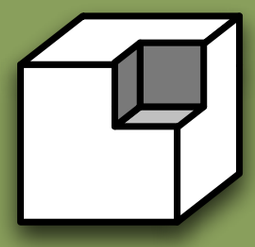


Cubes

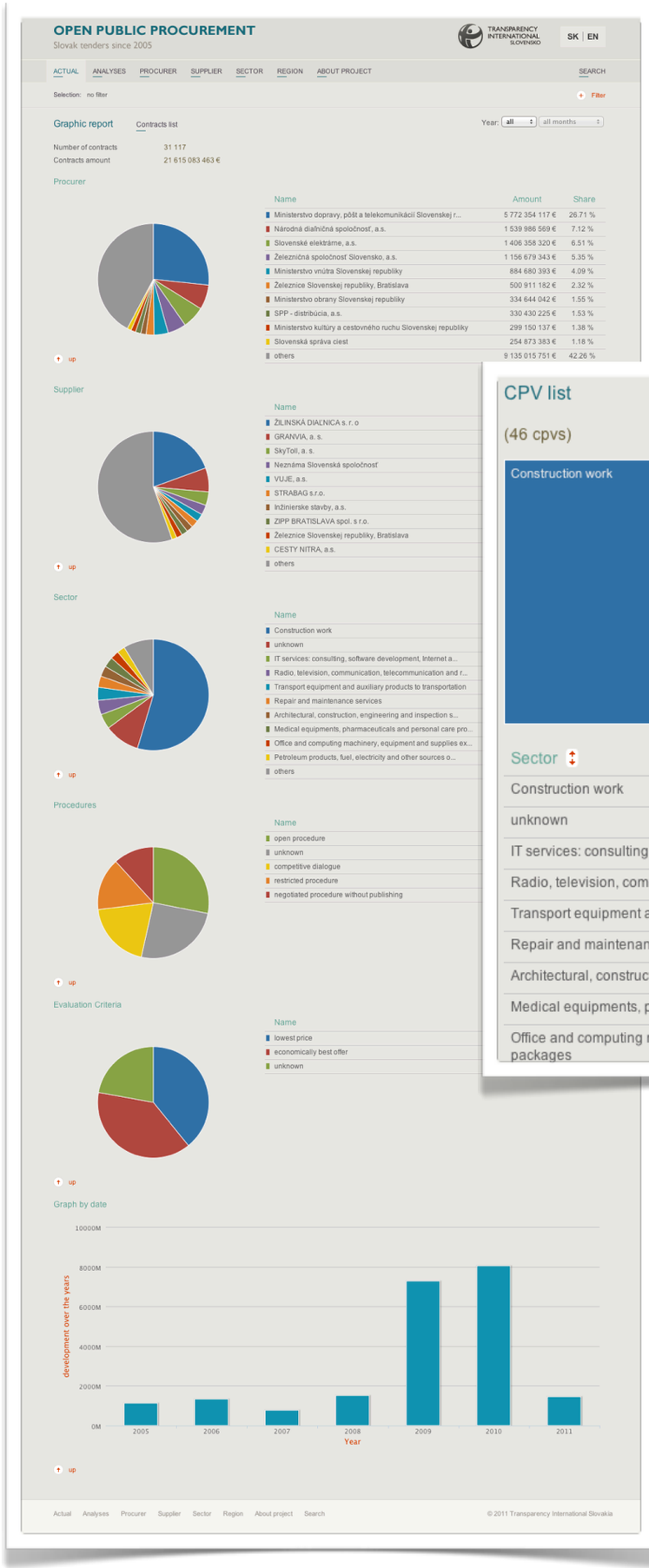
light-weight OLAP

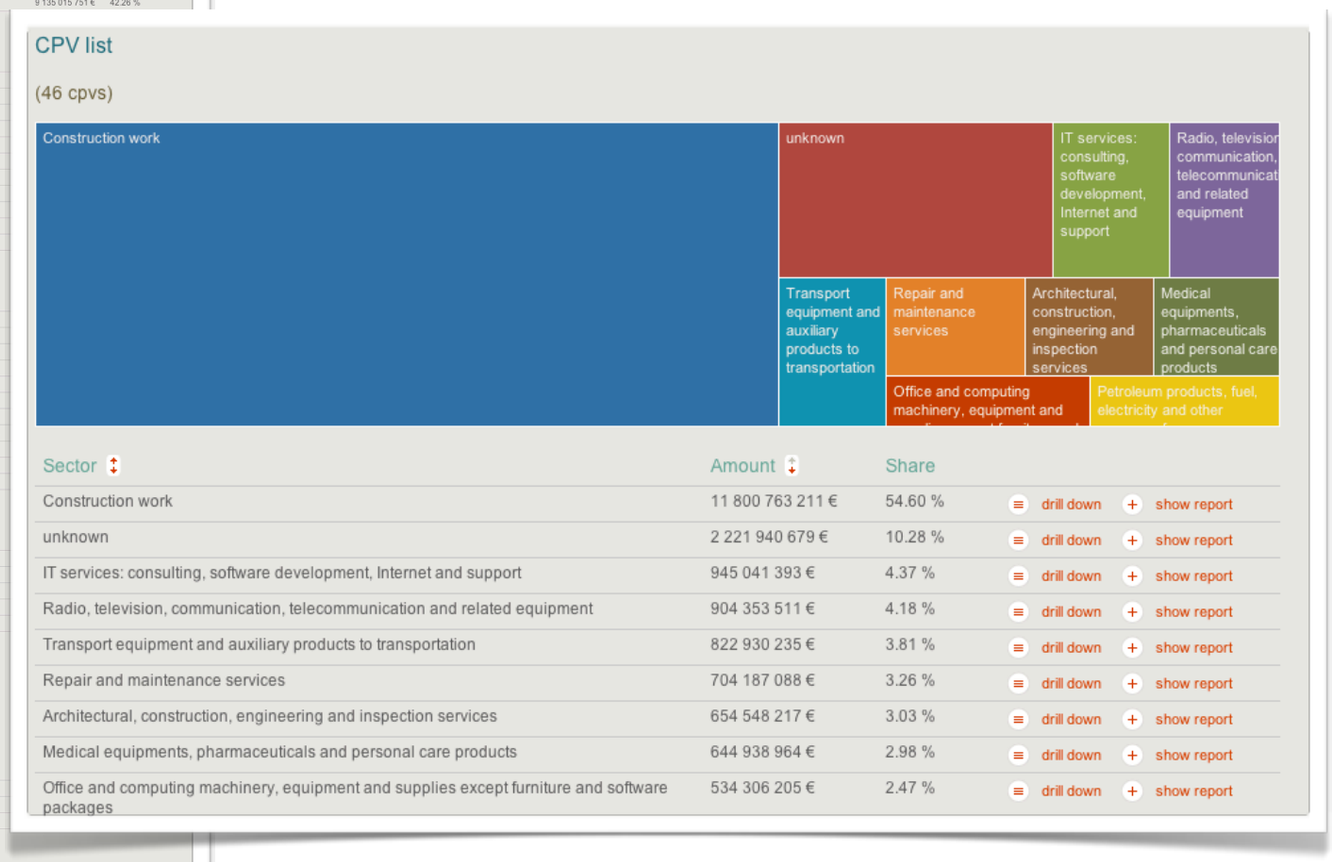
Stefan Urbanek ▪ @Stiivi ▪ stefan.urbanek@gmail.com ▪ July 2012

Overview

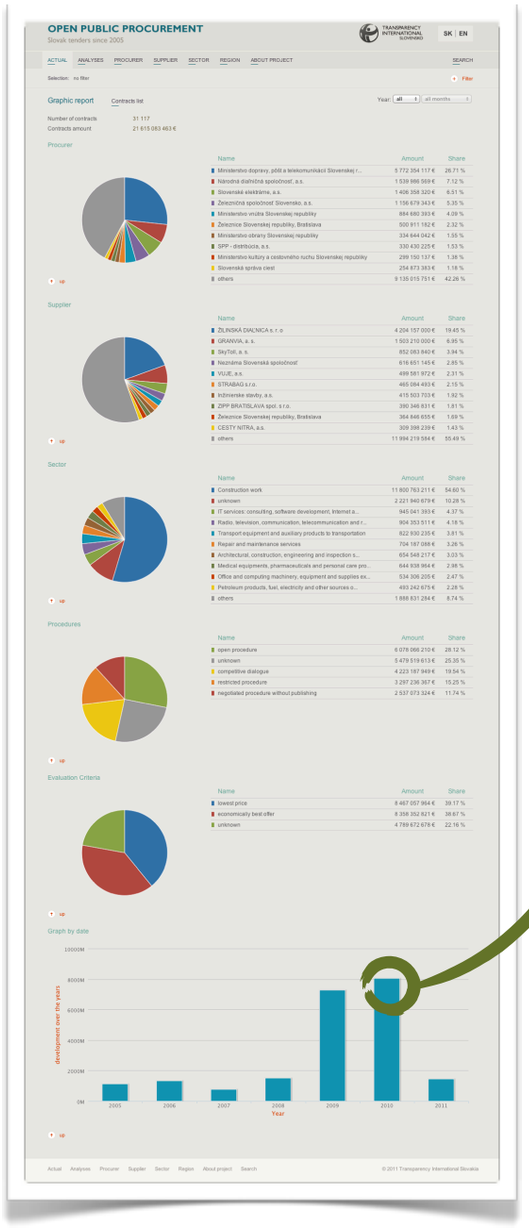
- purpose
- analytical modelling and OLAP
- slicing and dicing
- OLAP server
- SQL backend

