



Eiropas Sociālā fonda projekts

“Datorzinātnes pielietojumi un tās saiknes ar kvantu fiziku”

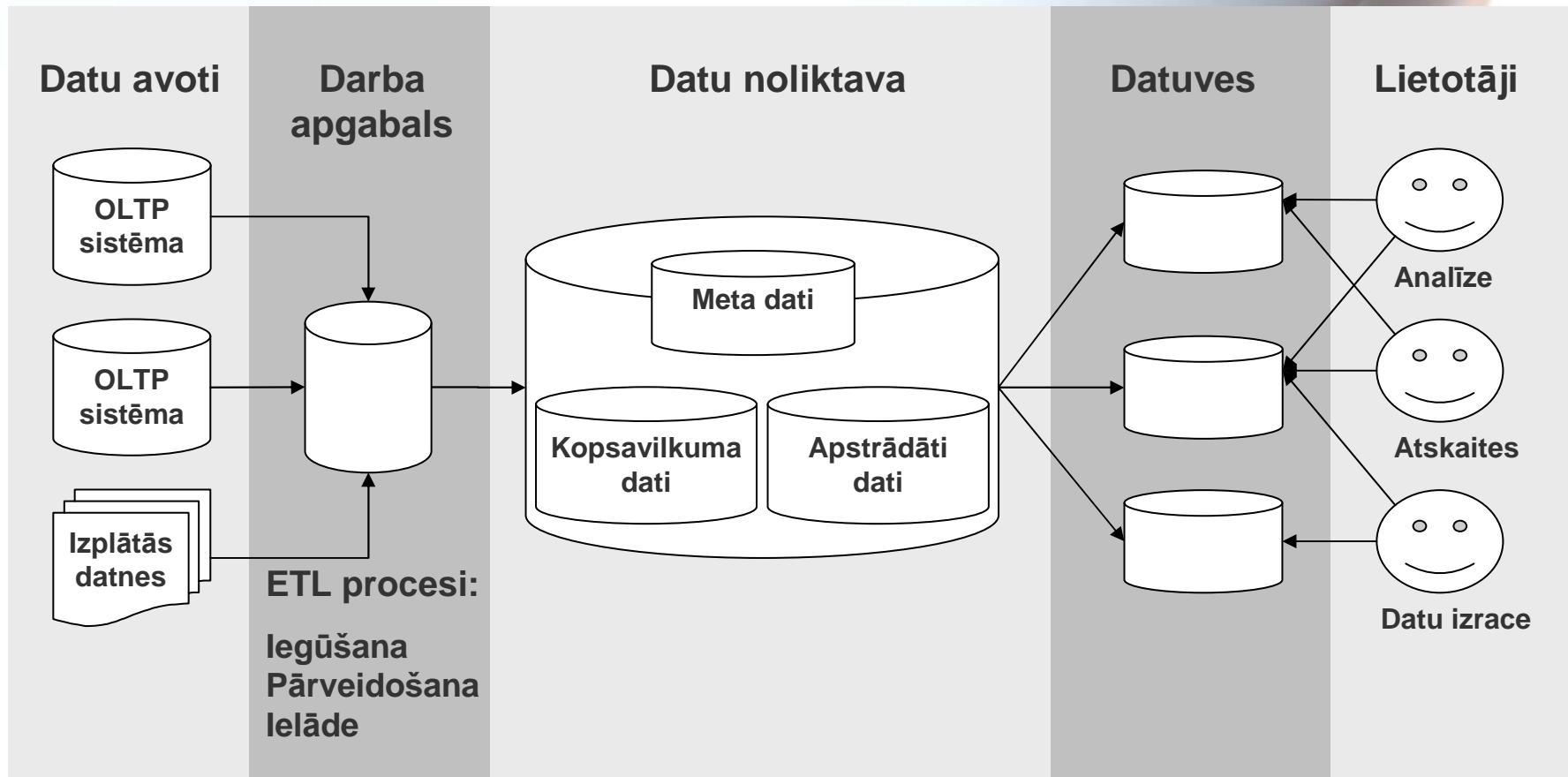
Nr.2009/0216/1DP/1.1.1.2.0/09/APIA/VIAA/044

Datu noliktavas evolūcijas problēmas

Darja Solodovņikova

Seminārs 01.06.2010.

