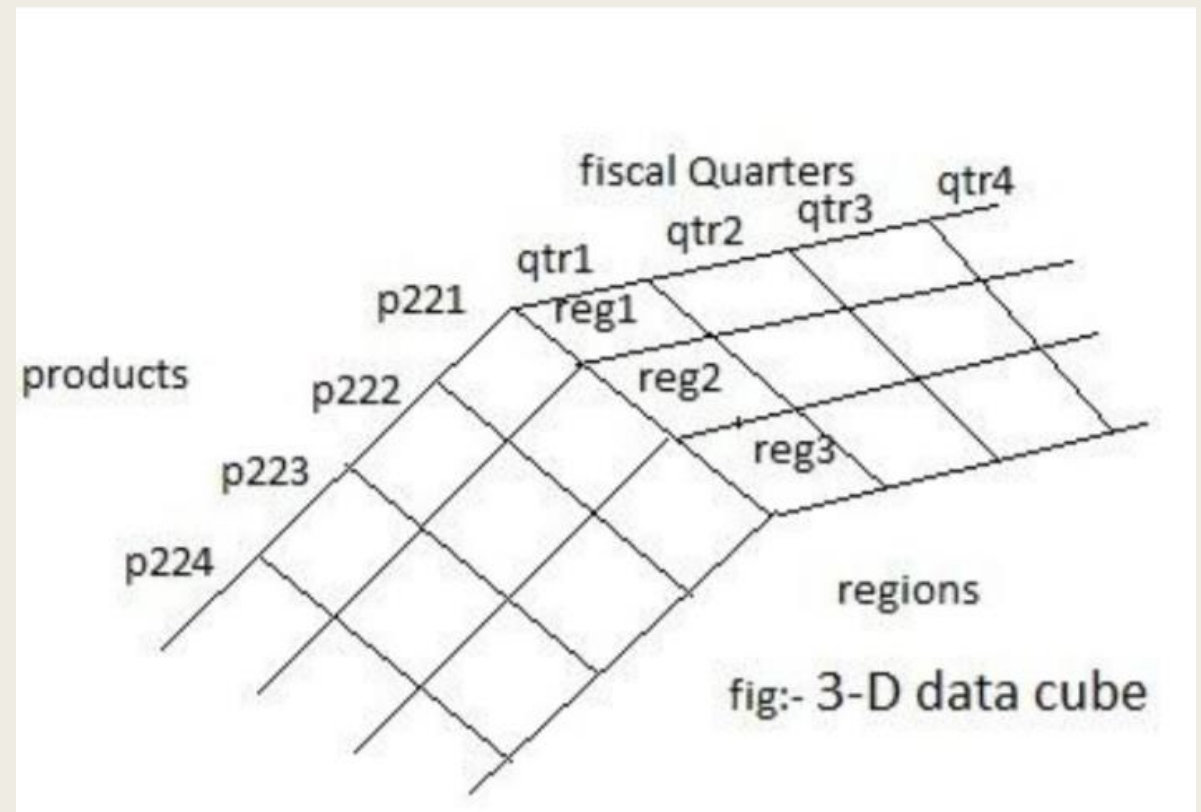
ред и колонна или продукты и
регионы.



Три измерения (величины)

■ Пример:

регион, продукты и фискальные
тримесечия.



Предимства

- Промяната от една йерархия на величини към друга е лесно осъществима в куба с данни чрез техника, наречена “pivoting” (известна като ротация).