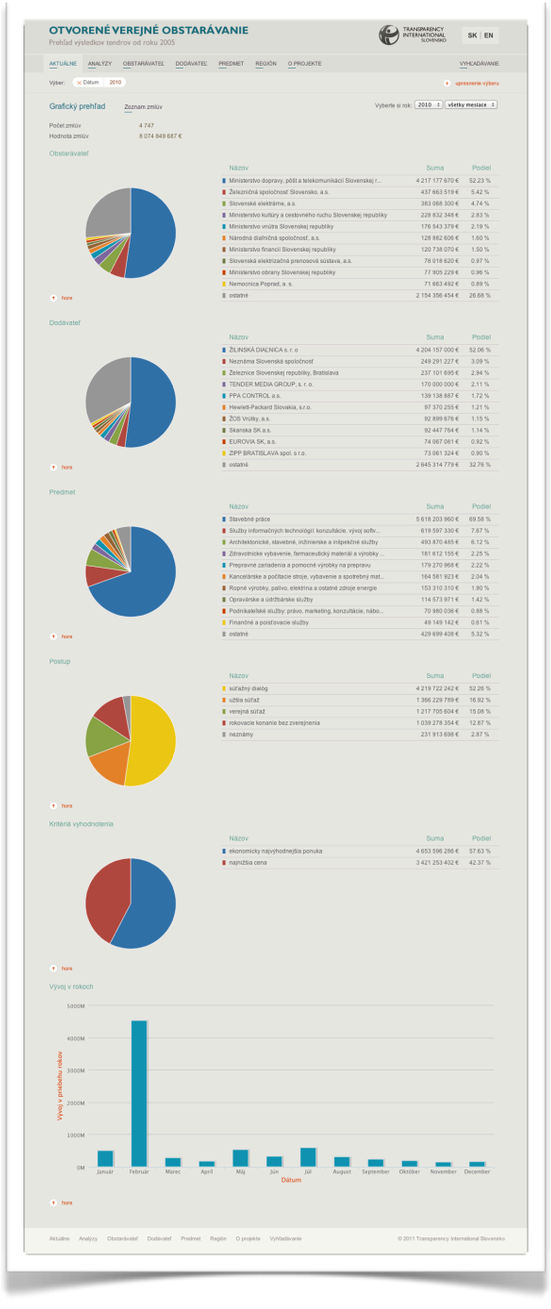
analytical data modelling lightweight





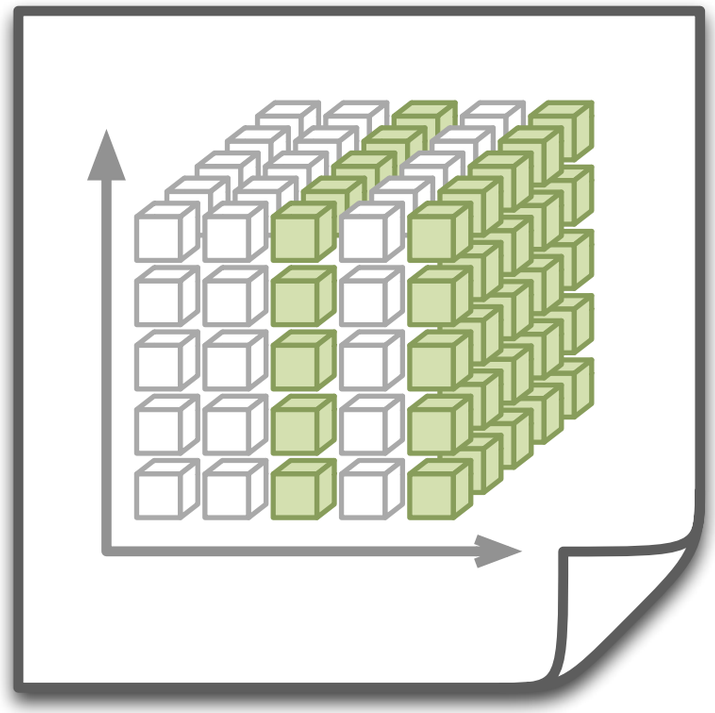
http://tendre.sme.sk

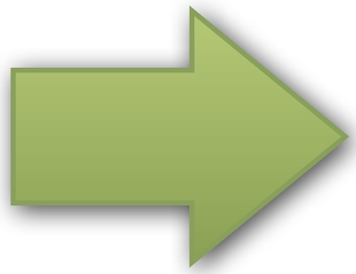


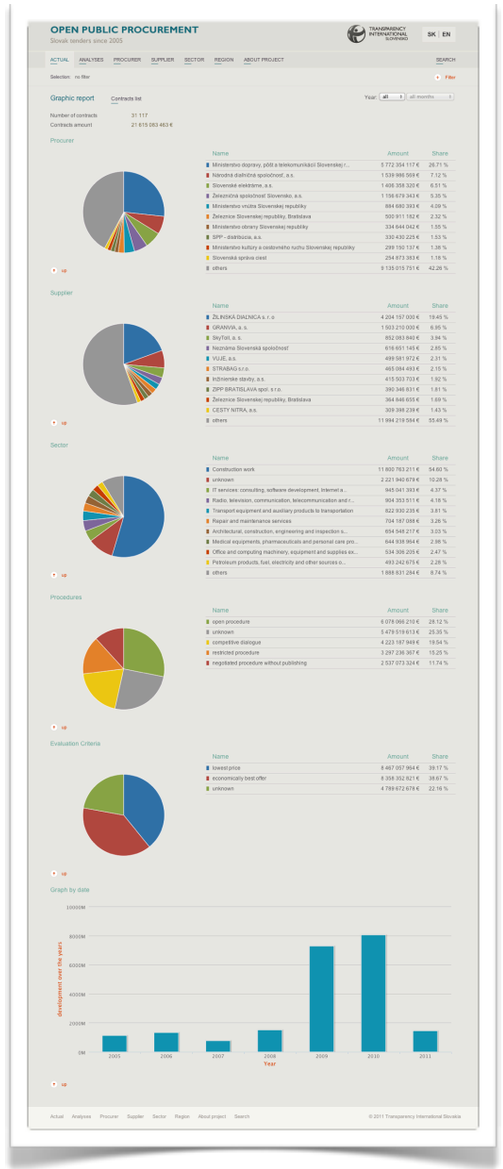


aggregation browsing

slicing and dicing





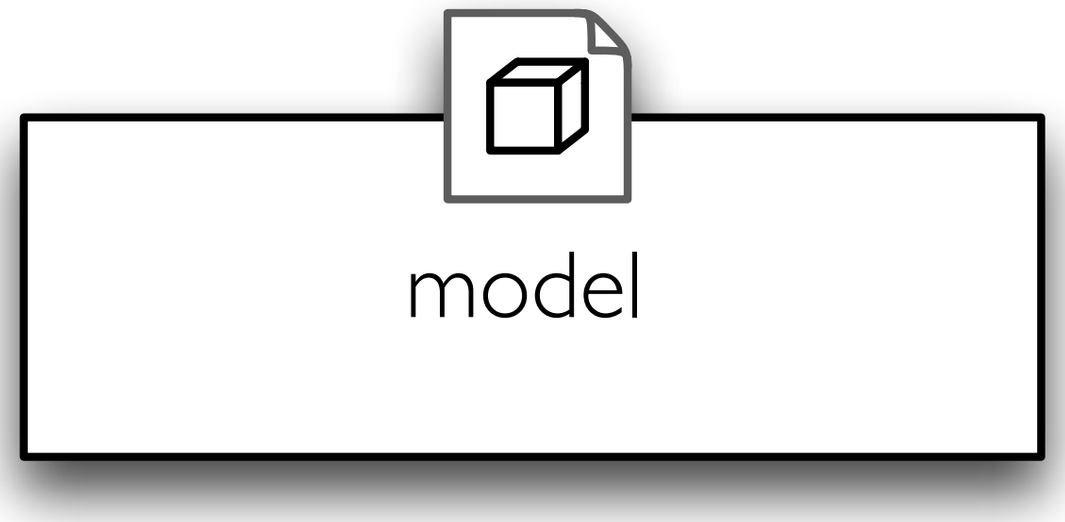


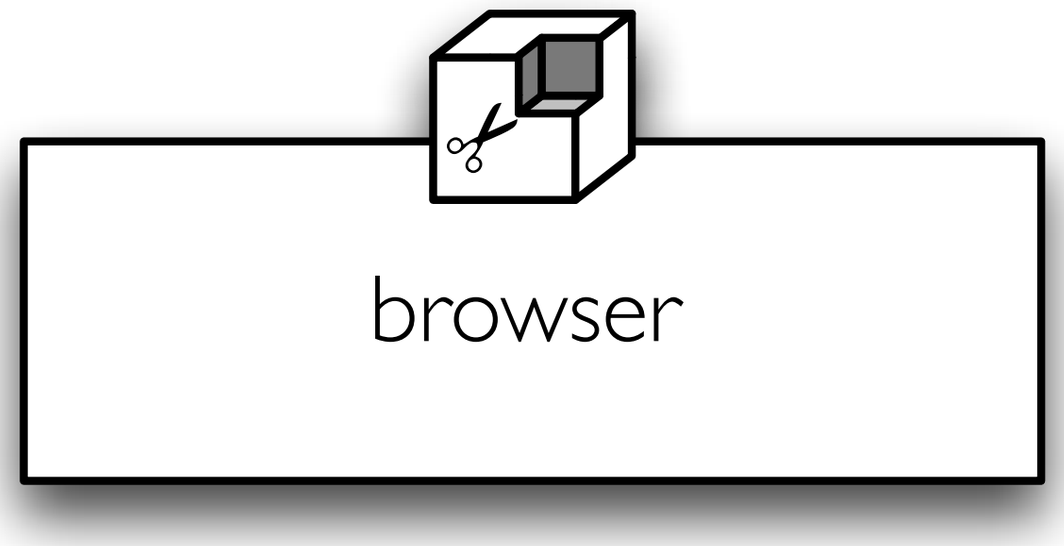
modelling

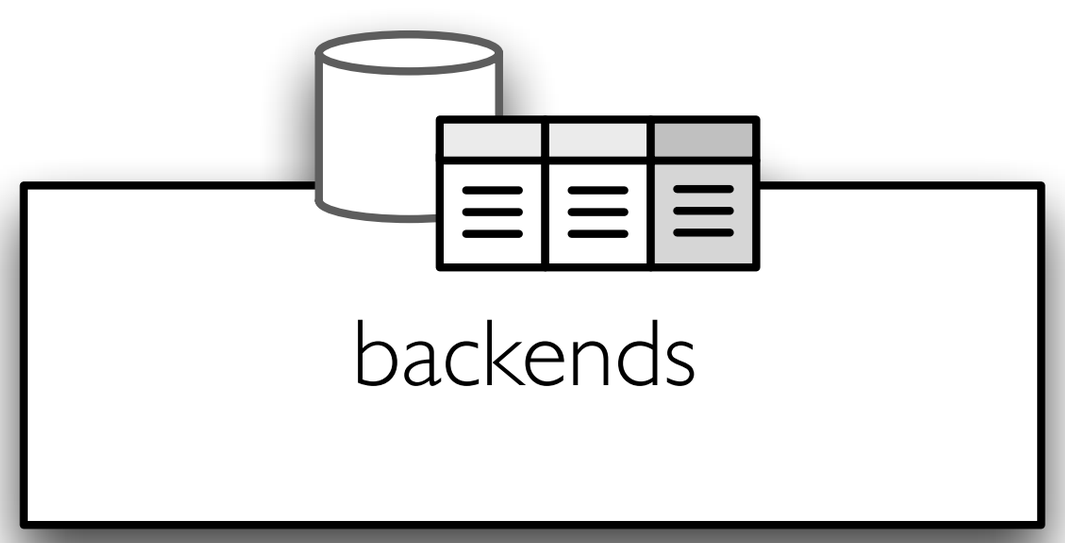
reporting

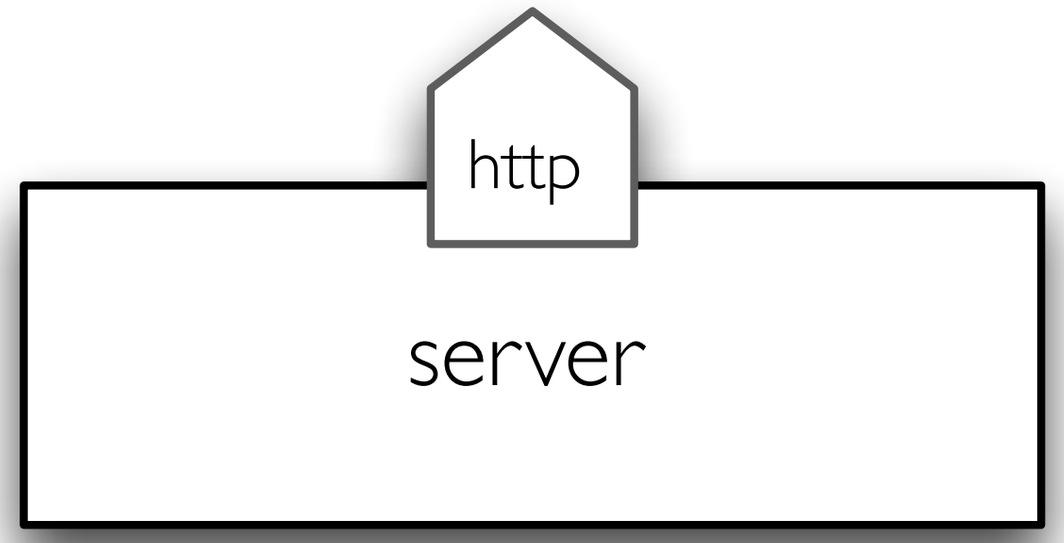
aggregation browsing

Architecture


model

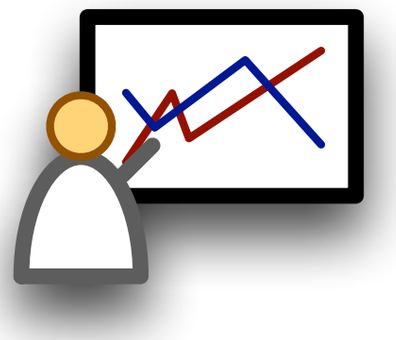

browser


backends