Datu noliktava



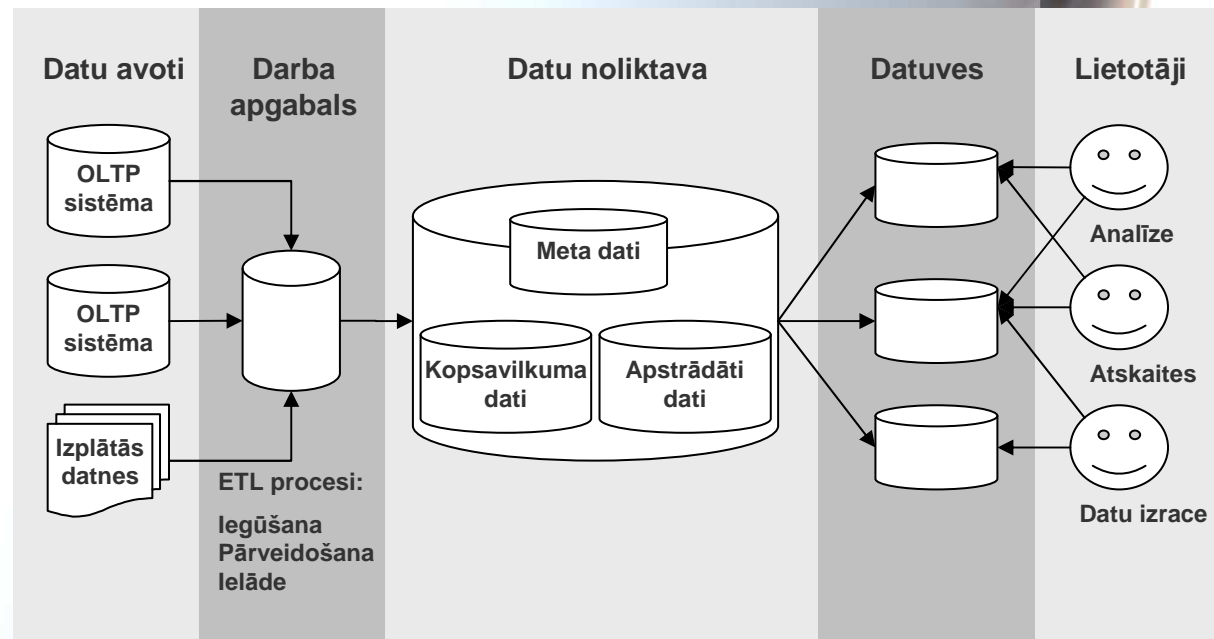
Datu noliktavas evolūcija un sekas

- **Izmaiņas**

- Datu avotos
- Biznesa prasībās

- **Sekas**

- Izmaiņas datu noliktavas shēmās
- Kļūdainie datu iegūšanas, pārveidošanas un ielādes (ETL) procesi
- Kļūdainas atskaites uz datu noliktavas shēmu



Datu noliktavas evolūcija – risinājumi

- **Izmaiņas biznesa prasībās**

- **Izmaiņas dimensijās** [Hurtado u.c. 1999.], [Mendelzon u.c. 2000.], [Bentayeb u.c. 2008]
- **Evolūcijas operācijas** [Blaschka 2000.], [Kaas u.c. 2004.], [Banerjee u.c. 2009]

