

# OLAP, *easy* to use

Brian J. Cardiff

# Application Data

- nice, a lot
- normalized
- indices

# Desire

- Know what our data says?
  - Average, Total, Count
  - Categorize them
  - Among time
- Print or Web pages
- Inspect by iterative navigation

# Nowadays

- Create some view at app
- Write custom ugly filtering controls
- Draw tables, export .csv
- Nice graphics
- and ... wait until de page loads :-(
- New report? Start all over. Time lost :-(

OLAP time

# Quickly: how to use it

- Get an OLAP server
- What are the aspects you are interested?
- Simplify and tidy up your data
- Setup a cube (yes, a CUBE)
- Play

# Demo: Patients

- Our nice app schema
  - **Patients** :: Name, Gender, Age, Registration Date, Location
  - **Visits** :: Patient\_id, Date
  - **Status** :: Visit\_id, Status

# Interested in ...

- There are patients
- A patient is active, dead, lost or transferred
- The location is important
- We will want to study age distribution

# Not interested in ...

- Name
- Visits (but yes their status)

# Data simplified

- Patients
  - Gender, Age, Registration Date,
  - Location,
  - Status, Status Date
- Making Active status explicit instead of implicit

# OLAP terms

- We analyze facts (patients)
- Setup dimension i.e. from which point we will want to view the facts
- Setup measures i.e. how facts will be appreciated: count, avg(age)

Let's play a bit

# We just notice

- Dimensions are awesome
  - Ranged dimension: Age
  - Tree structure: Geo, Time
- End users can use it, if they know excel

# Why this work?

- Pre-calculated measures
- Per each coordinate
- defined by dimension
- over the facts table
  
- It doesn't depends on #patients !

# Walkthrough

- Data source
- Dimensions, tables for dimensions
- Bind fact table to dimensions
- Deploy, load cube

# I'm a programmer

- Remember T-SQL for Relational DBs?
- SAME SAME, BUT DIFFERENT
- Welcome MDX for OLAP DBs
  - MultiDimensional eXpressions

# MDX

SELECT

[Age] ON COLUMNS,

[Gender] ON ROWS

FROM [Patients]

# MDX

SELECT

CROSSJOIN(

{[Measures].[AvgAge]} ,

[Age].[Adult].children) ON COLUMNS,

[Gender].children ON ROWS

FROM [Patients]

# Kit

- Google Refine (sanitization, big picture)
- icCube (community server)
- MDX Introduction <http://goo.gl/ICJDU>
- Excel (awesome pivot tables)
- Wabbit (reporting)
- libs: mondrian, etc.

# Pros & Cons

- Appropriate technology
  - Faster, Flexible
  - More things to learn
  - ...
- ETL process. Incremental?
  - Preprocessing time
  - More things to learn
  - ...

Questions? Thoughts?

# Thanks!

Brian J. Cardiff

@bcardiff