

CMSC724: Extensible Systems

Amol Deshpande

University of Maryland, College Park

March 13, 2008

Extensible Systems

- Need to support new data types/operators
 - GIS, CAD, document, time-series data, image databases
 - ...
 - More recently: XML
 - Simple XML can be easily supported using OR facilities
 - See “What goes around comes around”; Chapter 1, Redbook
- Trying to fit everything into SQL can be tricky
 - Unintuitive
 - Hard to write queries
 - Need support for new types of queries
 - Can't optimize or index etc..

Extensible Systems

- Object-oriented vs Object-relational
- Object-oriented
 - Shrink the impedance mismatch for app programmers
 - Semantically richer data models
 - New language features (e.g. complex programming)
 - Niche Market
- Object-relational
 - Storage and querying over complex data types

	Simple Data	Complex Data
Query	RDBMS	ORDBMS
No Query	File System	OODBMS

Postgres

- INGRES-ADT → Post-INGRES
- Query language: “quel”
- Ability to add new data type. Type defined by:
 - storage size
 - input and output methods (functions to convert to-and-from strings)
 - other methods
- Ability to add new operators
 - input/output types, precedence etc. . .
- Methods called via “function pointers”
- Dynamic linking vs static linking
- Trusted code ???

Postgres

- PostgreSQL
 - Extensibility in PostgreSQL
 - Types
 - User-defined Functions
 - Indexes and GiST

- Access Methods
 - B+-Trees can be used for any object that supports a set of functions
 - Similarly, hashing can be used if $H(\text{key})$ is supported
 - Requirements, different data types, selectivities all stored in relations
 - GiST !
- Interactions with transaction management
 - The hardest part for access methods
 - Recall that concurrency/recovery highly customized for B+-Trees
 - Somewhat okay if you only physical logging
 - Otherwise, can expose some of the API

- Query Processing
 - must be able to compute selectivities
 - need to know if we can use sort-merge join, hash join
 - figure out which access methods can be used for executing a query
 - rules to specify these things
 - fairly easy to make dynamic programming (Selinger et al) work with it

ORDBMS

- Postgres → Illustra → Informix → IBM
 - [Datablades](#): Packages that supported everything for a new data type
 - From IBM webpage: C-ISAM, Text, Geospatial data, Image data, Spatial data, Time-series, Video, Web
- Postgres → PostgreSQL (around 1995)
 - Named so because it supported SQL (as opposed to only Quel)
 - Postgres was BSD license
- Very successful
- Most database systems support it in one form or other
 - SQL:99 has support for it