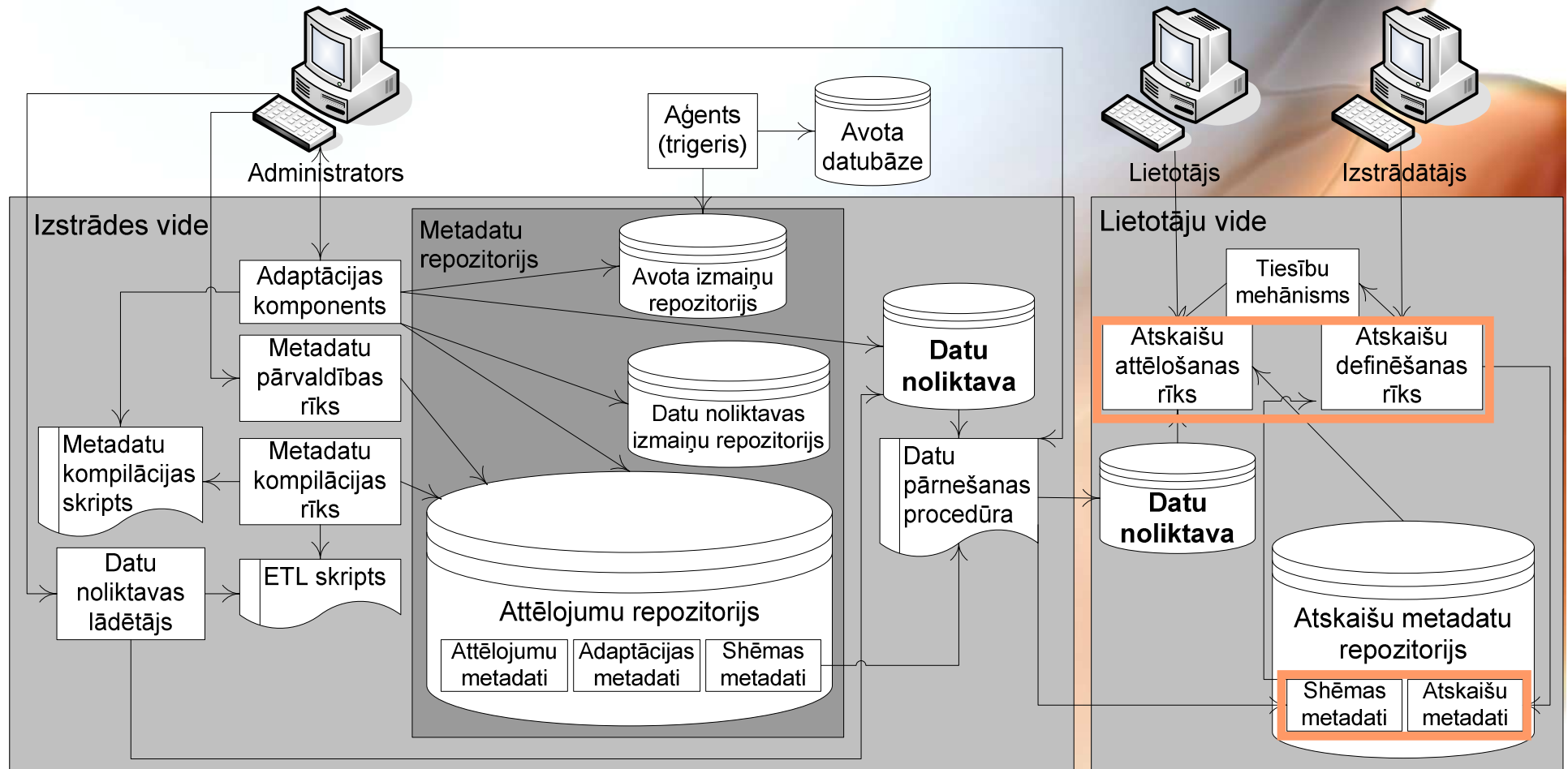
- **Izmaiņas datu avotos**

- **Shēmu transformācijas un attēlojumi** [Marotta 2000.], [McBrien u.c. 2003.], [Velegrakis u.c. 2003.]
- **Materializētu skatu pārdefinēšana** [Bellahsene 2002.], [Rundensteiner u.c. 2000.]

- **Datu noliktavas shēmas versijas**

- **Dimensiju versijas** [Body u.c. 2003.]
- **Daudzversiju daudzdimensiju shēmas** [Shahzad u.c. 2005.]
- **Shēmu palielinājumi** [Golfarelli u.c. 2006.]
- **Metadatu atbalsts un vaicājumi** [Morzy u.c. 2004.], [Wrembel u.c. 2005.]

