

## **Seminario: Fundamentos de Data Warehousing**

**Objetivo:** Presentar los principales conceptos en Data Warehousing y OLAP, y los temas de investigación y problemas abiertos en el área. Interesar a los participantes a integrar proyectos de investigación en el tema.

**Formato:** La materia tendrá formato de seminario. En la primera mitad del curso se presentarán los fundamentos de OLAP y Data Warehousing. En la segunda mitad, cada participante del curso presentará un paper relevante a elegir en conjunto con la cátedra, y liderará una discusión sobre el mismo, con apoyo de la cátedra. Durante el curso se deberá realizar un proyecto, a definir con la cátedra.

**Aprobación:** Para aprobar el curso, se evaluará la presentación realizada, la participación en clase, y se requerirá la aprobación del proyecto y de un examen final y una asistencia mínima del 75%.

**Requisitos:** CC42A ó CC55A

**Profesores:** Dr. Carlos Hurtado, Dr. Alejandro Vaisman

### **Horario Propuesto:**

### **Programa:**

**1. Introducción.** Presentación de los temas a tratar. Historia y evolución de los sistemas de data warehousing. Descripción de los principales problemas en Data Warehousing y OLAP.

**2. OLAP y Data Warehousing.** Diferencias entre OLAP y OLTP. Distintas arquitecturas. Procesos: data cleaning, loading, sumarización y utilización. Implementaciones: ROLAP, MOLAP y HOLAP.

**2. Modelo de Datos.** Distintos modelos de datos en OLAP. Los modelos estrella y snowflake. Jerarquías. Formas normales multidimensionales. Diseño del Data Warehouse. **3. Procesamiento de consultas en OLAP.**

Lenguajes de consulta para OLAP. Materialización de vistas: algoritmos; selección del conjunto de vistas a materializar; distintos tipos de funciones de agregación; mantenimiento de vistas materializadas; utilización de vistas materializadas en el procesamiento de consultas. Indexación de Data Warehouses: star-joins e índices bitmap. Sumarizabilidad de datos.

**4. Sistemas comerciales y académicos.** Características y limitaciones. Comparación entre distintos sistemas comerciales. Proyectos: ADMS, H2O, DWQ.

**5. Tópicos avanzados en Data Warehousing & OLAP.** Evolución y versionado de Data Warehouses. Updates en OLAP. Temporal Data Warehousing. Data Warehousing Distribuido. Manejo de información incompleta. P2P OLAP. Calidad de Datos – Análisis de Requerimientos en Data Warehousing & OLAP. Selección de fuentes de información. Spatio-Temporal OLAP.

Bibliografía y Referencias

1. S. Agarwal, R. Agrawal, P.M. Deshpande, A. Gupta, J.F. Naughton, R. Ramakrishnan, S. Sarawagi. On the Computation of Multidimensional Aggregates. Proc. VLDB '96, 506-521.
2. Body, M., M. Miguel, Y. Bedard, A. Tchounikine. Handling Evolutions in Multidimensional Structures. ICDE 2003, Bangalore, India.
3. L. Cabibbo, R. Torlone. Querying Multidimensional Databases. Proc. 6th DBPL Workshop, 1997, 253-269.
4. C.Y. Chan, Y.E. Ioannidis. Bitmap Index Design and Evaluation. Proc. ACM SIGMOD '98, 355-366.
5. S. Chaudhuri, U. Dayal. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record 26(1), March 1999.
6. E.F. Codd, S.B. Codd, C.T. Salley. Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. Technical report, 1993.
7. Data Warehousing Online. <http://datawarehousingonline.com>
8. C. Dyreson. Information Retrieval from an Incomplete Data Cube. Proc. VLDB '96, 532-543.
9. J. Gray, A. Bosworth, A. Layman, H. Piaves. Data Cube: a Relational Aggregation Operator Generalizing Group-By, Cross-Tabs and Subtotals. Int'l Conf. on Data Engineering '96.
10. Larry Greenfield, Data Warehousing Information Center. <http://dwinfocenter.com>
11. A. Gupta, I. S. Mumick. Maintenance of Materialized Views: Problems, Techniques, and Applications. Data Engineering Bulletin, June 1995.
12. H. Gupta, I.S. Mumick. Selection of Views to Materialize Under a Maintenance-Time Constraint. International Conference on Database Theory. Jerusalem, Israel, January 1999.
13. M. Gyssens, L.V.S. Lakshmanan. A Foundation for Multi-Dimensional Databases. Proc. VLDB '97, 106-115.
14. V. Harinarayan, A. Rajaraman, J.D. Ullman. Implementing Data Cubes Efficiently. Proc. ACM SIGMOD '96, 205-216.
15. Heum-Geun Kang, Chin-Wan Chung. Exploiting Versions for On-line Data Warehouse Maintenance in MOLAP Servers. 742-753. VLDB 2002, Hong-Kong.
16. C. Hurtado, A. Mendelzon. Reasoning About Summarizability in Heterogeneous Multidimensional Schemas. Proc. ICDT'01.
17. C. Hurtado, A. Mendelzon, A. Vaisman. Maintaining Data Cubes under Dimension Updates. Proc IEEE/ICDE '99.
18. C. Hurtado, A. Mendelzon, A. Vaisman. Updating OLAP Dimensions. DOLAP'99, Kansas City, November 1999, 60-66.