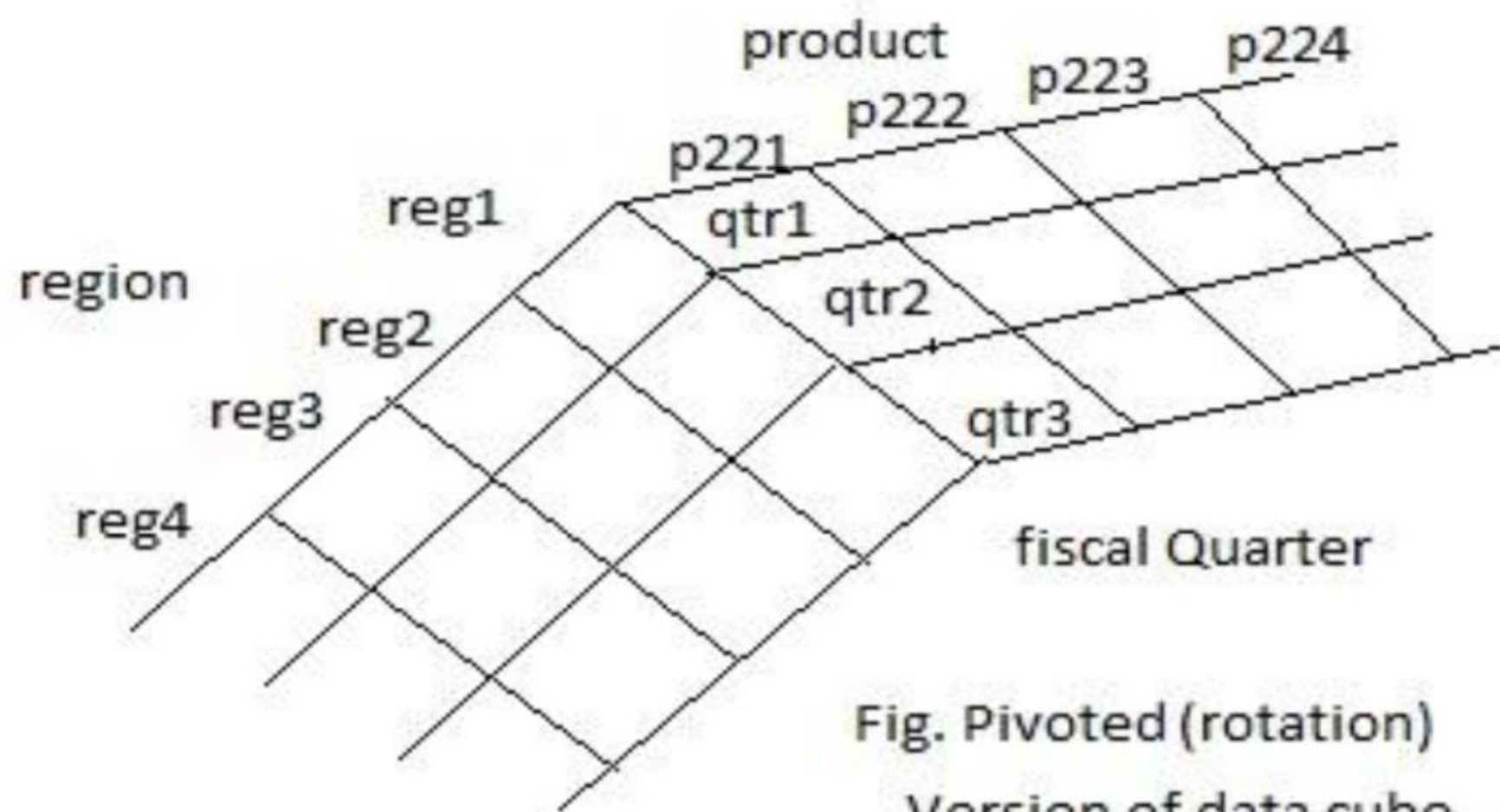


Fig. Pivoted (rotation)
Version of data cube

- Тези типове модели се прилагат към йерархичен изглед, като Roll-up Display и Drill Down Display.
- Roll-up Display
 - Когато се изпълнява “Roll-up” чрез намаляване на величината, едно или повече измерения се премахват от куба с величини.
 - с техниката “Roll” потребителя може да се отдалечи, за да види обобщено ниво на данните.
 - Навигационната пътека се определя от йерархията с величини.

■ Drill Down Display

- *Това е обратното на “Roll-up”.*
- *Навигация от по-малко подробни данни до по-подробни данни.*
- *Може да се изпълни и чрез добавяне на ново измерение към куб.*

Да се представят примери на модели описани чрез кубове.