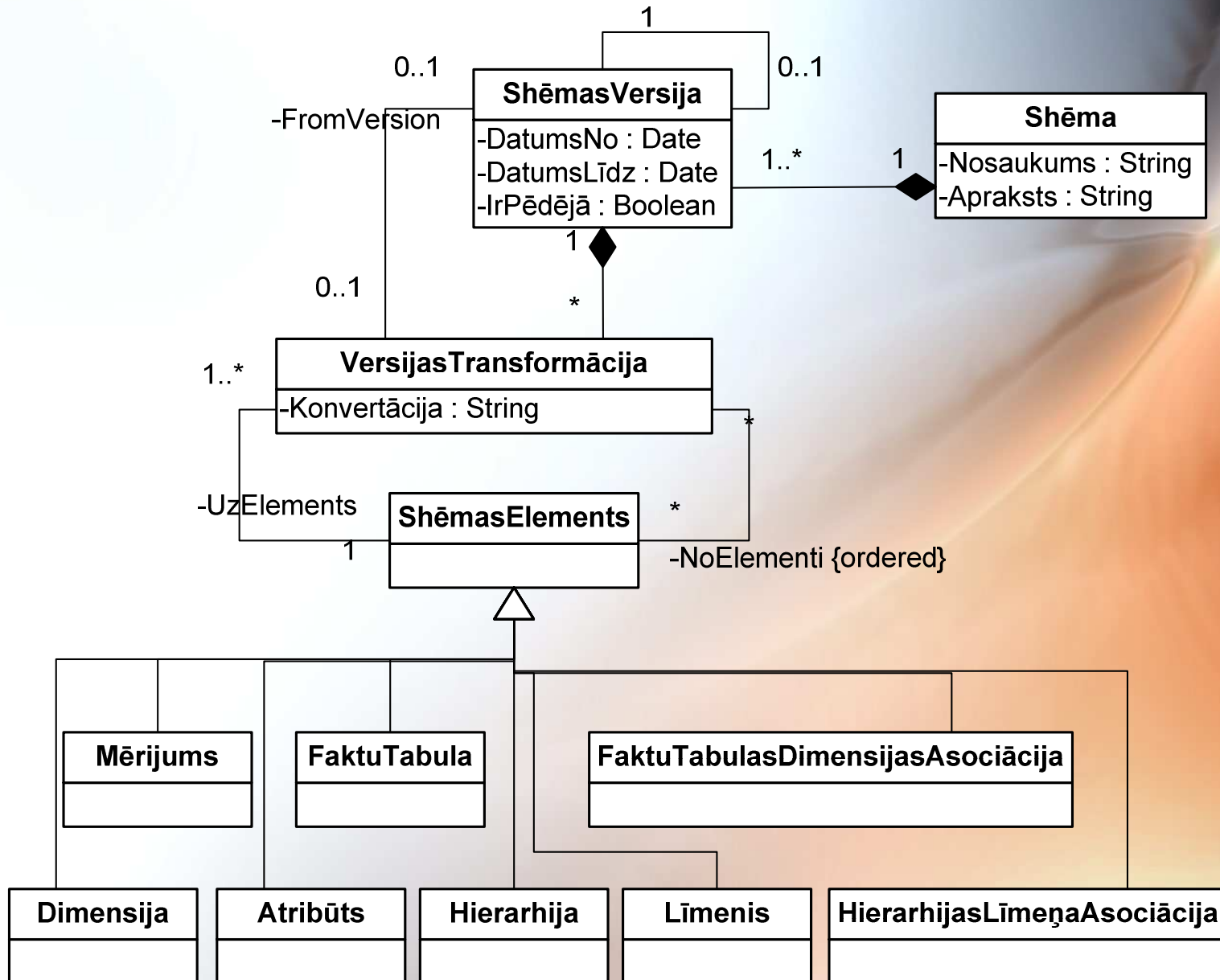
Datu noliktavas evolūcijas arhitektūra



Datu noliktavas metadati evolūcijas atbalstam

- **Par pamatu datu noliktavas metamodelim tika izmantots Common Warehouse Metamodel (CWM).**
- **CWM metamodelis tika paplašināts, lai atbalstītu datu noliktavas shēmas versijas.**
- **Shēmas metadati:**
 - **Fiziskais modelis**
 - Datu noliktavas realizācija relāciju datubāzē
 - CWM Relational pakotne
 - **Loģiskais modelis**
 - Daudzdimensiju datu noliktavas shēmas versijas
 - Sasaiste starp versijām caur versiju transformācijām
 - CWM OLAP pakotne papildināta ar shēmas versiju atbalstu
 - **Sasaiste – CWM Transformation pakotne**
- **Atskaišu metadati:**
 - Izveidoto atskaišu struktūra
 - CWM metadati nav pietiekami – jaunais atskaišu metadatu modelis