19. N. Huyn. Efficient View Self-Maintenance. Proc. ACM Workshop on Materialized Views: Techniques and Applications, June 1996.
20. M. Jarke, Y. Vassiliou. Data Warehouse Quality Design: A Review of the DWQ Project. Invited Paper. Proc. 2nd Conference on Information Quality. Massachusetts Institute of Technology, Cambridge, 1997.
21. Kimball, R. The Data Warehouse Toolkit. John Wiley, 1996.
22. Kimball, R. Intelligent Enterprise. <http://rkimball.com/html/articles.html>
23. W. Lehner, J. Albrecht, H. Wedekind. Multidimensional Normal Forms. 10th International Conference on Scientific and Statistical Data Management (SSDBM'98), Capri, Italy, July 1-3, 1998.
24. N. Mamoulis, S. Bakiras, P. Kalnis. Evaluation of top-k OLAP Queries Using Aggregate R-trees, International Symposium on Spatial and Temporal Databases (SSTD), 2005.
25. A. Mendelzon, A. Vaisman. Temporal Queries in OLAP. Proc. VLDB 2000, Cairo, September 2000.
26. M. Minuto, A. Vaisman. Efficient Intensional Redefinition of Aggregation Hierarchies in Multidimensional Databases. DOLAP 2001. Atlanta, Ga. November 2001.
27. Mauricio Minuto Espil, Alejandro A. Vaisman, Leonardo Terribile. Revising Data Cubes with Exceptions: a Rule-based Perspective. DMDW, 2002
28. Mauricio Minuto Espil, Alejandro A. Vaisman. Revising Aggregation Hierarchies in OLAP: a Rule-based Approach. Data & Knowledge Engineering 45(2): 225-256, 2003
29. M. Minuto Espil, A. Vaisman. Aggregate queries in peer-to-peer OLAP. DOLAP 2004,
30. I.S. Mumick, D. Quass, B.S. Mumick. Maintenance of data cubes and summary tables in a warehouse. Proc. ACM SIGMOD '97, 100-111.
31. T. Palpanas, R. Sidle, R. Cochrane, H. Pirahesh. Incremental Maintenance for Non-Distributive Aggregate Functions. 802-813. Proc. VLDB 2002, Hong-Kong.
32. C.-S. Park, M. H. Kim, Y.-J. Lee. Rewriting OLAP Queries Using Materialized Views and Dimension Hierarchies in Data Warehouses. Proc. of ICDE'01.
33. T. Pedersen, C. S. Jensen. Multidimensional Data Modeling for Complex Data. Proc. ICDE' 99.
34. D. Quass. Maintenance Expressions for Views with Aggregation. Proc. ACM Workshop on Materialized Views: Techniques and Applications, June 1996.

35. F. Rao, L. Zhang, X. L. Yu, Y. Li, Y. Chen. Spatial Hierarchy and OLAP-Favored Search in Spatial Data Warehouse. DOLAP, 2003.
36. N. Roussopoulos, C.M. Chen, S. Kelley, A. Delis, Y. Papakonstantinou. The Maryland ADMS Project: Views R Us. IEEE Data Engineering Bulletin, 18(2):19-28, June 1995.
37. Sathe, S. Sarawagi. Intelligent Rollups in Multidimensional OLAP Data. Proc. of VLDB'01.
38. A. Shoshani, OLAP and Statistical Databases: Similarities and Differences, in Proc. ACM PODS '97, 185-196.
39. A. Shukla, P.M. Deshpande, J.F. Naughton, K. Ramasamy. Storage Estimation for Multidimensional Aggregates in the Presence of Hierarchies. Proc. VLDB '96, 522-531.
40. A.Vaisman. Data Warehousing and OLAP: a Survey. Technical Report TR 15-98. FCEN-UBA.
41. A. Vaisman. From Model 201 to Bitmap Indices. Technical Report TR 16-98. FCEN-UBA
42. A. Vaisman, A. Mendelzon. A Temporal Query Language for OLAP: Implementation and a Case Study. Proc. DBPL 2001. Rome, Italy, September 2001.
43. Alberto O. Mendelzon, Alejandro A. Vaisman. Time in Multidimensional Databases. Multidimensional Databases: Problems and Solutions, Chapter 7. Idea Group, 2003
44. Alejandro A. Vaisman, Alberto O. Mendelzon, Walter Ruaro, Sergio G. Cymerman. Supporting Dimension Updates in an OLAP Server. Information Systems. 29(2): 165-185, 2004
45. A. Vaisman. Requirements Elicitation For Decision Support Systems: a Data Quality Approach. ICEIS 2006.
46. J. Widom. Research Problems in Data Warehousing. International Conf. On Information and Knowledge Management '95.
47. A. Witkowski, S. Bellamkonda, T. Bozkaya, G. Dorman, N. Folkert, A. Gupta, L. Shen, S. Subramanian. Spreadsheets in RDBMS for OLAP. SIGMOD Conference 2003, San Diego, CA (Best Paper Award).
48. M.C. Wu, A.P. Buchmann. Research Issues in Data Warehousing. BTW'97, Ulm, March, 1997.
49. M.C. Wu, A.P. Buchmann. Encoded Bitmap Indexing for Data Warehouses. Proc. ICDE '98, 220-230.
50. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom. View Maintenance in a Warehousing Environment. Proc. ACM SIGMOD '95, 316-327.