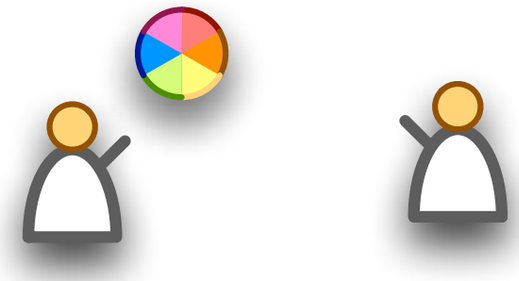

http
server

Logical Model

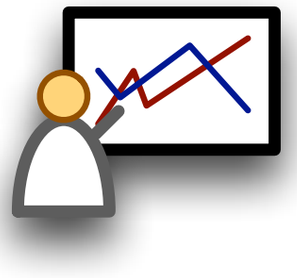
multidimensional, analytical



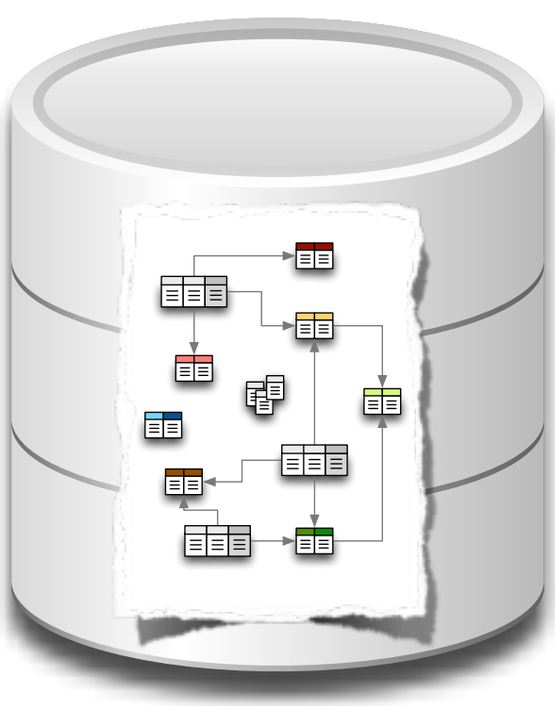
business/analyst's point of view

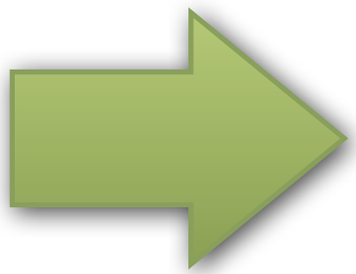


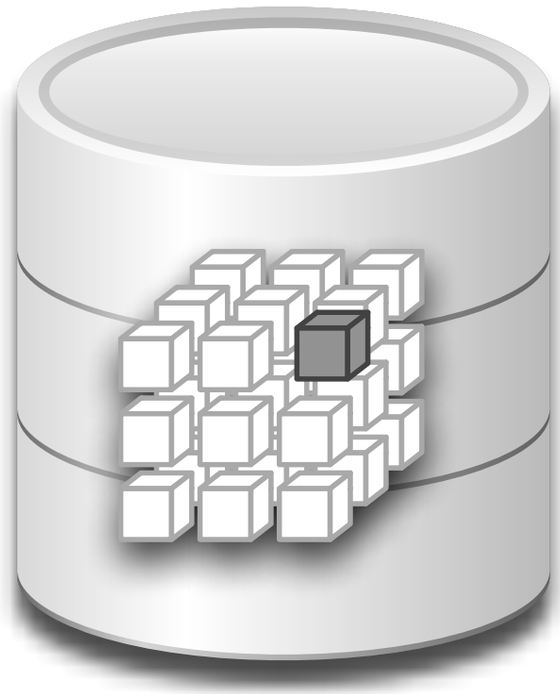
transactions
OLTP



analysis
OLAP





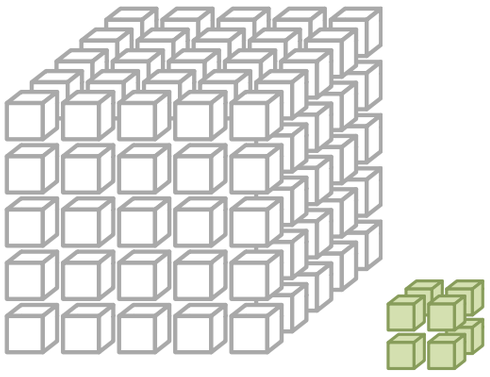


application (operational) data

analytical data

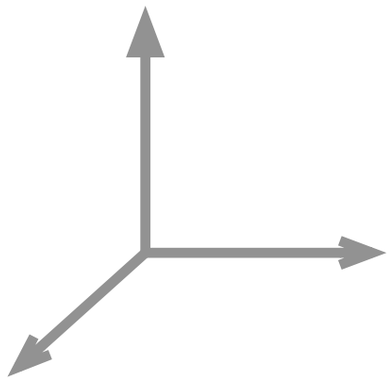
Model

```
{
  “name” = “My Model”
  “description” = ....

  “cubes” = [...]
  “dimensions” = [...]
}
```



