

Online Analytical Processing & Business Intelligence

DATA, INFORMACE, ZNALOSTI - ZOPAKOVÁNÍ

Data

- hodnota schopná přenosu, uchování, interpretace či zpracování
- z hlediska IT jde o ***hodnoty*** různých ***datových typů***
- data sama o sobě ***nemají sémantiku*** (význam), jsou to věty nějakého formálního jazyka určující ***syntaxi***
- hodnoty dat obvykle udávají ***stav*** nějakého systému

Informace

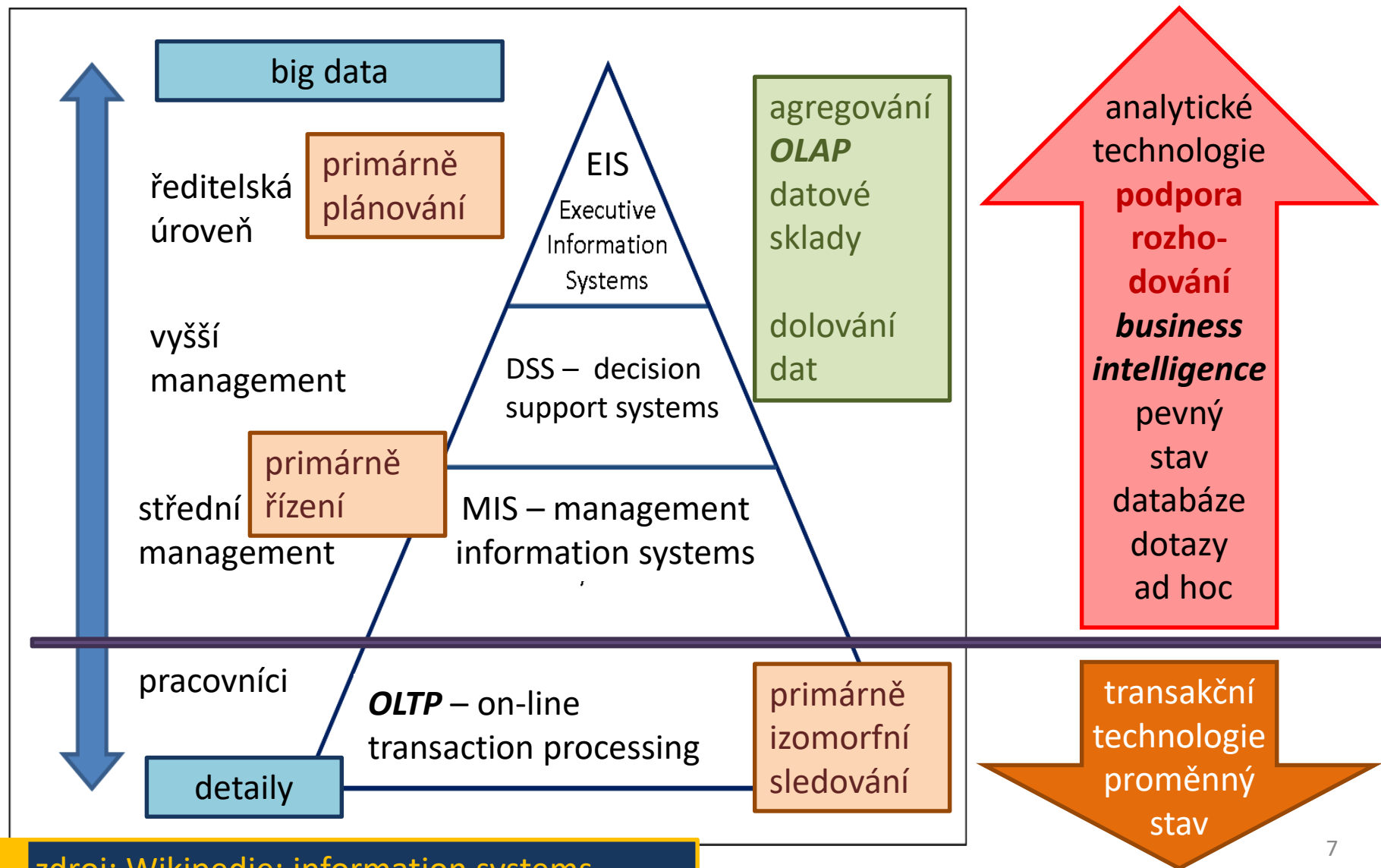
- **informace** jsou interpretovaná **data** lidmi nebo strojově
- po interpretaci mají **sémantiku** (význam)
- transformaci dat na informace provádí **uživatel** nebo **specializovaný systém**
- je nezbytné zajistit shodnou interpretaci dat u všech uživatelů informace
 - vzdělání, školení, zavedení konvencí, **ontologie**

Znalosti

- informace zařazená do souvislostí (často s redukováným množstvím dat)
- interpretace je však ještě hůře definovatelná, neboť může jít o agregace informací
- znalosti chápeme často jako ***sekundární odvozené informace***
- některé informační systémy se zabývají pouze ***informacemi (transakční - OLTP)***, některé pracují se ***znalostmi a velkými daty (pro podporu rozhodování a plánování – OLAP)***
- ***znalosti*** jsou často získávány operacemi nad ***velkými daty*** (agregace, data mining, apod.)

DATOVÉ SKLADY A ON-LINE ANALYTICAL PROCESSING (OLAP) BUSINESS INTELLIGENCE

Pyramidové schéma



Motivace, příklady

- manager potřebuje vědět, kterým klientům může nabídnout po telefonu úvěrovou kartu, u kterých klientů je vysoké riziko odchodu ke konkurenci
- manager potřebuje znát vývoj tržeb za posledních třicet dní v členění dle regionů a produktů či jak se liší skutečný výkon společnosti od plánovaného

Pojem *business intelligence*

- procesy, technologie a nástroje potřebné k ***přetvoření dat a informací do znalostí*** pro podporu rozhodování na různých úrovních
- ***Vstup***: velké objemy (big data) primárních (produkčních) dat
- ***Výstup***: znalosti, které lze využít v procesu rozhodování

Prostředky *business intelligence*

1. *Datové sklady (data warehouses)*

- systém převodu a uložení dat pro analýzu, definování formálního modelu dat

2. *OLAP (On-line Analytical Processing)*

- rozhraní pro manipulaci s modelem a zpřístupnění výsledků uživateli a dalším aplikacím

3. *Data Mining (dolování dat)*

viz specializovaný kurz Získávání znalostí z databází

DATOVÉ SKLADY (DATA WAREHOUSES)

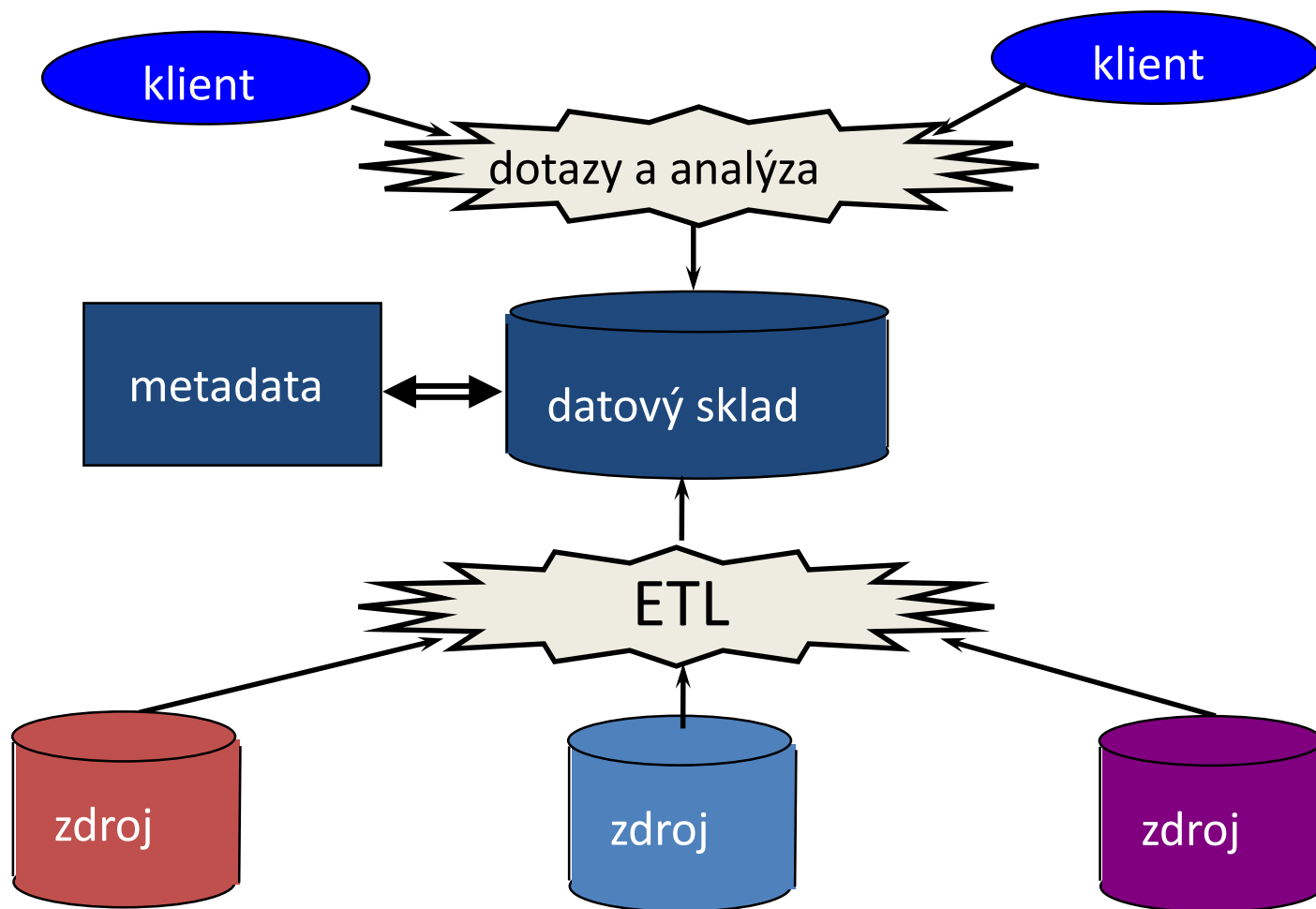
Pojem datového skladu - slovně

- Podnikově strukturovaný depozitář ***subjektově orientovaných, integrovaných, časově proměnlivých, historických dat*** použitých na získávání ***znalostí*** a podporu rozhodování
- obsahuje operační i agregovaná data.

Pojem datového skladu

- ***Datovým skladem*** nazýváme technologii
 - natažení (extrakce a transformation)
 - uložení (loading) a
 - poskytovánídat pro ***podporu rozhodování prováděnou analýzou informací a vytvářením znalostí***
- je typicky provozován ***odděleně*** od základní ***operační databáze*** (též ***databáze detailů*** nebo ***produkční databáze***)

Rámcová architektura datového skladu



Nevhodnost produkčních databází jako datových skladů

- slouží především pro ***ukládání*** primárních detailních dat ***izomorfního modelu (modelování reálného systému)***
- výsledkem vesměs pevného počtu dotazů jsou především data explicitně uvedená v databázi bez dalších agregačních úprav
- výhodné především pro ***jednoduché transakce*** (vkládání, mazání...) - OLTP, naopak ***nevhodné pro složitější analýzu velkých dat***

Nevhodnost produkčních databází jako datových skladů

- obvykle ***decentralizovanost*** systémů OLTP
 - data potřebná pro analýzu jsou většinou ***uložena v různých heterogenních DB na různých serverech***, není většinou k dispozici integrovaný zdroj údajů a je složité tato data integrovat a homogenizovat
- ***nehomogenní struktura údajů*** – různé názvy vlastností, datové typy...
- nevhodnost technologie pro analýzy pro standardní výpočtové prostředky a modely

Nevhodnost produkčních databází jako datových skladů

- degradace výpočetního výkonu databázového stroje – neustále ***se opakujícími stejnými agregačními výpočty***
- případně nejsou uchovávány historické údaje – uchovávají se zpravidla ***pouze aktuální data***

Vhodný model = vícerozměrnost

- aby bylo možno provádět komplexní analýzu a vizualizaci, jsou data v datovém skladu typicky modelována ***multidimenzionálně***.
- jiný datový model, nežli relační
- klade důraz na ***strukturu pro budoucí dotazy ad hoc vyžadující agregační a statistické výpočty***

Multidimenzionální databáze

- slouží jako platforma pro získání **agregovaných údajů**
- výpočty, které by se případně opakovaně prováděly, **mohou být spočteny předem a uloženy (materializovány)** z důvodu rychlého přístupu k agregovaným datům – zapojení učících algoritmů
- redundance zde není tak podstatným problémem (data jsou **read-only, nevzniká problém udržování konzistence a vícenásobného přístupu**)

DEFINICE MULTIDIMENZIONÁLNÍHO MODELU

XML for Analysis

XMLA

XMLA úvod

- ***XML for Analysis*** (zkráceně ***XMLA***)
- Průmyslový standard pro přístup k datům v analytických systémech jako je OLAP a data mining.
- je založen na průmyslových standardech XML, SOAP and HTTP.
- je udržováno společností *XMLA Council* se členy Microsoft, Hyperion a SAS , kteří jsou oficiálními zakladateli.

XMLA historie

- XMLA bylo navrženo společností Microsoft jako následník systému OLE DB for OLAP v dubnu 2000.
- V lednu 2001 byl přibrán XMLA navržený společností Hyperion.
- Verze 1.0 byla zveřejněna v dubnu 2001
- V září 2001 byl ustaven XMLA Council.
- V dubnu 2002 byla přibrána SAS jako zakládající člen XMLA Council.
- Během času přebralo tento standard více než 25 společností.

XMLA použití

- XMLA je vhodná syntaxe pro zápis součástí multidimenzionálního modelu
- Budeme používat pro příklady jak definice multidimenzionální kostky, tak pro operace nad ní

XMLA API

- XMLA sestává pouze ze 2 metod SOAP methods.
- *Discover*
- *Execute*

XMLA Discover

- Navržena pro modelování všech zjišťovacích metod z OLE DB
- Zahrnuje schémata jako ***rowset***, ***properties***, ***keywords***, ale taky ***kostku*** atd.
- Umožňuje definovat, jak a co bude zkoumáno, tak možná omezení nebo vlastnosti

XMLA Execute

- Má 2 parametry:
- **Command** – Příkaz, který má být proveden. Může být v jazycích **MDX**, DMX nebo SQL.
- **Properties** - seznam v jazyce XML sestávající z vlastností jako je Timeout, Catalog name atd.
- Výsledkem příkazu **Execute** může být *Multidimensional Dataset* nebo *Tabular Rowset*.

Implementace XMLA

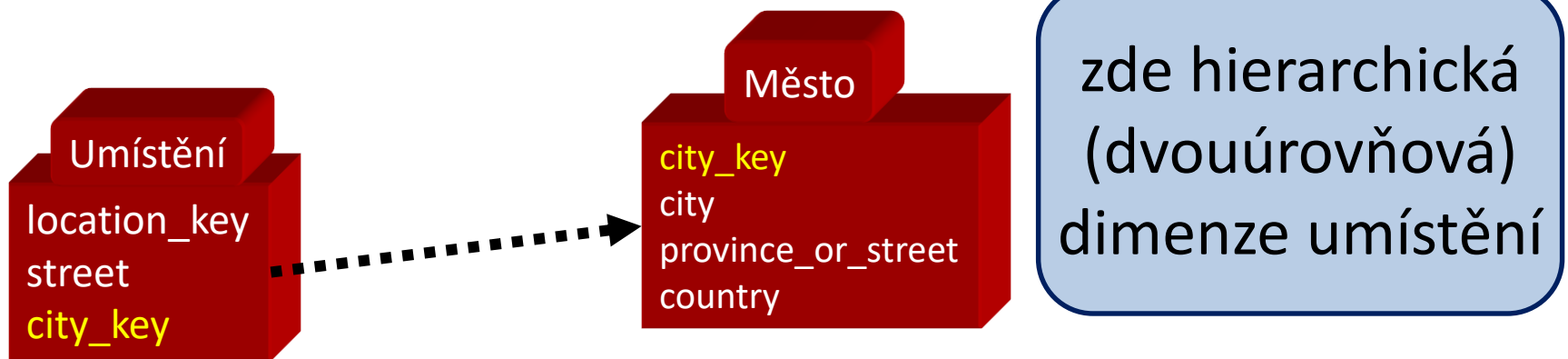
- dostupný systém *icCube*
- budeme používat pro příklady zejména v multidimenzionálním dotazovacím jazyce MDX.

MULTIDIMENZIONÁLNÍ MODEL

DIMENZE

Dimenze

- ***Dimenze*** je ***uspořádatelná*** množina hodnot ***diskrétního*** základního typu (integer, výčet, čas) nebo množina jejich struktur ***hierarchicky organizovaných***



Dimenze příklady

jednorozměrné

čas

leden

únor

březen

duben

...

prosinec

katalog

kabáty

kalhoty

bundy

čepice

...

trička

místo

Brno

Praha

Ostrava

Plzeň

...

Olomouc

hierarchické

čas

2000

2001

2002

...

2015

leden

únor

...

prosinec

leden

únor

...

prosinec

leden

únor

...

prosinec

Příklad definice dimenze Geography v icCube (hierarchie Geo a Economy)

▼ **D** Geography

▼ **H** Geo

▼ **LS** Levels

LA All-Level

L Continent

L Country

L City

▼ **AD** All Regions

▼ **M** America

▶ **M** Canada

▼ **M** United States

M Los Angeles

M New York

M San Francisco

▶ **M** Mexico

▶ **M** Venezuela

▶ **M** Europe

▼ **H** Economy

▼ **LS** Levels

LA All-Level

L Partnership

L Country

▼ **AD** All

▼ **M** NAFTA

M Canada

M United States

M Mexico

M Venezuela

▼ **M** EU

M France

M Spain

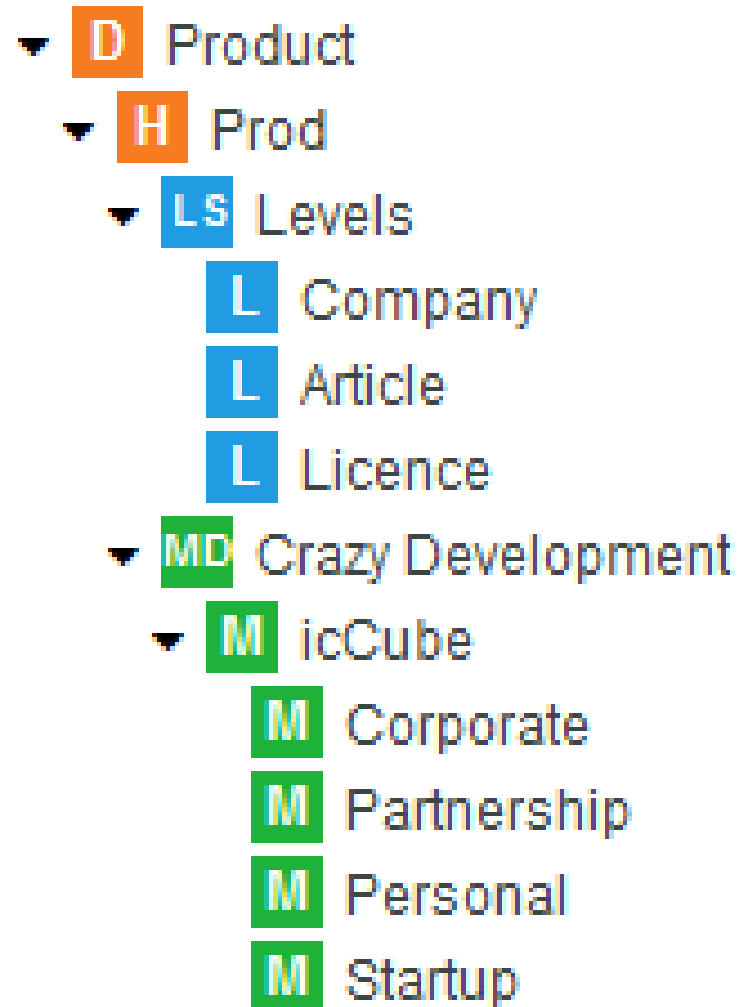
▼ **M** None

M Switzerland

Příklad – Geography v Excelu

[Geo]. [Continent]	[Geo].[Country]. [Name]	[Geo]. [Country]. [Key]	[Geo]. [City]	[Economy]. [Partnership]	[Economy]. [Country]
America	Canada	CA	Quebec	NAFTA	Canada
America	Canada	CA	Toronto	NAFTA	Canada
America	United States	US	Los Angeles	NAFTA	United States
America	United States	US	New York	NAFTA	United States
America	United States	US	San Francisco	NAFTA	United States
America	Mexico	ME	Mexico	NAFTA	Mexico
America	Venezuela	VE	Caracas	NAFTA	Venezuela
Europe	France	FR	Paris	EU	France
Europe	Spain	ES	Barcelona	EU	Spain
Europe	Spain	ES	Madrid	EU	Spain
Europe	Spain	ES	Valencia	EU	Spain
Europe	Switzerland	CH	Geneva	None	Switzerland
Europe	Switzerland	CH	Lausanne	None	Switzerland
Europe	Switzerland	CH	Zurich	None	Switzerland

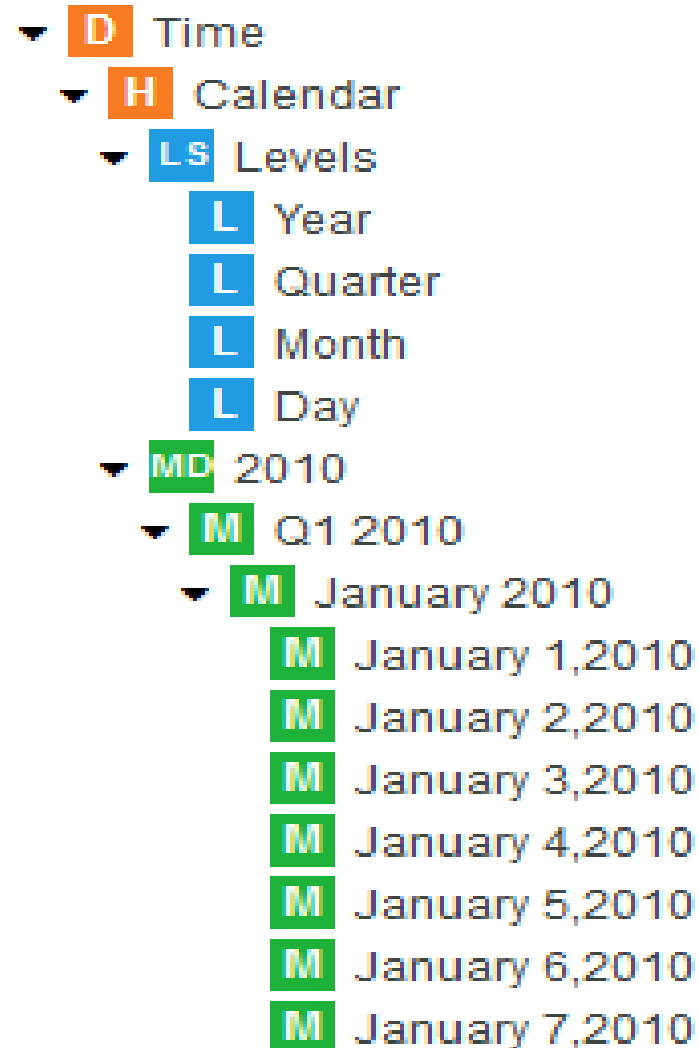
Příklad definice dimenze Product hierarchie Prod pro icCube



Příklad Product v Excelu

[Prod].[Company]	[Prod]. [Article]	[Prod]. [Licence]
Crazy Development	icCube	Corporate
Crazy Development	icCube	Partnership
Crazy Development	icCube	Personal
Crazy Development	icCube	Startup

Příklad definice dimenze Time hierarchie Calendar pro icCube



Příklad Time v Excelu

[Calendar].[Year]	[Calendar].[Quarter]	[Calendar].[Month]	[Calendar].[Day]	[Calendar].[Day].[Key]
2010Q1 2010		January 2010	January 1,2010	January 1, 2010
2010Q1 2010		January 2010	January 2,2010	January 2, 2010
2010Q1 2010		January 2010	January 3,2010	January 3, 2010
2010Q1 2010		January 2010	January 4,2010	January 4, 2010
2010Q1 2010		January 2010	January 5,2010	January 5, 2010
2010Q1 2010		January 2010	January 6,2010	January 6, 2010
2010Q1 2010		January 2010	January 7,2010	January 7, 2010
2010Q1 2010		January 2010	January 8,2010	January 8, 2010
2010Q1 2010		January 2010	January 9,2010	January 9, 2010
2010Q1 2010		January 2010	January 10,2010	January 10, 2010
2010Q1 2010		January 2010	January 11,2010	January 11, 2010
2010Q1 2010		January 2010	January 12,2010	January 12, 2010
2010Q1 2010		January 2010	January 13,2010	January 13, 2010
2010Q1 2010		January 2010	January 14,2010	January 14, 2010
2010Q1 2010		January 2010	January 15,2010	January 15, 2010
2010Q1 2010		January 2010	January 16,2010	January 16, 2010

není to celé atd.

Definice multidimenzionální kostky

- Necht' existuje ***uspořádaná množina n dimenzí*** $\{D_1, D_2, D_3, \dots, D_n\}$
- *tj. existuje relace uspořádání $R (<)$ nad množinou dimenzí*

čas

katalog

místo

Definice multidimenzionální kostky

- *počet různých prvků* relace R je $n!$ (počet **permutací** nad n , počet různých uspořádání)
- je to také **počet stěn n -dimenzionální kostky**. Proto má **3D kostka na vrhcáby** 6, tj. $3!$ stěn a dvoudimenzionální čtverec 2, tj. $2!$ stěn.

1	čas	katalog	místo
2	čas	místo	katalog
3	katalog	čas	místo

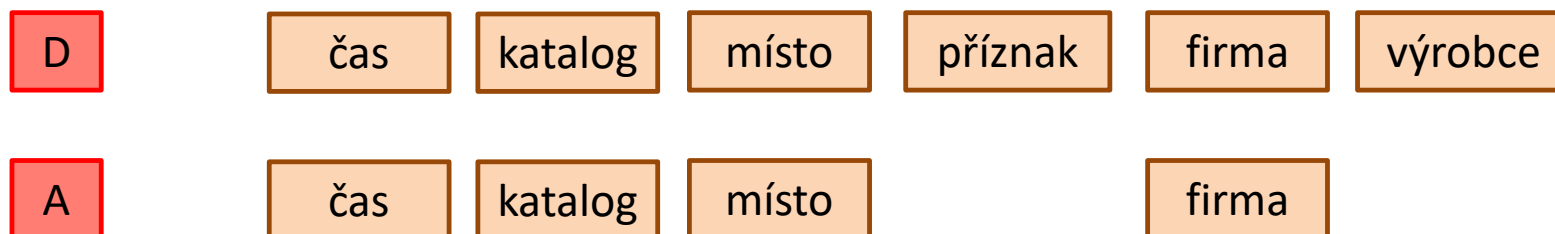
4	katalog	místo	čas
5	místo	katalog	čas
6	místo	čas	katalog

Definice multidimezionální kostky

- Necht' existuje ***uspořádaná podmnožina aktivních dimenzí*** $\{A_1, A_2, A_3, \dots, A_m\}$, kde $m \leq n$,
 $A_1 = D_{i1}, A_2 = D_{i2}, A_3 = D_{i3}, \dots, A_m = D_{im}$ a
 $D_{i1} < D_{i2} < D_{i3} < \dots < D_{im}$

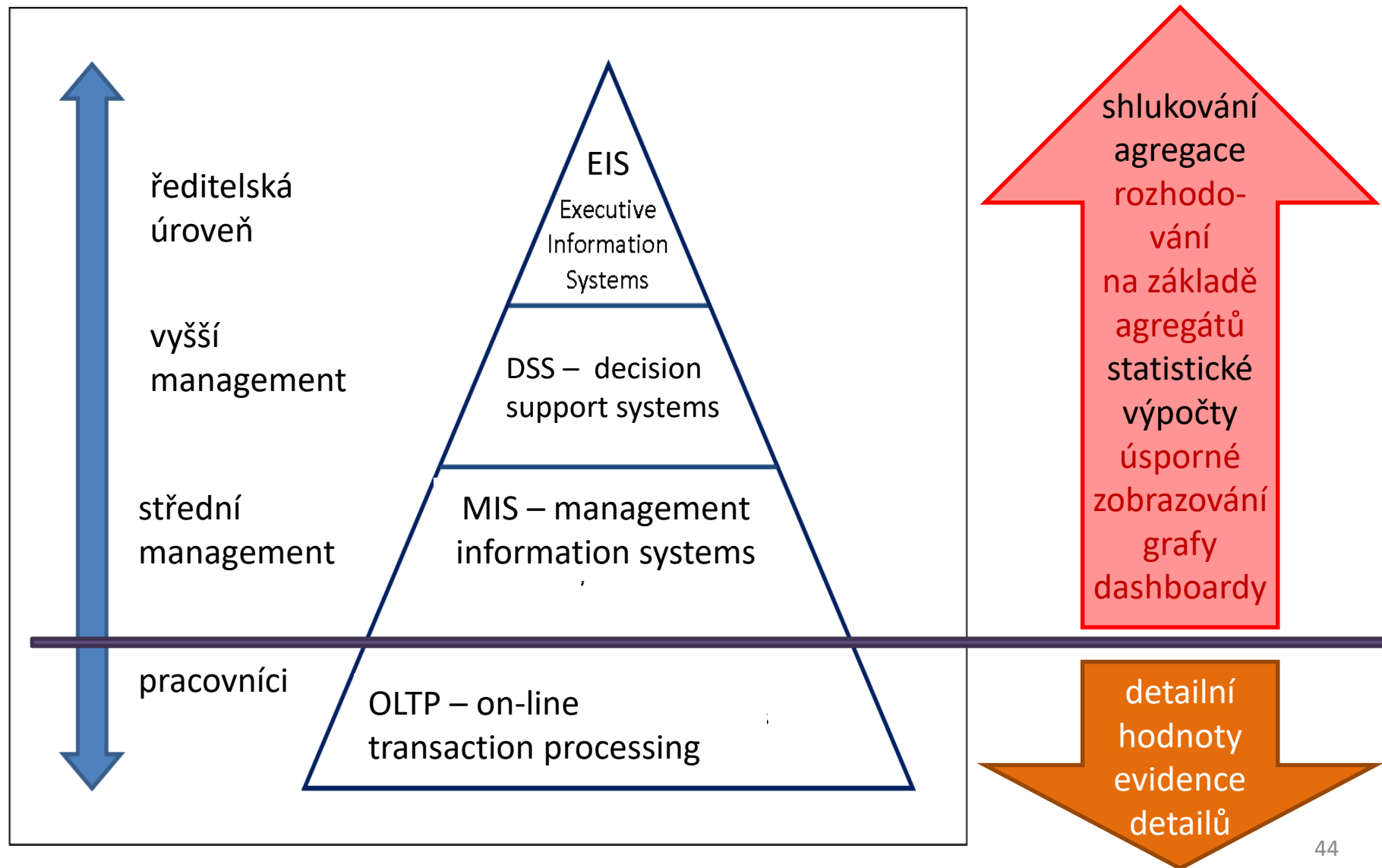
Definice multidimezionální kostky

- Aktivní dimenze budou uspořádány stejně jako v původní množině všech dimenzí



ZOPAKOVÁNÍ POJMŮ AGREGACE A AGREGAČNÍCH FUNKCÍ

Shlukování, agregace



Agregace

- ***agregace*** (z lat. *ad-*, *při-* a *grex, gregis*, stádo) znamená připojení, přičlenění, případně také ***shrnutí či shlukování***.
- ***agregační funkce*** – shlukující ***několik (množinu obvykle číselných) údajů*** do ***jediné hodnoty*** jistou ***agregační funkcí***
- ***agregát*** – výsledek agregace, spojení několika strojů, hodnot a pod.