Loģiskā līmeņa metadati – versiju metadati



Evolūcijas atbalsts

- **Fiziskās izmaiņas**

- Darbojas ar datubāzes objektiem (tabulām, kolonām)
- Rezultātā tiek mainīti gan loģiskā, gan fiziskā līmeņa metadati
- Piemēri: jaunais dimensijas atribūts, dimensijas atribūta datu tipa maiņa, dimensijas atribūta dzēšana, jaunā faktu tabula u.c.

- **Loģiskās izmaiņas**

- Maina shēmas metadatus
- Rezultātā tiek mainīti loģiska līmeņa metadati vai var tikt pievienotas atslēgas un atslēgu kolonas
- Piemēri: dimensijas piesaistīšana faktu tabulai, jaunā dimensijas hierarhija, dimensijas hierarhijas dzēšana u.c.

Vaicājuma veidošana

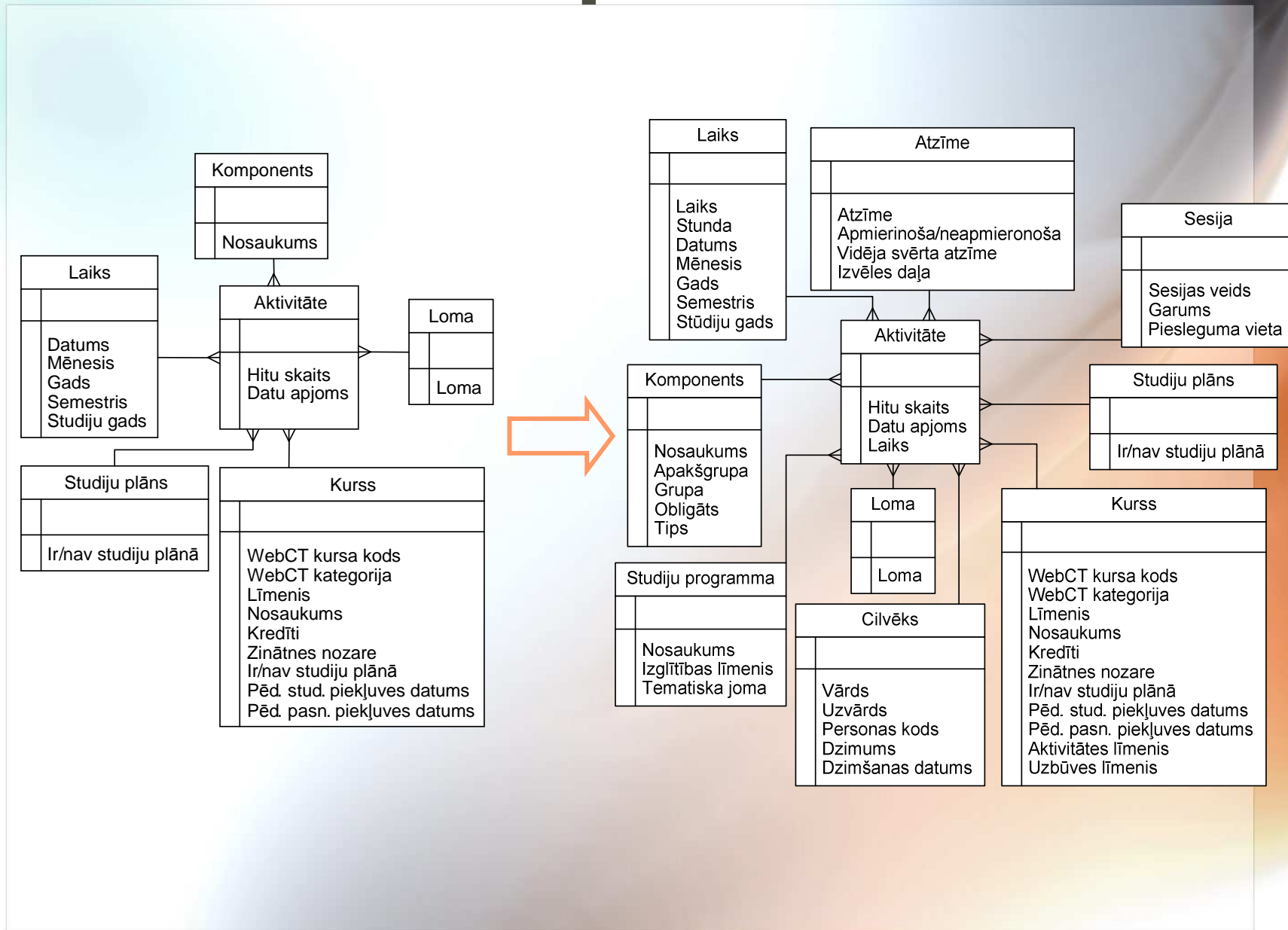
```
SELECT      {DISTINCT} <izvēlēto elementu saraksts>  
FROM        <kolonu kopu saraksts>  
{WHERE     ( <savienojumu saraksts> ) }  
{AND       ( <nosacījumu saraksts> ) }  
{GROUP BY  <grupēšanas elementu saraksts> }  
{HAVING    <grupēšanas nosacījumu saraksts> }  
{ORDER BY  <kārtošanas elementu saraksts> }
```