cubes

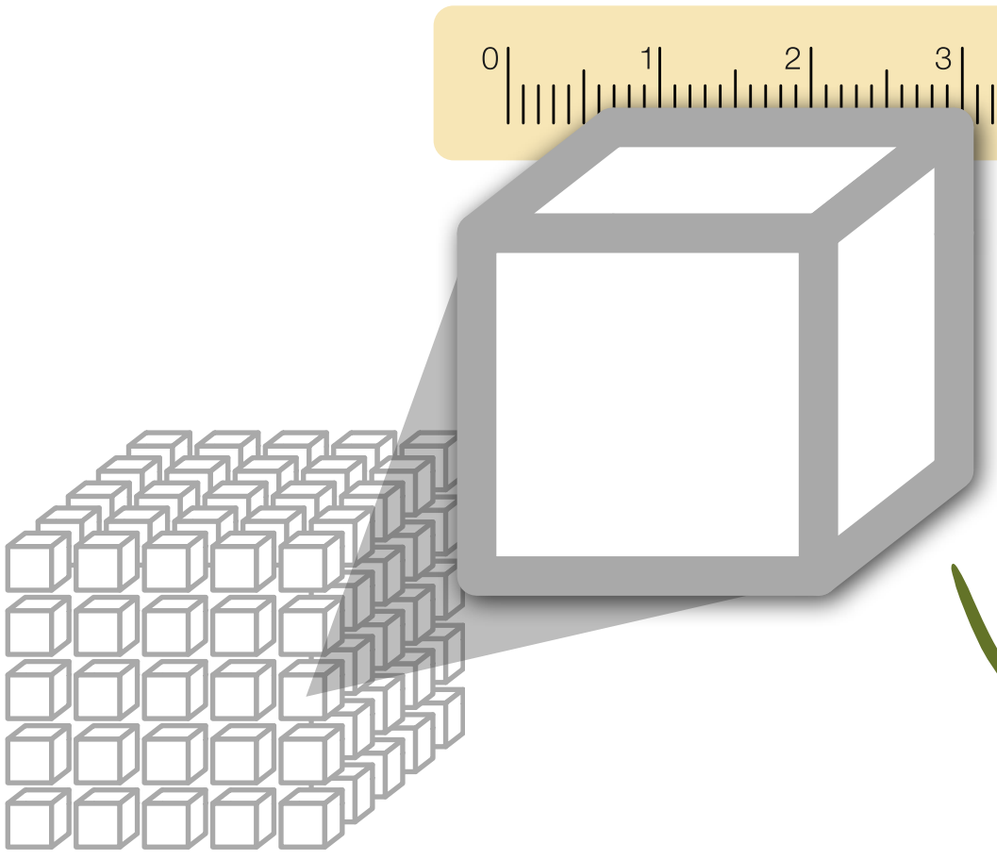
measures



dimensions

levels, attributes, hierarchy

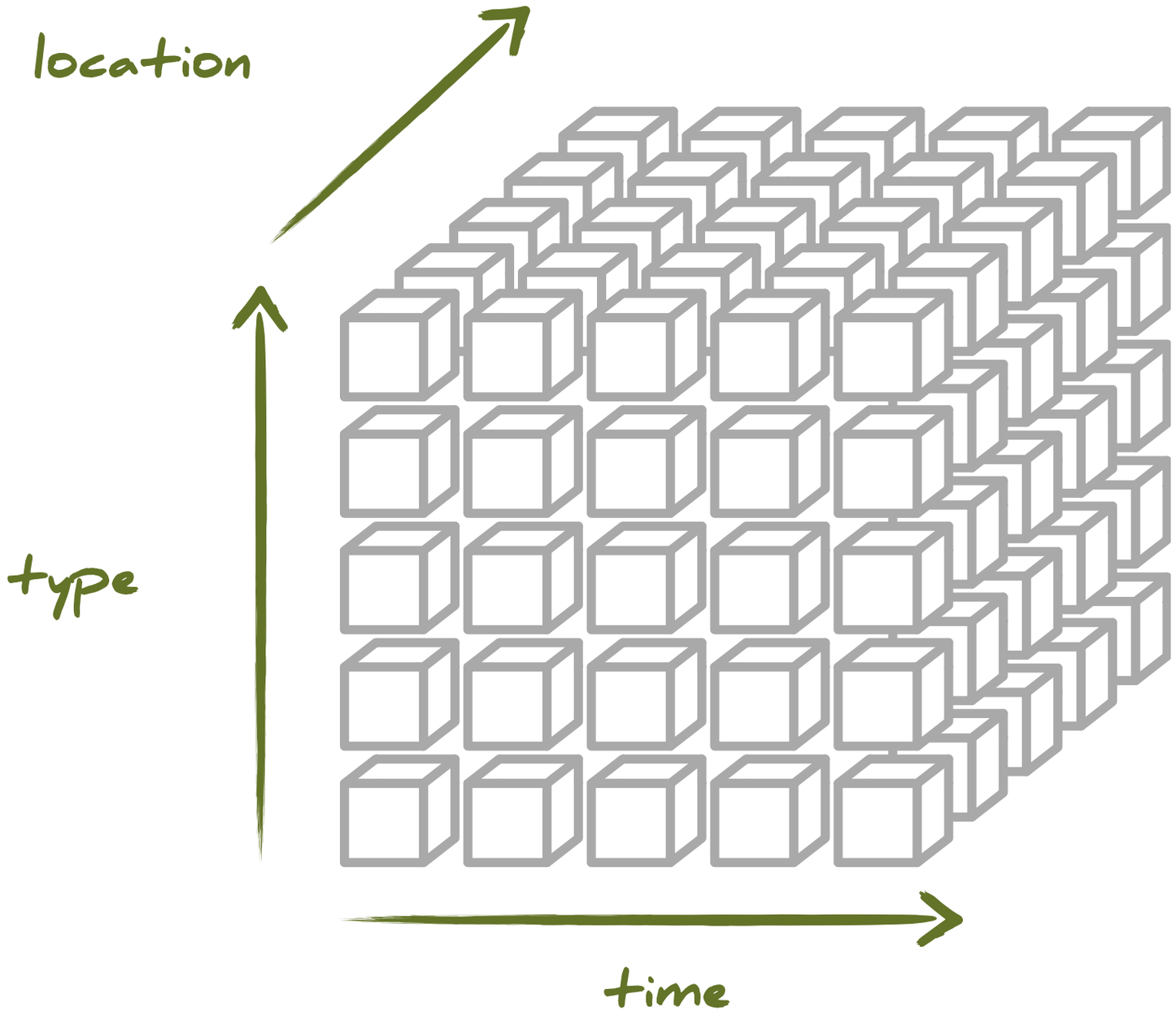
Facts



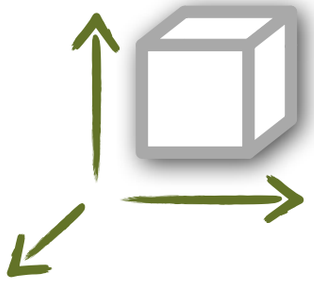
measurable

fact data cell

most detailed information



dimensions



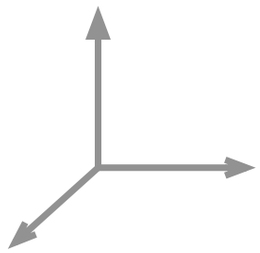
Dimension

- provide **context** for facts
- used to **filter** queries or reports
- control **scope of aggregation** of facts

Hierarchy

2010 MAY 1ST

levels



Dimension

- levels and attributes
- hierarchy*
- key attributes
- label attributes

```
“dimensions” = [
  {
    “name”:”date”,
    “levels”: ...
    “hierarchy”: ...
  },
  ...
]
```

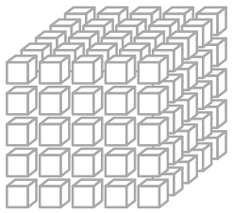
*partial support for multiple hierarchies

Sector	Amount	Share		
Construction work	11 800 763 211 €	54.60 %	drill down	show report
unknown	2 221 940 679 €	10.28 %	drill down	show report
IT services: consulting, software development, Internet and support	945 041 393 €	4.37 %	drill down	show report
Radio, television, communication, telecommunication and related equipment	904 353 511 €	4.18 %	drill down	show report
Transport equipment and auxiliary products to transportation	822 930 235 €	3.81 %	drill down	show report
Repair and maintenance services	704 187 088 €	3.26 %	drill down	show report
Architectural, construction, engineering and inspection services	654 548 217 €	3.03 %	drill down	show report
Medical equipments, pharmaceuticals and personal care products	644 938 964 €	2.98 %	drill down	show report
Office and computing machinery, equipment and supplies except furniture and software packages	534 306 205 €	2.47 %	drill down	show report

label attribute

key attribute

for links to slices



Cube

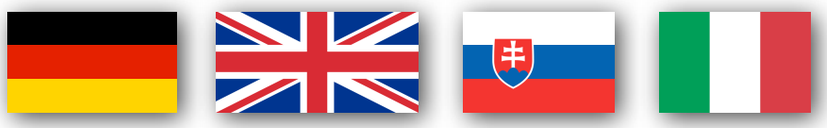
- dimensions
- measures

```
“cubes” = [
  {
    “name”:”contracts”,
    “dimensions”: [ “date”,
                    “category” ]
    “measures”: [
      {
        “name”: “amount”,
        “label”: “Contract Amount”,
        “aggregations”: [“sum”]
      }
    ]
  },
  ...
]
```

*partial support for multiple hierarchies

localizable

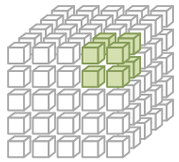
model and attributes



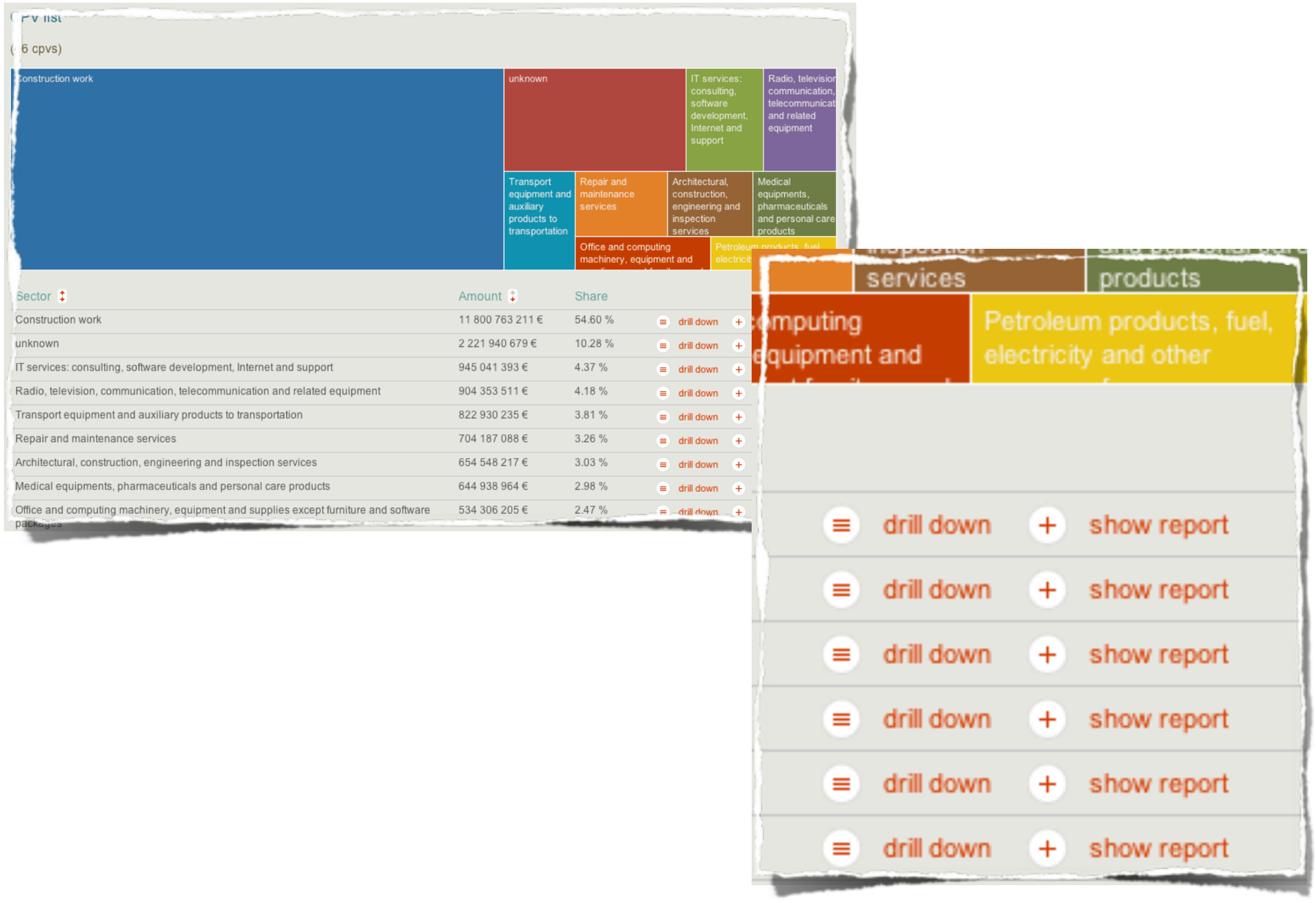
```
"attributes": [
    {
     "name":"group",
     "label": "Group code"
    },
    {
     "name":"group_label",
     "label": "Group",
     "locales": ["en", "sk"]
    }
]
```

