Data Catalogs

Aims:

At the end of this lecture, you should be able to understand what a data catalog is and how it can be organized and used.

Overview

- 1. Introduction to data catalogs
- 2. Catalogs for relational DBMSs
- 3. Using a catalog
- 4. System catalog in Oracle
- 5. Data dictionary

Introduction to Data Catalogs

- Data catalogs: description of databases
- Contain metadata: "data about data"
  - external schemas
  - conceptual schemas
  - internal schema
  - mappings between the schemas
  - information for specific DBMS components

Information stored in the data catalog

- For each table: table name, cardinality, file name, structure and size
- For each attribute: attribute name and type
- For each index: index name, structure and cardinality, table name, indexing attribute(s),
  - index height and range
- For each view: view name and definition (query)
- Accounting info
- Authorization info

Catalogs for Relational DBMSs

- Contain descriptions of various relational database objects
  - Relations, attributes, domains, constraints, external views, storage structures and indexes, information about authorization, security, owners of relations and similar)
- A catalog is stored as a set of read-only relations and views (system tables)

Using a catalog

- DDL and SDL compilers
- Query and DML parser
- Query and DML compiler convert high–level queries and DML statements into low–level file access commands
- Query and DML optimizer determine the best way to execute a query or a DML command by accessing the internal schema
- Authorization and security checking is done by the DBA who has special privileged commands to update the catalog
- External–to–conceptual mapping of queries and DML commands is done by accessing the catalog description of the view

System catalog in Oracle

- The catalog consists of
  - Base tables
  - User-accessible views
    - USER views (Objects owned by user)
    - ALL views (Objects to which user has privileges)
    - DBA views (All objects)

Conceptual Schema Information

- ALL_CATALOG(owner, table_name, table_type)
- ALL_OBJECTS(object_name, object_id, object_type, created, . . .)
- USER_CATALOG(table_name, table_type)
- USER_TAB_COLUMNS(table_name, column_name, data_type, . . . , nullable, . . . , num_distinct, low_value, high_value, density, num_nulls, . . . , last_analyzed, global_stats, user_stats, avg_col_length)
- USER_CONSTRAINTS(owner, table_name, constraint_name, constraint_type, search_condition, . . .)

Example QueryY of Metadata

- SELECT *
  FROM ALL_CATALOG
  WHERE OWNER='TANJA';

Might lead to further queries and analysis:

- ANALYZE TABLE movie
  COMPUTE STATISTICS;

Schema Information

- External schema information:
  - USER_VIEWS(view_name, text_length, text, . . .)
- Internal schema information:
  - USER_TABLES(table_name, tablespace name, . . . , pct_free, initial_extent, num_rows, blocks, empty_blocks, avg_row_length)
  - USER_INDEXES(index_name, index_type, table_owner, table_name, table_type, uniqueness, compression, . . . , blevel, leaf blocks, distinct_keys, avg_leaf_blocks_per_key, avg_data_blocks_per_key)

Example query: SELECT * from USER_TABLES;