1. Izvēlēto elementu analīze un izmantoto tabulu noteikšana
2. Savienojumu analīze
3. Nosacījumu kopas ģenerēšana
4. Grupēšana un nosacījumu būvēšana agregātfunkcijām
5. Lietotāja tiesību ierobežojumu pievienošana
6. Vaicājumu vienkāršošana un optimizācija
7. Versiju analīze

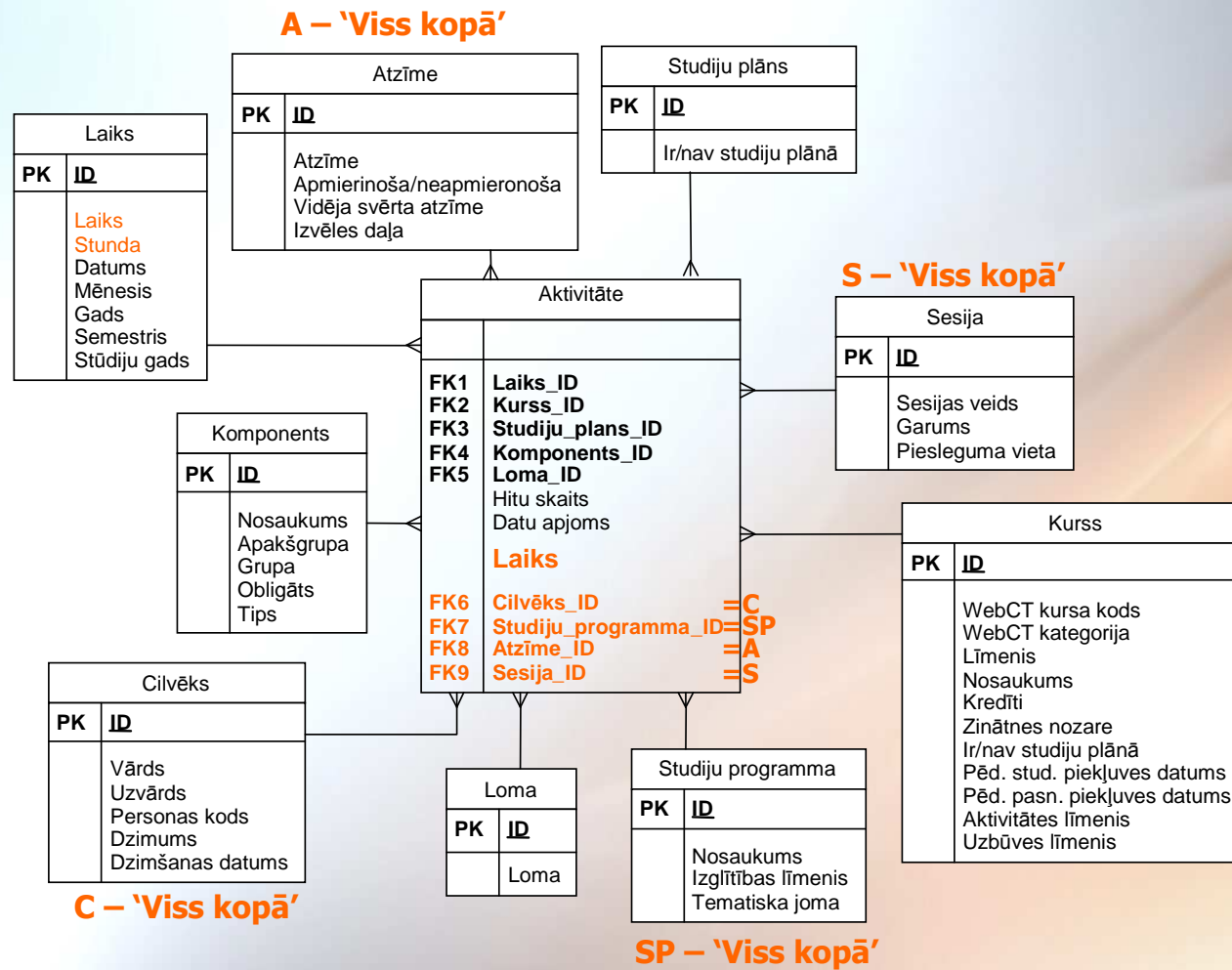
Vaicājumu veidošana – Versiju analīze

- Katram atribūtam un mērījumam tiek atrastas versijas, kad tas bija spēkā.
- Tiek atfiltrētas tas versijas, kas nebija spēkā atskaites laika periodā.
- Ja paliek tikai viena versija, tad vaicājuma rezultāti tiek attēloti saskaņā ar šo versiju.
- Ja paliek vairākas versijas, tad atkarībā no shēmas elementu piederības versijām lietotājam tiek piedāvāts:
 - Attēlot atskaiti saskaņā ar vienu no versijām
 - Attēlot elementus no vairākām versijām vienā atskaitē
 - Attēlot atskaiti atsevišķi katrai versijai

Pielietošanas piemērs



Datu bāze un fiziskie metadati



Loģiskie metadati

- 2 shēmas versijas V_1 un V_2 ;