Aggregation Browser

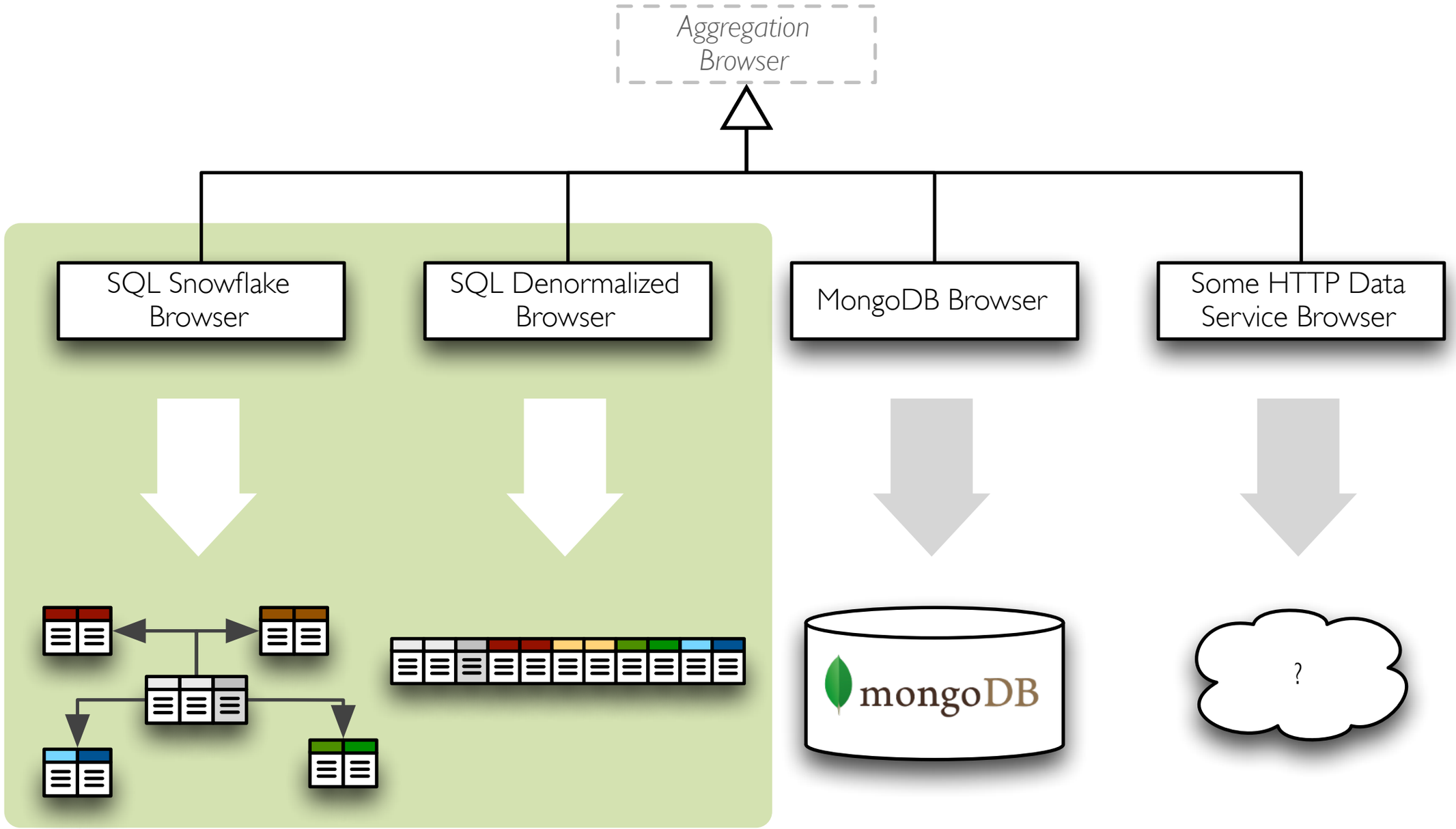
∑ measures



Name	Amount	Share
Construction work	11 800 763 211 €	54.60 %
unknown	2 221 940 679 €	10.28 %
IT services: consulting, software development, Internet a...	945 041 393 €	4.37 %
Radio, television, communication, telecommunication and r...	904 353 511 €	4.18 %
Transport equipment and auxiliary products to transportation	822 930 235 €	3.81 %
Repair and maintenance services	704 187 088 €	3.26 %
Architectural, construction, engineering and inspection s...	654 548 217 €	3.03 %
Medical equipments, pharmaceuticals and personal care pro...	644 938 964 €	2.98 %
Office and computing machinery, equipment and supplies ex...	534 306 205 €	2.47 %
Petroleum products, fuel, electricity and other sources o...	493 242 675 €	2.28 %
others	1 888 831 284 €	8.74 %

