PowerDB: A Coordinator for a Database Cluster

OHO Lecture
Torsten Grabs
ETH Zurich
The PowerDB Approach

- Cluster of databases
  - Cluster of off-the-shelf PCs
  - Each PC runs a commercially available RDBMS
  - Fast Ethernet Connection (100 Mbit/s)

- The “Scale-out” vision
  - Adding new nodes for higher performance
  - More nodes allow to increase parallelism
The PowerDB architecture

Invoke

Application Service
parallel

Invoke

Application Service
parallel

Invoke

Application Service
parallel

Invoke

Application Service
parallel

SQL

DBMS
DB

SQL

DBMS
DB

SQL

DBMS
DB

SQL

DBMS
DB

Invoke

Application Service
parallel
Transaction management

Alternatives to implement transactional guarantees:
- two-phase commit (2PC) distributed transactions
- composite transaction management
The PowerDB cluster

128 Nodes

1 GHz P-III CPU
256 MB Memory
100 MBit Network
2x 18 GB SCSI
Windows 2000
SQL Server 2000
The PowerDB-XML approach

- XML Retrieval
- XML Deletion
- XML Insertion
- XML Retrieval

Frontend

parallel

Coordinator

parallel

parallel

parallel

parallel

Side Table Executor

Document Text Executor

IR Executor

Side Table Executor

Document Text Executor

IR Executor

DBMS

DB

DB
Physical data organization

- Document and materializations at same node
- Physical data organization: node groups
  - Category assignment
  - Node group striping
  - Node group replication

Documents:

- C1: T1, T2, T3, T4
- C2: A1, A2, A3

Node groups:

- Node group 1: DBS1, DBS2
  - T1, T3
  - T2, T4
  - Striping

- Node group 2: DBS3, DBS4
  - T1, T3
  - T2, T4
  - Striping

- Node group 3: DBS5, DBS6
  - A1, A3
  - A2
  - Striping

Replicas of category 1:

- T1, T3

Category 2:

- T2, T4
Scalability with striping

Scalability with workload (30,30)

Throughput (requests per sec) vs. cluster size for VS retrieval (30,30) and VS insertion (30,30).
Data warehouse scenario

- Problems we looked at:
  - "up-to-dateness" of data?
  - performance
PowerDB OLAP approach

OLTP Clients

Cluster of Databases

OLAP Clients

OLTP requests

Coordination Middleware

OLAP requests

"OLTP node"

"OLAP nodes"

Insert
Delete
Update

OLTP requests

Notification/ Activation

refresh tx.

OLAP queries
update-subtransaction $t_{1CO}$ commits on OLTP node $c_o$

first query of a read-only transaction $t_3$ with freshness limit 1 is scheduled

C₁ must first reach this ’up-to-date state’ of C₀

=> refresh-subtransaction for C₁ start first

serialization order: $t_1 \rightarrow t_3 \rightarrow t_2 \rightarrow t_4$
Trading freshness for MRT (TPC-R)

FAS allows to trade freshness of data for query response time; Freshness 1 yields about 60% higher MRT.
Future work for students

• Results so far are convincing
  – IR support by DB cluster
  – XML services on DB cluster
  – New method of replication management and freshness
  – Cache Aware Routing

• DA/SAs: XML document management
  – PowerDB-XML Control Center
  – Computing vector-spaces for IR on-the-fly
  – Contact: Torsten Grabs <grabs@inf.ethz.ch>

• DA/SAs: OLTP/OLAP
  – Replication and partitioning
  – Second layer optimizer
  – Physical db cluster design wizard
  – Contact:
    • Fuat Akal <akal@inf.ethz.ch>
    • Dr. Uwe Röhm <roehm@inf.ethz.ch>


