

COMP 2003

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

- 1. DEPARTAMENTO: Computación.
- 2. CUATRIMESTRE: Segundo 2003
- 3. ASIGNATURA: Fundamentos de Data Warehousing
- 4. CARRERA: Licenciatura en Ciencias de la Computación
- 5. CARÁCTER DE LA MATERIA: Optativa
- 6. NUMERO DE CÓDIGO DE CARRERA: 18
- 7. NUMERO DE CÓDIGO DE MATERIA: C
- 8. PUNTAJE: 2
- 9. PLAN DE ESTUDIOS AÑO: 1993
- 10. DURACIÓN DE LA MATERIA: Cuatrimestral
- 11. HORAS DE CLASE SEMANAL:

a)TEÓRICAS/PRACTICAS: 4	c)PROBLEMAS	HS. -----
b)LABORATORIO: -----	d)SEMINARIOS	HS -----

- 12. CARGA HORARIA TOTAL: 4
- 13. ASIGNATURAS CORRELATIVAS: Algoritmos y Estructuras de Datos III, Algebra I, Analisis, Logica y Computabilidad, Bases de Datos .
- 14. FORMA DE EVALUACIÓN: Examen oral
- 15. PROGRAMA Y BIBLIOGRAFÍA:

FECHA:17/6/03.....



Firma del Profesor

Alejandro VAISMAN

Firma del Director

Dr. Ricardo Rodriguez
Director Adjunto
Dpto. de Computación
FCE y N - UBA

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.

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

Correlativas requeridas: Bases de datos con final aprobado antes de comenzar el curso.

Profesor: Dr. Alejandro Vaisman

Horario Propuesto: Martes 17 a 20 hs

Cantidad de clases: 12 (teórico-prácticas)

Puntaje: 2 puntos

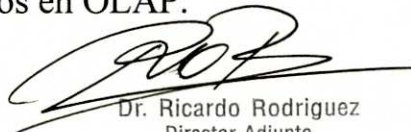
Nota: dada la característica del curso, se propone limitar el número de participantes a 15.

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 uso. 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. Calidad de datos en OLAP.



Dr. Ricardo Rodriguez
Director Adjunto
Dpto. de Computación
FCE y N - UBA


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 académicos: ADMS, H2O, DWQ.

5. Topics avanzados en Data Warehousing & OLAP. Evolución y versionado de data warehouses. Updates en OLAP. Temporal data warehousing. Data warehousing distribuído. Manejo de información incompleta.

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 and Y.E. Ioannidis, Bitmap Index Design and Evaluation, Proc. ACM SIGMOD '98, 355-366.
5. S. Chaudhuri and U. Dayal, An Overview of Data Warehousing and OLAP Technology, ACM SIGMOD Record 26(1), March 1999.
6. E.F. Codd, S.B. Codd, and 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, and H. Pirahesh, 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>


Dr. Ricardo Rodriguez
Director Adjunto
Dpto. de Computación
FCE y N - UBA

11. A. Gupta, and I. S. Mumick. Maintenance of Materialized Views: Problems, Techniques, and Applications, Data Engineering Bulletin, June 1995.
12. H. Gupta and 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 and L.V.S. Lakshmanan, A foundation for multi-dimensional databases, Proc. VLDB '97, 106-115.
14. V. Harinarayan, A. Rajaraman, and 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. In 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. Artículos varios de *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. A. Mendelzon, A. Vaisman. Temporal Queries in OLAP. In Proc. VLDB 2000, Cairo, September 2000.
25. M. Minuto, A. Vaisman. Efficient Intensional Redefinition of Aggregation Hierarchies in Multidimensional Databases. DOLAP 2001. Atlanta, Ga. November 2001.
26. I.S. Mumick, D. Quass, and B.S. Mumick, Maintenance of data cubes and summary tables in a warehouse, Proc. ACM SIGMOD '97, 100-111.
27. T. Palpanas, R. Sidle, R. Cochrane, H. Pirahesh. Incremental Maintenance for Non-Distributive Aggregate Functions. 802-813. VLDB 2002, Hong-Kong.



Dr. Ricardo Rodriguez
 Director Adjunto
 Dpto. de Computación
 FCE y N - UBA

28. C.-S. Park, M. H. Kim, and Y.-J. Lee. Rewriting OLAP Queries Using Materialized Views and Dimension Hierarchies in Data Warehouses. In Proc. of ICDE'01.
29. T. Pedersen and C. S. Jensen, Multidimensional Data Modeling for Complex Data. In Proc. ICDE' 99.
30. D. Quass. Maintenance Expressions for Views with Aggregation, Proc. ACM Workshop on Materialized Views: Techniques and Applications, June 1996.
31. N. Roussopoulos, C.M. Chen, S. Kelley, A. Delis, and Y. Papakonstantinou, The Maryland ADMS Project: Views R Us. IEEE Data Engineering Bulletin, 18(2):19-28, June 1995.
32. Sathe and S. Sarawagi. Intelligent Rollups in multidimensional OLAP data Proc. of VLDB'01.
33. A. Shoshani, OLAP and Statistical Databases: Similarities and Differences, in Proc. ACM PODS '97, 185-196.
34. A. Shukla, P.M. Deshpande, J.F. Naughton, and K. Ramasamy, Storage Estimation for Multidimensional Aggregates in the Presence of Hierarchies, Proc. VLDB '96, 522-531.
35. A.Vaisman. Data Warehousing and OLAP: A Survey. Technical Report TR 15-98. FCEN-UBA.
36. A. Vaisman. From Model 201 to Bitmap Indices. Technical Report TR 16-98. FCEN-UBA
37. A. Vaisman, A. Mendelzon. A Temporal Query Language for OLAP: Implementation and a Case Study. In Proc. DBPL 2001. Rome, Italy, September 2001.
38. J. Widom, Research Problems in Data Warehousing, Int'l Conf. on Information and Knowledge Management '95.
39. 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).
40. M.C. Wu, A.P. Buchmann, Research Issues in Data Warehousing, BTW'97, Ulm, March, 1997.
41. M.C. Wu, A.P. Buchmann, Encoded Bitmap Indexing for Data Warehouses, Proc. ICDE '98, 220-230.
42. Y. Zhuge, H. Garcia-Molina, J. Hammer, and J. Widom. View Maintenance in a Warehousing Environment, Proc. ACM SIGMOD '95, 316-327.



Dr. Ricardo Rodriguez
Director Adjunto
Dpto. de Computación
FCE y N - UBA