Agregační funkce

- **Agregační funkce** jsou funkce, které shlukují (různým výpočtem) dohromady množiny hodnot do jediné hodnoty
- Obecně známé jsou:
- **počet**
- **součet**
- aritmetický **průměr** (nadále je aritmetický průměr a průměr synonymum)
- **maximum**
- **minimum**
- Méně známé jsou:
- **medián**
- **modus**

Modus

- Modus je hodnota, která se vyskytuje v množině hodnot nejčastěji.

Medián

- prostřední hodnota oddělující horní a dolní polovinu uspořádané datové množiny

Porovnání průměru, mediánu a modu

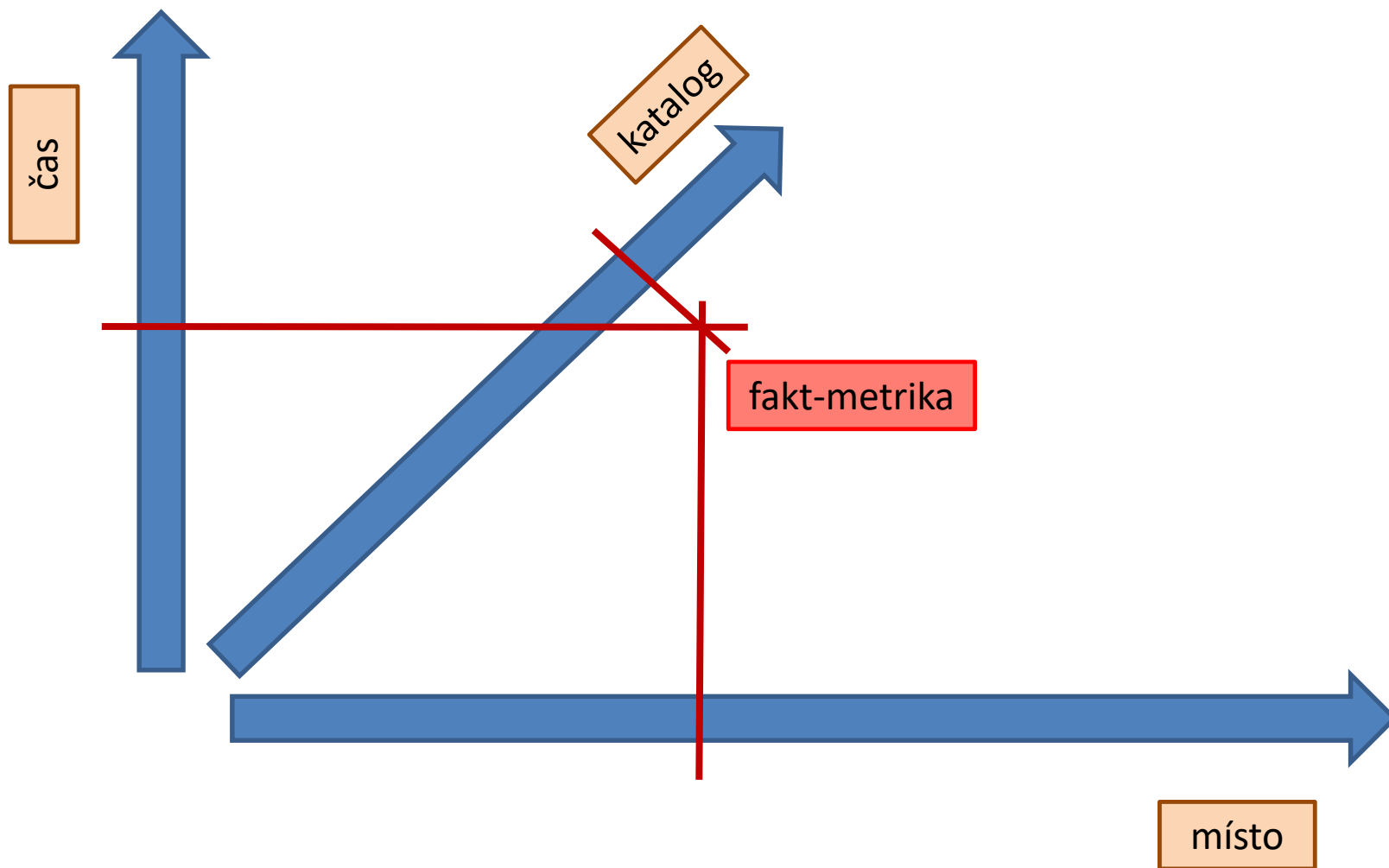
{ 1, 2, 2, 3, 4, 7, 9 }

funkce	popis	výpočet	výsledek
<i>průměr</i>	součet hodnot dělený počtem	$(1+2+2+3+4+7+9) / 7$	4
<i>medián</i>	prostřední hodnota oddělující horní a dolní polovinu uspořádané datové množiny	1, 2, 2, 3 , 4, 7, 9	3
<i>modus</i>	nejfrekventovanější hodnota v datové množině	1, 2, 2 , 3, 4, 7, 9	2

Definice multidimezionální kostky

- **Multidimenzionální kostka** je funkce $g_m (A_1 \times A_2 \times A_3 \times \dots \times A_m) = F$, kde $f \in F$ nazýváme **fakt (míra, measure)** a $A_1 \times A_2 \times A_3 \times \dots \times A_m$ je **kartézský součin** dimenzí.
- **Fakt (míra, measure)** je libovolná **agregovatelná** hodnota (lze ji sčítat, průměrovat, řetězit apod.), tedy existují kostky počtů, aritmetických průměrů, součtů apod.

Definice multidimenzionální kostky



Příklad definice míry Amount



Příklad míry Amount v Excelu

[Calendar].[Day]. [Key]	[Geo]. [City]	[Prod]. [Licence]	[Measures]. [Amount]
May 12, 2010	Madrid	Personal	1
May 13, 2010	Barcelona	Personal	2
May 14, 2010	Paris	Personal	4
May 15, 2010	Lausanne	Personal	8
May 16, 2010	Lausanne	Corporate	16
May 17, 2010	Lausanne	Partnership	32
May 18, 2010	Zurich	Partnership	64
May 19, 2010	Geneva	Corporate	128
May 20, 2010	New York	Corporate	256
May 21, 2010	New York	Corporate	512

ZOPAKOVÁNÍ - SVAZ

Částečně uspořádaná množina

- ***Částečně uspořádaná množina*** se skládá z množiny S a relace částečného uspořádání \leq . Obvykle ji, ale současně někdy také její graf, označujeme $[S; \leq]$.

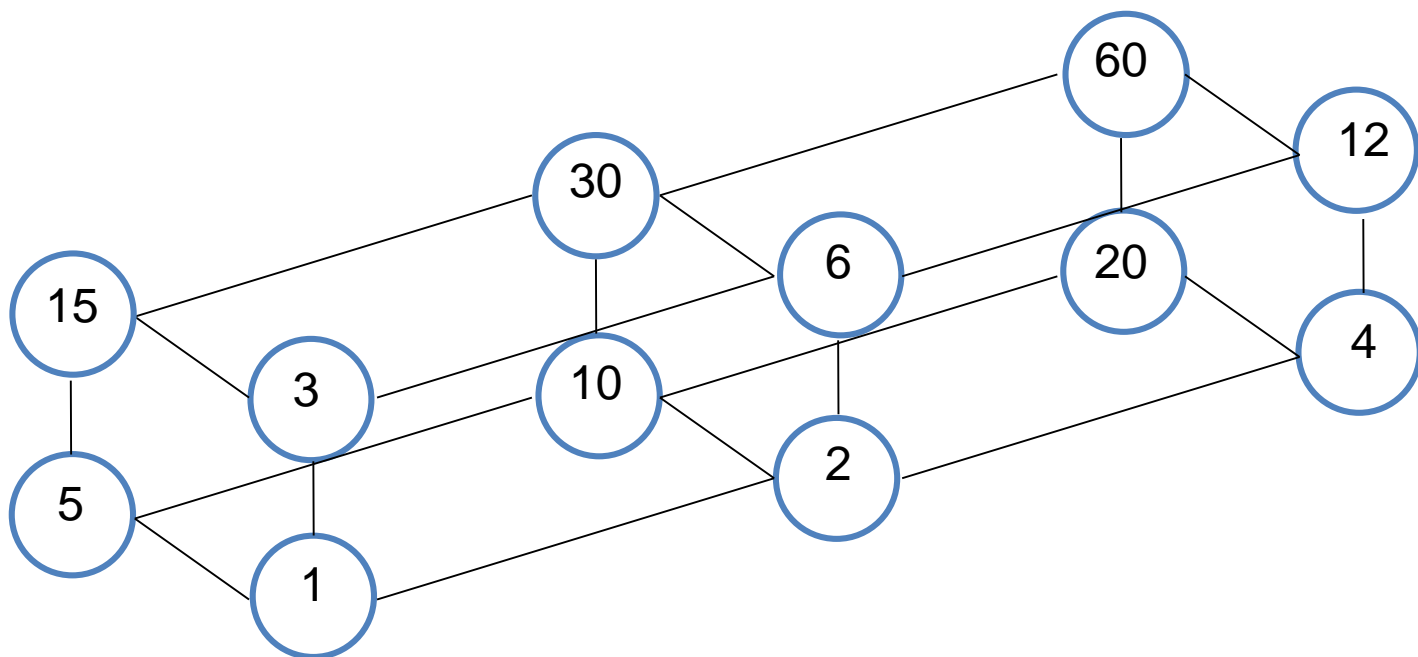
Přímý předchůdce, přímý následník

- Necht' $[S; \leq]$ je částečně uspořádaná množina. Pak prvek a nazveme **bezprostředním** (nebo také **přímým**) **předchůdcem** prvku b , pokud $a \leq b$ a neexistuje takový prvek c , pro který by platilo $a \leq c \leq b$. Relaci bezprostředního (přímého) předchůdce označujeme symbolem $<$. Inverzní relaci k nazýváme **bezprostředním** (**přímým**) **následníkem**.

Přímý předchůdce, přímý následník

- Relace bezprostředního předchůdce není ***ani reflexivní, ani symetrická, ani tranzitivní***.
- Relace přímého předchůdce je také podmnožinou původní relace částečného uspořádání.

Hasseův diagram



Množina $A = \{ 1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60 \}$ všech dělitelů čísla 60 je částečně uspořádaná podle dělitelnosti

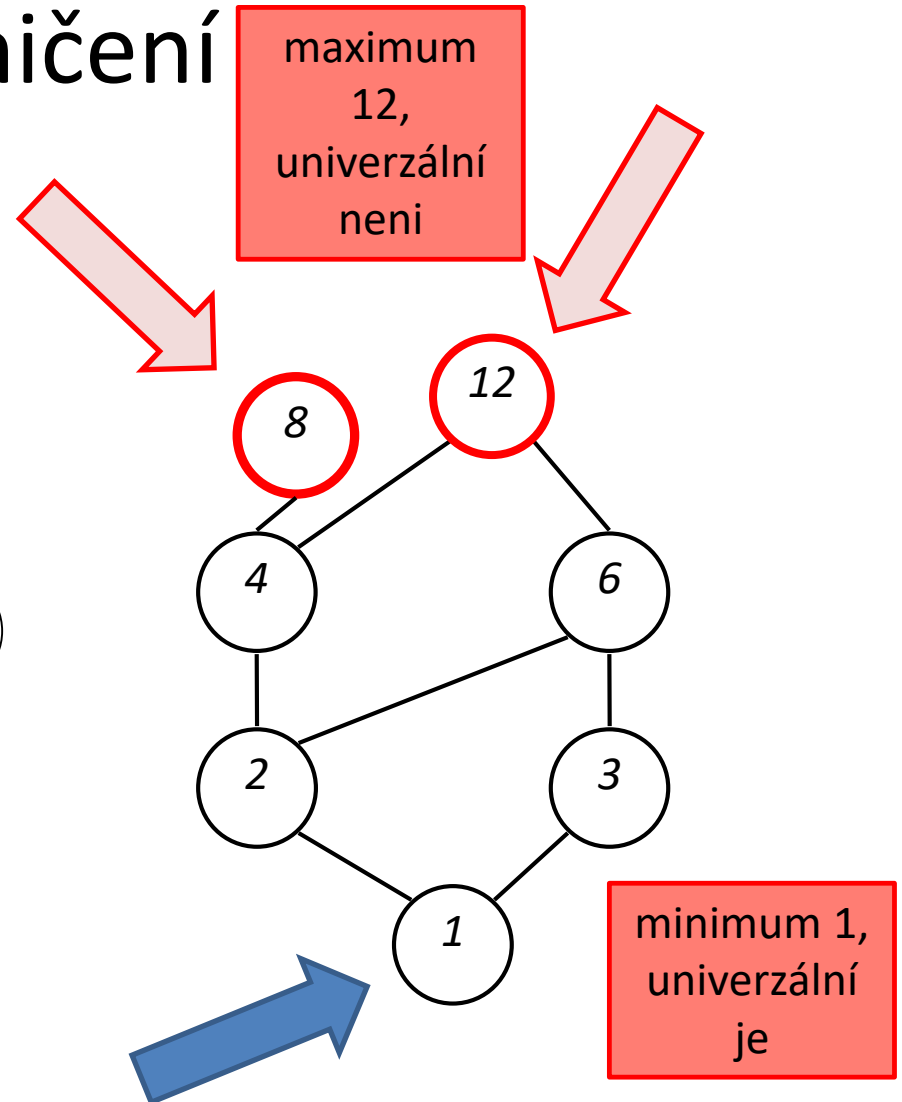
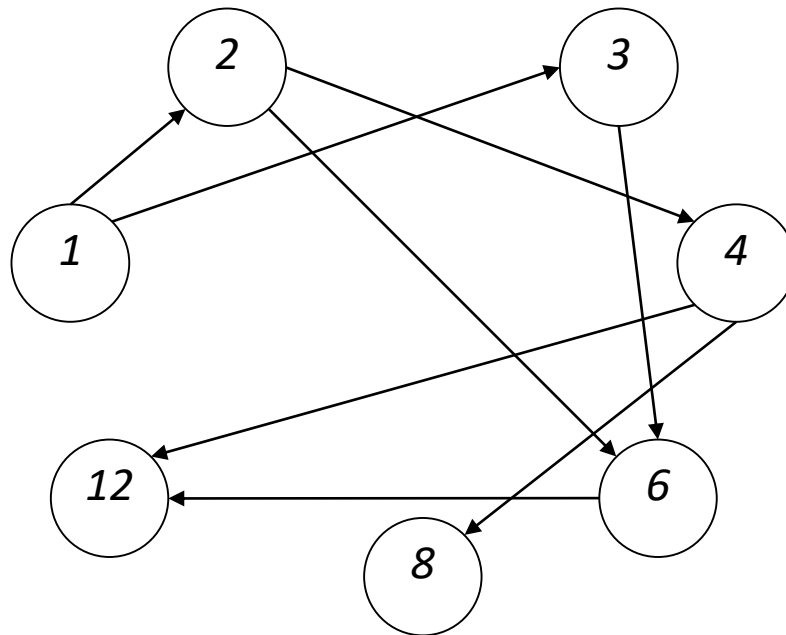
Maximální a minimální

- Prvek $\max \in [S; \leq]$ se nazývá **největší** (**maximální**), pokud neexistuje prvek $a \in [S; \leq]$ různý od \max , pro který by platilo platí $\max \leq a$. Podobně se pro inverzní relaci \geq definuje **nejmenší** (**minimální**) prvek \min .

Univerzální horní a dolní ohraničení

- Prvek $I \in [S; \leq]$ se nazývá ***univerzální horní ohraničení***, pokud pro všechny $a \in [S; \leq]$ platí $a \leq I$. Podobně se pro inverzní relaci \geq definuje ***univerzální dolní ohraničení*** prvek O .

Příklad univerzálního dolního ohraničení



Univerzální horní a dolní ohraničení

- Konečná částečně uspořádaná množina má vždy maximální a minimální prvek. Nemusí ovšem mít horní, resp. dolní univerzální ohraničení.
- částečně uspořádaná množina z předchozího obrázku Hasseova diagramu má univerzální dolní ohraničení (tj. prvek 1), ale nemá horní univerzální ohraničení (ale má maximální prvek 12).

Spojení

- Necht' a, b jsou prvky částečně uspořádané množiny $a, b \in [S; \leq]$. ***Spojením*** prvků nazýváme prvek c , pro který platí $a \leq c$ a $b \leq c$ a neexistuje prvek $x \in (S - \{c\})$ takový, že $a \leq x \leq c$ a také $b \leq x \leq c$. Pokud dva prvky mají ***jediné spojení***, hovoříme o ***nejmenším horním ohraničení*** a označujeme jej $a \vee b$.

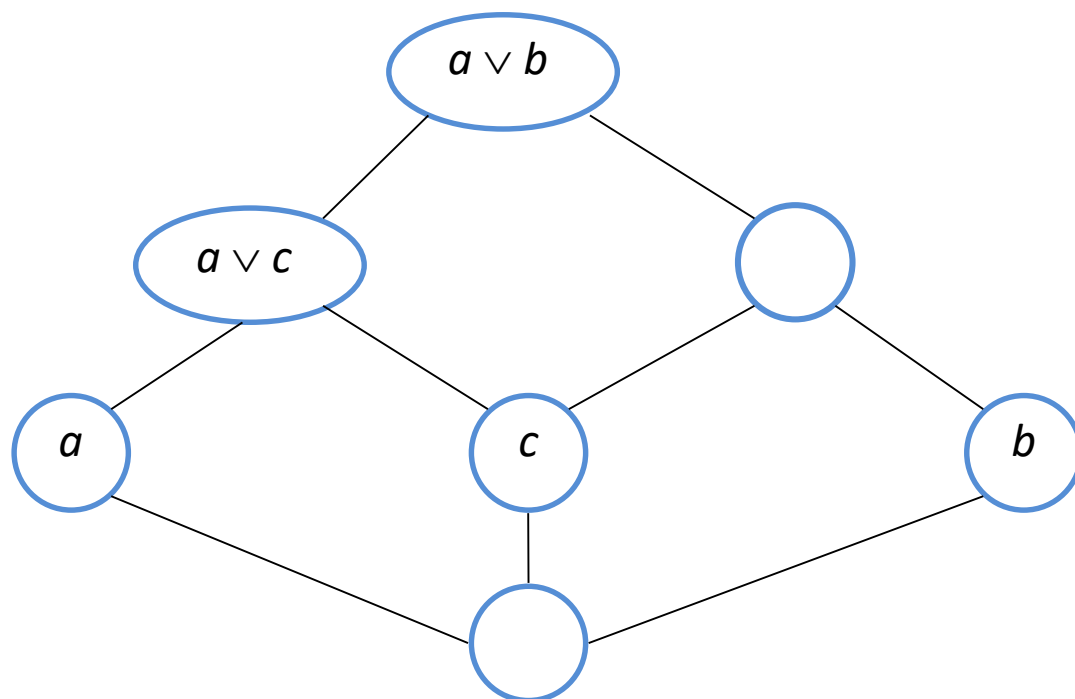
Průsek

- Necht' a, b jsou prvky částečně uspořádané množiny $a, b \in [S; \leq]$. **Průsekem** prvků nazýváme prvek d , pro který platí $d \leq a$ a $d \leq b$ a neexistuje prvek $x \in (S - \{d\})$ takový, že $d \leq x \leq a$ a také $d \leq x \leq b$. Pokud dva prvky mají **jediný průsek**, hovoříme o **největším dolním ohraničení** a označujeme jej $a \wedge b$.

Spojení a průsek

- Spojení e dvou libovolných prvků a a b z konečné uspořádané množiny $[S; \leq]$ zjistím poměrně snadno z Hasseova diagramu.
- Skutečnost, že $a \leq e$ znamená, že existuje řetězec hran vedoucích vzhůru z a do e .
- To stejné platí i pro $b \leq e$. e je potom společným prvkem ležícím na obou řetězcích takový, že žádný jiný nemá takovou vlastnost.
- Z toho také pochází i název spojení. Spojení dvou prvků a a b je takový vrchol, který poprvé spojí řetězce hran vedoucí vzhůru vycházející z prvků a a b .

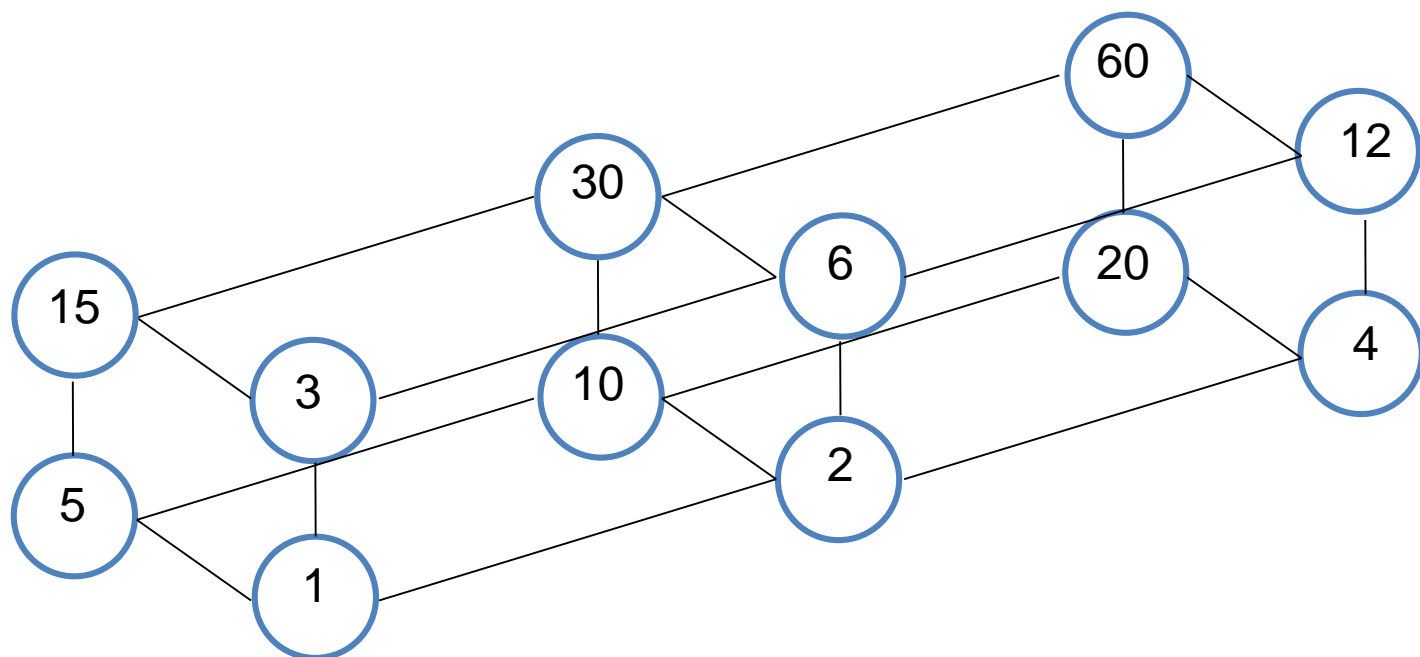
Spojení pomocí Hasseova diagramu



Svaz

- ***Svaz*** (anglicky ***lattice***, což je anglicky mřížka a souvisí to s podobou Hasseova diagramu pro svaz) je
- ***částečně uspořádaná množina*** $[S; \leq]$
- každé dva prvky mají ***jediné spojení a jediný průsek***.
- svaz označujeme $[S, \vee, \wedge]$.

Příklad svazu dělitelnosti



Příklad svazu

- každá dvojice prvků má jediné spojení a jediný průsek.
- množina $A = \{ 1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60 \}$ společně se spojením ***nejmenší společný násobek*** a průsekem ***největší společný dělitel*** je ***svazem***.

ZPĚT K MULTIDIMENZIONÁLNÍMU MODELU

Definice multidimezionální kostky

- ***Podkostky (kuboidy)*** odvozené od jednoho uspořádání dimenzí tvoří ***poset*** aneb ***částečně uspořádanou množinu***.

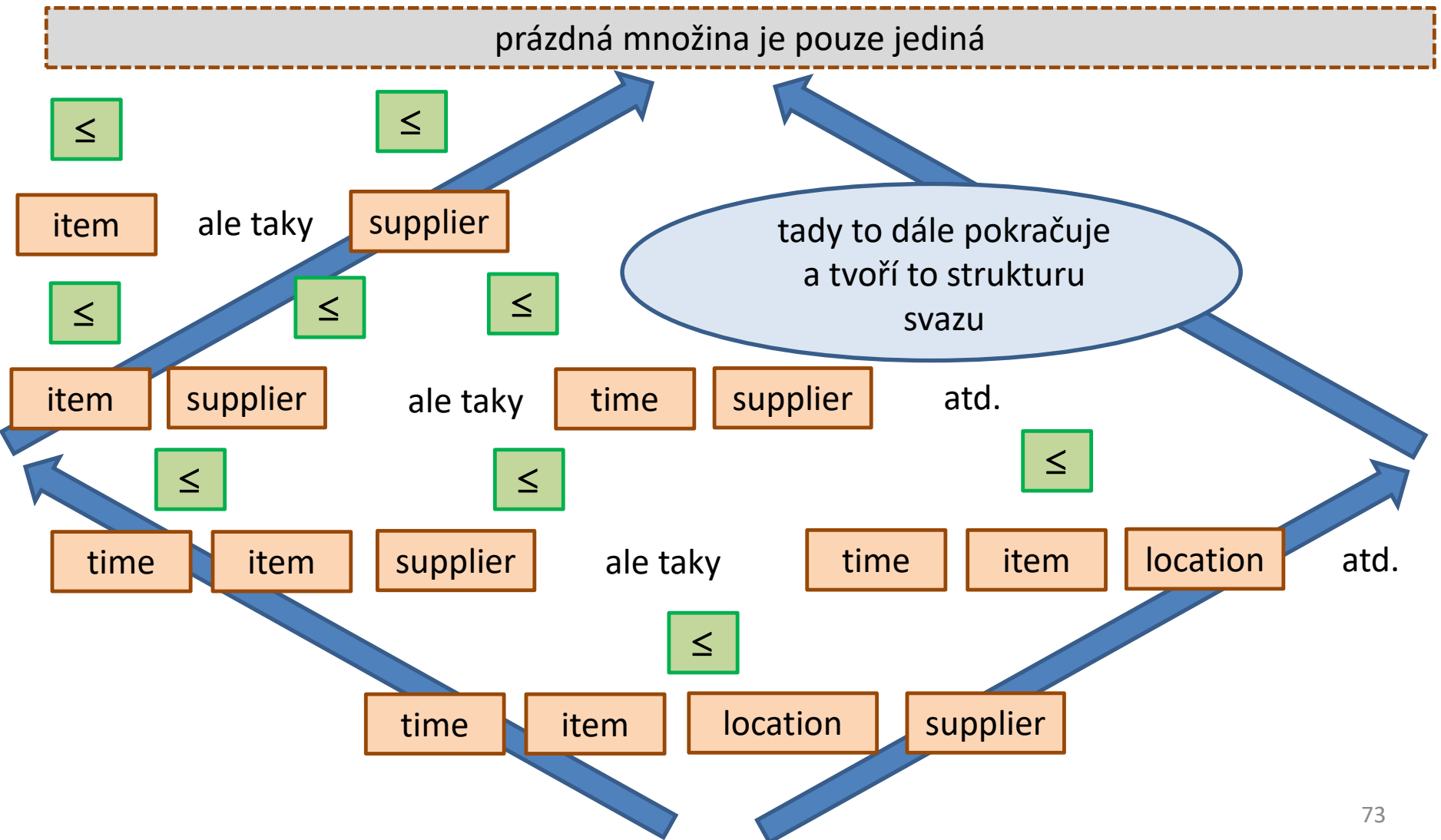
Částečně uspořádaná množina podkostek

- Poset nad podkostkami kostky \mathbf{g} je definován následovně:

$$\mathbf{g}_m (A_1 \times A_2 \times A_3 \times \dots \times A_i \times \dots \times A_m) \\ \leq \\ \mathbf{g}_{m-1} (A_1 \times A_2 \times A_3 \times \dots \times A_{m-1})$$

- Podkostka je přímo větší, pokud má o jednu méně dimenzí a ostatní jsou uspořádány stejně v obou podkostkách

Příklad podkostek a relace \leq



Částečně uspořádaná množina podkostek

- ***Multidimenzionální podkostky*** pro jedno uspořádání dimenzí a jednu agregační funkci tvoří ***svaz***

