

Integrating Clustering Techniques and OLAP Methodologies: The ClustCube Approach

Alfredo Cuzzocrea, Paolo Serafino

ICAR-CNR, University of Calabria

Venice, Italy - June 25, 2012

Objectives:

- Mine *complex* database objects extracted from *distributed* database settings.
- Combine mining techniques (Clustering) and OLAP.
- Mine clustered objects in a *multidimensional* and *multi-resolution* fashion.

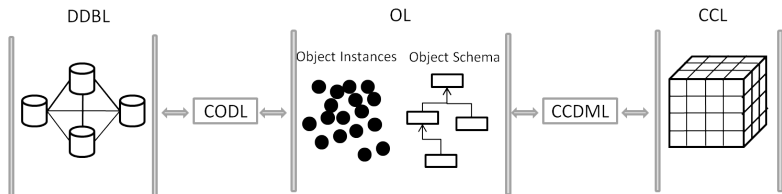
Objectives:

- Mine *complex* database objects extracted from *distributed* database settings.
- Combine mining techniques (Clustering) and OLAP.
- Mine clustered objects in a *multidimensional* and *multi-resolution* fashion.

Main Contribution:

- ClustCube, a novel computational paradigm for clustering complex database objects extracted from distributed database settings via well-understood OLAP technology.
- *Effective* and *efficient* algorithms for computing ClustCube cubes that are capable of significantly reducing computational efforts with respect to traditional approaches.

ClustCube Logic Architecture

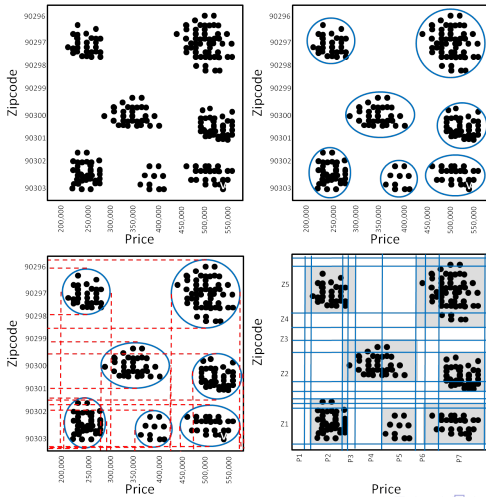


- **Distributed DataBase Layer (DDBL):** where the target distributed database is located.
- **Complex Object Definition Layer (CODL):** primitives for defining and managing complex objects.
- **Object Layer (OL):** where complex objects are located.
- **ClustCube Definition & Management Layer (CCDML):** primitives for defining and managing ClustCube cubes.
- **ClustCube Layer (CCL):** stores the final ClustCube cubes.

- ClustCube cells *store clusters of complex objects* instead of SQL-based aggregations.
- Complex objects at OL are clustered by the CCDML Layer on the basis of analysis/mining tasks defined by the administrator of the analysis/mining process.
- Objects are clustered by means of a generic clustering algorithm \mathcal{A} (depending on the characteristics of the input data).
- \mathcal{A} employs a distance function d_{CODL} defined over the fields of complex objects

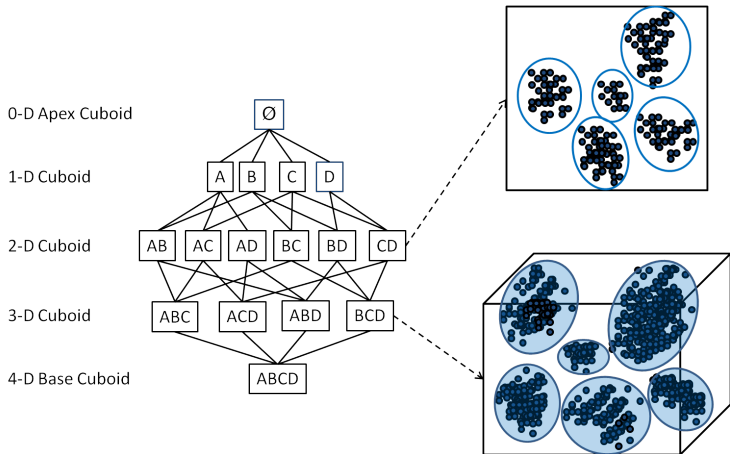
ClustCube Data Model

- Data cell C may contain a *whole cluster* or a *sub-cluster*.
- *Irregular partitions* of the target object domain.



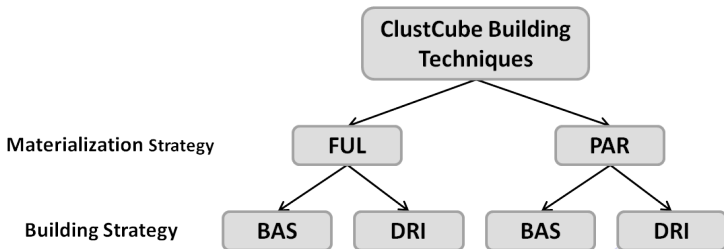
ClustCube Cuboid Lattice

- N -dimensional ClustCube cubes are equipped with a cuboid lattice \mathcal{L} , which is a hierarchical structure composed of 2^{N-1} cuboids.



ClustCube Building Techniques

- **Materialization Strategy:** which cuboids to materialize?
 - 1 **Full Materialization (FUL):** all cuboids in \mathcal{L} are materialized.
 - 2 **Partial Materialization (PAR):** a sub-set of the 2^{N-1} cuboids of \mathcal{L} is materialized.
- **Building Strategy:** how to compute cuboids?
 - 1 **Baseline (BAS):** each cuboid C_i in \mathcal{L} , is re-computed from the scratch (i.e., by applying algorithm \mathcal{A}).
 - 2 **Drill-down (DRIL):** cuboids at level l of \mathcal{L} are computed from cuboids at level $l - 1$.



Thanks for your attention!!!

Any question?