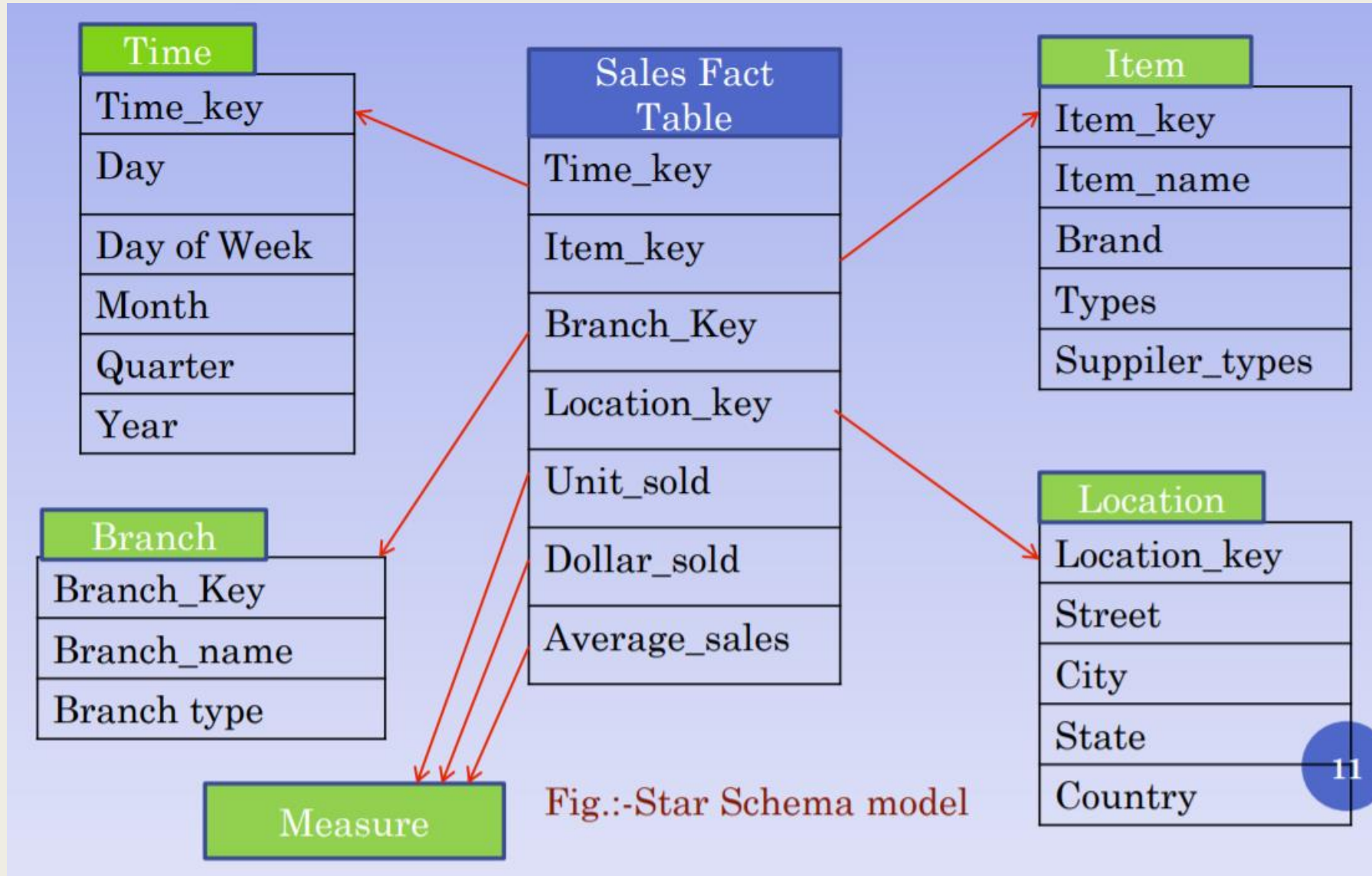
Кубове данни

Модел на схема звезда (Star schema)

- В компютърните науки, схема звезда е най-простият стил на “data mart schemata” и е най-широко използваният подход за разработване на складове за данни и матрици с данни за величини. Схемата звезда се състои от една таблица с факти, отнасящи се до произволен брой таблици с величини. Звездната схема е важен специален случай на “snowflake schema” и е по-ефективна за обработка на по-прости заявки.