Svaz podkostek

- Co je to **svaz** viz také

zdroj: studijní opora: Matematické základy modelování informačních systémů

- **Univerzální horní ohraničení** - vrcholová podkostka **all**
- **Univerzální dolní ohraničení** - základní kostka

$$g_n(D_1 \times D_2 \times D_3 \times \dots \times D_n)$$

Svaz podkostek kostky g_n

- **Průsek** \wedge

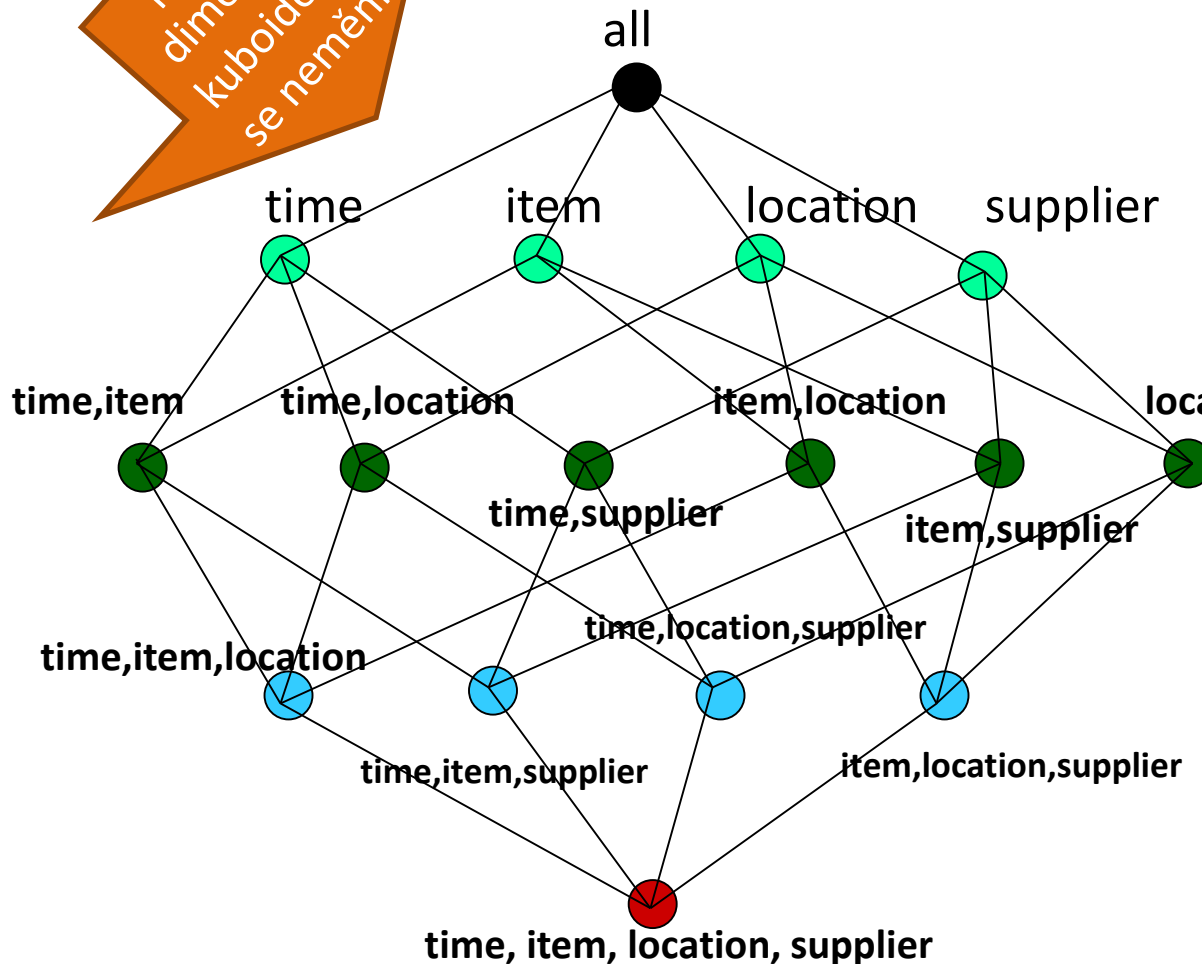
$$\begin{aligned} &g_m(A_1 \times A_2 \times A_3 \times \dots \times A_i \times \dots \times A_m)^\wedge \\ &g_m(A_1 \times A_2 \times A_3 \times \dots \times A_j \times \dots \times A_m) = \\ &g_{m+1}(A_1 \times A_2 \times A_3 \times \dots \times A_i \times \dots \times A_j \times \dots \times A_{m+1}) \end{aligned}$$

- **Spojení** \vee

$$\begin{aligned} &g_m(A_1 \times A_2 \times A_3 \times \dots \times A_i \times \dots \times A_m)^\vee \\ &g_m(A_1 \times A_2 \times A_3 \times \dots \times A_j \times \dots \times A_m) = \\ &g_{m-1}(A_1 \times A_2 \times A_3 \times \dots \times A_{m-1}) \end{aligned}$$

Kostka jako svaz kuboidů

pořadí
dimenzí v
kuboidech
se nemění



0D (vrcholový) kuboid

1D kuboidy

2D kuboidy

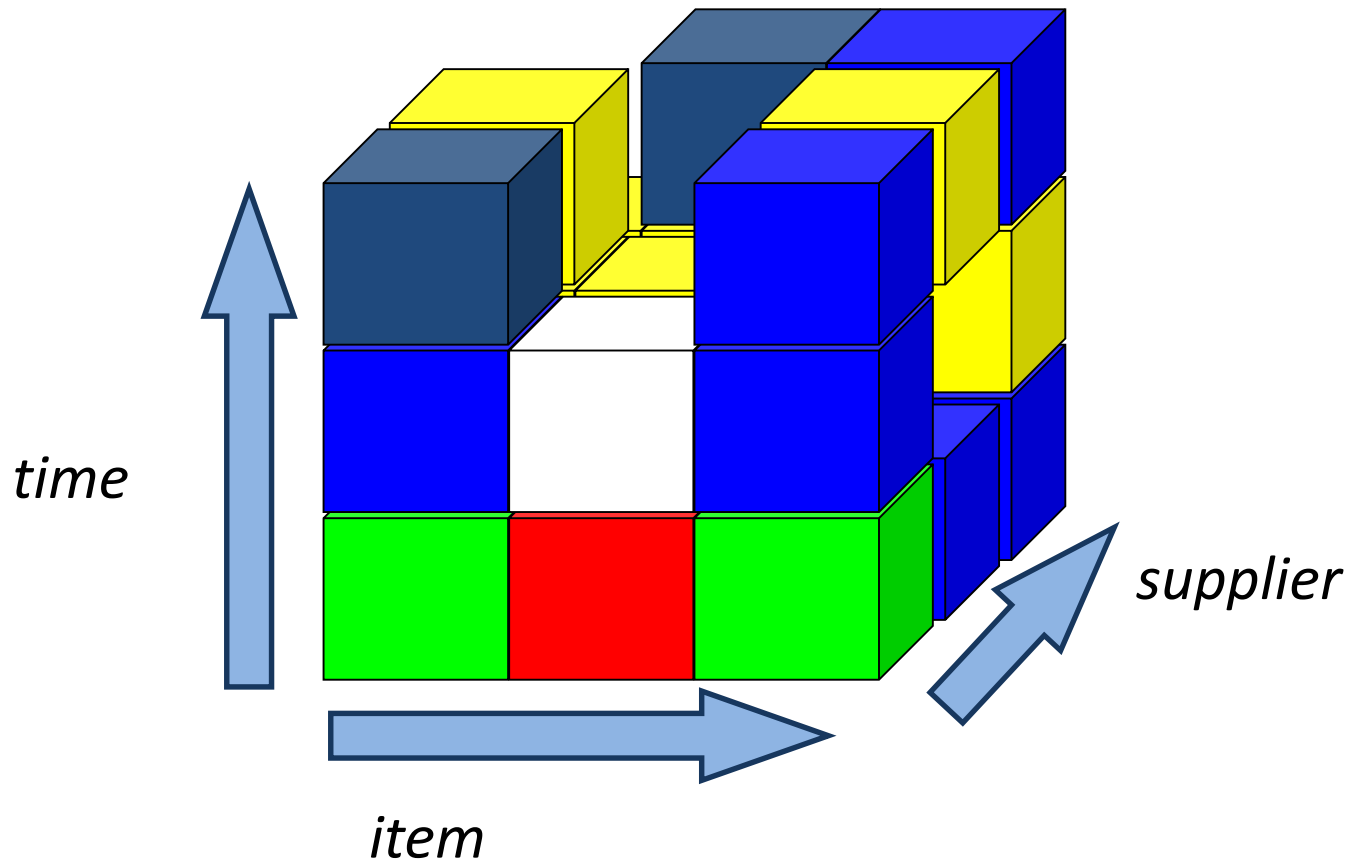
3D kuboidy

4D (základní) kuboid

Model multidimenzionální kostky

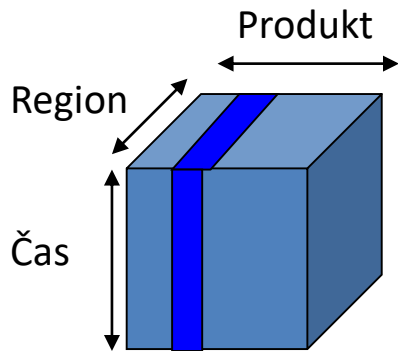
- Každá dimenze A_i má ***možnou kardinalitu*** danou *datovým typem dimenze* (i nekonečnou) a ***skutečnou kardinalitu*** k_i (tj. skutečný počet prvků dimenze daný ***existujícími ohodnocenými fakty***)
- Proto bude výsledná kostka "***vykotlaná***", tj. ***řádká n -rozměrná matice***.

Multidimenzionální kostka jako řídká matice

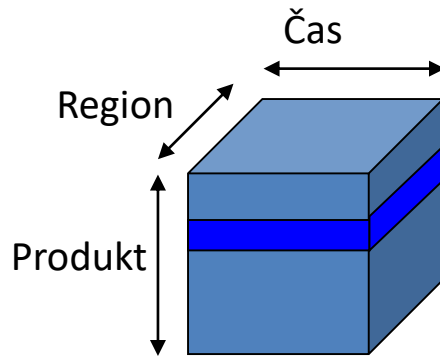


3D kostka - příklad 3! otočení

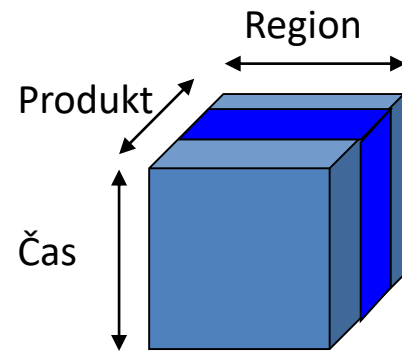
- Princip multidimenzionální kostky



Analýza údajů
pro určitý produkt-
produkt, region, čas



Analýza pro určité
časové období -
čas, region, produkt



Analýza údajů podle
regionálních kritérií -
region, produkt, čas

je možný i produkt, čas, region - čas, produkt, region a region, čas, produkt

Typy agregačních výpočtů-distributivní

- ***Distributivní***: Pokud můžeme kostku rozdělit na části, pro něž se dané funkce vypočte zvlášť a poté jednotlivé výsledky spočítat dohromady, a neovlivní to výsledek funkce, pak je funkce distributivní
- např. počet, ***součet***, min, max
- ***součet*** je distributivní, tj. celkový součet je součtem dílčích součtů

Typy agregačních výpočtů - algebraické

- ***Algebraické***: jde o výsledek funkce, která má M parametrů, z nichž každý může být získán výpočtem distributivní funkce.
- Příkladem je průměr, který lze spočítat jako součet/počet
- ***průměr*** je algebraický, tj. celkový průměr ***není obecně průměrem dílčích průměrů***

Kdo tomu nevěří

- Platí to, pouze ***jsou-li počty vzorků v obou množinách stejné***
- {4,5,6}, součet=15, průměr=15/**3**=5
- {100,200,300}, součet=600,
průměr=600/**3**=200
- {5,200}, součet=205, průměr=205/2=**102.5**
- {4,5,6,100,200,300}, součet=615,
průměr=615/6=**102.5**

Kdo tomu nevěří

- nejsou-li stejné:
- $\{1,2,3,4,5,6\}$, součet=21, průměr=21/**6**=3.5
- $\{100,200,300\}$, součet=600,
průměr=600/**3**=200
- $\{3.5,200\}$, součet=203.5,
průměr=203.5/2=**101.75**
- $\{1,2,3,4,5,6,100,200,300\}$, součet=621,
průměr=621/9=**69**

Typy agregačních výpočtů - holistické

- ***Holistické (celostní)***: pro tyto jednotky neexistuje algebraické funkce,
- např. zde patří funkce pro zjištění nejčastěji se vyskytující položky

Typy agregačních výpočtů

- Budeme se zabývat hlavně ***distributivními*** (počet, součet) a
- ***algebraickými*** agregáty (aritmetický průměr).

Neagregované - detailní hodnoty

- Předpokládejme, že existuje n dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$
- Každá z n dimenzí má **skutečnou kardinalitu** k_i prvků pro všechna $1 \dots i \dots n$.
- Fakt dané kostky pro agregační funkci **agr**, **detail** $(D_1 \times D_2 \times D_3 \times \dots \times D_n) = F$, kterou je existující hodnota na průsečíku hodnot dimenzí $(d_1, d_2, d_3, \dots, d_i)$ nazveme detailní hodnotou nebo **detailem**.

Kostky počet, součet a průměr

- Uvažujme nyní výpočet funkcí (multidimenzionálních kostek)
- *počet*,
- *součet*,
- *průměr*
- pro uspořádanou podmnožinu aktivních dimenzí $\{A_1, A_2, A_3, \dots, A_m\}$ a dané uspořádání dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$.

První krok agregace detailů

- Pro dané uspořádání dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$ platí pro jednotlivé podkostky ***počet_i***, ***součet_i*** a ***průměr_i*** ($d_1, d_2, d_3, \dots, d_i$) a detailní hodnoty ***h*** ohodnocené hodnotami všech dimenzí ($d_1, d_2, d_3, \dots, d_n$):
- ***součet_n***($d_1, d_2, d_3, \dots, d_n$) = Σ ***detail*** ($d_1, d_2, d_3, \dots, d_n$)
- ***počet_n***($d_1, d_2, d_3, \dots, d_n$) = Σ ***1*** přes všechny hodnoty ***detail*** ($d_1, d_2, d_3, \dots, d_n$) s týmiž hodnotami dimenzí
- ***průměr_n***($d_1, d_2, d_3, \dots, d_n$) = $\text{***součet_n***($d_1, d_2, d_3, \dots, d_n$) / ***počet_n***($d_1, d_2, d_3, \dots, d_n$)}$

Podkostky

- Potom pro dané uspořádání $n > m \geq 0$ dimenzí $\{A_1, A_2, A_3, \dots, A_m\}$ platí pro příslušné podkostky **počet**_m, **součet**_m a **průměr**_m a jejich přímé následníky $m+1$ v částečném uspořádání $\{A_1, A_2, A_3, \dots, A_i, \dots, A_m\}$:
- **součet**_m($d_1, d_2, d_3, \dots, d_m$) = Σ **součet**_{m+1}($d_1, d_2, d_3, \dots, d_i, \dots, d_m$) přes všechny hodnoty z D_i
- **počet**_m($d_1, d_2, d_3, \dots, d_m$) = Σ **počet**_{m+1}($d_1, d_2, d_3, \dots, d_i, \dots, d_m$) přes všechny hodnoty z D_i
- **průměr**_m($d_1, d_2, d_3, \dots, d_m$) = **součet**_m($d_1, d_2, d_3, \dots, d_m$) / **počet**_m($d_1, d_2, d_3, \dots, d_m$)

Příklad detailních hodnot

hodnoty dimenzí seřazeny např. vzestupně podle
uspořádání

time	item	location	fakt
22.6.	párek	Brno	4
22.6.	párek	Brno	9
22.6.	párek	Praha	21
22.6.	rohlík	Brno	12
22.6.	rohlík	Brno	4
22.6.	rohlík	Brno	3
22.6.	rohlík	Praha	16
22.6.	rohlík	Praha	6
23.6.	rohlík	Brno	5
23.6.	rohlík	Praha	14

celkový
součet
se
nemění

94

více
hodnot
faktů pro
stejně
hodnoty
dimenzí

První krok pro kostku *součet*

time	item	location	fakt
22.6.	párek	Brno	13
22.6.	párek	Praha	21
22.6.	rohlík	Brno	19
22.6.	rohlík	Praha	22
23.6.	rohlík	Brno	5
23.6.	rohlík	Praha	14
			94

- není více hodnot faktů pro stejné hodnoty dimenzí
- nebylo dimenzí

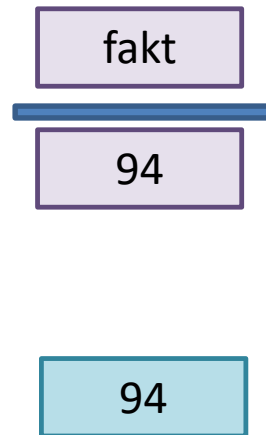
Podkostka *součet* bez dimenze location

time	item	fakt
22.6.	párek	34
22.6.	rohlík	41
23.6.	rohlík	19
		94

Podkostka *součet* bez dimenze item

time	fakt
22.6.	75
23.6.	19
94	

Podkostka *součet* bez dimenzí



při nulovém počtu dimenzí
jde o celkový agregát

Jiné pořadí dimenzí v prvním kroku

location	time	item	fakt
Brno	22.6.	párek	13
Brno	22.6.	rohlík	19
Brno	23.6.	rohlík	5
Praha	22.6.	párek	21
Praha	22.6.	rohlík	22
Praha	23.6.	rohlík	14

94

Podkostka *součet* bez dimenze item

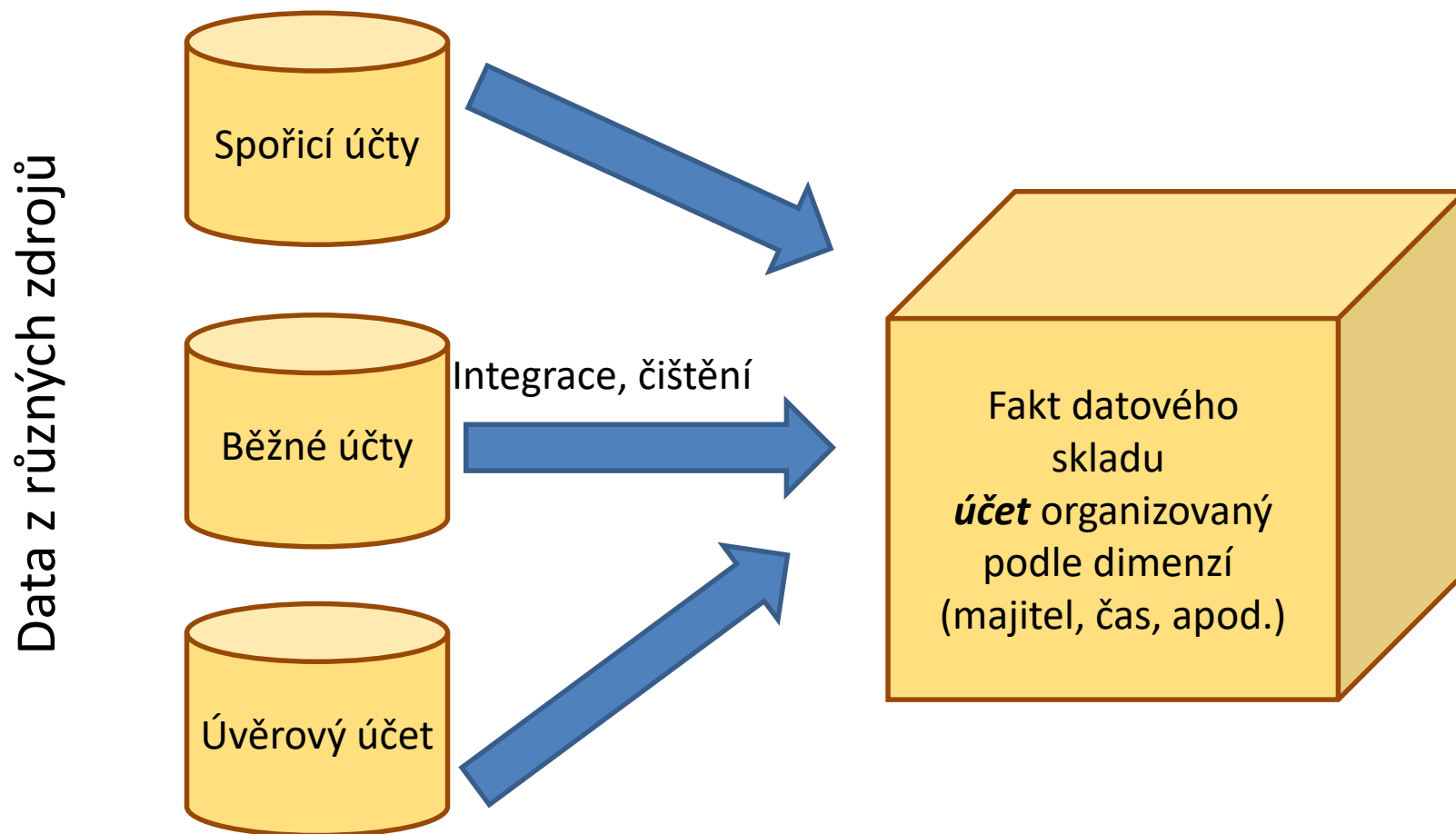
location	time	fakt
Brno	22.6.	32
Brno	23.6.	5
Praha	22.6.	43
Praha	23.6.	14
		94

VLASTNOSTI DATOVÉHO SKLADU

Orientace podle dimenzí

- Fakta jsou zapisována podle vlastností vyjádřeného průsečíkem ***n-dimenzí***

Integrace příklad



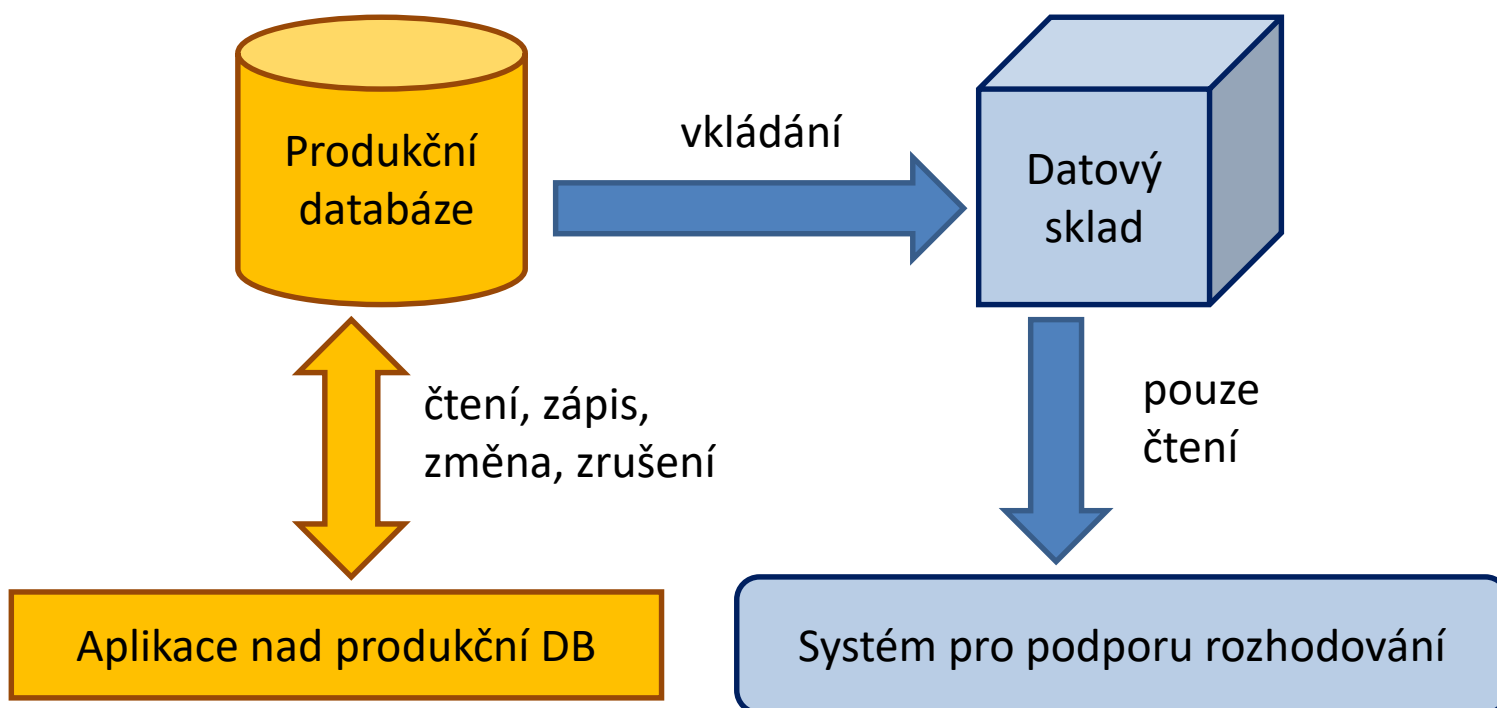
Integrace

- data týkající se konkrétního předmětu se ukládají ***pouze jednou***
 - jednotná terminologie, jednotky veličin
- vytvořen případně ***integrací*** několika heterogenních zdrojů dat - relační databáze, textové soubory, on-line transakce
 - problém nekonzistentních zdrojů dat
- nutnost ***úpravy, vyčištění a sjednocení*** (integrace) vstupních dat
 - je nutné ověřit konzistenci v pojmenování proměnných, jejich struktury a jednotkách pro různé zdroje dat

Čas

- klíčový atribut (*lineární, uspořádaný a spojitý*), vhodný na spojité grafy
- Časový horizont datového skladu je zpravidla podstatně delší než u produkční databáze
 - Produkční databáze: pouze současně aktuální data
 - Data v datovém skladu: poskytují informace z historické perspektivy (např. posledních 5-10 let)
- Každá klíčová struktura v datovém skladu
 - obsahuje čas, explicitně nebo implicitně
 - klíč u produkčních dat nemusí vždy obsahovat čas

Neměnnost



Neměnnost

- zpravidla ***fyzicky oddělené*** uložení dat transformovaných z produkčních databází
- v datových skladech se data ***nemění***, manipulace s daty je tedy jednodušší.
 - dva typy operací: ***vkládání*** dat a ***čtení*** dat
 - optimalizace a normalizace ztrácí smysl
 - ***nepotřebuje se*** zpracování ***transakcí***, zotavení, mechanismy ***pro řízení souběžného přístupu***

Produkční databáze x **datový sklad**

- ***Uživatelé a orientace systému:***
 - běžný uživatel x ***manager***
- ***Datový obsah:***
 - současná, detailní x ***historická, sloučená***
- ***Návrh databáze:***
 - ER model + aplikace x ***multidimenzionální kostka, dimenze, fakta***
- ***Pohled na data:***
 - aktuální, lokální x ***agregovaný***
- ***Přístupové vzory:***
 - jednoduchá aktualizace x ***read-only, komplexní dotazy***

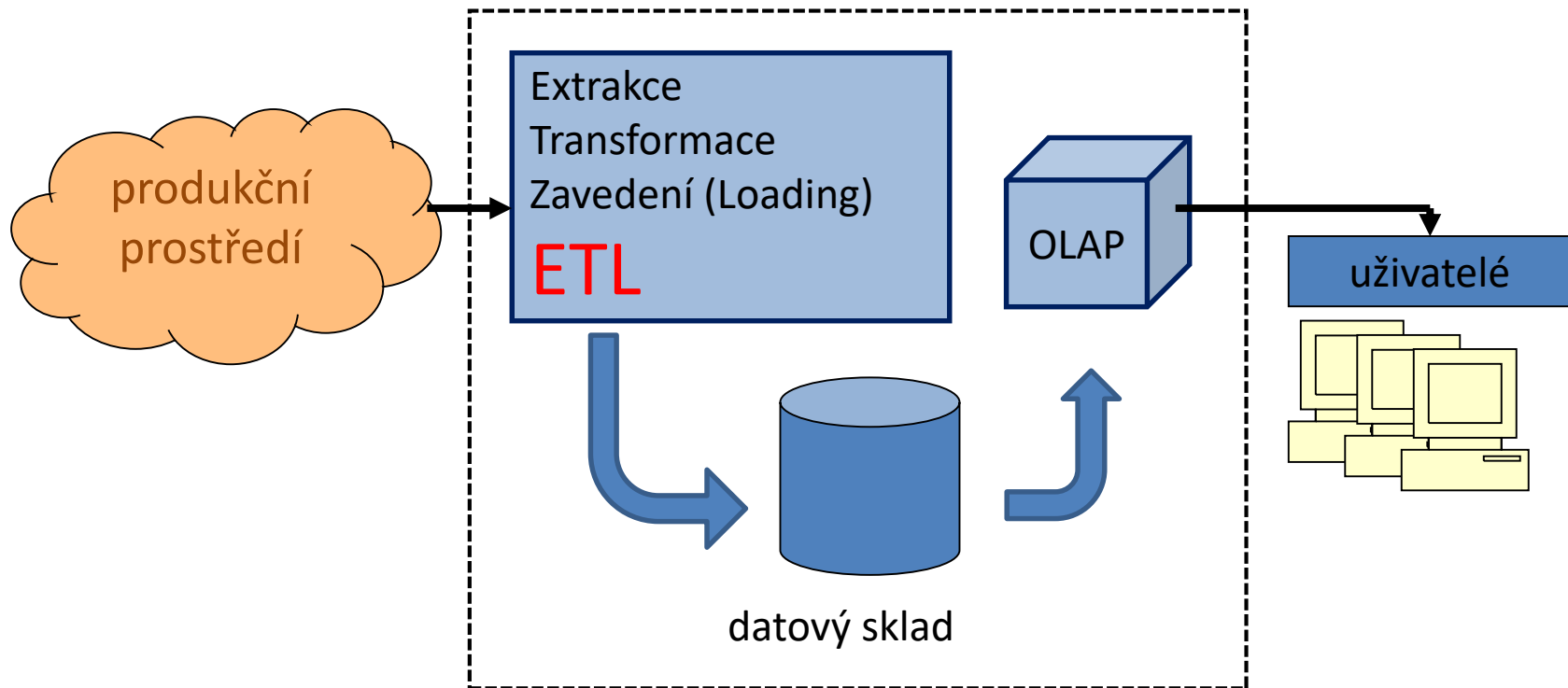
Porovnání vlastností

Vlastnost	Produkční DB	Datový sklad
Čas odezvy	ms - s	s - hod
Operace	DML, např. SQL	Jen čtení, např. MDX
Původ dat	30 – 60 dní	Snímky za čas. úsek
Organizace dat	Podle aplikace	Podle dimenzí
Velikost	Malá až velká	big data
Zdroje dat	operační, interní	operační, interní, externí
Činnosti	Transakce (OLTP)	Analýza (OLAP)

Shrnutí požadavků na datový sklad

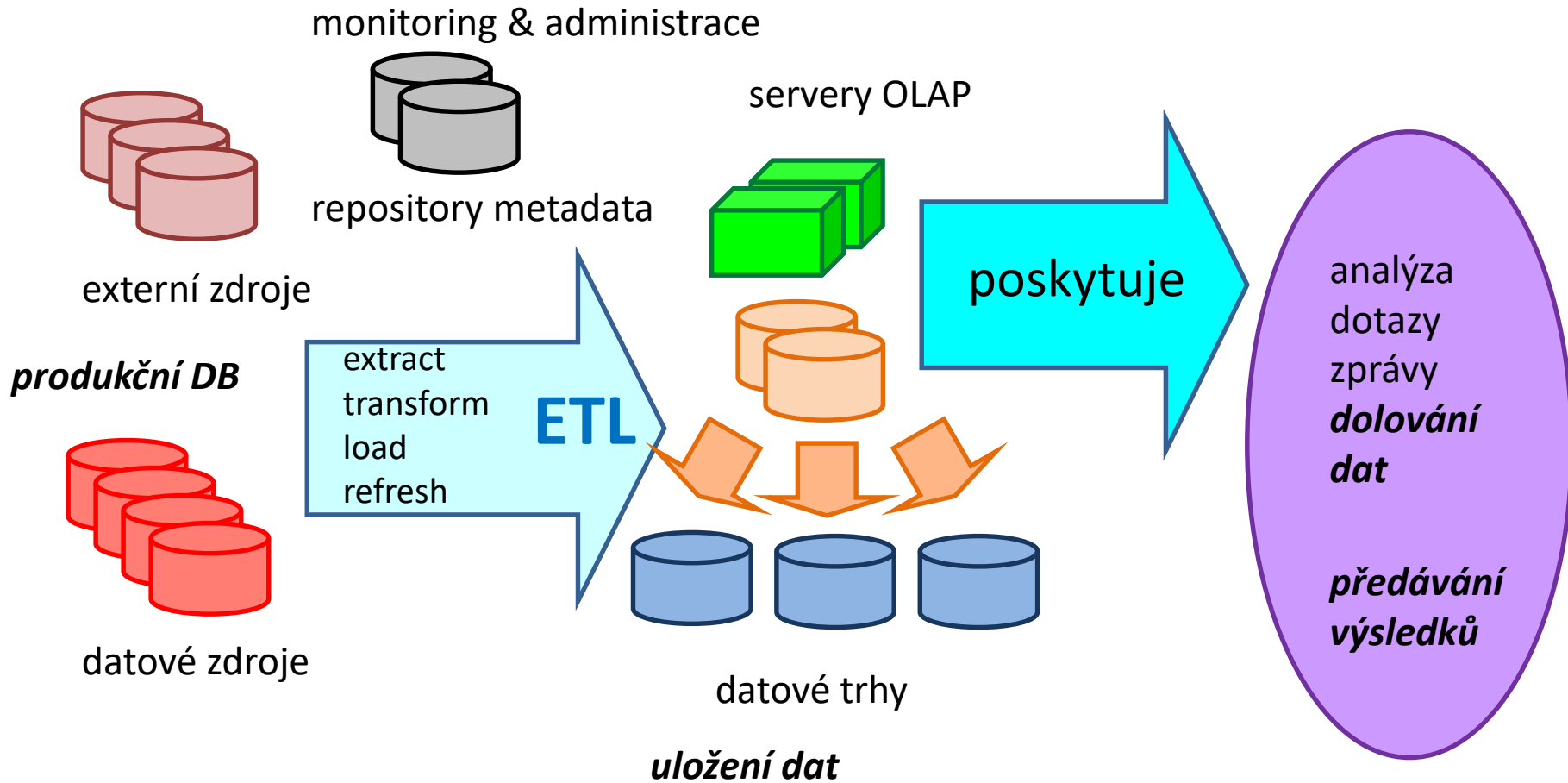
- schopnost ***agregace***
- databáze navržená pro ***analytické*** dotazy
- možnost ***integrovat*** data z více aplikací
- častá operace ***čtení*** z databáze
- nahrání dat po určité časové periodě
- možnost využití ***současných i historických dat***

Celkové schéma datového skladu



Získání údajů -> úprava a zavedení do datového skladu
-> analýza -> zpřístupnění uživatelům

Architektura datového skladu podrobnější



Architektura datového skladu

- 3 hlavní části
- ***Získání dat***
 - Zdrojová data + místo přípravy dat (***ETL***)
- ***Uložení dat (datová kostka)***
 - Datový sklad + datové trhy + uložení metadat
- ***Předávání výsledků***
 - přes model multidimenzionální databáze samotné získání informací (***OLAP***, data mining, tiskové sestavy atd.)