- Nemainīgi elementi tiek piesaistīti abām versijām;
- Dimensijas Cilvēks, Sesija, Atzīme un Studiju programma un atbilstošas faktu-tabulu-dimensiju asociācijas un atribūti Laiks un Stunda ir piesaistīti tikai otrajai versijai;
- Versijā V_2 tika izveidoti jauni mērījumi "Hitu skaits 2. versija", "Datu apjoms 2. versija" un "Laiks"
- Versiju transformācijas no V_2 uz V_1 :

No	Uz	Uz objektu	Konvertācija
V_2	V_1	Hitu skaits	SUM(Hitu_skaits 2. versija)
V_2	V_1	Datu_apjoms	SUM(Datu_apjoms 2. versija)

Atskaite

- **Pasniedzēju atskaite: reižu skaits, kad studenti lietoja katru konkrēto kursa komponentu (klikšķu skaits)**
- **Rīka demonstrācija**

Turpmākais darbs

- **Atskaišu rīks:**

- Rīka ieviešana Latvijas Universitātē,
- Piedāvātas pieejas dziļākais novērtējums, balstoties uz dažiem citiem datu noliktavu projektiem,
- Testēšana un uzlabošana
 - sarežģītāko versijas transformāciju definēšana,
 - ģenerēto vaicājumu tālāka optimizācija,
 - atskaišu daudzveidības paplašināšana,
 - atskaišu personalizācija.