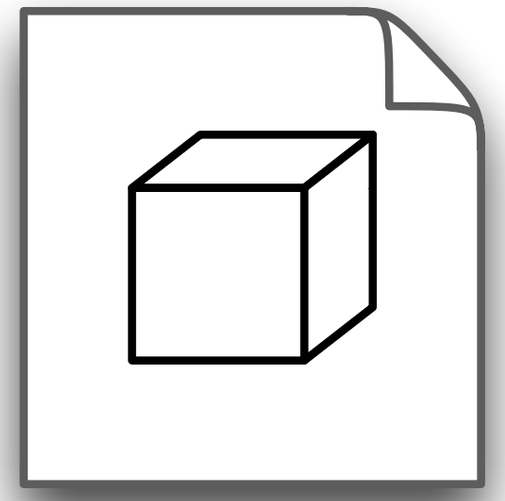


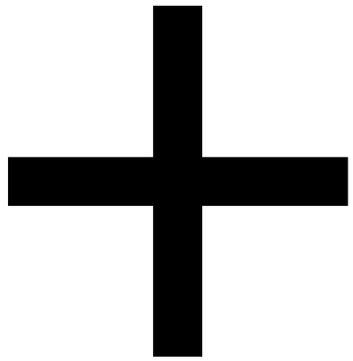
get more details

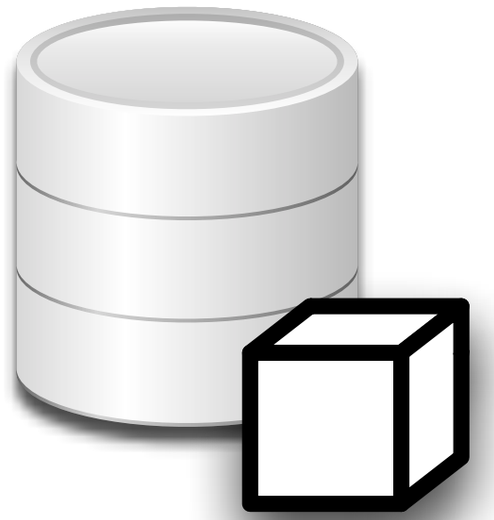


"batteries" that are included

Browser Workspace





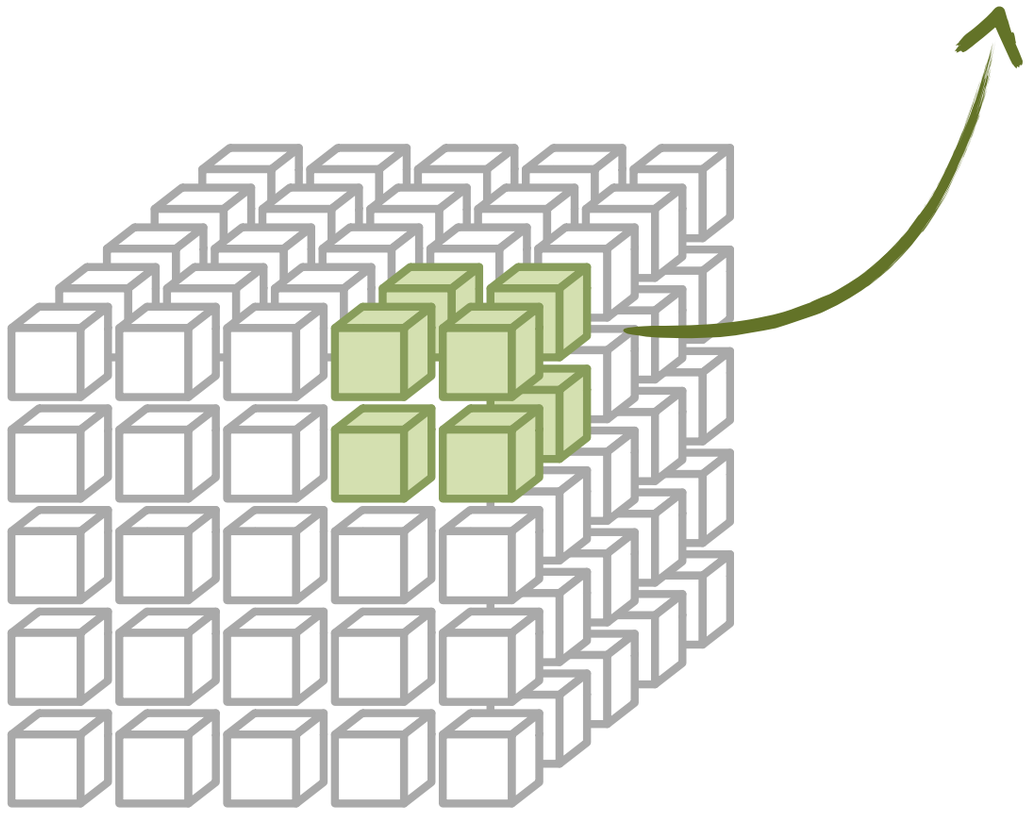


logical model

data

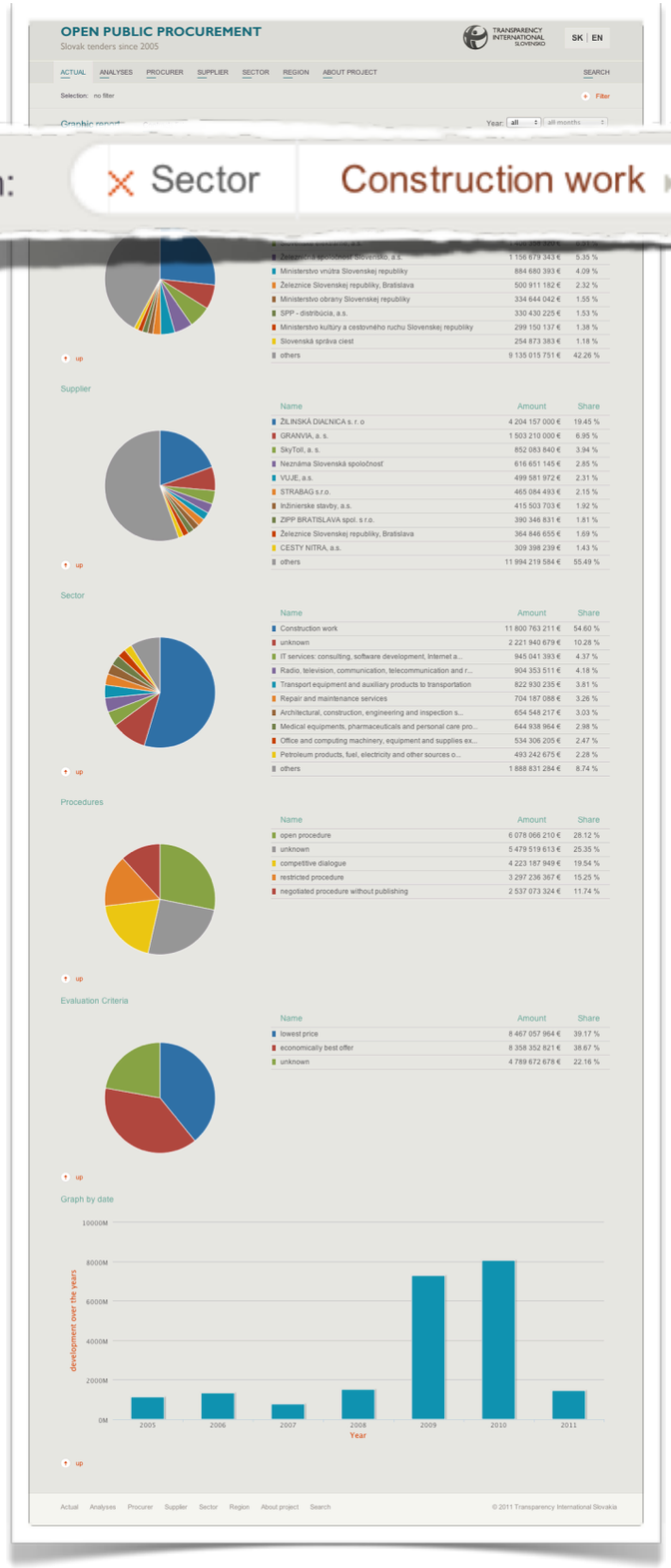
Cell

context of interest

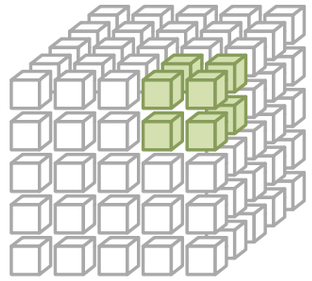


cell

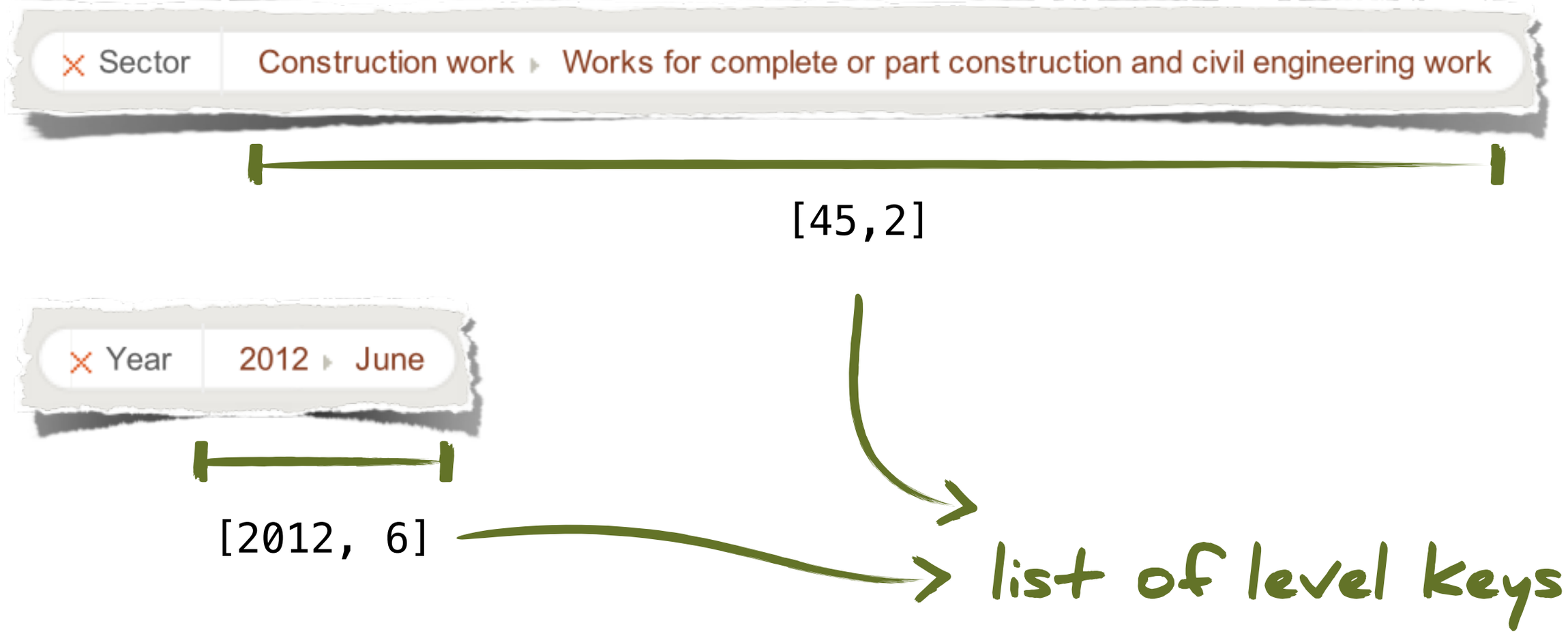
Selection: × Sector | Construction work ▸ Works for complete or part construction and civil engineering work | × Year | 2012 ▸ June

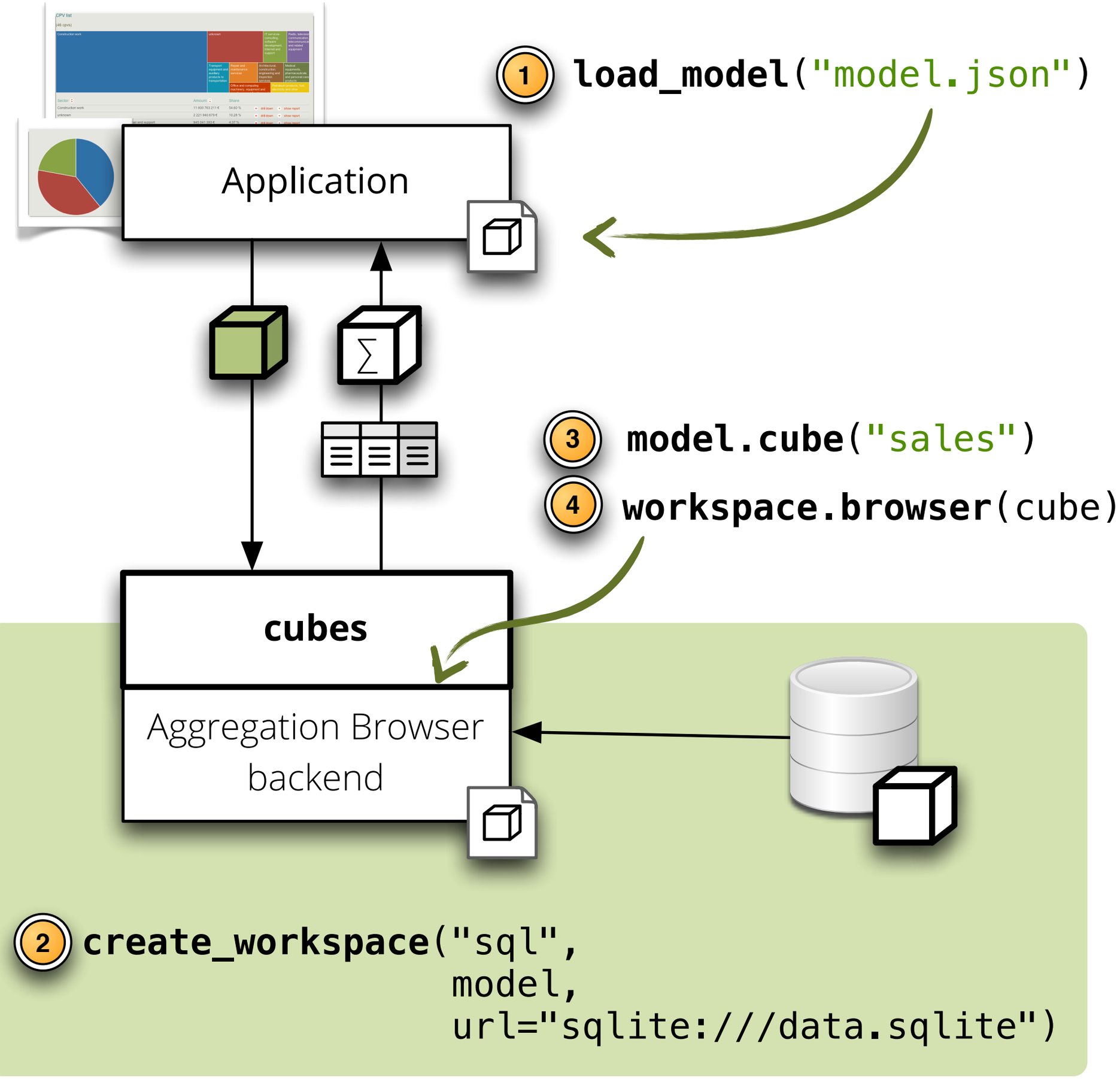


cell



Path

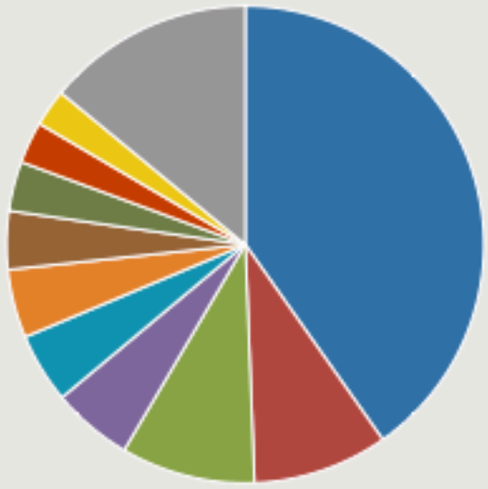

Sector
Construction work
Works for complete or part construction and civil engineering work
[45,2]
Year
2012
June
[2012, 6]
list of level keys