Získání dat

- ***extrahování dat*** z mnoha produkčních databází a externích zdrojů
- čištění, transformování a integrování těchto dat - ***Extraction, Transformation, Loading - ETL***
- ***periodické doplňování*** datového skladu tak, aby odrážel změny v produkčních databázích a přenos dat z datového skladu, nejčastěji do pomalejší archivní paměti.

Uložení dat

- k hlavnímu datovému skladu je přidruženo několik ***datových trhů (data marts)*** různých dílčích uživatelů.
- data ve skladu a trhu jsou uložena a spravována jedním nebo více skladovými servery, které prezentují multidimenzionální pohled na data OLAPu a/nebo přímo různým koncovým prostředkům

Uložení dat

- Jde o oddělené skladiště pro uložení velkého množství především historických dat
- Je navrženo ***pro analýzu***, ne pro rychlý přístup k datům
- ***read-only***
- Musí být přístupné pro více druhů nástrojů – odpovídající rozhraní

Předávání výsledků

- OLAP
- a jiné prostředky
- dotazovací,
- generátory zpráv,
- analytické a
- prostředky ***dolování dat (data mining)***.

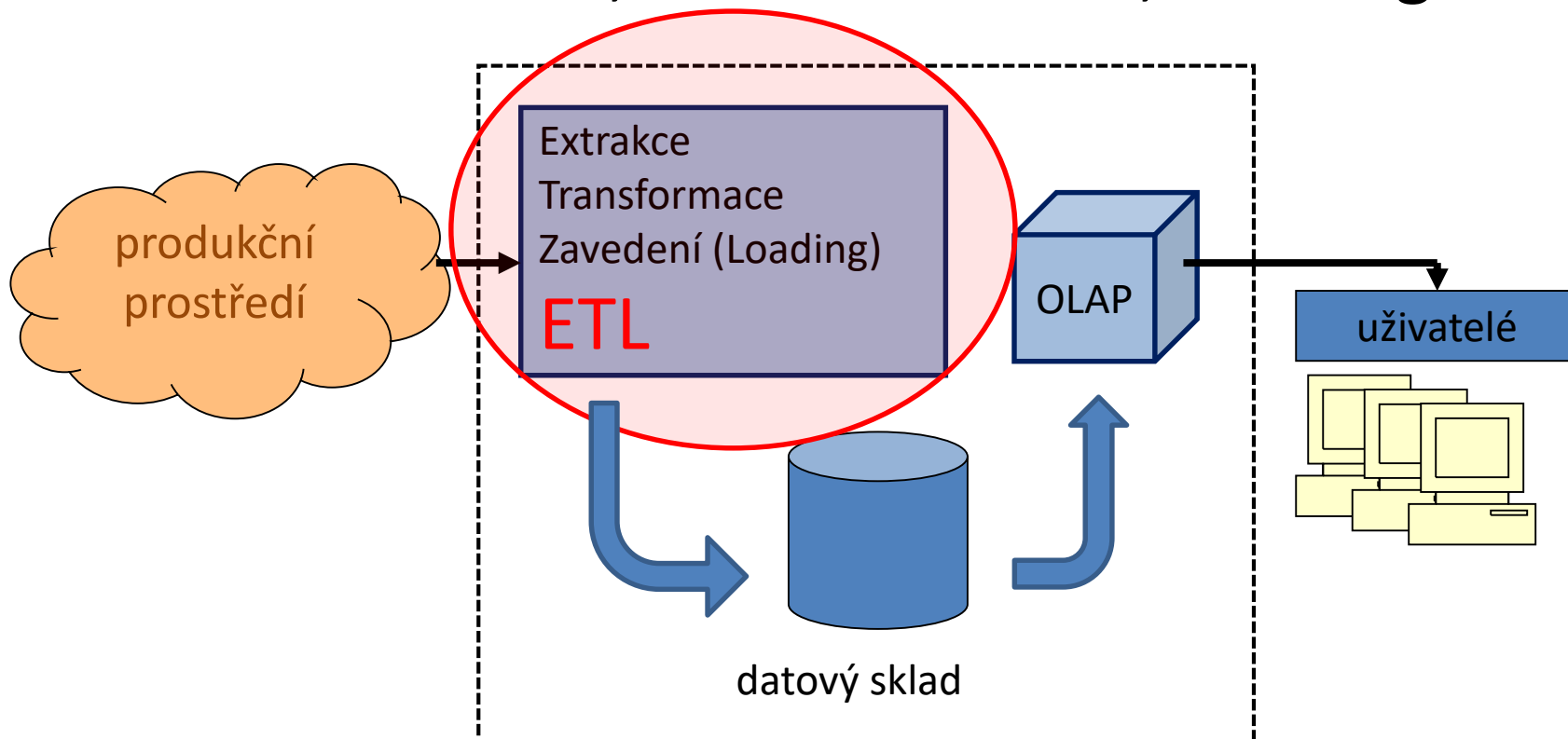
Předávání výsledků

- Poskytuje informace pro různé uživatele
 - ***Začínající uživatelé***: tiskové sestavy, jednoduché dotazy
 - ***Běžní uživatelé***: statistická analýza, různá zobrazení dat, předdefinované dotazy
 - ***Pokročilí uživatelé***: provádění multidimenzionální analýzy, formulace vlastních OLAP dotazů, používání data miningu

ZÍSKÁNÍ DAT - EXTRACTION, TRANSFORMATION, LOADING

Příprava údajů – etapa ETL

- Klíčová úloha správy datového skladu
- ETL = Extraction, Transformation, Loading



Typy zdrojových dat

- ***Produkční data***

- Data získaná z různých produkčních DB podniku pomocí dotazů

- ***Interní data***

- Data uložená v privátních souborech (zpravidla Excel XLS) zaměstnanců organizace

- ***Externí data***

- Data z různých zdrojů, která mohou být pro organizaci užitečná (webové služby)

Etapa ETL

- ***Extrakce*** – výběr dat různými metodami
- ***Transformace*** – ověření, čištění, integrace a časové označení dat
- ***Loading*** – přesun dat do datového skladu
- ***Hlavní cíl***: integrace údajů

Hlavní úkoly ETL procesu

- ***Určit zdroje dat***, interní i externí
- Příprava ***mapování*** mezi zdrojovými a cílovými daty
- Stanovení ***metadat*** pro extrakci dat
- Určit ***pravidla pro transformaci a čištění*** dat
- Plán pro ***agregaci*** tabulek
- Návrh oblasti přípravy dat
- Napsat procedury pro nahrávání dat

Extrakce

- **Zdroj:** Data z nehomogenního produkčního prostředí, popř. z archivních dat
- Různé možnosti extrakce
 - **Periodická extrakce** – z interních zdrojů
 - **Občasná extrakce** – z externích zdrojů (např. Internet)
 - **První extrakce** – provádí se především z archivních dat

Extrakce - součástí procesu

- **Identifikace zdrojů** (struktury a aplikace)
- Stanovení **metody extrakce** pro každý zdroj
 - manuální – napíše si sám SQL příkazy
 - s využitím nástrojů
- **Frekvence extrakcí** pro každý zdroj – většinou se liší pro různé zdroje
- Stanovení **časového okna** pro extrakci – kdy ji provádět
- **Paralelní x sériová** extrakce pro jednotlivé zdroje dat
- Zpracování **výjimek** při extrakci

Postup při identifikaci zdrojů

- Výpis všech ***datových položek potřebných v tabulce faktů***
- Výpis všech ***dimenzí***
- Pro každou cílovou položku najít zdroj a jeho položku
- Je-li více zdrojů pro jednu cílovou položku, vybrat preferovaný zdroj
- Identifikace vícenásobných zdrojů pro jeden cíl – stanovení ***konsolidačních pravidel***
- Identifikace vícenásobných cílů na jeden zdroj – stanovení ***dělicích pravidel***
- Určení ***implicitních hodnot***
- Zjištění ***chybějících hodnot ve zdrojových datech***

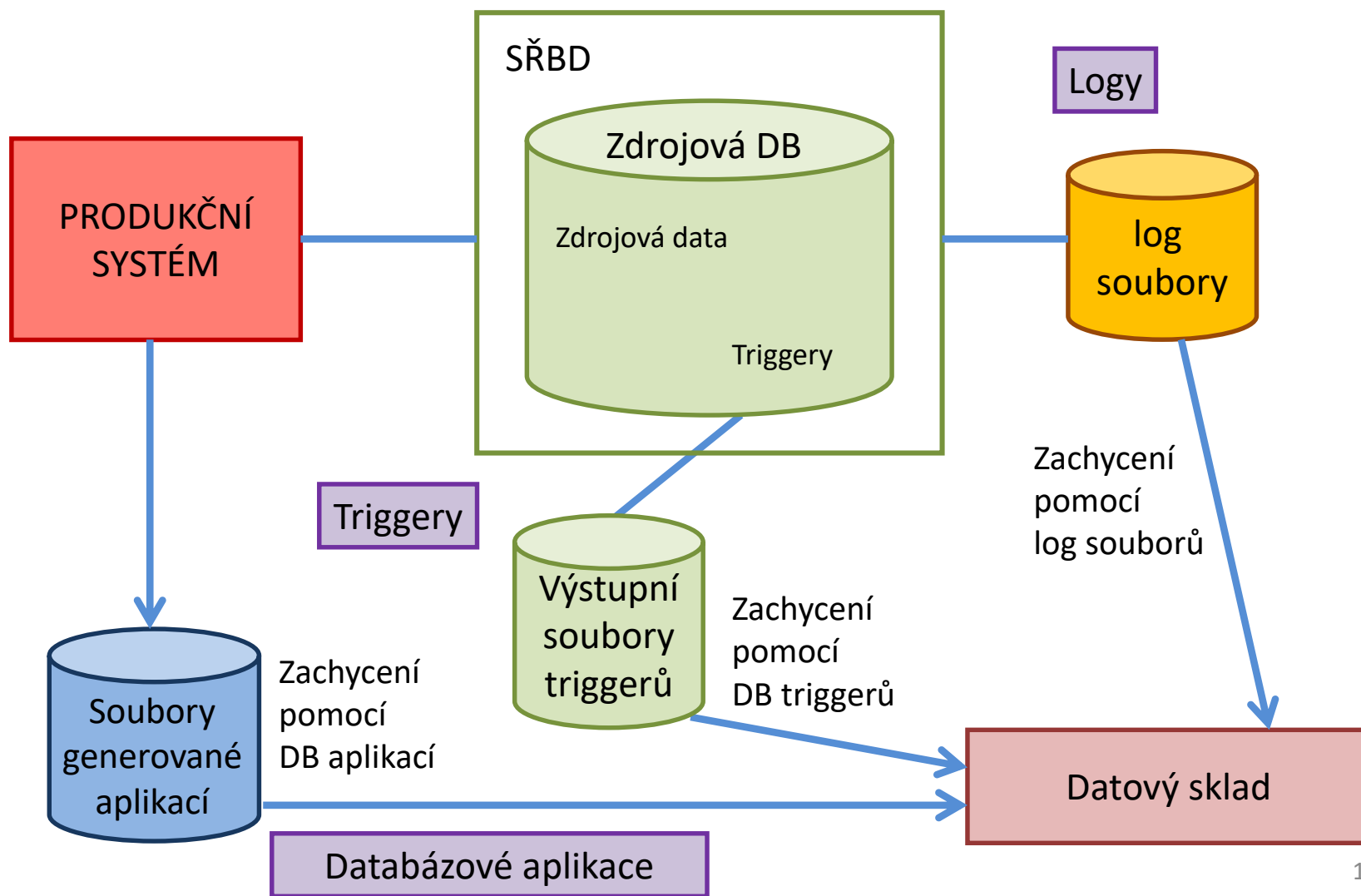
Metody extrakce

- Metoda extrakce statických dat
 - Vytvoření obrazu zdrojové databáze na výstupu
 - Používá se při iniciálním nahrávání dat do skladu
- Metody extrakce při aktualizaci dat
 - Metody *přímé extrakce*
 - Metody *odložené extrakce*

Přímá extrakce

- Liší se způsobem zachycení změn v DB od posledního nahrání
 - Zachycení pomocí **log souborů** (vytvořených databází)
 - Zachycení pomocí databázových **triggerů**
 - Při každé změně se spustí trigger, který zapíše změnu do souboru
 - Zachycení pomocí samotných **databázových aplikací**
 - Editace aplikace tak, aby ukládala záznamy o provedených změnách v DB

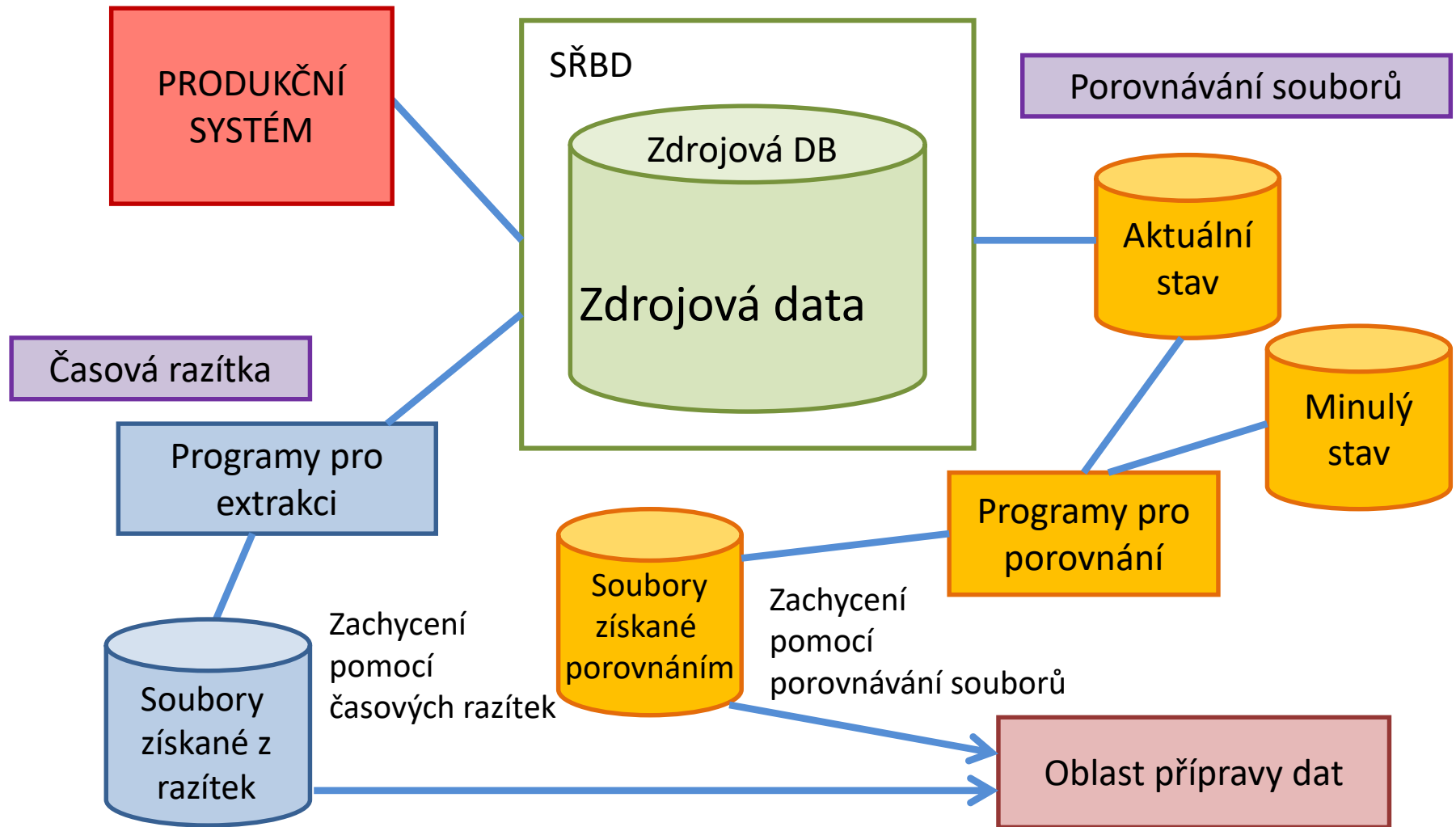
Přímá extrakce



Odložená extrakce

- Nezachycují změny při jejich vzniku, ale až při nahrávání se ***porovnává zdrojová a cílová DB***
 - Zachycení pomocí ***časových razítek***
 - Razítky jsou označeny záznamy, které byly přidány nebo editovány – ty se pak při nahrávání dat naleznou (problém s mazáním)
 - Zachycení pomocí ***porovnávání souborů***
 - Vytvoří se soubor s kopií dat ve stavu současném a včerejším, pak se soubory porovnají (velmi neefektivní)

Odložená extrakce



Transformace

- Cílem je zvýšit kvalitu vstupních dat a zvýšit jejich použitelnost pro cílového uživatele
- Někdy je kvalita vstupních dat velmi proměnlivá -> čištění dat (odstranění nekvalitních dat)
- Často je potřeba odstranit tzv. ***anomálie***, které v klasických databázích běžně vznikají

Anomálie

- Příklady anomálií:
 - Přejít z MS-DOSu na Windows – např. kódování češtiny
 - Lidský faktor – různé překlady, pravopisné chyby
- Potřeba rozdělení složených atributů na atomické

Časté problémy

- ***Nejednoznačnost údajů***
 - Např. různě uložená informace o pohlaví zákazníka (M, muž, Muž atd...)
- ***Chybějící hodnoty***
 - Tyto hodnoty je potřeba doplnit, popř. ignorovat nebo označit nějakým příznakem
- ***Duplicitní hodnoty***
 - Většinou není příliš velký problém je odstranit, někdy je to však časově náročné

Transformace

- ***Konvence názvů pojmů a objektů***
 - je nutné sjednotit terminologii používanou různými zdroji dat
- ***Různé peněžní měny***
 - problém vznikne např. při přechodu z CZK na Euro
- ***Formáty čísel a textových řetězců***
 - použití různých datových typů pro ukládání čísel (např. řetězce)

Transformace

- ***Referenční integrita***
 - Neustálé změny v reálném světě zkreslují data – např. i po zrušení oddělení firmy zůstanou v DB údaje o jeho zaměstnancích
- ***Chybějící datum***
 - Časový aspekt je v datových skladech velmi důležitý, ve vstupních datech však čas často chybí, často je nutné jej doplnit

Loading

- Přesun údajů a jejich uložení do tabulek datového skladu
- Pokud možno by měl probíhat automatizovaně
- Rozlišujeme podle periody přesunů...
 - to závisí především na požadavcích aplikace
 - většinou jde o časově náročnou operaci, především u iniciálního přenosu

3 typy loadingu (nahrávání)

- ***Iniciální nahrávání***
 - Nahrávání všech dat do prázdného skladu
- ***Inkrementální nahrávání***
 - promítnutí změn v DB do datového skladu (provádí se periodicky)
- ***Přepis dat***
 - kompletní smazání obsahu skladu a nahrání aktuálních dat

Módy nahrávání dat

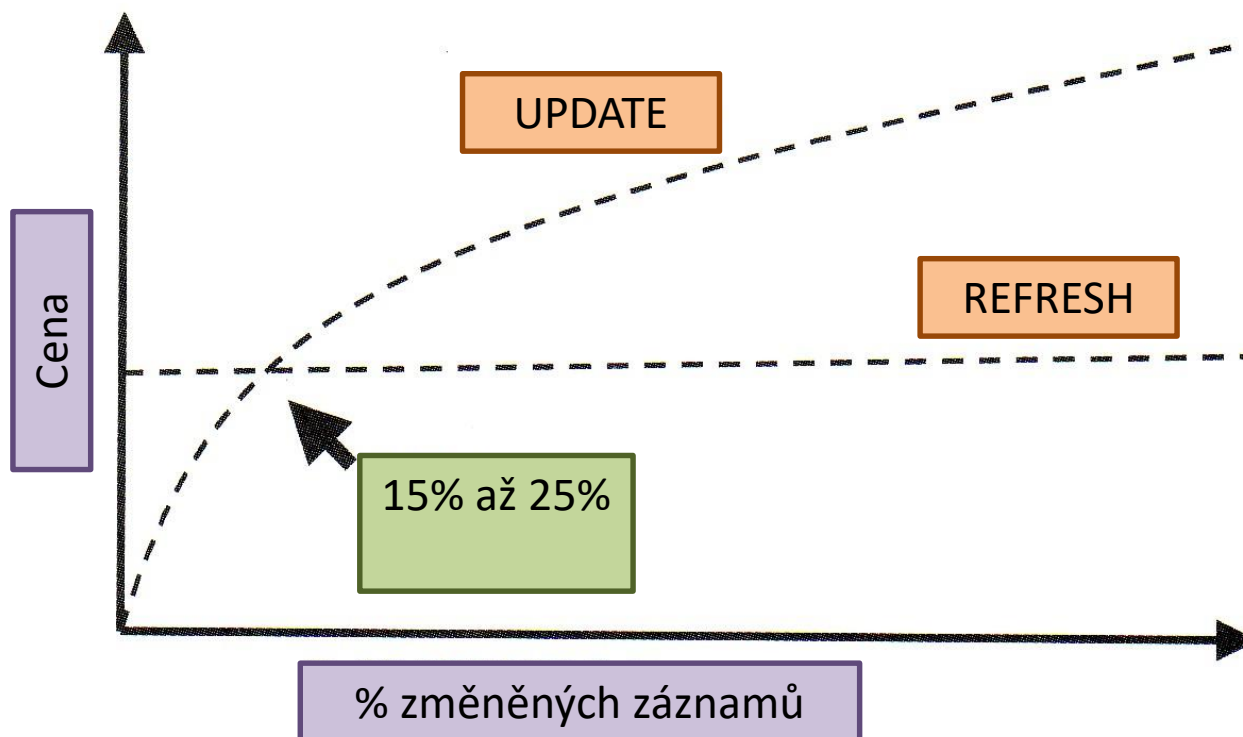
- ***Nahrání (Load)***
 - Pokud cílová tabulka obsahuje data, pak jsou smazána a nahrazena aktuálními
- ***Přidání (Append)***
 - Přidání nových dat ke stávajícím, při duplicitě může uživatel zvolit další postup
- ***Destruktivní sloučení***
 - Stejně jako přidání, při stejných klíčích se přepíše hodnota daného řádku
- ***Konstruktivní sloučení***
 - Při stejných klíčích se přidá nový prvek a označí se jako nový, starý v datovém skladu zůstane

Refresh x Update

Po počátečním naplnění jsou data v datovém skladu udržována v aktuálním stavu pomocí:

REFRESH - kompletní natažení ve specifikovaném intervalu

UPDATE - aplikace inkrementálních změn



Architektury serveru OLAP

- ***Multidimenzionální OLAP (MOLAP)***
 - Multidimenzionální paměťový stroj založený na polích (array) - techniky řídkých matic
 - Rychlé indexování předzpracovaných agregovaných dat
- ***Relační OLAP (ROLAP)***
 - Užívá relační nebo rozšířená relační SŘBD
 - Zahrnují optimalizaci back-endu SŘBD, implementaci agregační navigační logiky a dodatečných pomůcek a služeb
 - Velká možnost škálování
- ***Hybridní OLAP (HOLAP)***
 - Uživatelská flexibilita kombinující obě předchozí metody
- ***Specializovaný SQL server***
 - specializovaná podpora pro SQL dotazy nad schémata hvězda/sněhová vločka

Multidimezionální OLAP (MOLAP)

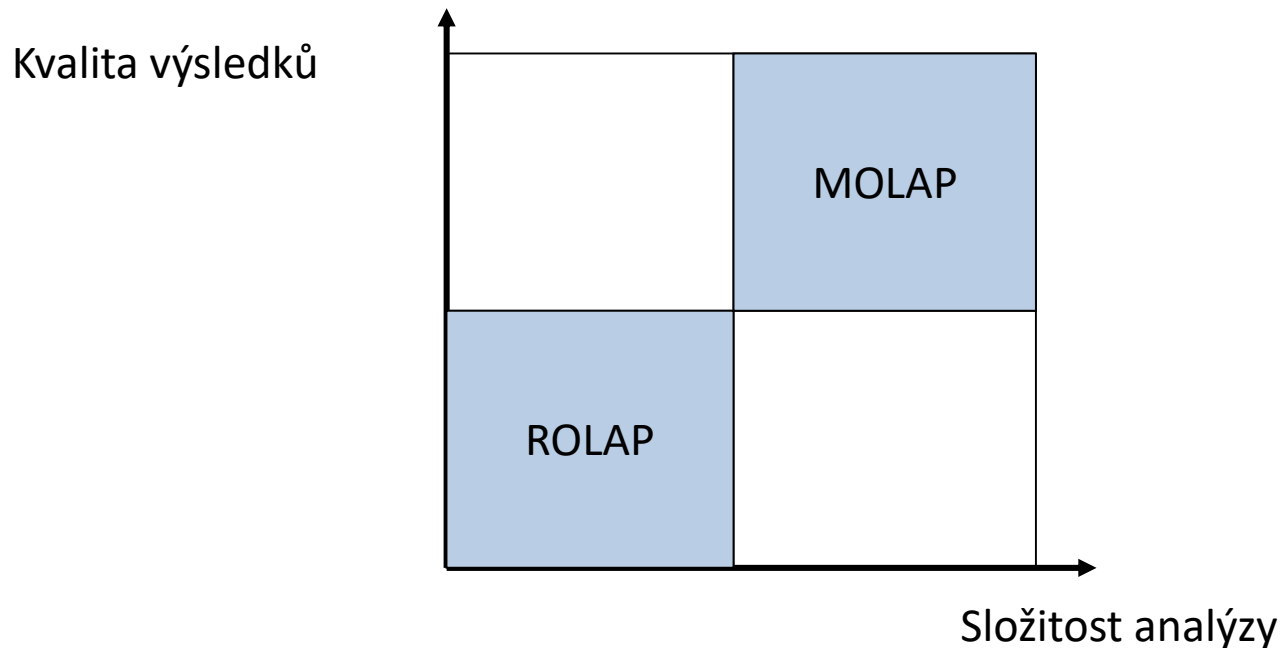
- Data se ukládají do vlastních datových struktur
- Databáze konstruována pro rychlé vyhledávání údajů
- **Výhoda:** maximální výkon
- **Nevýhoda:** redundance údajů, velké prostorové nároky

Relační databázový OLAP (ROLAP)

- Údaje jsou získávány z ***relačních tabulek***, jsou uživateli předkládány jako multidimenzionální pohled
- Data jsou uložena jako záznamy relační tabulky
- Žádná redundance

ROLAP x MOLAP

- Záleží na důležitosti analýzy
- Záleží na složitosti dotazů uživatelů



ROLAP x MOLAP - uložení dat

- **ROLAP**

- Data **uložena** v relačních tabulkách
- Možnost získat detailní i agregovaná data
- Velké datové prostory
- Veškerý přístup k datům prostřednictvím datového skladu

- **MOLAP**

- Data **neuložena** v relačních tabulkách
- Různá agregovaná data uložena v multidimenz. databázi
- Průměrně velké datové prostory
- Přístup jak do MDDB (sumy), tak do datového skladu (detailní data)

ROLAP x MOLAP - technologie

- ***ROLAP***

- Použití komplexních SQL dotazů k získání dat ze skladu
- Datové kostky jsou vytvářeny za běhu
- Multidimenzionální pohledy vytváří prezentační vrstva

- ***MOLAP***

- Vytváření datových kostek předem
- Použití technologie pro ukládání multidim. dat v polích, ne tabulkách
- Technologie pro zpracování řídkých matic

ROLAP x MOLAP - funkce a vlastnosti

- ***ROLAP***

- Známé prostředí a možnost použití známých nástrojů
- Limitované použití komplexních analýz
- Omezené použití operace drill-across

- ***MOLAP***

- Rychlejší přístup
- Velká knihovna funkcí pro komplexní analýzu
- Snadná analýza bez ohledu na počet dimenzí
- Rozšiřující prostor pro operace drill-down a slice-and-dice

Hybridní OLAP (HOLAP)

- Kombinace MOLAP a ROLAP
- Data ***v relačních tabulkách***, agregace se ukládají do multidimenzionálních struktur
- Využití multidimezionální cache

KONCEPTUÁLNÍ NÁVRH DATOVÉHO SKLADU

Fakta (míry) a dimenze

- Navrhujeme v relačním modelu - tabulky
- Každá datová kostka obsahuje 2 typy údajů – ***fakta (míry) a dimenze***
- ***Fakta***
 - Největší relace v DB, zpravidla jen jedna
 - Obsahuje numerické měrné jednotky obchodování
 - V kombinaci s relacemi dimenzí tvoří základní schémata

