Seismo Project

ADT team 2010

- Goal formulation
- Database schema
- Loading mechanism
- Query processing
- Application interface

- Challenges and open issues

Goal formulation

- To create a full functional data delivery system for earthquake event data
- To demonstrate we can load TB of data efficiently in the DBMS
- To evaluate the queries for indexing, summarization
- How to facilitate datamining in the DBMS

Database schema

- Canonical representation
- Partial time series

MSEED reference

```
-- all records in the mseed files correspond to a row in the catalog
CREATE TABLE mseedCatalog (
Mseed       int,                -- Vault file id
Seqno       int,                -- SEED record sequence number, should be between 0 and 999999
        PRIMARY KEY (mseed,seqno),
dataquality char,               -- Data record indicator, should be 'D=data unknown qual',
                                -- 'R=raw no quality', 'Q= quality controlled' or 'M'
Network     varchar(11),        -- Network
Station     varchar(11),        -- Station
Location    varchar(11),        -- Location
Channel     varchar(11),        -- Channel
starttime   timestamp,          -- Record start time, the time of the first sample, as a high precision epoch
    time
Samplerate  double,             -- Nominal sample rate (Hz)
Samplecnt   int,                -- Number of samples in record
Sampletype      string,         -- storage type in mseed record
Minval      float,              -- statistics for search later
Maxval      float
);
```

```
-- mseed data volumns may appear in different formats
-- we try to postpone them, assuming the optimizer can
   guide JIT.
CREATE TABLE mseedVolumn (
time   timestamp,
mseed     int,
adata  varchar(20),
idata   int,
fdata   float,
ddata double
);
```

Alternatives

- Controlling the database size
  - Hierarchical organisation of chunks
  - Vertical partitioning by station
  - Single consolidated time line

  - Represent a limited domain as strings
  - Expand the data to full scale

Loading

- Bulk loading
- Lazy loading

Query processing

- Detect possible contraint violations
- Q1 Select all events bounded by time and space (a subcube)
- Q2 select all related events by wave propagation prediction
- Q3 find gaps
- Q4 shifting patterns, deviances from earlier propagation experiences

A shaking world movie

- Show on a map how a wave propagates and how it is picked up by remote sensors

Challenges and open issues

- Bulk loading with temporal merge
- Controlling the size or not?