- **Tālākie pētījumi datu noliktavu jomā:**

- Politikas, saskaņā ar kurām var automātiski vai automatizēti veidot versijas transformācijas;
- Datu izmaiņu pētīšana (piemēram, datu semantikā)
 - Kā aprakstīt dažādu datu semantiku metadatos?
 - Kā noteikt šāda veida izmaiņas datu noliktavā?
 - Kā tās apstrādāt?
 - Vai varētu veidot datu versijas?

Literatūra

- Banerjee S., Davis, K.C. *Modeling Data Warehouse Schema Evolution over Extended Hierarchy Semantics*. Journal on Data Semantics XIII, LNCS, vol. 5530, Springer, Heidelberg, 2009, pp. 72-96.
- Bentayeb F., Favre C., Boussaid O. *A User-driven Data Warehouse Evolution Approach for Concurrent Personalized Analysis Needs*, Integrated Computer-Aided Engineering, IOS Press, Volume 15, Number 1, 2008, pp. 21-36
- Blaschka, M.: *FIESTA: A Framework for Schema Evolution in Multidimensional Databases*, PhD thesis, Technische Universität München, Germany, (2000)
- Blaschka, M., Sapia, C., Höfling, G.: *On Schema Evolution in Multidimensional Databases*. In: LNCS, Vol. 1676, Springer-Verlag, (1999) 153–164
- Bellahsene, Z.: *Schema Evolution in Data Warehouses*. Knowl. Inf. Syst. 4, 283–304 (2002)
- Body, M., Miquel, M., Bedard, Y., Tchounikine, A.: *Handling Evolutions in Multidimensional Structures*. In: Proc. of the 19th Intl. Conference on Data Engineering, Bangalore, India, (2003)
- Golfarelli, M., Lechtenböcker, J., Rizzi, S., Vossen, G.: *Schema versioning in data warehouses: Enabling cross-version querying via schema augmentation*. Data Knowl. Eng. 59(2), 435–459 (2006)
- Hurtado, C.A., Mendelzon, A.O., Vaisman, A.A.: *Maintaining Data Cubes under Dimension Updates*. In: Proc. of the 15th Intl. Conference on Data Engineering, Washington, DC, (1999) 346-357
- Kaas, C.E., Pedersen, T.B., Rasmussen, B.D.: *Schema Evolution for Stars and Snowflakes*. In: Proc. of the 6th Intl. Conference ICEIS, Porto, Portugal, (2004) 425-433
- Marotta, A.: *Data Warehouse Design and Maintenance through Schema Transformations*. Master thesis, Universidad de la República Uruguay (2000).
- McBrien, P., Poulouvasilis, A.: *Data Integration by Bi-Directional Schema Transformation Rules*. In: Proc. of 19th Intl. Conference ICDE, Bangalore, India (2003) 227–238
- Mendelzon, A.O., Vaisman, A.A.: *Temporal Queries in OLAP*. In: Proc. of the 26th Intl. Conference VLDB, Cairo, (2000) 242-253
- Morzy, T., Wrembel, R.: *On Querying Versions of Multiversion Data Warehouse*. In: Proc. of the 7th ACM Intl. Workshop DOLAP, Washington, DC, USA, (2004) 92–101
- Rundensteiner, E.A., Koeller, A., Zhang, X.: *Maintaining Data Warehouses over Changing Information Sources*. Commun. ACM. 43(6), 57–62 (2000)
- Shahzad, M.K., Nasir, J.A., Pasha, M.A.: *CEV-DW: Creation and Evolution of Versions in Data Warehouse*. Asian Journal of Information Technology. 4(10), 910–917 (2005)
- Velegrakis, Y., Miller R.J., Popa, L.: *Mapping Adaptation under Evolving Schemas*. In: Proc. of 29th Intl. Conference VLDB, Berlin, (2003)584-595
- Wrembel, R., Bebel, B.: *Metadata Management in a Multiversion Data Warehouse*. In: LNCS, Vol. 3761, Springer-Verlag, (2005) 1347-1364

The background features a soft gradient from light blue on the left to a warm orange on the right. A metallic, reflective sphere is partially visible in the upper right corner, with its surface reflecting the surrounding colors. The overall aesthetic is clean and modern.

Jautājumi?