Fakta a dimenze

- ***Dimenze (číselníky)***
 - logicky nebo hierarchicky uspořádané údaje
 - textové popisy obchodování
 - jsou menší a nemění se tak často
 - nejčastěji: časové, geografické a produktové dimenze (stromové struktury)

Kontinent
Země
Územní celek
Město

Druh produktu
Kategorie
Subkategorie
Název produktu

Rok
Kvartál
Měsíc
Týden

Konceptuální modelování kostky

- **Schéma hvězdy** - *jedna tabulka faktů* je ve středu připojena k množině relací dimenzí
- **Schéma sněhové vločky** - zjemnění schématu hvězdy, kde existuje *hierarchie dimenzí* normalizovaná do množiny navázaných relací dimenzí
- **Konstelace faktů** - více relací faktů, které sdílejí relace dimenzí - je možné chápat jako kolekci hvězd, proto se také někdy nazývá *schéma galaxie*

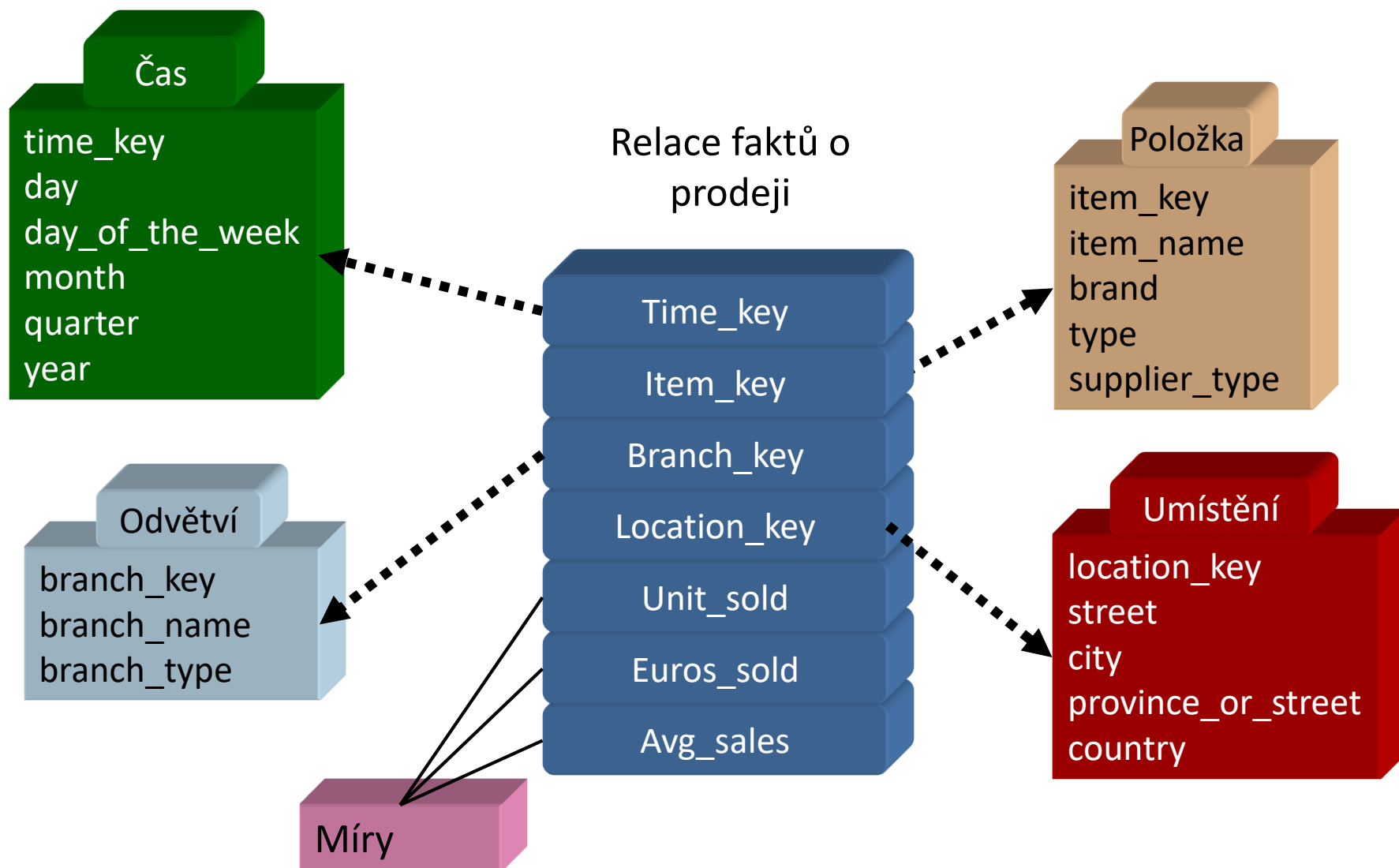
Schéma hvězdy - star

- Každá relace dimenze sestává z množin, které odpovídají hodnotám dimenze.
- Hvězdicové schéma ***neposkytuje explicitně podporu pro hierarchii atributů.***
- Lze obejít organizačně

Schéma hvězdy - star

- Relace faktů obsahuje cizí klíče do relací dimenzí, ty se vztahují k jejím primárním klíčům
- Snadno pochopitelné
- Relace dimenzí jsou však nejsou normalizované, je to tedy poměrně pomalé

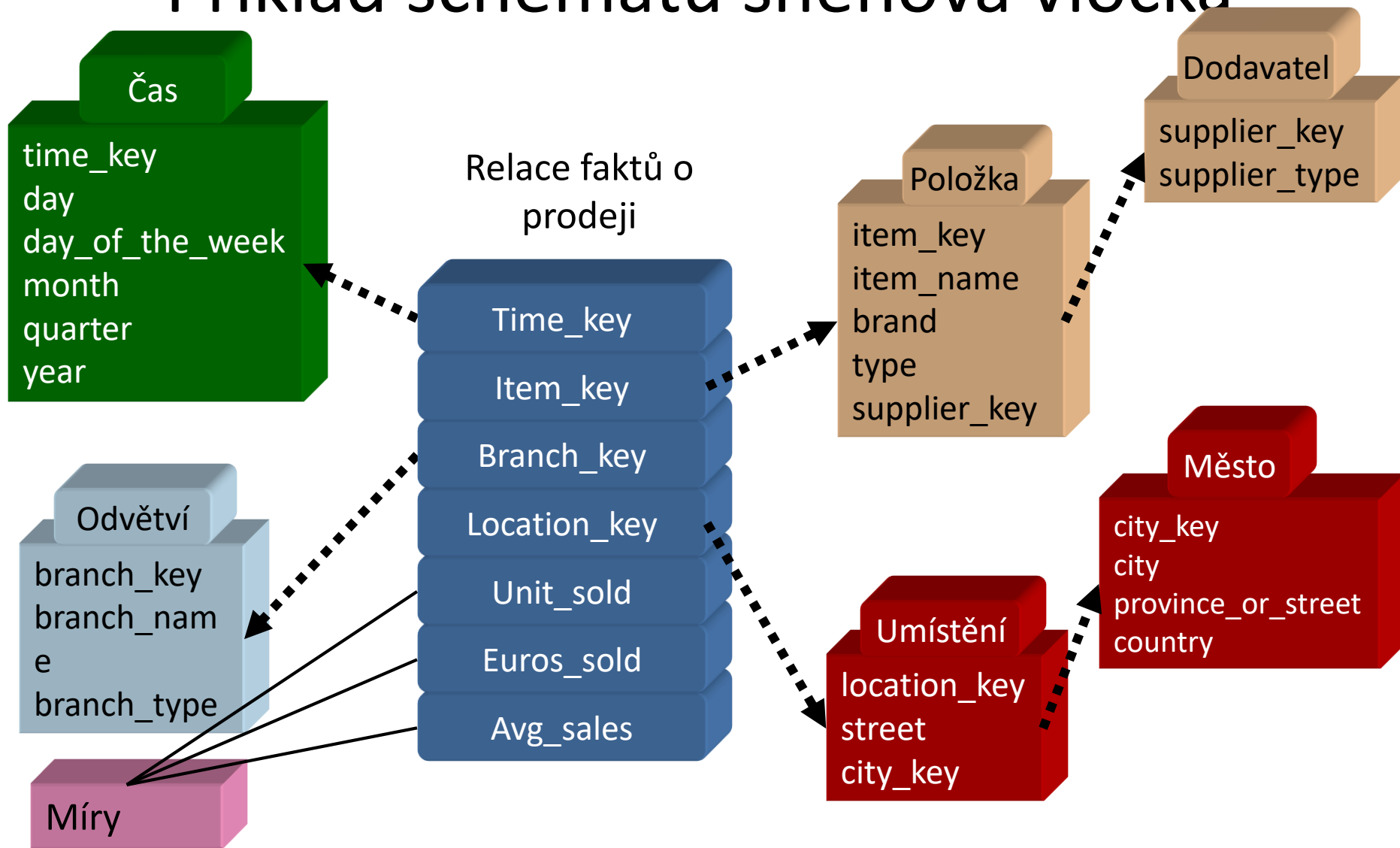
Příklad schématu hvězda - star



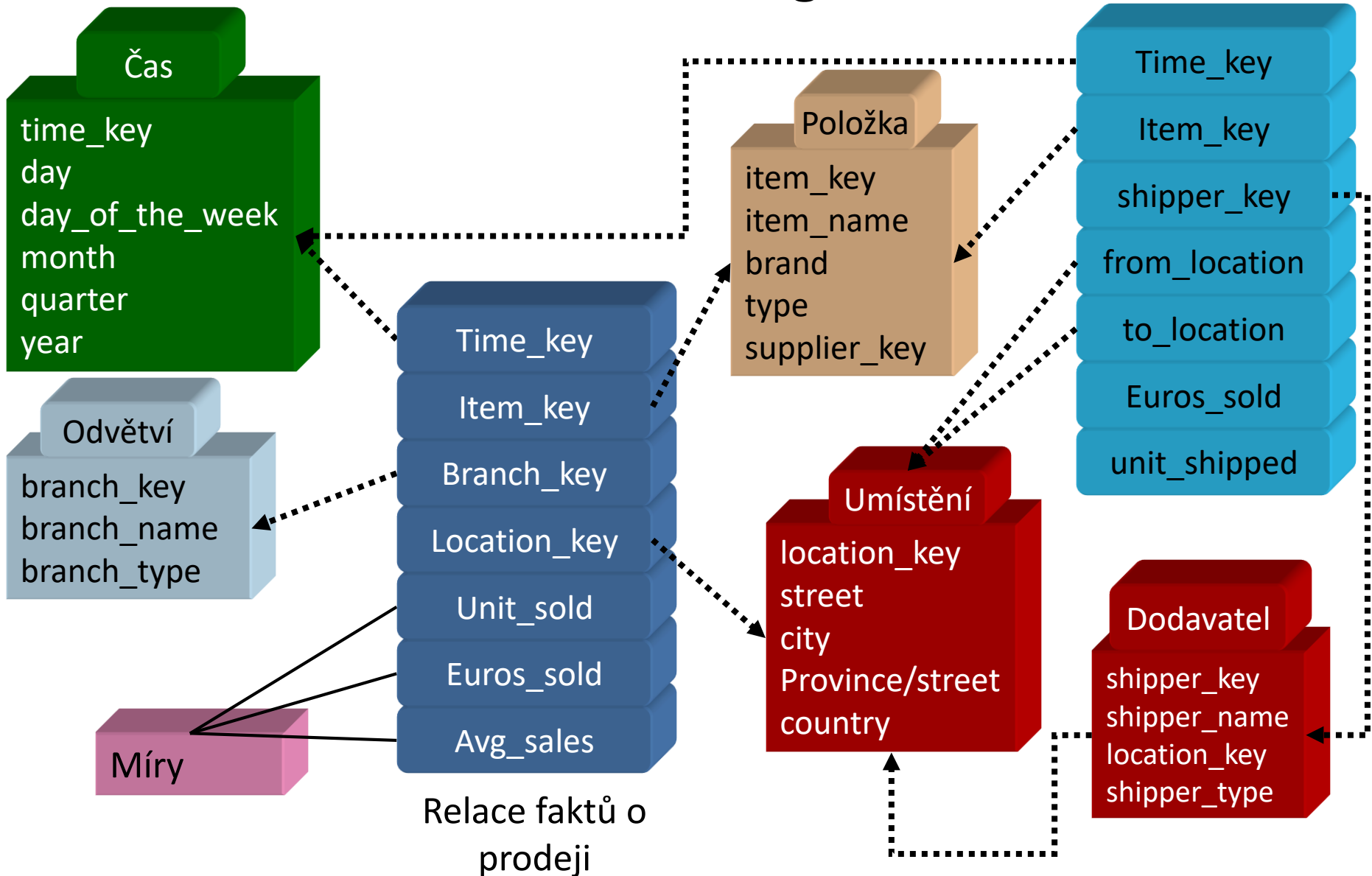
Sněhová vločka - snowflake

- ***Schéma sněhové vločky*** poskytuje zjemnění hvězdicového schématu tak, že ***hierarchie dimenzí*** je explicitně reprezentována normalizováním relací dimenzí. To vede k výhodné údržbě relací dimenzí.
- Nicméně nenormalizovaná struktura dimenzionální reze hvězdicovém schématu může být mnohem vhodnější pro procházení dimenzemi.

Příklad schématu sněhová vločka



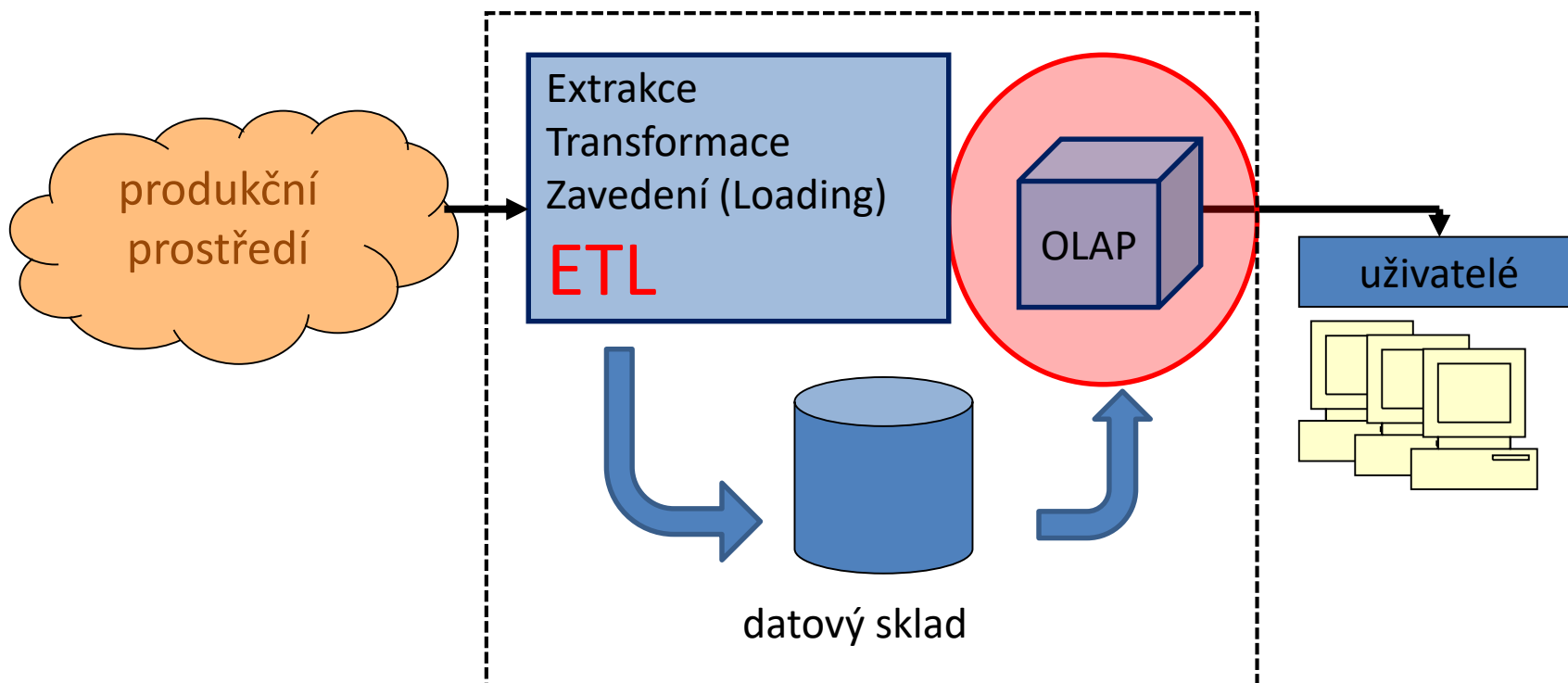
Příklad konstelace faktů - galaxie



OLAP

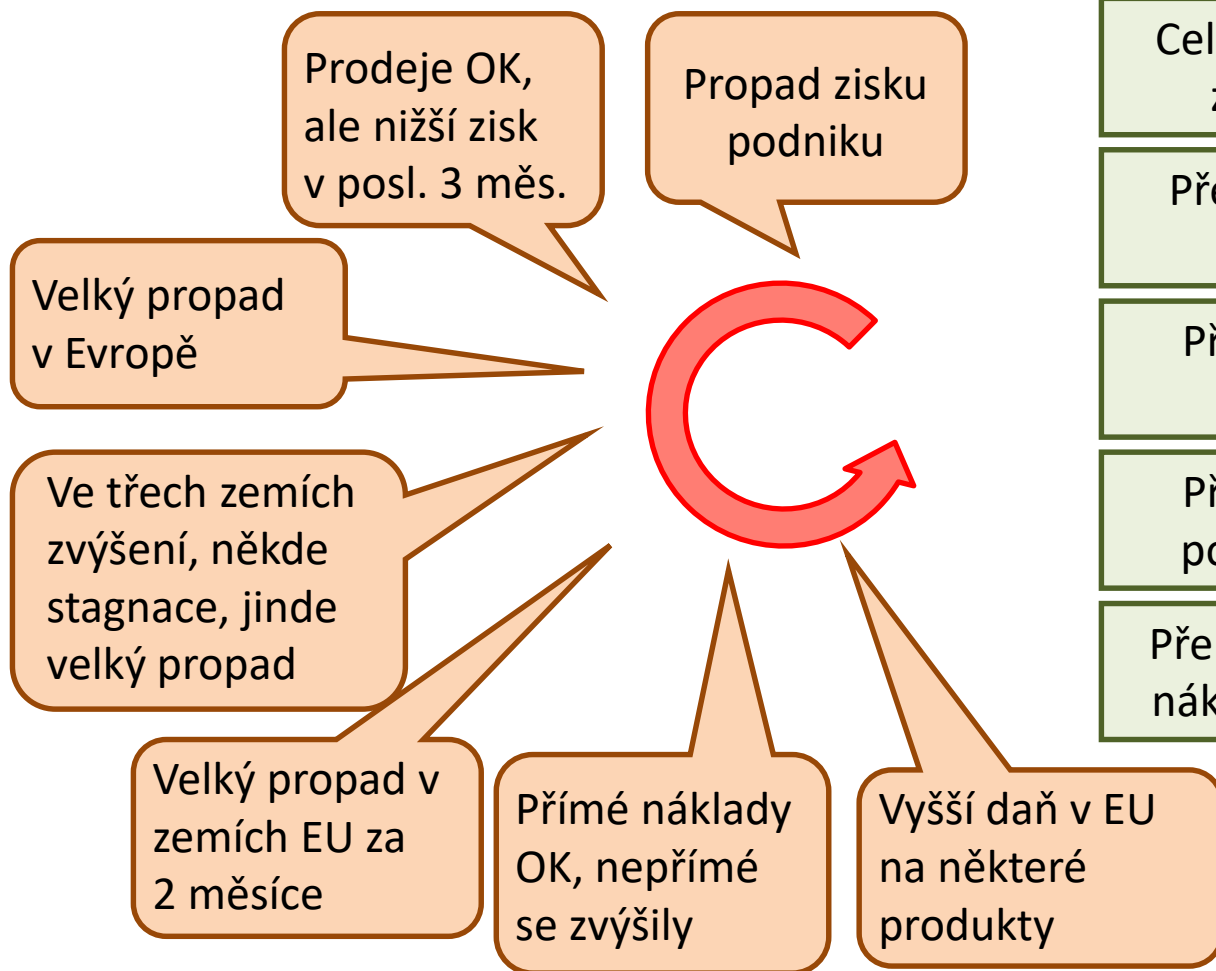
Analýza OLAP

- Analytické zpracování údajů v datovém skladu podle uživatelských dotazů



Příklad komplexní analýzy

Myšlenkový pochod při analýze



Sekvence dotazů při této analýze

Celosvětové měsíční prodeje za posledních 5 měsíců

Přehled měsíčních prodejů po regionech

Přehled prodejů v Evropě po zemích

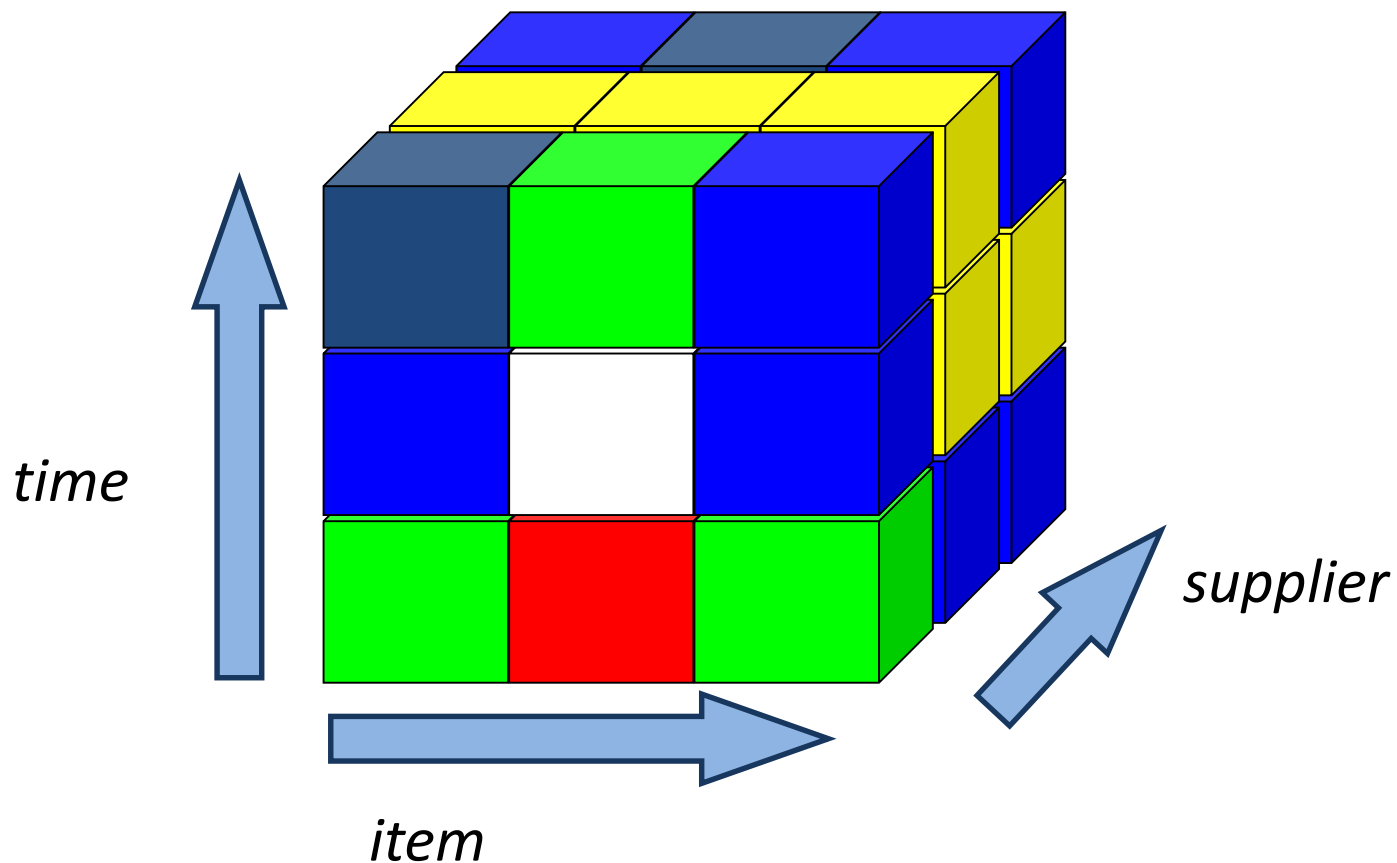
Přehled prodejů v Evropě po zemích, po produktech

Přehled přímých a nepřímých nákladů v evropských zemích

Požadavky na OLAP systémy (příklady)

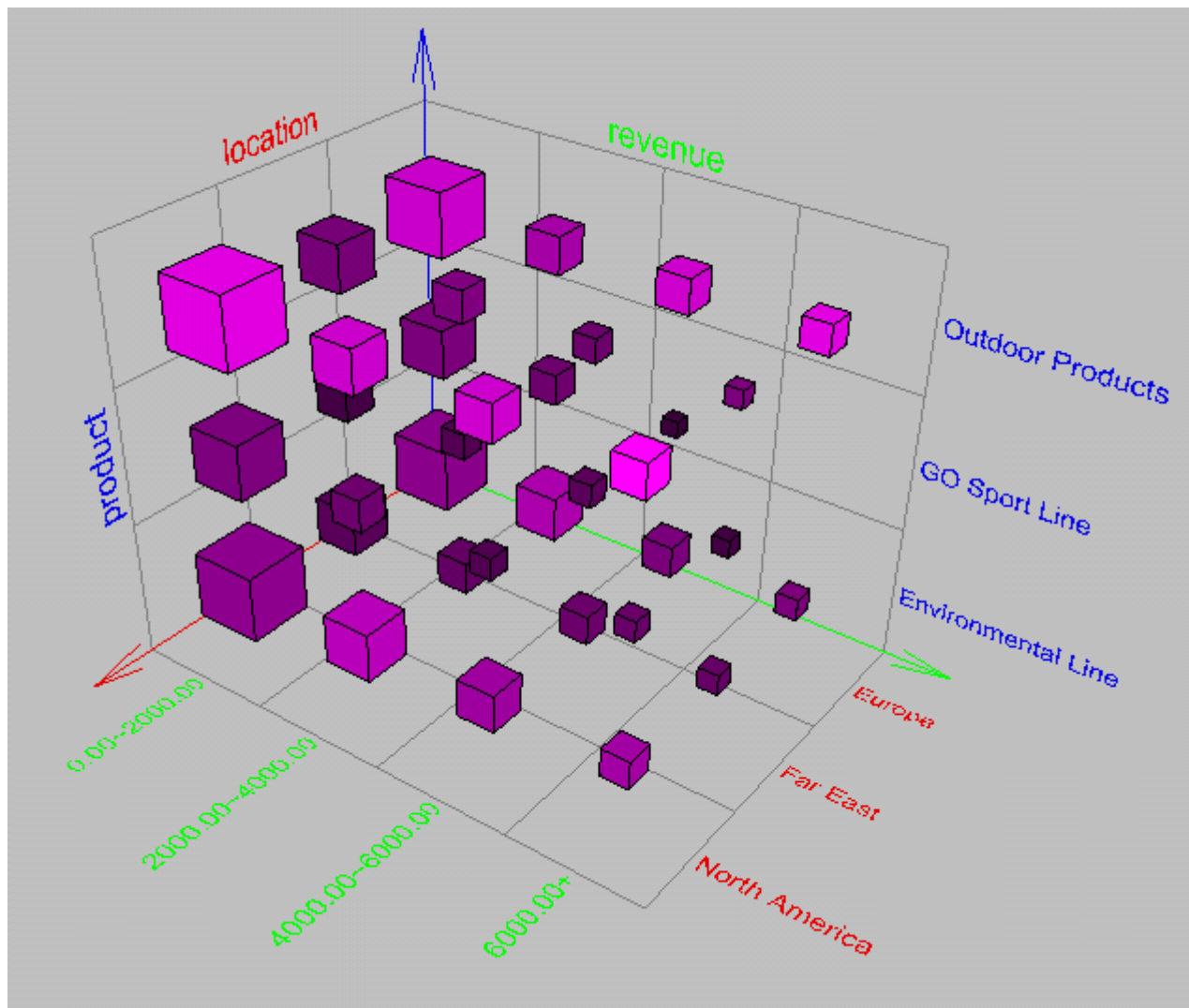
- Poskytování **agregačních** funkcí podle hierarchií
- Možnost **detailního pohledu - zooming** na data
- Jednoduché kalkulace, např. výpočet zisku (prodeje – náklady)
- Sdílení kalkulací za účelem procentuálního vyjádření vzhledem k celku
- Algebraické rovnice pro výpočet klíčových indikátorů
- Přenos průměrů a procentuálních vyjádření
- Analýza trendů statistickými metodami

Datová kostka 3D příklad



OPERACE NAD DATOVOU KOSTKOU

Prohlížení datové kostky

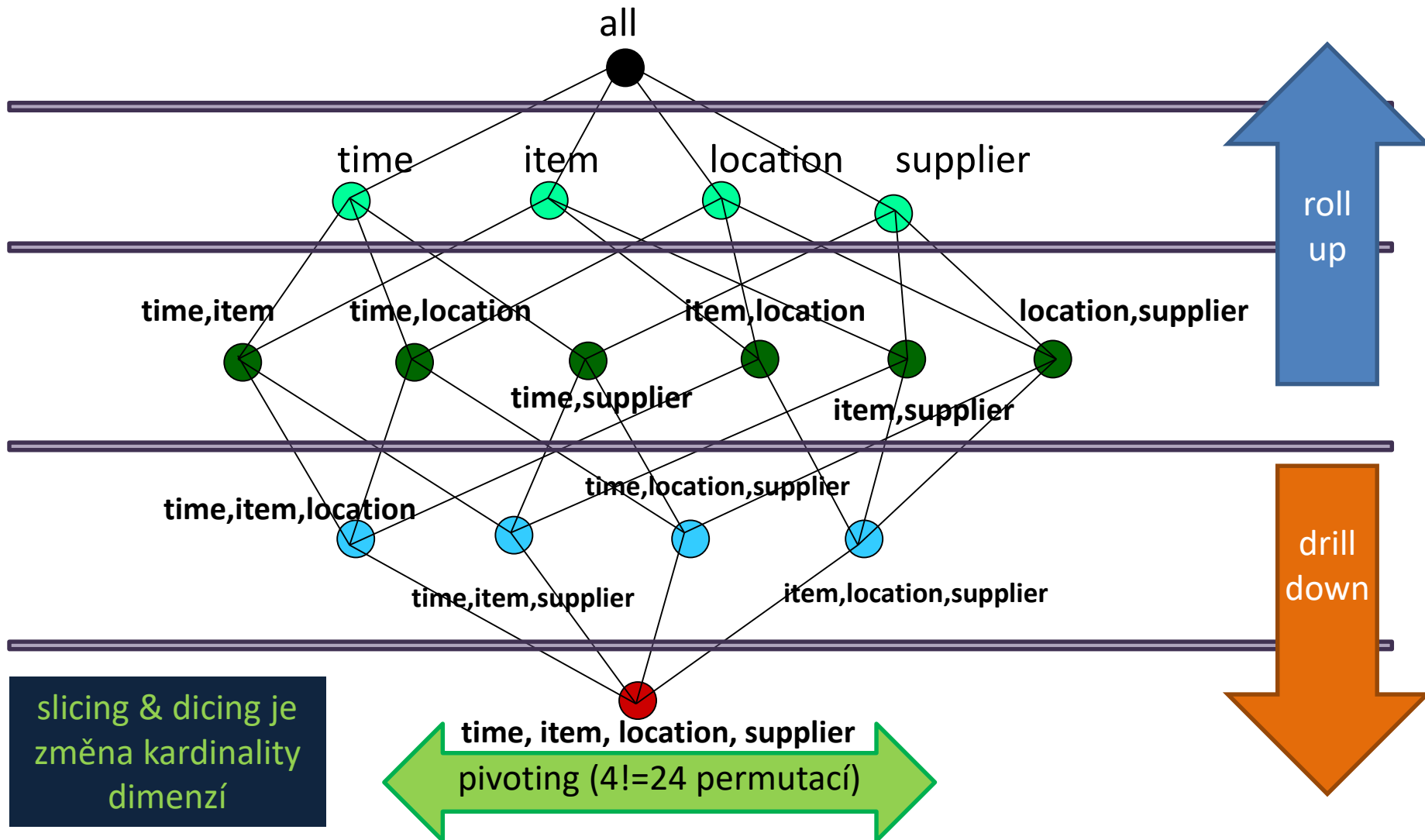


- vizualizace
- operace
- interaktivní manipulace

Operace nad datovou kostkou

- ***roll-up*** – (vyrolování) vzrůst úrovně agregace
- ***drill-down*** (zavrtání) – snížení úrovně agregace a zvýšení detailu
- ***pivoting*** (přetočení) – změna relace R pro uspořádání dimenzí
- ***slicing & dicing*** (seřiznutí) – výběr projekce

