**Just-In-Time Data Virtualization: Lightweight Data Management with ViDa**
Manos Karpathiotakis, Ioannis Alagiannis, Thomas Heinis, Miguel Branco, Anastasia Ailamaki

### Big Data blocks Data Analysis
- Exponential growth of data volume
- Increasing data variety
- Data movement regulations

**Goal:**
Make data analysis independent of data properties (size, format, location, etc.)

### The ViDa Approach: Data Virtualization

**ViDa Architecture**

**Query Language:**
- Monoid comprehensions
- Can query / transform
  - Tables
  - Hierarchies
  - Arrays

```sql
SELECT r.age
FROM Patients p
JOIN BrainRegions r
ON (p.id = r.id)
WHERE r.amygdala.Vol > 0.2
```

**Just-In-Time Executor:**
- Adapt access paths to data
- Adapt engine to incoming queries
- Adapt internal structures to workload

**Execution engine generated at runtime:**
Generate “hard-coded” operators

**ViDa in Action: Discovering Disease Signatures**

**Human Brain Project:**
**ViDa vs. State-of-the-art**

- Size: > 40000 patients
- Formats: CSV & JSON
- Queries: 150 x 3way joins

**Challenge:**
Physical integration & diverse queries

**Competitive with state-of-the-art by**
- Not loading / transforming data
- Adapting to raw files and queries
- Exploiting locality