- Тя се нарича схема звезда, тъй като Е-Р диаграмата ѝ прилича на звезда, с точки (таблици), излъчващи се от централната таблица.
- Заявките на звезда представляват обединение на фактически таблици с таблици измерения.
- Всяка таблица измерение се свързва с таблица факти посредством връзката първичен към външен ключ, но различните таблици измерения не се обединяват помежду си.
- Типичната фактологична таблица съдържа ключ и мярка.

Схема звезда



```
SELECT
    P.Brand,
    S.Country AS Countries,
    SUM(F.Units_Sold)
FROM Fact_Sales F
INNER JOIN Dim_Time D      ON (F.Date_Id = D.Id)
INNER JOIN Dim_Location S  ON (F.Store_Id = S.Id)
INNER JOIN Dim_Item P      ON (F.Product_Id = P.Id)

WHERE D.Year = 1997 AND P.Product_Category = 'tv'

GROUP BY P.Brand, S.Country
```

Схема звезда заявка

Предимства

- По-прости заявки – обединяването на таблици в схемата звезда представлява по-елементарна операция от обединяването на таблици в силно нормализиран релационен модел.
- Опростена бизнес логика за отчитане - в сравнение със силно нормализираните схеми, схемата звезда представя лесна и елементарна структура за създаване на отчети.

- Засилване на производителността на заявките - схемите звезда могат да осигурят подобрения на ефективността на приложенията за отчитане (но само за четене), в сравнение със силно нормализираните схеми.
- Бързи агрегации - по-опростените заявки към схема звезда могат да доведат до подобрена производителност на операциите за агрегиране.

Недостатъци

- Основният недостатък на схемата звезда е, че целостта на данните не се прилага добре, тъй като схемата е в силно денормализирано състояние. Еднократните вмъквания и актуализации могат да доведат до аномалии в данните, които нормализираните схеми са предназначени да избегнат. Най-общо казано, схемата звезда се зарежда с данни в силно контролиран вид чрез периодично обработване, за да се компенсира липсата на защита, осигурявана от нормализирането.