Operace nad 4D datovou kostkou



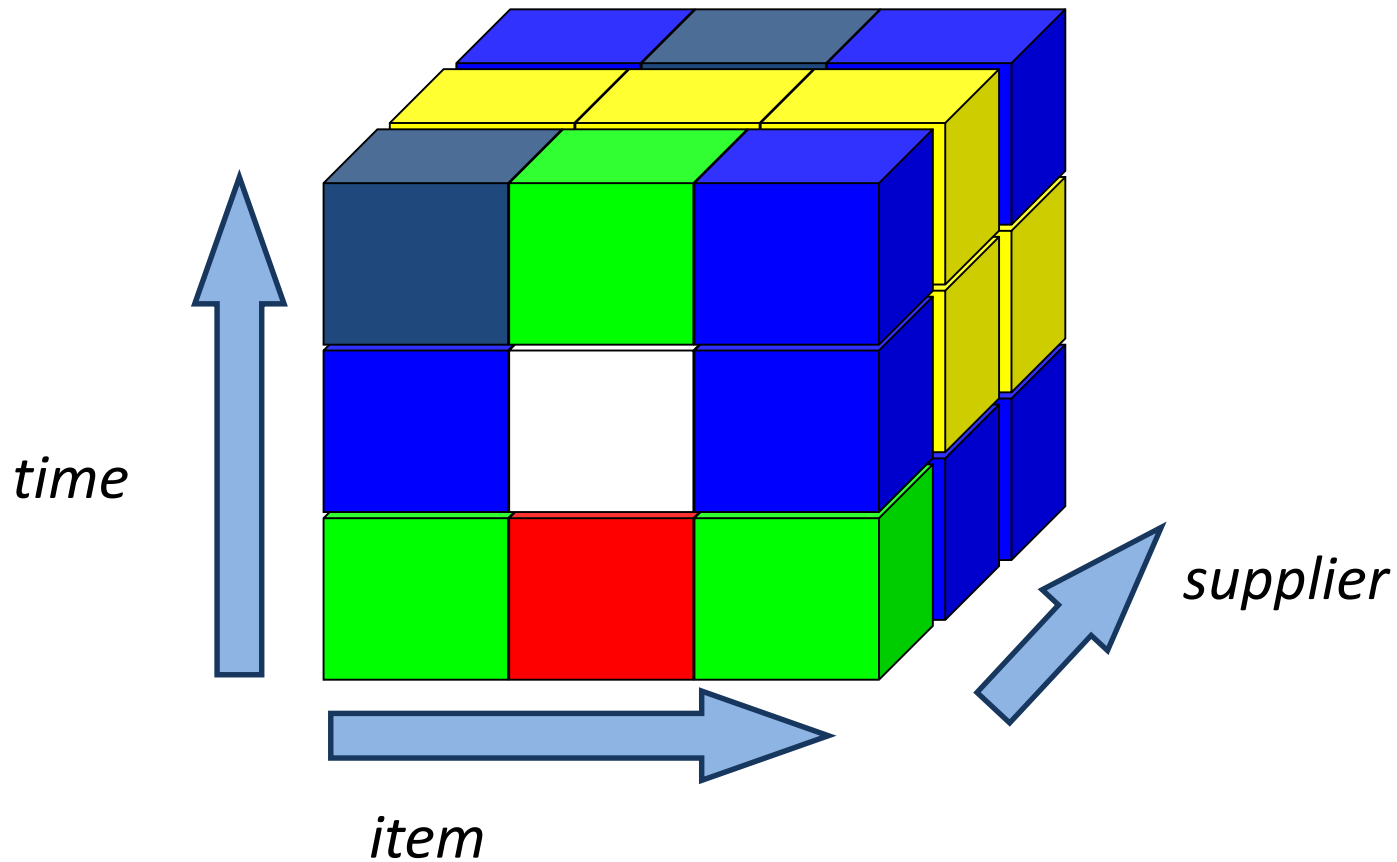
Roll-up

- ***Posun o jednu úroveň výše v uspořádání kuboidů***
- ***Vstup:*** uspořádaná množina m aktivních dimenzí $\{A_1, A_2, A_3, \dots, A_i, \dots, A_{m-1}, A_m\}$, kde $m \geq 1$
- ***Výstup:*** uspořádaná množina $m-1$ aktivních dimenzí $\{A_1, A_2, A_3, \dots, A_m\}$, tj. A_i bylo deaktivováno.
- Nejčastěji se deaktivuje nejmenší dimenze A_m , tj. z $\{A_1, A_2, A_3, \dots, A_m\}$ vznikne $\{A_1, A_2, A_3, \dots, A_{m-1}\}$

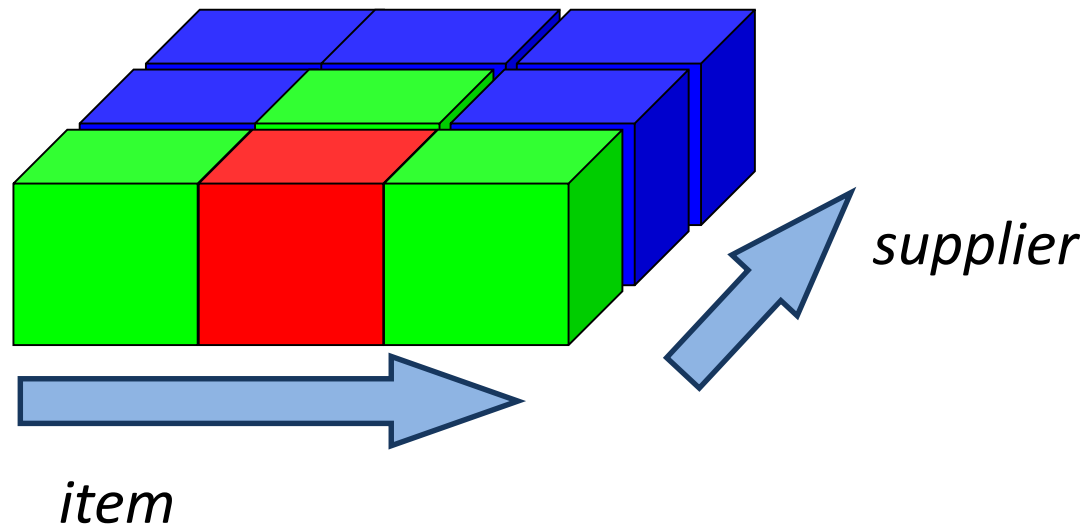
Drill-down

- **Posun o jednu úroveň níže v uspořádání kuboidů**
- **Vstup:** uspořádaná množina m viditelných dimenzí $\{A_1, A_2, A_3, \dots, A_m\}$, kde $m \leq n$
- **Výstup:** uspořádaná množina $m+1$ aktivních dimenzí $\{A_1, A_2, A_3, \dots, A_i, \dots, A_{m-1}, A_m\}$, kde A_i je vybrána z neaktivních D_i tak, aby platila definice aktivních dimenzí
- Nejčastější variantou je přidání nejmenší neaktivní dimenze na konec, tedy vznikne $\{A_1, A_2, A_3, \dots, A_m, A_{m+1}\}$
- Pro **$m=n$** bude výsledkem **detail** všech hodnot
- Je možné provést i **zavrtání do základního kuboidu**, neboť i jeho fakta jsou agregací detailních hodnot pro všechny aktivní dimenze

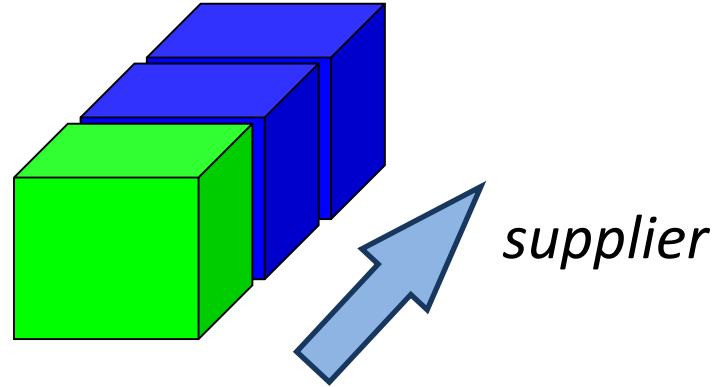
Před operací roll-up



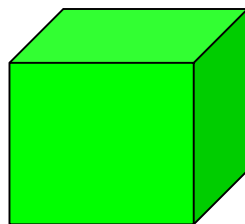
Po operaci roll-up (bez dimenze *time*)



Po operaci roll-up (bez dimenze *item*)

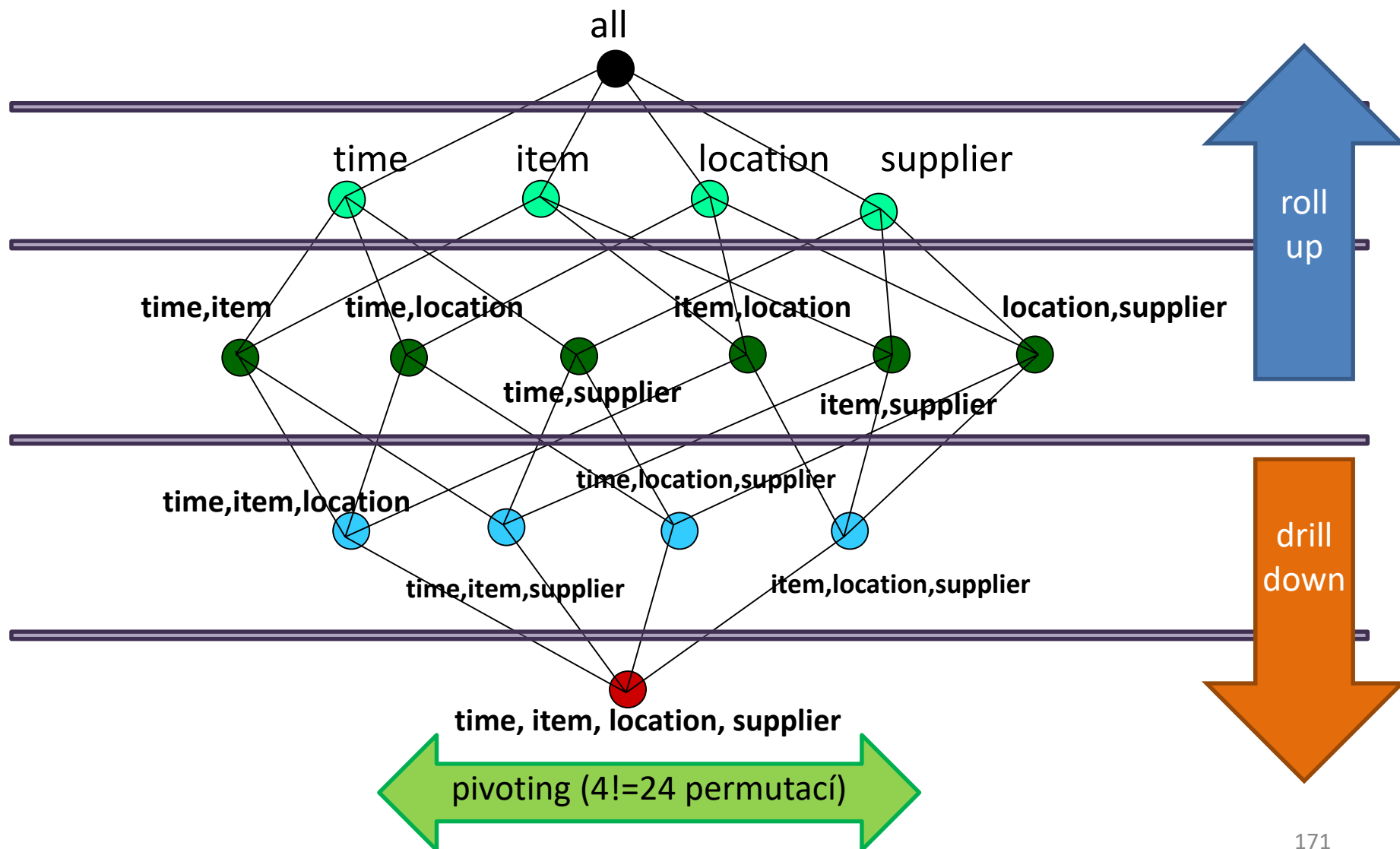


Po operaci roll-up (bez dimenzí)

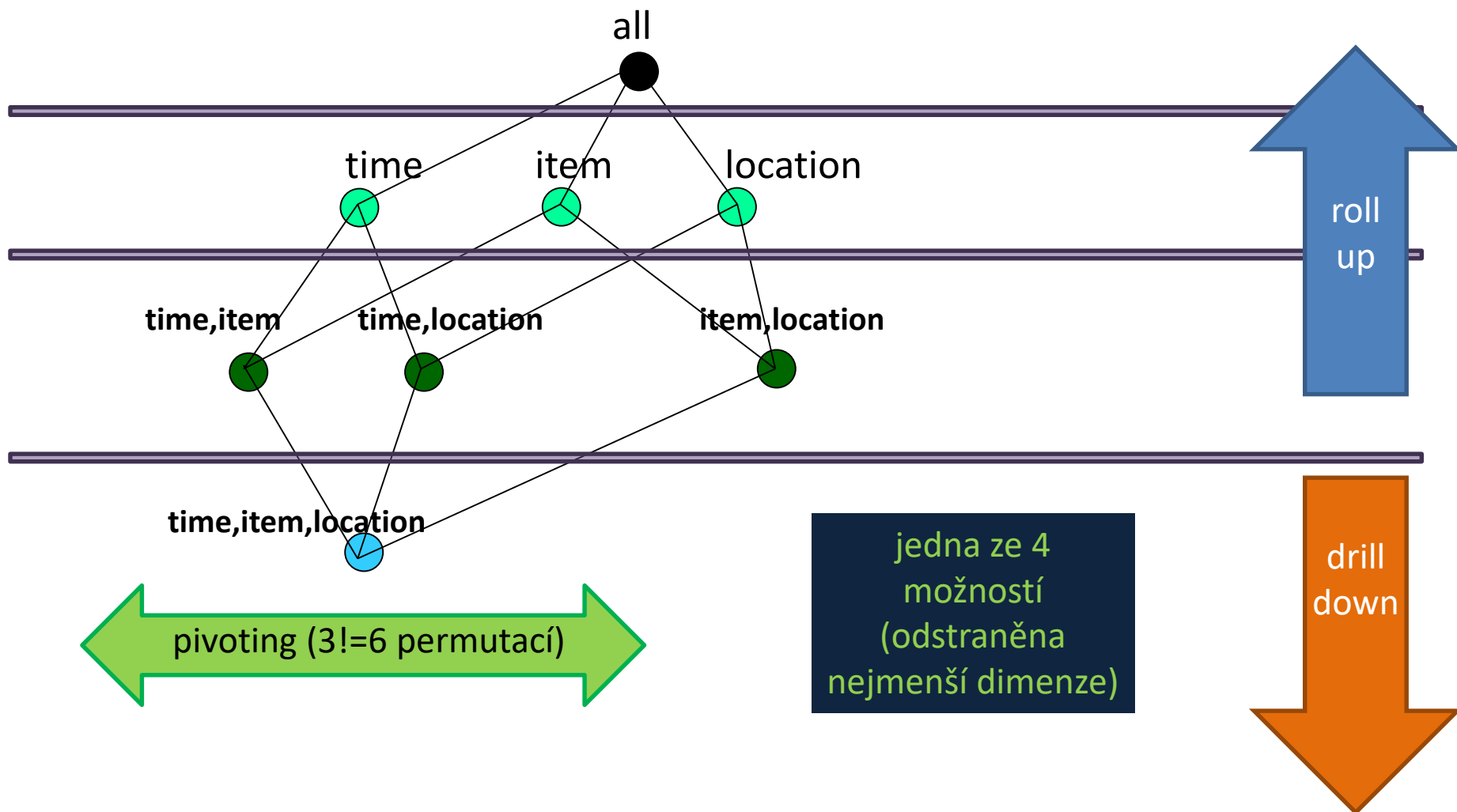


pokud existuje definovaná hodnota agregační funkce,
musí existovat i kostička představující agregovaný fakt

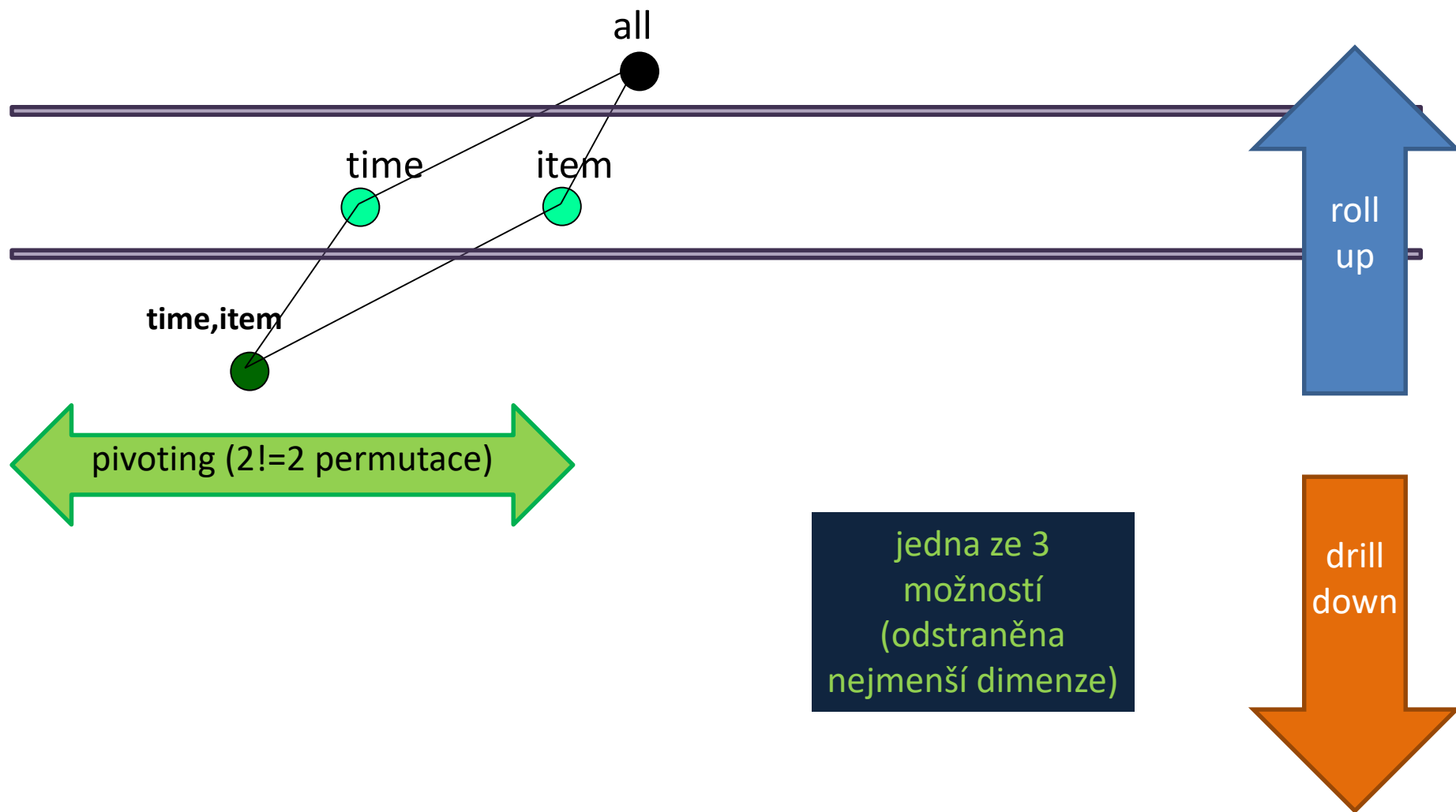
Roll-up nad 4D datovou kostkou



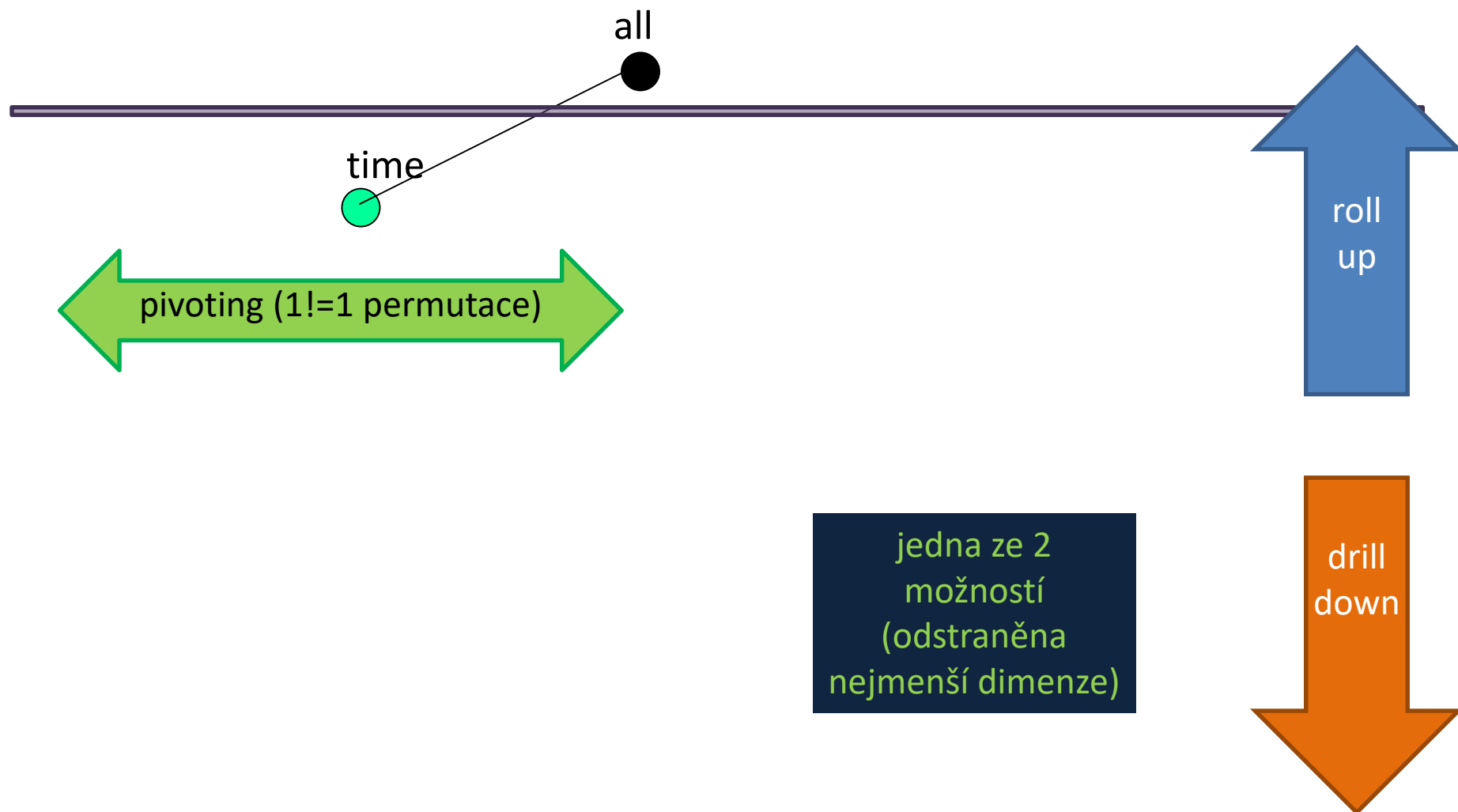
Vyrolovaná datová kostka 3D



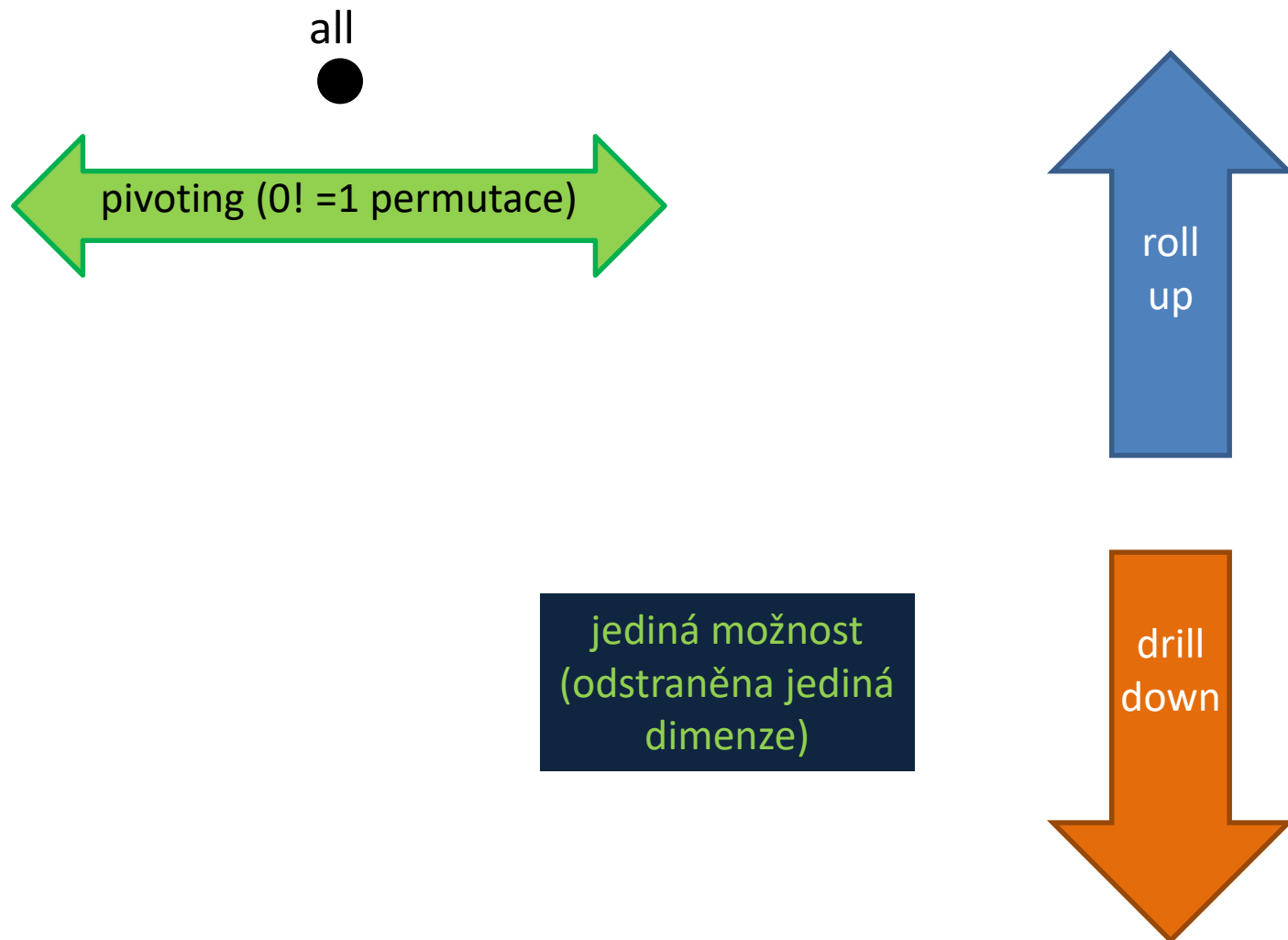
Vyrolovaná datová kostka 2D



Vyrolovaná datová kostka 1D



Vyrolovaná datová kostka 0D



Příklad použití drill-down

LINE	TOTAL SALES
Clothing	\$12,836,450
Electronics	\$16,068,300
Video	\$21,262,190
Kitchen	\$17,704,400
Appliances	\$19,600,800
Total	\$87,472,140