CPV list
Application
cubes
Aggregation Browser
backend
1 load_model("model.json")
3 model.cube("sales")
4 workspace.browser(cube)
2 create_workspace("sql",
model,
url="sqlite:///data.sqlite")

summary

Number of contracts	3 945
Contracts amount	2 163 664 086 €

Sector



Name	Amount
Construction work	870 210 815 €
Petroleum products, fuel, electricity and other sources o...	197 994 104 €
Medical equipments, pharmaceuticals and personal care pro...	196 990 697 €
Research and development services and related consultancy...	117 876 331 €
Transport equipment and auxiliary products to transportation	102 853 103 €
IT services: consulting, software development, Internet a...	100 416 319 €
Office and computing machinery, equipment and supplies ex...	85 520 037 €
Laboratory, optical and precision equipments (excl. glasses)	72 537 778 €
Repair and maintenance services	61 406 771 €
Hotel, restaurant and retail trade services	54 509 473 €

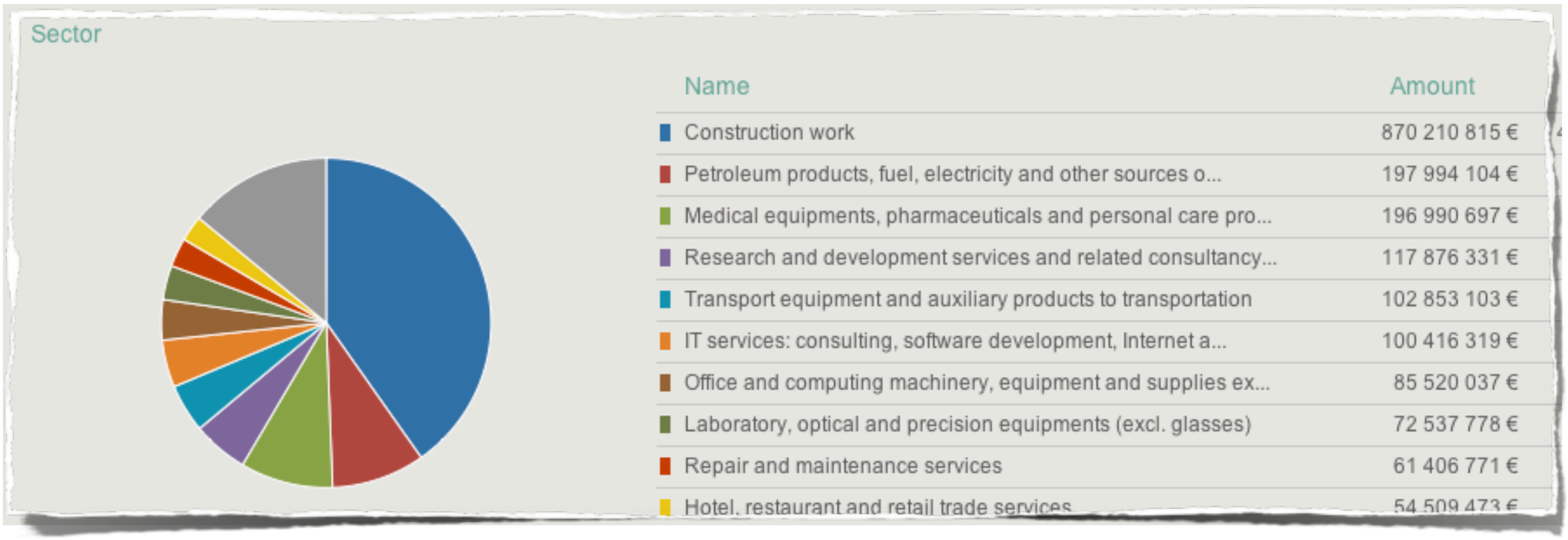
drill-down

```
browser.aggregate(🏠 cell)
```

Number of contracts	3 945
Contracts amount	2 163 664 086 €

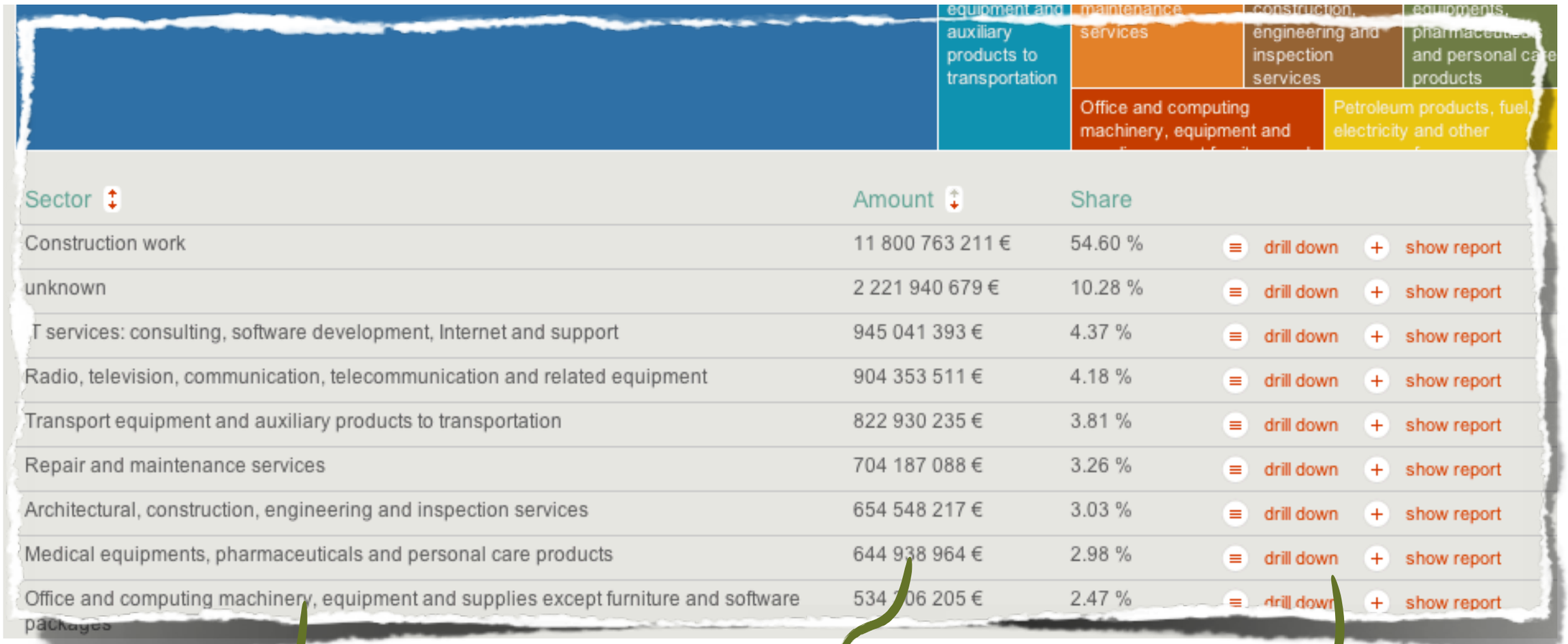
summary

```
browser.aggregate(cell,
                  drilldown=["sector"])
```



drill-down

```
for row in result.drilldown:
```

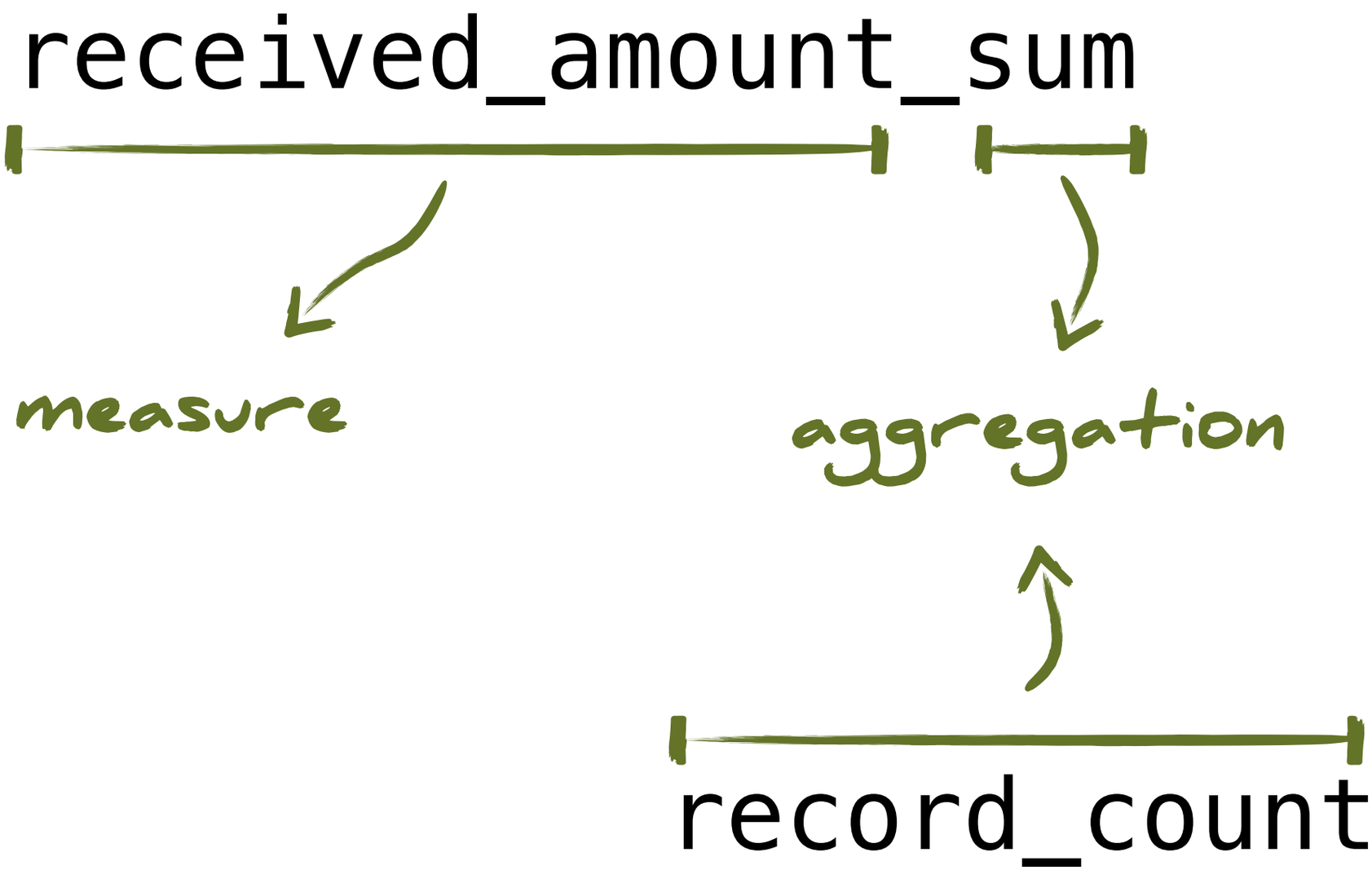


Sector	Amount	Share		
Construction work	11 800 763 211 €	54.60 %	drill down	show report
unknown	2 221 940 679 €	10.28 %	drill down	show report
T services: consulting, software development, Internet and support	945 041 393 €	4.37 %	drill down	show report
Radio, television, communication, telecommunication and related equipment	904 353 511 €	4.18 %	drill down	show report
Transport equipment and auxiliary products to transportation	822 930 235 €	3.81 %	drill down	show report
Repair and maintenance services	704 187 088 €	3.26 %	drill down	show report
Architectural, construction, engineering and inspection services	654 548 217 €	3.03 %	drill down	show report
Medical equipments, pharmaceuticals and personal care products	644 938 964 €	2.98 %	drill down	show report
Office and computing machinery, equipment and supplies except furniture and software packages	534 706 205 €	2.47 %	drill down	show report

```
row["amount_sum"]
```

```
row[ label_attribute]
```

```
row[ key]
```