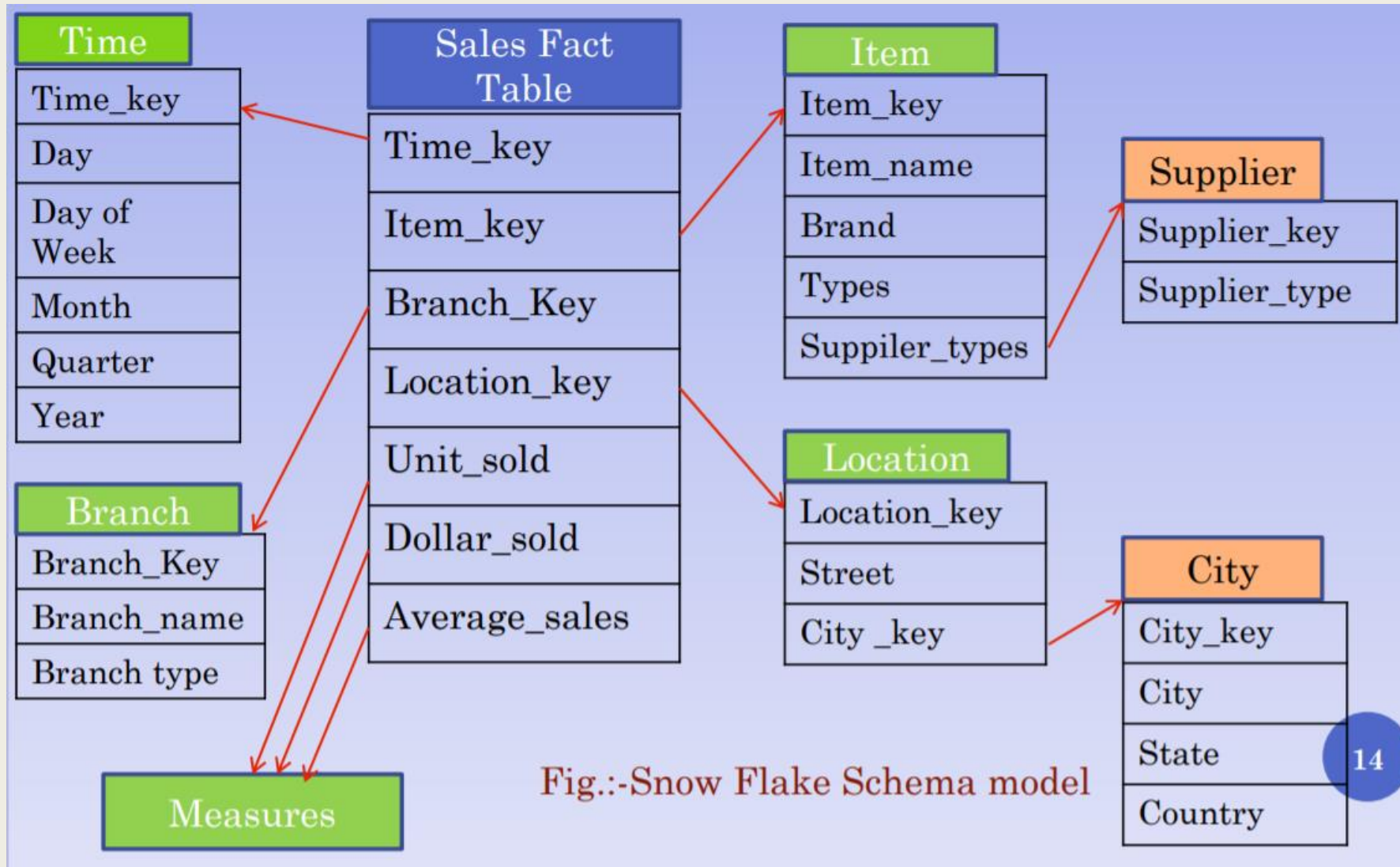
Да се представят примери на модели описани чрез схема звезда.

Схема звезда

Модел на схема снежинка (Snowflake schema)

- Схемата за снежинка представлява централизирана фактологична таблица, която е свързана с множество измерения.
- Схемата снежинка е малко по-различна от схемата звезда. Таблиците измерения от схемата звезда се организират в йерархия чрез нормализация.
- Ефектът на схемата снежинка засяга само таблиците величини, а не таблиците факти.

Схема снежинка




```
SELECT B.Id
INNER JOIN Dim_Product_Category C ON P. B.Brand,
      SUM(F.Units_Sold)
FROM Fact_Sales F
INNER JOIN Dim_Time D      ON F.Date_Id = D.Id
INNER JOIN Dim_Location S  ON F.Store_Id = S.Id
INNER JOIN Dim_Item P      ON F.Product_Id = P.Id
INNER JOIN Dim_Brand B     ON P.Brand_Id Product_Category_Id = C.Id

WHERE D.Year = 1997 AND C.Product_Category = 'tv'

GROUP BY  B.Brand
```

Схема снежинка заявка

Предимства

- Някои OLAP инструменти за моделиране на многомерни бази данни са оптимизирани за схема снежинка.
- Нормализирането на атрибутите води до спестявания на място при съхранение.

Недостатъци

- Основният недостатък на схемата за снежинката е, че допълнителните нива на нормализиране на атрибутите увеличават сложността на заявките за извличане.

Да се представят примери на модели описани чрез схема снежинка.