1

aktivní
dimenze je
produkt

2

zaktivizována
dimenze čas

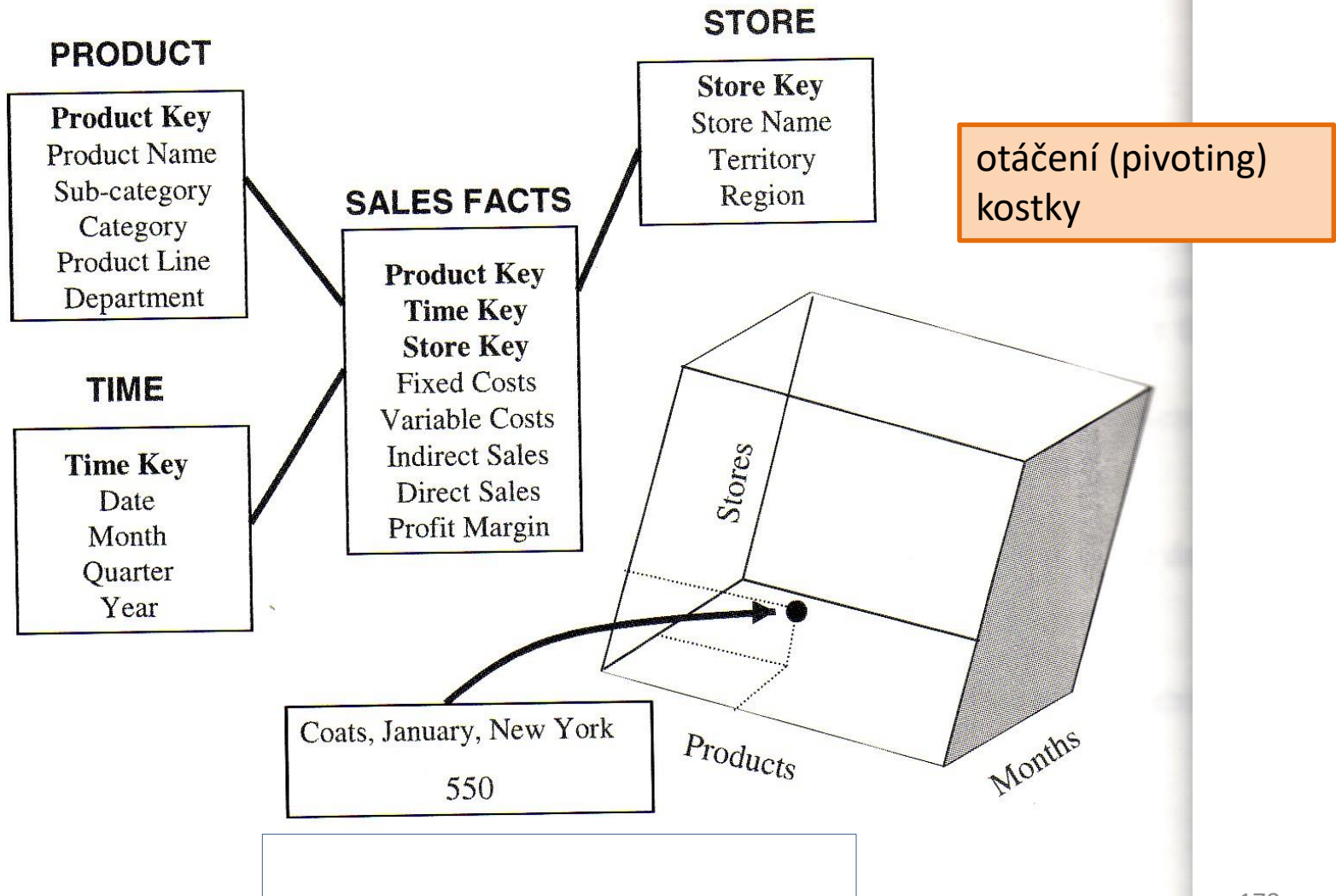
LINE	1998	1999	2000	TOTAL
Clothing	\$3,457,000	\$3,590,050	\$5,789,400	\$12,836,450
Electronics	\$5,894,800	\$4,078,900	\$6,094,600	\$16,068,300
Video	\$7,198,700	\$6,057,890	\$8,005,600	\$21,262,190
Kitchen	\$4,875,400	\$5,894,500	\$6,934,500	\$17,704,400
Appliances	\$5,947,300	\$6,104,500	\$7,549,000	\$19,600,800
Total	\$27,373,200	\$25,725,840	\$34,373,100	\$87,472,140

mírou je
objem
prodejů

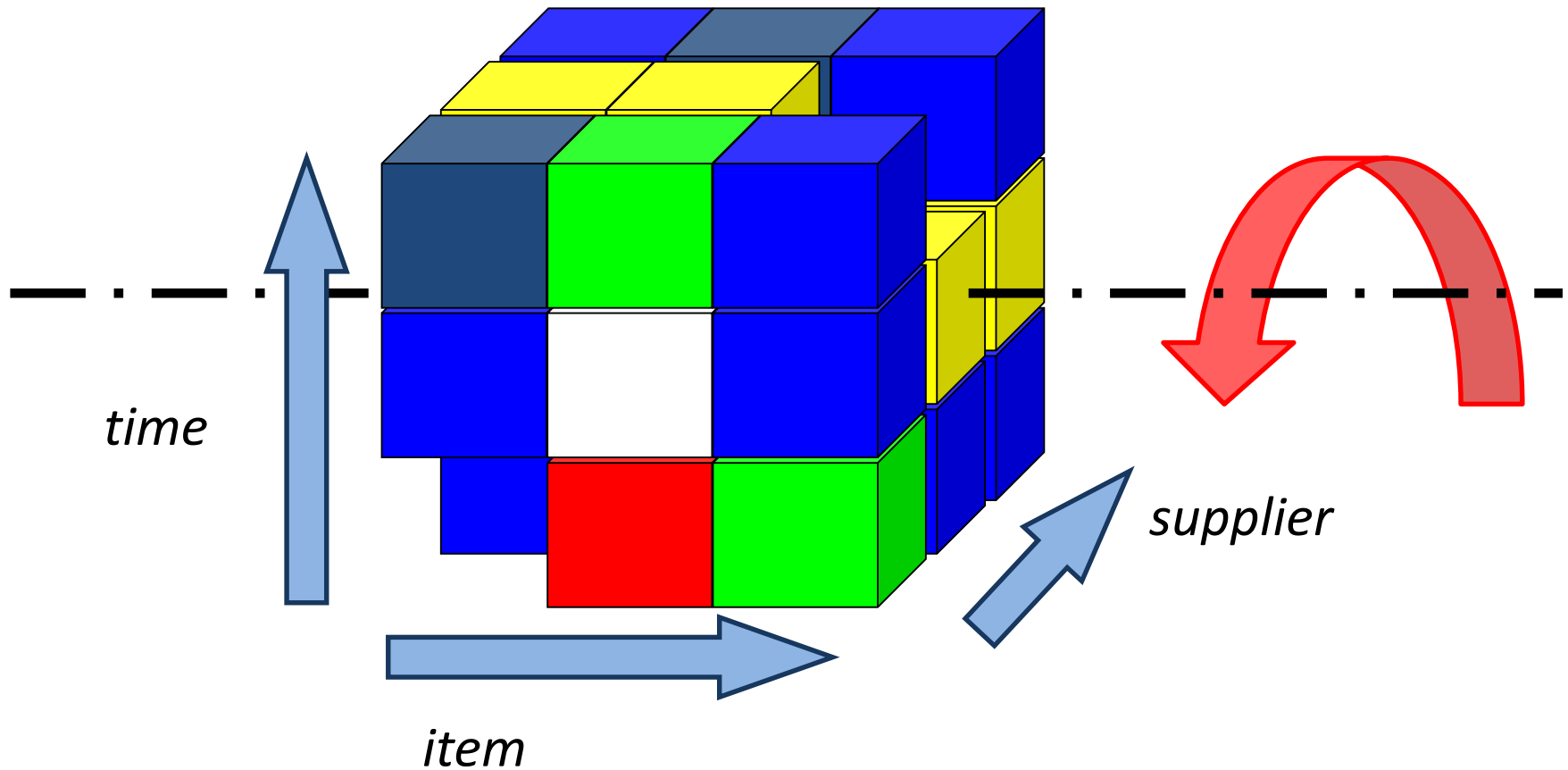
Pivoting

- ***Změna uspořádání R nad stejnou množinou dimenzí***
- ***Vstup:*** uspořádaná množina dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$, kde uspořádání je jedna z možných relací R , kterých je $n!$
- ***Výstup:*** uspořádaná množina dimenzí $\{D_{x1}, D_{x2}, D_{x3}, \dots, D_{xn}\}$ a obecně jiná relace uspořádání R (jedna z možných $n!$)
- Jde o ***otočení jedné ze stěn kostky k sobě***, proto ***pivoting***.

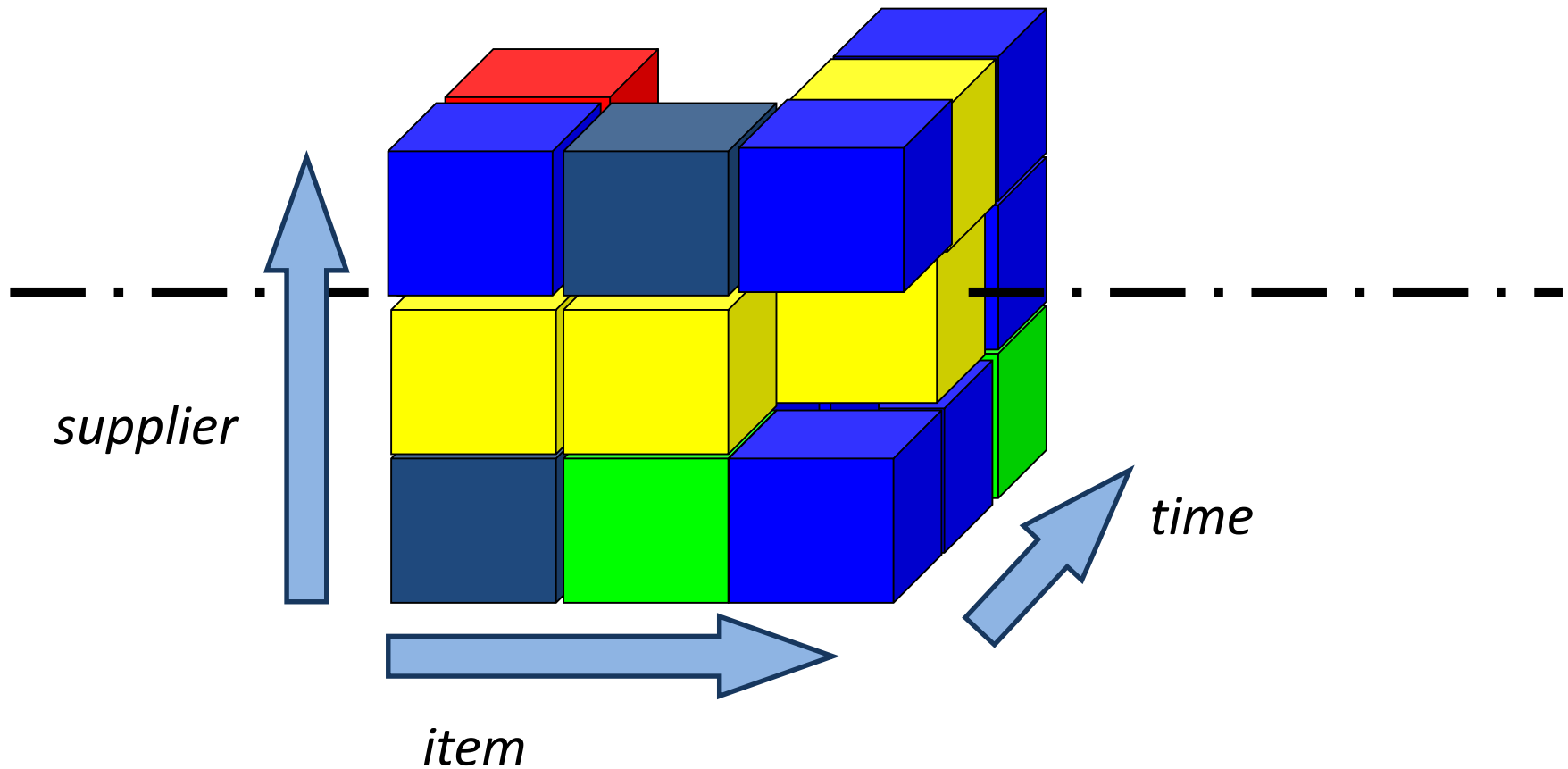
Příklad otáčení



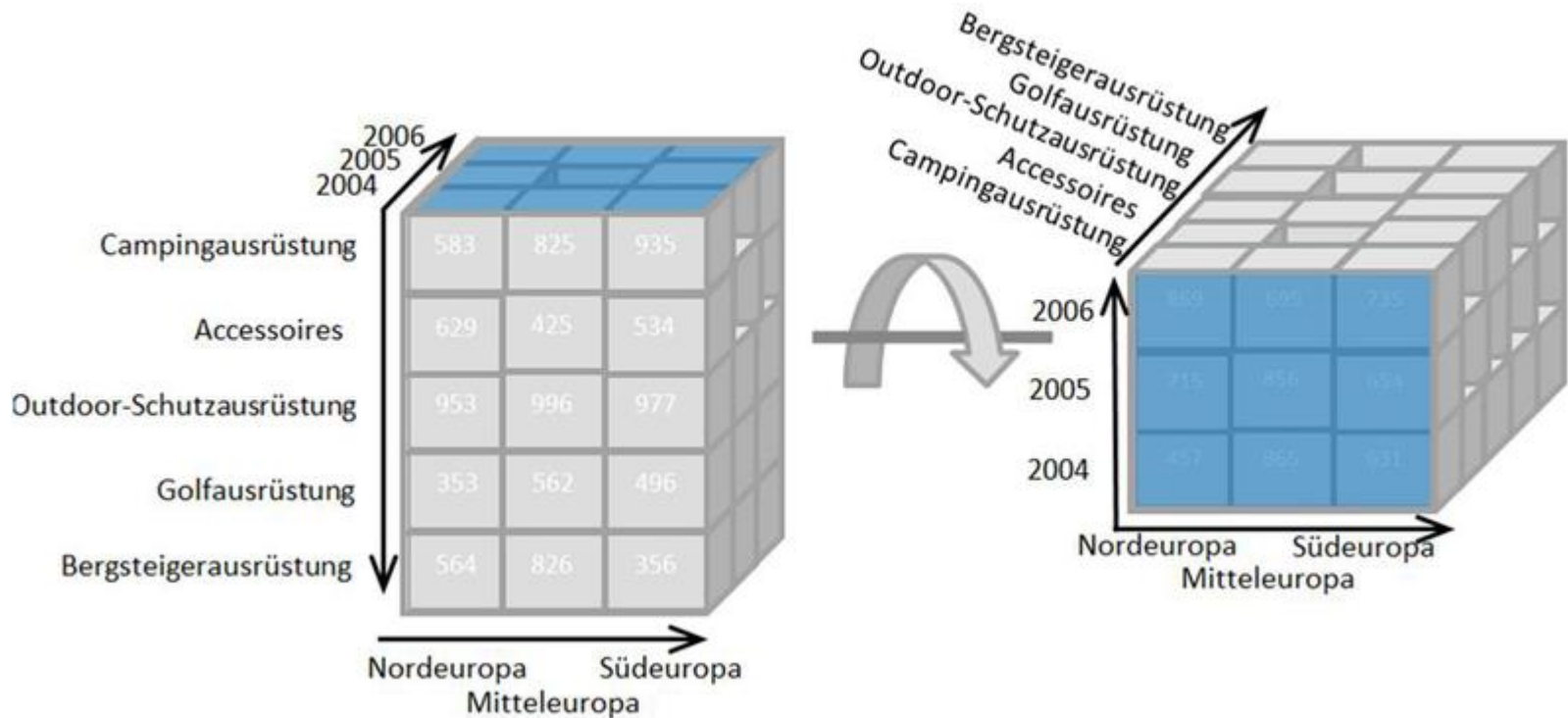
Před operací pivoting



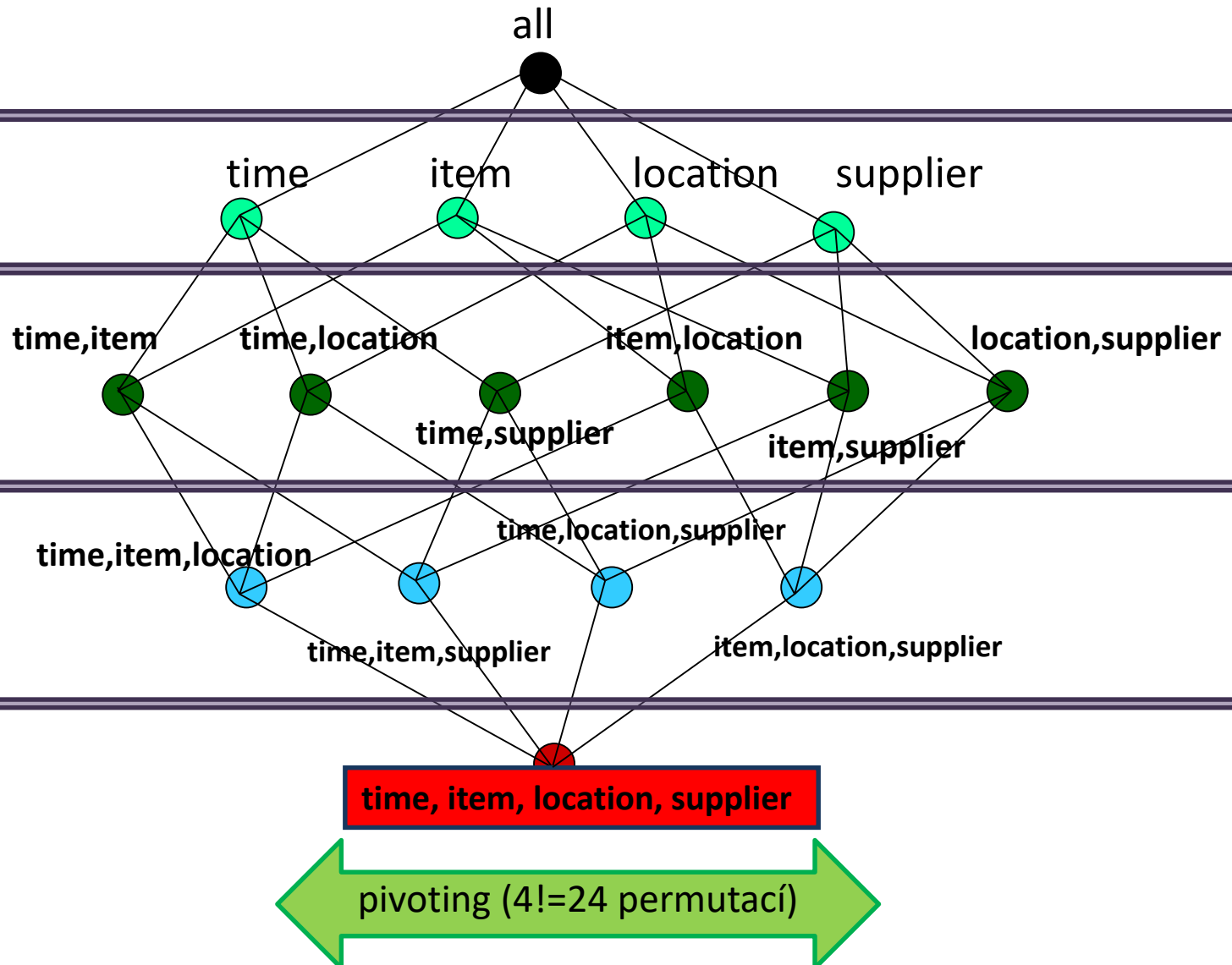
Po operaci pivoting



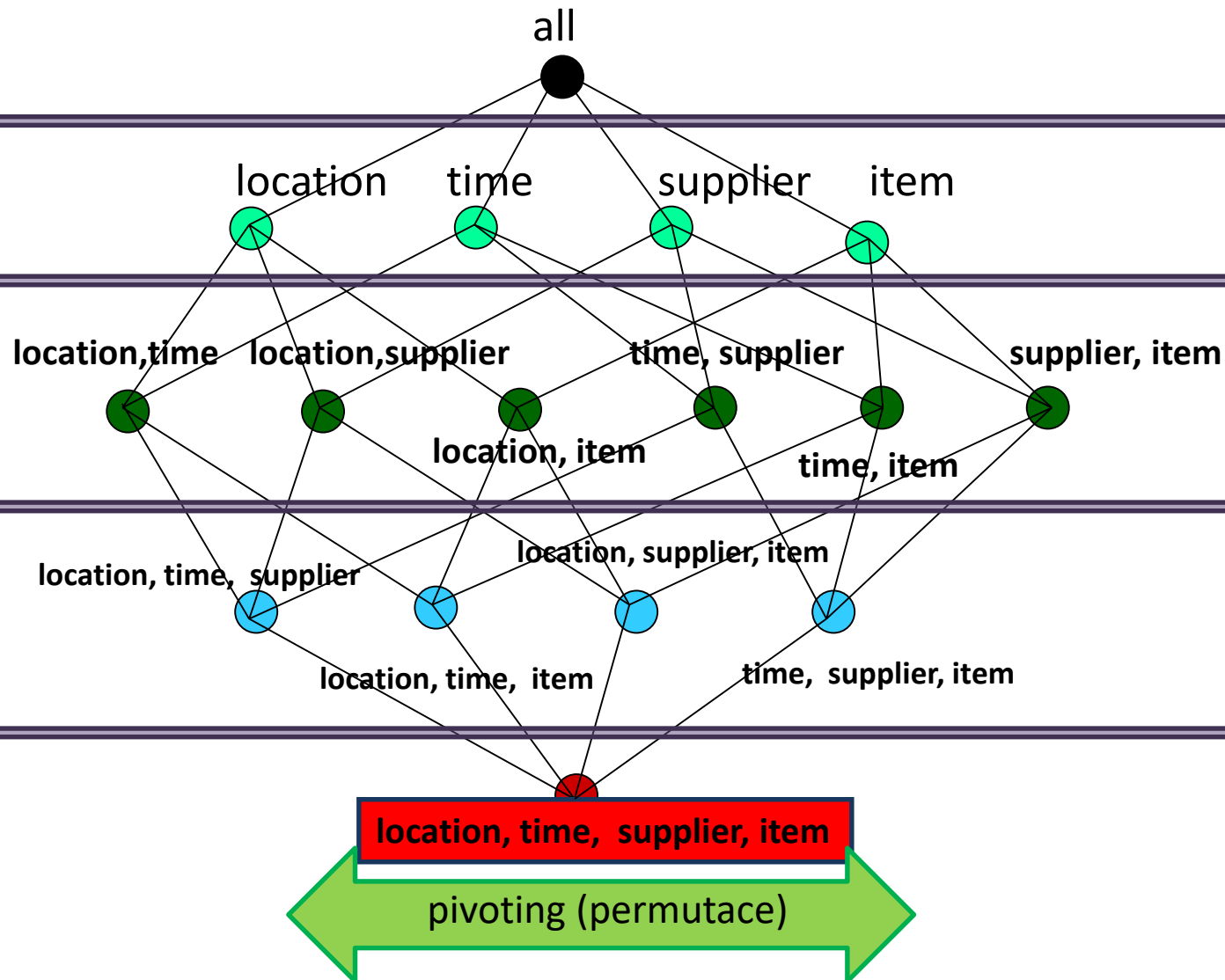
Pivoting



Původní relace R



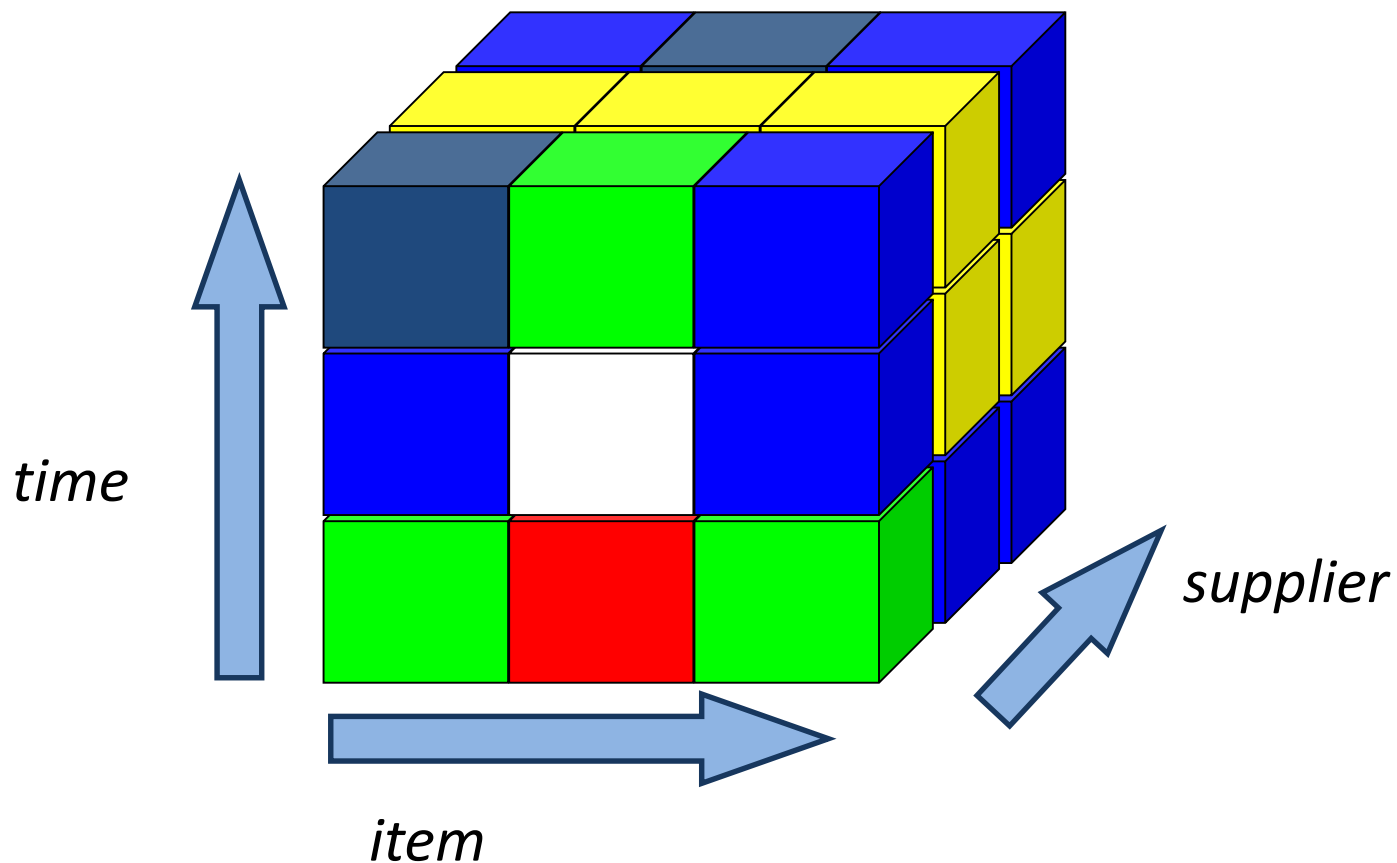
Otočená datová kostka



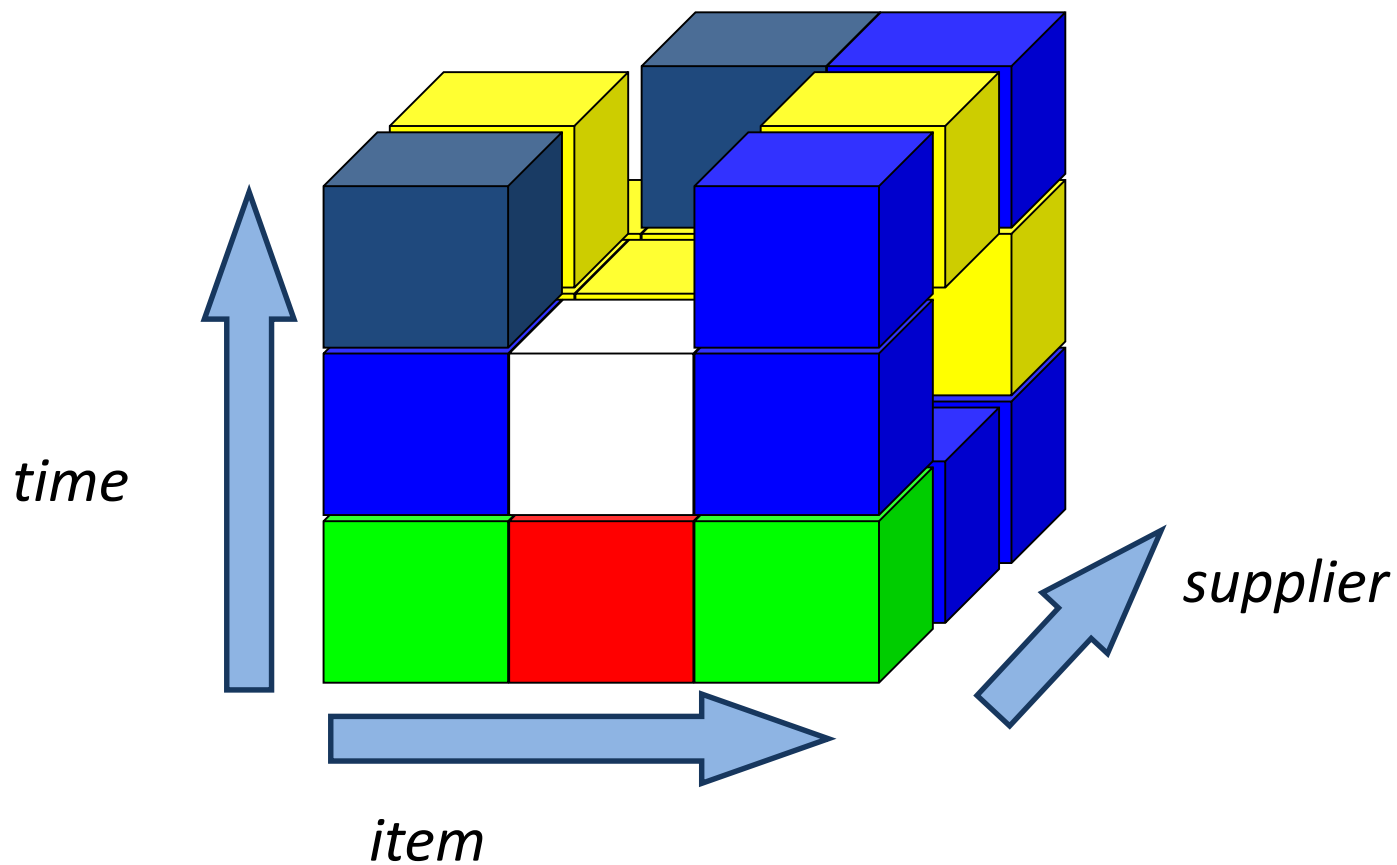
Slicing and Dicing

- **Změna skutečné kardinality jedné nebo více dimenzí**
- **Vstup:** uspořádaná množina dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$, kde $k_1, k_2, k_3, \dots, k_i, \dots, k_n$ jsou skutečné kardinality jednotlivých dimenzí
- **Výstup:** uspořádaná množina dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$ a obecně jiné $k_1, k_2, k_3, \dots, l_i, \dots, k_n$, kde $k_i \neq l_i$
- Změnu lze provést výběrem, nastavením **filtru** ve **tvaru predikátu** apod.
- Výsledek ovlivňují i **filtry neaktivních dimenzí**

Před operací slicing and dicing



Po operaci provedené výběrem



VIZUALIZACE MULTIDIMENZIONÁLNÍ KOSTKY A OLAP OPERACÍ NA 2D PRŮMĚTNĚ

Kolekce struktur

- Nelze zobrazovat $m > 3$, ale obecně stejně zobrazujeme na 2D průmětnu
- Funkce $g(A_1 \times A_2 \times A_3 \times \dots \times A_m) = F$ je vlastně kolekcí struktur, kterou vizualizujeme **tabulkou**

collection of
structure

properties

A_1 : jednoduchý datový typ₁

A_2 : jednoduchý datový typ₂

...