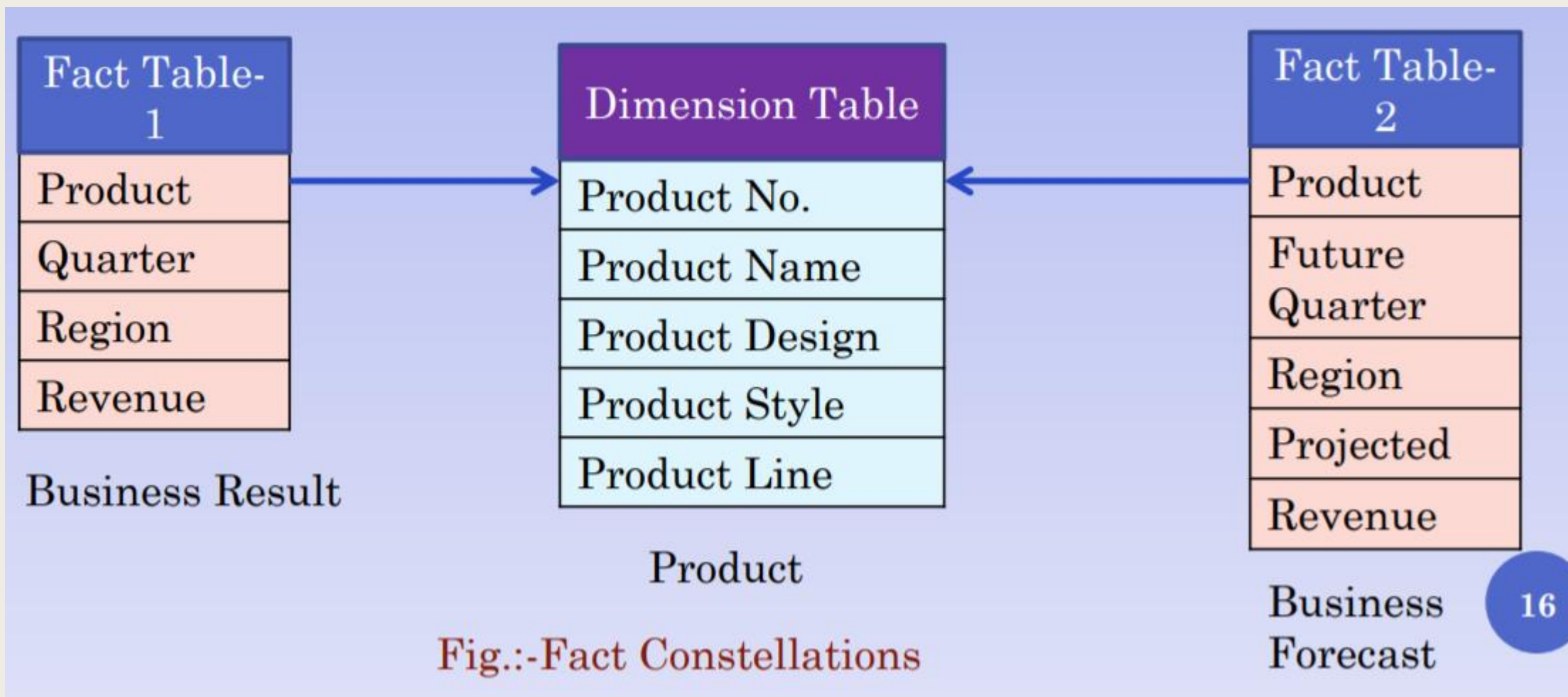
Схема снежинка

Фактически съзвездия (Fact constellations)

- Фактически съзвездия, или съзвездието "факт", е мярка за онлайн аналитична обработка, която е съвкупност от множество фактически таблици, споделящи таблици измерения, разглеждани като колекция от схеми звезди.
- Представява подобрение на схемата звезда.

- Схемата за фактическо съзвездие има множество фактически таблици. Тя е известна още като галактическа схема. Това е широко използвана схема и е по-сложна от схемата звезда и схемата снежинка.

Фактически съзвездия



Да се дадат примери на модели описани чрез фактически съзвездия.

Фактически съзвездия

ВЪПРОСИ ?