A_m : jednoduchý datový typ_m

F : agregovatelný datový typ

end structure

Tabulka

D:\DATAFIX\Vema\EKOS1700 FAK\Vema TH								
Soubor Úpravy Stránka								
VMMI0102 - Fakturační případy vydané								
Servis Viditelnost zpět Náhled tisku Přechody Verze								
Největší dlužníci								
Výběr	1-12	< x >40000000	< x <31.10	< x 1	< x	< x	< x >0.00	< x
		x ▲ Firma	x ▲ Splatnost	x ▲ Nákl. stř. nejvyšší	x ▲ Nákl. stř. vyšší	x ▲ Nákl. stř. základní	x Celkem	x ▼ Zbývá
<input type="checkbox"/>	1	45860951-Kristian s.r.o.	24.06.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	61 244.00	61 244.00
<input type="checkbox"/>	2	42035709-MEDIK s.r.o.	20.05.01	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	50 473.90	50 473.90
<input type="checkbox"/>	3	46218487-BALLSTREET	04.03.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	59 187.78	9 467.78
<input type="checkbox"/>	4	46471219-Johansson s.r.o.	18.04.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	7 870.45	7 870.45
<input type="checkbox"/>	5	47952425-A B S tech.činnosti	22.04.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	5 390.00	5 390.00
<input type="checkbox"/>	6	42587912-BSIO a.s.	30.03.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	3 856.70	3 856.70
<input type="checkbox"/>	7	47526869-AKO s.r.o.	30.03.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	18 100.00	2 850.00
<input type="checkbox"/>	8	99990001-Kutálek Petr, Ing.	15.01.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	1 428.00	1 428.00
<input type="checkbox"/>	9	47935469-Computer office	06.02.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	1 439.60	707.60
<input type="checkbox"/>	10	45872463-Renovace a.s.	07.04.04	1-Brno	3-výroba	31-výroba nábytku	121.40	121.40
<input type="checkbox"/>	11	47952425-A B S tech.činnosti	20.03.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	16.50	16.50
<input type="checkbox"/>	12	49385719-Ekolab s.r.o.	30.03.04	1-Brno	9-rozvahové pracoviště	91-rozvahové pracoviště	70 759.90	0.90
	12						279 888.23	143 427.23
Ladění++								

Tabulka

- pokud máme více (k) funkcí g se stejnými dimenzemi (*galaktické schéma*), lze je vizualizovat v jediné tabulce

collection of
structure

properties

A_1 : jednoduchý datový typ₁

A_2 : jednoduchý datový typ₂

...

A_m : jednoduchý datový typ_m

F_1 : fakt- agregovatelný datový typ₁

F_2 : fakt- agregovatelný datový typ₂

...

F_k : fakt- agregovatelný datový typ_k

end structure

Tabulka

- Lze provádět všechny 4 operace:
- ***roll-up*** – ubírání dimenzionálních sloupců zprava i obecně
- ***drill-down*** – přidávání dimenzionálních sloupců napravo i obecně
- ***pivoting*** – změna pořadí dimenzionálních sloupců
- ***slicing & dicing*** – filtry, případně výběr ad hoc např. zaškrtáváním

Sloupce, řádky a stránky

- omezený způsob vhodný pouze pro $\leq 3D$ kostky

Příklad

- Příklad: zjistit prodeje jednotlivých položek v jednotlivých měsících roku v jednotlivých obchodech
- Výsledek ve kompaktním tvaru kontingenční tabulky s jediným faktem

Store: New York		Products					
<u>PAGES</u> : STORE dimension		<u>COLUMNS</u> : PRODUCT dimension					
<u>ROWS</u> : TIME dimension		Hats	Coats	Jackets	Dresses	Shirts	Slacks
	Jan	200	550	350	500	520	490
	Feb	210	480	390	510	530	500
	Mar	190	480	380	480	500	470
	Apr	190	430	350	490	510	480
	May	160	530	320	530	550	520
	Jun	150	450	310	540	560	330
	Jul	130	480	270	550	570	250
	Aug	140	570	250	650	670	230
	Sep	160	470	240	630	650	210
	Oct	170	480	260	610	630	250
	Nov	180	520	280	680	700	260
Months	Dec	200	560	320	750	770	310

Další příklady dotazů OLAP

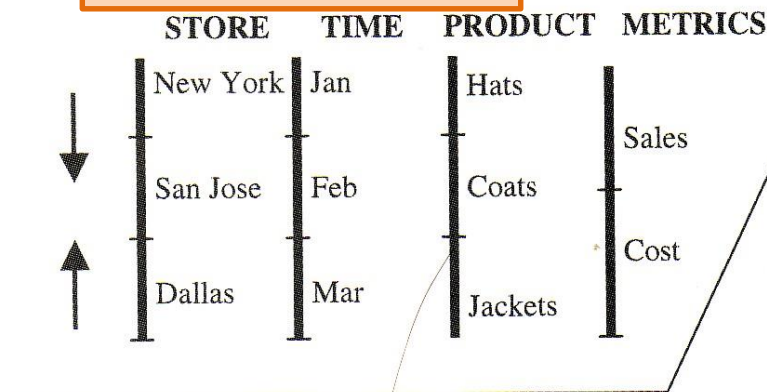
- ***Celkové prodeje všech produktů za 5 let***
 - Řádky: roky 2000, 1999, 1998 atd.
 - Sloupce: součet prodejů pro všechny produkty
 - Stránky: jedna stránka pro každý obchod
- ***Porovnej prodeje všech produktů a obchodů mezi roky 1999 a 2000***
 - Řádky: roky 2000, 1999, průměry, rozdíly
 - Sloupce: jeden sloupec pro každý produkt
 - Stránka: pouze jedna – pro všechny obchody
- ***Porovnej totéž, ale pouze u zlevněných produktů***
 - Řádky: roky 2000, 1999, 1998 atd.
 - Sloupce: jeden sloupec pro jeden produkt, ale pouze vybrané produkty
 - Stránky: jedna stránka pro každý obchod

Vícedimenzionální hyperkostky

- **Problém:** více než tři dimenze
- Uložení v datovém skladu není problém, dimenzí může být libovolný počet
- Problém je se zobrazením výsledků OLAP analýzy – nevystačíme **s řádky, sloupci a stránkami**

Příklad – zobrazení 4 dimenzí

multidimenzionální
struktura



zobrazení ve 3D

PAGE: Store Dimension

ROWS: Time Dimension

COLUMNS:
sloupce
kombinované s
dvojtečkou (dosti
nesystematické)

New York Store

	Hats:Sales	Hats:Cost	Coats:Sales	Coats:Cost	Jackets:Sales	Jackets:Cost
Jan	450	350	550	450	500	400
Feb	380	280	460	360	400	320
Mar	400	310	480	410	450	400

Příklad – zobrazení 6 dimenzí

multidimenzionální
struktura

DEMO	PROMO	STORE	TIME	PRODUCT	METRICS
Life Style	Type	New York	Jan	Hats	Sales
Income	Coupon	San Jose	Feb	Coats	Cost

HOW DISPLAYED ON A PAGE

PAGE: Demographics &
Promotion Dimensions
combined

ROWS: Store & Time
Dimensions combined

COLUMNS: Product &
Metrics combined

Life Style : Coupon

		Hats Sales	Hats Cost	Coats Sales	Coats Cost
New York	Jan	220	170	270	220
	Feb	190	140	230	180
Boston	Jan	200	160	240	200
	Feb	180	130	220	170

přechod ke
kontingenci
tabulce

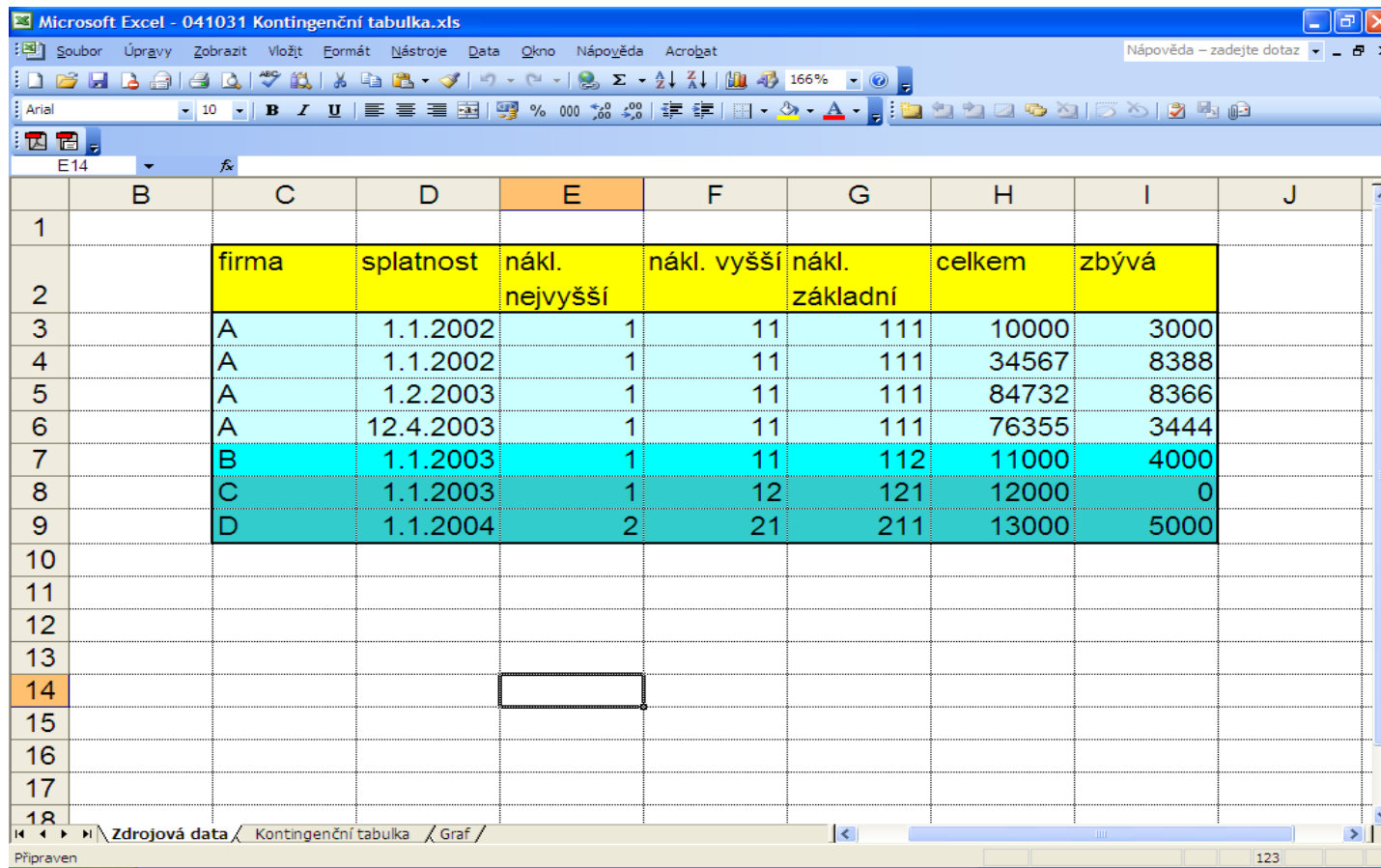
Kontingenční tabulka (Pivot table)

- řádky i sloupce jsou dimenze i hierarchické
- fakta jsou na průsečíku dimenzí
- fakt je proto v zásadě ***pouze jediný***, více faktů se v průsečíku nedá zobrazit
- řeší se to opakováním tabulky, záložkami nebo více sloupci, což se zdá méně přehledné

Kontingenční tabulka

- dimenze lze přesunovat a zaměňovat – pivoting
- zakrýváním hierarchie dimenzí lze zvyšovat a snižovat agregaci – roll-up, drill-down
- všechny sloupce lze zneviditelnit nebo nevyužívat – slice & dice
- prostorově méně náročná, avšak nejsou vidět všechny míry a hierarchie dimenzí může být nepřehledná

Prvotní data ve formě relace



Microsoft Excel - 041031 Kontingenční tabulka.xls

Nápověda – zadejte dotaz

166%

Arial 10

E14

	B	C	D	E	F	G	H	I	J
1									
2		firma	splatnost	nákl. nejvyšší	nákl. vyšší	nákl. základní	celkem	zbývá	
3		A	1.1.2002	1	11	111	10000	3000	
4		A	1.1.2002	1	11	111	34567	8388	
5		A	1.2.2003	1	11	111	84732	8366	
6		A	12.4.2003	1	11	111	76355	3444	
7		B	1.1.2003	1	11	112	11000	4000	
8		C	1.1.2003	1	12	121	12000	0	
9		D	1.1.2004	2	21	211	13000	5000	
10									
11									
12									
13									
14									
15									
16									
17									
18									

Zdrojová data Kontingenční tabulka Graf

Připraven 123

Kontingenční tabulka

Microsoft Excel - 041031 Kontingenční tabulka.xls

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda Acrobat

Nápověda - zadejte dotaz

Arial 10 B I U

E23

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3	Součet z celkem		nákl. nejvyšší	nákl. vyšší	nákl. základní							
4			1					Celkem z 1	2		Celkem z 2	Celkový součet
5			11		Celkem z 11	12	Celkem z 12		21		Celkem z 21	
6	splatnost	firma	111	112		121			211			
7	1.1.2002	A	44567		44567			44567				44567
8	Celkem z 1.1.2002		44567		44567			44567				44567
9	1.1.2003	B		11000	11000			11000				11000
10		C				12000	12000	12000				12000
11	Celkem z 1.1.2003			11000	11000	12000	12000	23000				23000
12	1.2.2003	A	84732		84732			84732				84732
13	Celkem z 1.2.2003		84732		84732			84732				84732
14	12.4.2003	A	76355		76355			76355				76355
15	Celkem z 12.4.2003		76355		76355			76355				76355
16	1.1.2004	D							13000	13000	13000	13000
17	Celkem z 1.1.2004								13000	13000	13000	13000
18	Celkový součet		205654	11000	216654	12000	12000	228654	13000	13000	13000	241654
19												
20												
21												
22												
23												
24												
25												
26												

Zdrojová data Kontingenční tabulka Graf

Připraven

123

Pivoting

- operace otočení provede přesun dimenzí mezi souřadnicemi x a y případně změní jejich pořadí v hierarchii

Pivoting

Microsoft Excel - 041031 Kontingenční tabulka.xls

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda Acrobat

Nápověda – zadejte dotaz

Arial 10

119

	A	B	C	D	E	F	G	H	I	J	K	L	M	N			
1																	
2																	
3	Součet z celkem			firma	splatnost												
4				A			Celkem z A	B		Celkem z B	C		Celkem z C	D		Celkem z D	Celkový součet
5	nákl. nejvyšší	nákl. vyšší	nákl. základní	1.1.2002	1.2.2003	12.4.2003		1.1.2003		1.1.2003		1.1.2004					
6	1	11	111	44567	84732	76355	205654	11000	11000								205654
7			112														11000
8		Celkem z 11		44567	84732	76355	205654	11000	11000								216654
9		12	121								12000	12000					12000
10		Celkem z 12									12000	12000					12000
11	Celkem z			44567	84732	76355	205654	11000	11000	12000	12000						228654
12	2	21	211									13000	13000				13000
13		Celkem z 21										13000	13000				13000
14	Celkem z											13000	13000				13000
15	Celkový součet			44567	84732	76355	205654	11000	11000	12000	12000	13000	13000				241654
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	

Zdrojová data Kontingenční tabulka Graf

Připraven

123

Roll-up a drill-down

- V kontingenční tabulce jsou operace zvětšování a zmenšování detailů, které postupně odstraňují nižší patra hierarchických klíčů a počítají zadanou operaci (nejčastěji součtový agregát)
- Operace ***drill-down*** je opačná k ***roll-up***, tj. zmenšuje detaily.

Roll-up a drill-down

Microsoft Excel - 041031 Kontingenční tabulka.xls

Nápověda – zadejte dotaz

Arial 10

	A	B	C	D	E	F	G	H	I
1									
2									
3	Součet z celkem			firma	splatnost				
4				A	B	C	D	Celkový součet	
5	nákl. nejvyšší	nákl. vyšší	nákl. základní						
6	1			205654	11000	12000		228654	
7	2						13000	13000	
8	Celkový součet			205654	11000	12000	13000	241654	
9									
10									
11									
12									
13									
14									

Zdrojová data Kontingenční tabulka Graf

Připraven

123

Slicing and dicing

- Je možné zadávat filtry nad dimenzemi

[stáhnout olkt a pak spustit](#)

GRAFY V MULTIDIMENZIONÁLNÍ KOSTCE

Grafické zobrazení

- K multidimenzionálnímu pohledu na hodnoty je velmi dobré přiřadit transformovaný pohled pomocí grafů

Jednoduchý graf (histogram)

- Kolekce dvojic (x-vodorovná, y-svislá)
- x – výčtový, číselný, rovnoměrný i nerovnoměrný
- y – číselný, měřitelný
- nutno redukovat počet dimenzí a faktů na 1

Koláčový (výsečový) graf - Pie

- Kolekce dvojic (x-označení, y-velikost úhlu výseče)
- x – výčtový, číselný, rovnoměrný i nerovnoměrný
- y – číselný **nezáporný**, úhel = $360 \cdot (\text{hodnota} / \text{součet})$
- Obvyklé vyjádření v procentech (y ve stupních)/3,6
- nutno redukovat počet dimenzí i faktů na 1

Násobný graf (histogram)

- Kolekce dvojic (x -vodorovná, (y_1, y_2, \dots, y_n) - svislá)
- x – výčtový, číselný, rovnoměrný i nerovnoměrný
- y_i – číselný, měřitelný
- nutnost redukovat počet dimenzí na 1, počet faktů není třeba omezovat

Operace kostky u grafů

- Operace nad kostkou silně omezeny:
- ***roll-up*** – neužívá se
- ***drill-down*** – speciální provedení bude diskutováno separátně
- ***pivoting*** – pouze jedna dimenze - nemá smysl
- ***slicing & dicing*** – filtry, případně výběr ad hoc např. zaškrtáváním

Kombinovaná operace drill-down u živých grafů

- ***živé grafy***
- předpokládáme uspořádání dimenzí $\{D_1, D_2, D_3, \dots, D_n\}$
- pokud klikneme v grafu na hodnotu viditelné dimenze D_i , která je jediná, nastaví se pro dimenzi D_i ***filtr rovnosti s kliknutou hodnotou (slicing)*** a aktuální dimenzí v grafu se stane dimenze D_{i+1} (***drill-down***). Lze provést pouze pro $i \leq n$.

Multi Dimensional eXpressions – dotazy pro multidimenzionální kostku

MDX (XMLA EXECUTE PRO MULTIDIMENZIONÁLNÍ KOSTKY)

Kostka Sales - dimenze Geography

- ▼ **D** Geography

- ▼ **H** Geo

- ▼ **LS** Levels

- LA** All-Level

- L** Continent

- L** Country

- L** City

- ▼ **AD** All Regions

- ▼ **M** America

- ▶ **M** Canada

- ▼ **M** United States

- M** Los Angeles

- M** New York

- M** San Francisco

- ▶ **M** Mexico

- ▶ **M** Venezuela

- ▶ **M** Europe

- ▼ **H** Economy

- ▼ **LS** Levels

- LA** All-Level

- L** Partnership

- L** Country

- ▼ **AD** All

- ▼ **M** NAFTA

- M** Canada

- M** United States

- M** Mexico

- M** Venezuela

- ▼ **M** EU

- M** France

- M** Spain

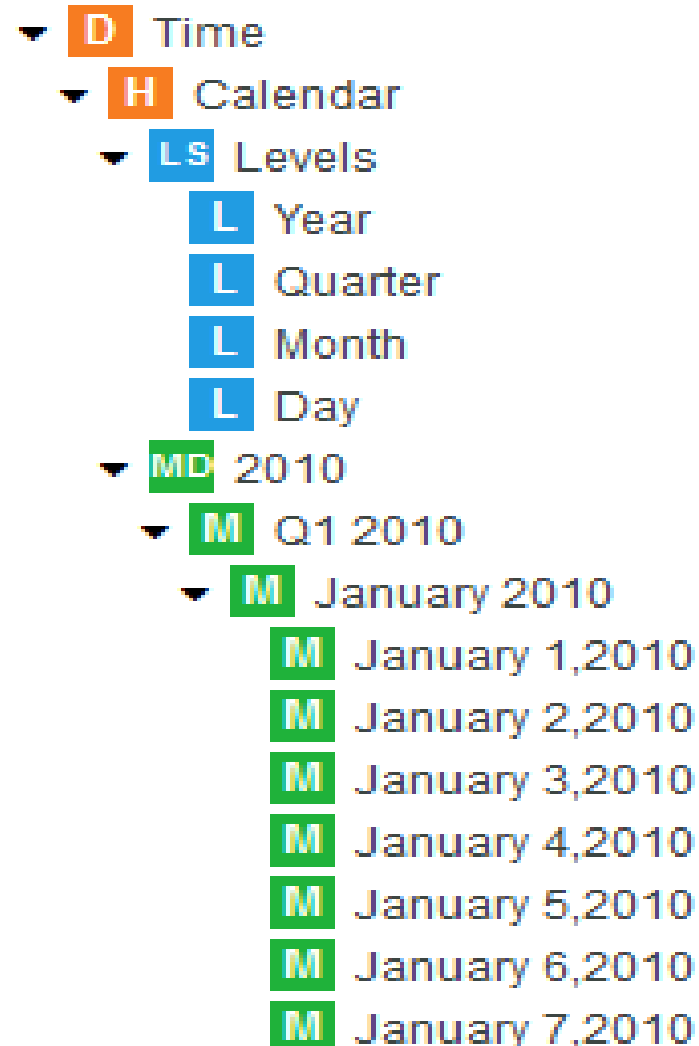
- ▼ **M** None

- M** Switzerland

Sales - dimenze Product

- ▼ **D** Product
 - ▼ **H** Prod
 - ▼ **LS** Levels
 - L** Company
 - L** Article
 - L** Licence
 - ▼ **MD** Crazy Development
 - ▼ **M** icCube
 - M** Corporate
 - M** Partnership
 - M** Personal
 - M** Startup

Sales - definisce dimensioni Time hierarchie Calendar

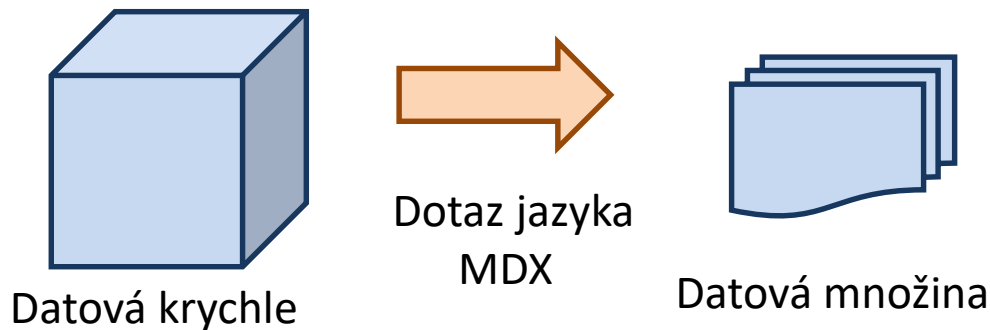


Sales - definice míry Amount



Jazyk MDX

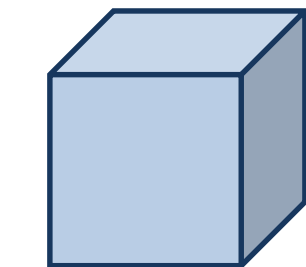
- MDX = **M**ulti**D**imensional **E**Xpressions
- Dotazovací jazyk pro navigaci v multidimenzionálních údajích
- Výsledek dotazu – zobrazitelná tabulka (bud' to 2D nebo některé náhradní techniky)



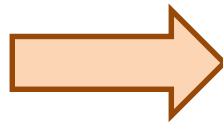
Jazyk MDX - příklad

- Příklad příkazu SELECT

```
SELECT <specifikace_osy> ON COLUMNS,  
      <specifikace_osy> ON ROWS  
FROM <krychle>  
[WHERE <filtr>]
```



Datová krychle



MDX

	Čechy	Slovensko
	Zisk	Zisk
Mléko	1299343	2145441
Pivo	4545784	1954001
Rohlíky	214147	774755
Víno	7474587	7847111

Množina (set)

- Více výsledných buněk v tabulce

```
{[William Wong],[Andrea Donlon]}  
{([Non-consumable],[USA]),([Beverages],[Mexico])}
```

- Definuje množinu více průsečíků dimenzí

Úplná agregace bez dimenzí

pouze sloupce

```
SELECT [Measures].[Amount] ON COLUMNS  
FROM [Sales]
```

- Výpis jedné hodnoty v jednom sloupci, celková utržená částka za zboží.

Amount
1023

```
SELECT  
    [Measures].[Amount] ON COLUMNS  
FROM  
    [Sales]
```

Amount

1023

Uspořádaná n-tice

pouze sloupce

```
SELECT ([Product].[Prod].[Licence].[Corporate],  
        [Geography].[Geo].[Continent].[America]) ON COLUMNS  
FROM [Sales]
```

- Výpis uspořádané n-tice v jednom sloupci, celková utržená částka za zboží.

Corporate
America
768

```
SELECT  
    ([Product].[Prod].[Licence].[Corporate],  
     [Geography].[Geo].[Continent].[America]) ON COLUMNS  
FROM  
    [Sales]
```

Corporate
America
768

Řádky i sloupce

```
SELECT {[Product].[Prod].[Licence].[Corporate],  
       [Product].[Prod].[Licence].[Partnership],  
       [Product].[Prod].[Licence].[Personal]} ON COLUMNS,  
       {[Geography].[Geo].[Country].[Spain],  
       [Geography].[Geo].[Country].[United States],  
       [Geography].[Geo].[Country].[Switzerland]} ON ROWS  
FROM [Sales]  
WHERE ([Time].[Calendar].[Year].[2010])
```

- Výpis tabulky s řádky i sloupci
- WHERE specifikuje omezení

	Corporate	Partnership	Personal
Spain			3
United States	768		
Switzerland	144	96	8

Řádky i sloupce

```
SELECT { [Product].[Prod].[Licence].[Corporate],  
        [Product].[Prod].[Licence].[Partnership],  
        [Product].[Prod].[Licence].[Personal]} ON COLUMNS,  
       { [Geography].[Geo].[Country].[Spain],  
        [Geography].[Geo].[Country].[United States],  
        [Geography].[Geo].[Country].[Switzerland]} ON ROWS  
FROM [Sales]  
WHERE ([Time].[Calendar].[Year].[2010])
```

RESULT

	<i>Corporate</i>	<i>Partnership</i>	<i>Personal</i>
→ <i>Spain</i>			3
→ <i>United States</i>	768		
→ <i>Switzerland</i>	144	96	8

Další dimenze

- Až 128
- COLUMNS, ROWS, PAGES, SECTIONS, CHAPTERS, AXIS (číslo)
- osy číslovány od 0
- $0=x$, $1=y$, $2=z$

Řádky a sloupce jako osy

```
SELECT {[Product].[Prod].[Licence].[Corporate],  
       [Product].[Prod].[Licence].[Partnership],  
       [Product].[Prod].[Licence].[Personal]} ON AXIS(0),  
       {[Geography].[Geo].[Country].[Spain],  
       [Geography].[Geo].[Country].[United States],  
       [Geography].[Geo].[Country].[Switzerland]} ON AXIS(1)  
FROM [Sales]  
WHERE ([Time].[Calendar].[Year].[2010])
```

- Výpis tabulky s řádky i sloupci
- WHERE specifikuje omezení

	Corporate	Partnership	Personal
Spain			3
United States	768		
Switzerland	144	96	8

Řádky a sloupce jako osy

```
SELECT { [Product].[Prod].[Licence].[Corporate],  
        [Product].[Prod].[Licence].[Partnership],  
        [Product].[Prod].[Licence].[Personal]} ON AXIS (0),  
       { [Geography].[Geo].[Country].[Spain],  
        [Geography].[Geo].[Country].[United States],  
        [Geography].[Geo].[Country].[Switzerland]} ON AXIS (1)  
FROM [Sales]  
WHERE ([Time].[Calendar].[Year].[2010])
```

RESULT

	<i>Corporate</i>	<i>Partnership</i>	<i>Personal</i>
→ <i>Spain</i>			3
→ <i>United States</i>	768		
→ <i>Switzerland</i>	144	96	8

Fakta

```
SELECT [Measures].MEMBERS ON COLUMNS,  
       [Product].MEMBERS ON ROWS  
FROM [Sales]
```

- Vrací fakta pro každý produkt se součtem na každé úrovni součtů

RESULT	
	<i>Amount</i>
▼ <i>Crazy Development</i>	1023
▼ <i>icCube</i>	1023
<i>Corporate</i>	912
<i>Partnership</i>	96
<i>Personal</i>	15
<i>Startup</i>	

Všetchny prvky MEMBERS

```
SELECT Measures.MEMBERS ON COLUMNS,  
    {[Store].[Store State].[CA],  
    [Store].[Store State].[WA]} ON ROWS  
FROM [Sales]
```

- Vrací fakta pro státy California a Washington

Funkce Children

```
SELECT Measures.MEMBERS ON COLUMNS,  
  {[Geography].[Geo].[Country].[United States].CHILDREN,  
   [Geography].[Geo].[Country].[Switzerland].CHILDREN} ON ROWS  
FROM [Sales]
```

- pracuje s hierarchií a pro sloupce užije všechny následníky států United States a Switzerland.
- podobně lze užít i ***další operace na stromem***

Funkce Children

```
SELECT Measures.MEMBERS ON COLUMNS,  
       { [Geography].[Geo].[Country].[United States].CHILDREN,  
         [Geography].[Geo].[Country].[Switzerland].CHILDREN } ON ROWS  
FROM [Sales]
```

RESULT

	<i>Amount</i>
<i>Los Angeles</i>	
<i>New York</i>	768
<i>San Francisco</i>	
<i>Geneva</i>	128
<i>Lausanne</i>	56
<i>Zurich</i>	64

From jako filtr - poddotazy

- Jiný způsob slice v klauzuli FROM
- Použití poddotazu

```
SELECT  
  [Measures].MEMBERS ON 0,  
  [Geography].[Geo].MEMBERS ON 1  
FROM  
  (SELECT [Time].[Calendar].[Year].[2010] ON 0 FROM [Sales])
```

From jako filtr - poddotazy

```
SELECT
    [Measures].Members ON 0,
    [Geography].[Geo].Members ON 1
FROM
    (SELECT [Time].[Calendar].[Year].[2010] ON 0 FROM [Sales])
```

	Amount
▼ All Regions	1023
▼ America	768
▼ Canada	
Quebec	
Toronto	
▼ Mexico	
Mexico	
▼ United States	768
Los Angeles	
New York	768
San Francisco	
▼ Venezuela	
Caracas	
▼ Europe	255
▼ France	4
Paris	4
▼ Spain	3
Barcelona	2
Madrid	1
Valencia	
▼ Switzerland	248
Geneva	128
Lausanne	56
Zurich	64

Efektivní výpočet datových kostek

- ***Materializace datové kostky***
- Materializace je ***vypočtení agregčních hodnot předem***
- Materializace každého kuboidu (***plná materializace***), žádného (***bez materializace***) nebo jen některých (***částečná materializace***)
- Výběr podkostek, které se budou materializovat
- Na základě velikosti, sdílení nebo frekvence přístupů

KONEC