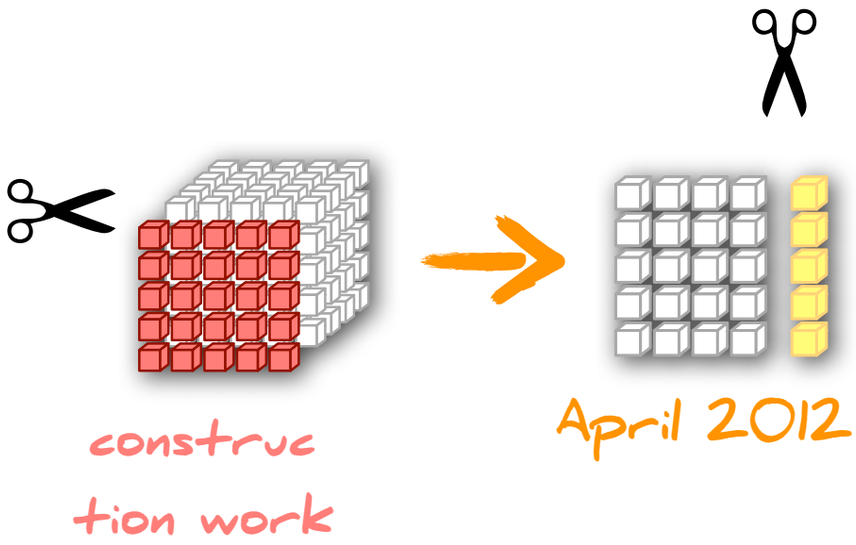

received_amount_sum
measure
aggregation
record_count

```
browser.facts(cell)
```

```
browser.values(cell, dimension)
```

```
browser.cell_details(cell)
```

Slicing and Dicing



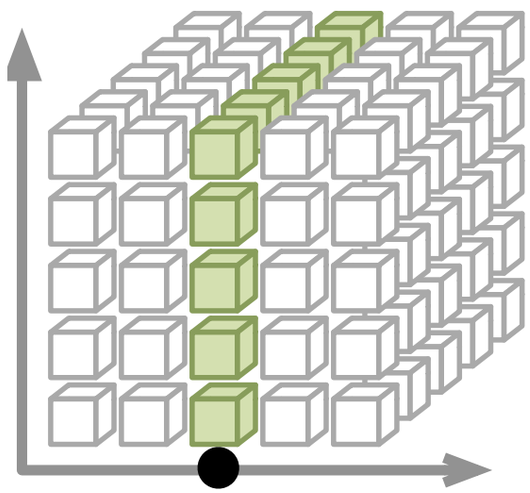
construction work
in april 2012

type

supplier

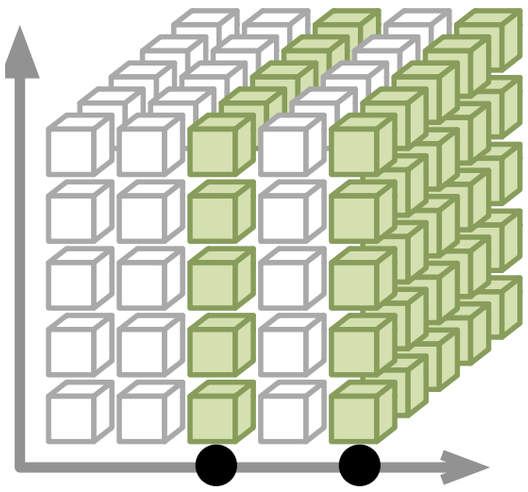
date

cut types



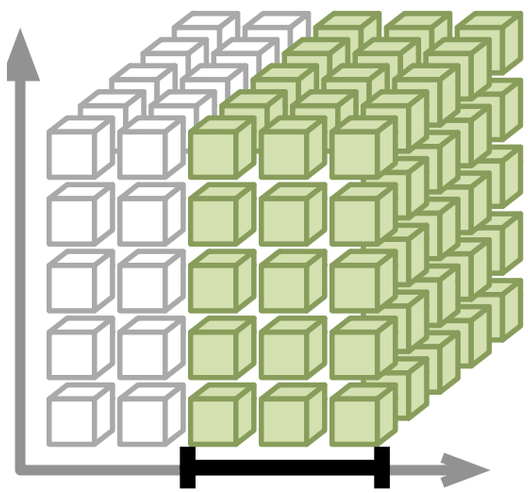
point

```
[2010]
```



set

```
[[2010,10],
 [2010,12]]
```

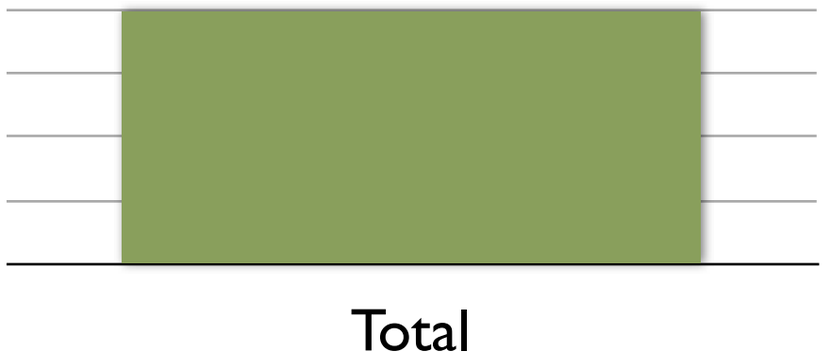


range

```
from=[2010,10]
to=[2010,12]
```

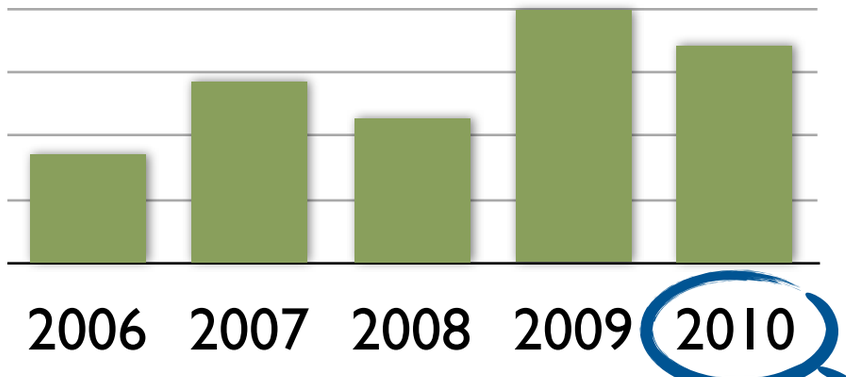
Implicit Hierarchy

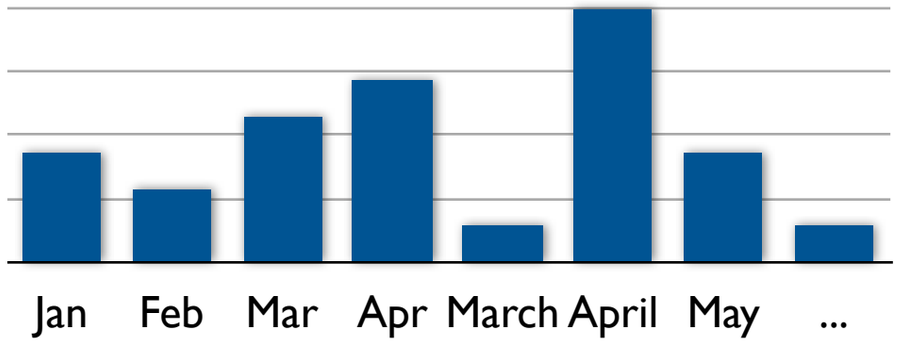
drilldown



whole cube

```
cell = Cell(cube)
browser.aggregate(cell)
```






```
cut = PointCut("date", [2010])
cell = cell.slice(cut)

browser.aggregate(cell,
                  drilldown=["date"])
```

```
browser.aggregate(cell,
                  drilldown=["date"])
```

Drill-down Level

```
drilldown = ["date"]
```

implicit: next from cell

```
drilldown = {"date": "month"}
```

explicit

Cross Table

experimental interface

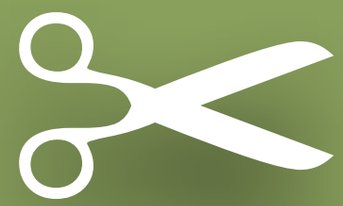
		2009	2010
Assets	Due from Banks	3044	1803
Assets	Investments	41012	36012
Assets	Loans Outstanding	103657	118104
Assets	Nonnegotiable	1202	1123
Assets	Other Assets	2247	3071
Assets	Other Receivables	984	811
Assets	Receivables	176	171
Assets	Securities	33	289
Equity	Capital Stock	11491	11492
Equity	Deferred Amounts	359	313
Equity	Other	−1683	−3043
Equity	Retained Earnings	29870	28793
Liabilities	Borrowings	110040	128577
Liabilities	Derivative Liabilities	115642	110418
Liabilities	Other	57	8
Liabilities	Other Liabilities	7321	5454
Liabilities	Sold or Lent	2323	998

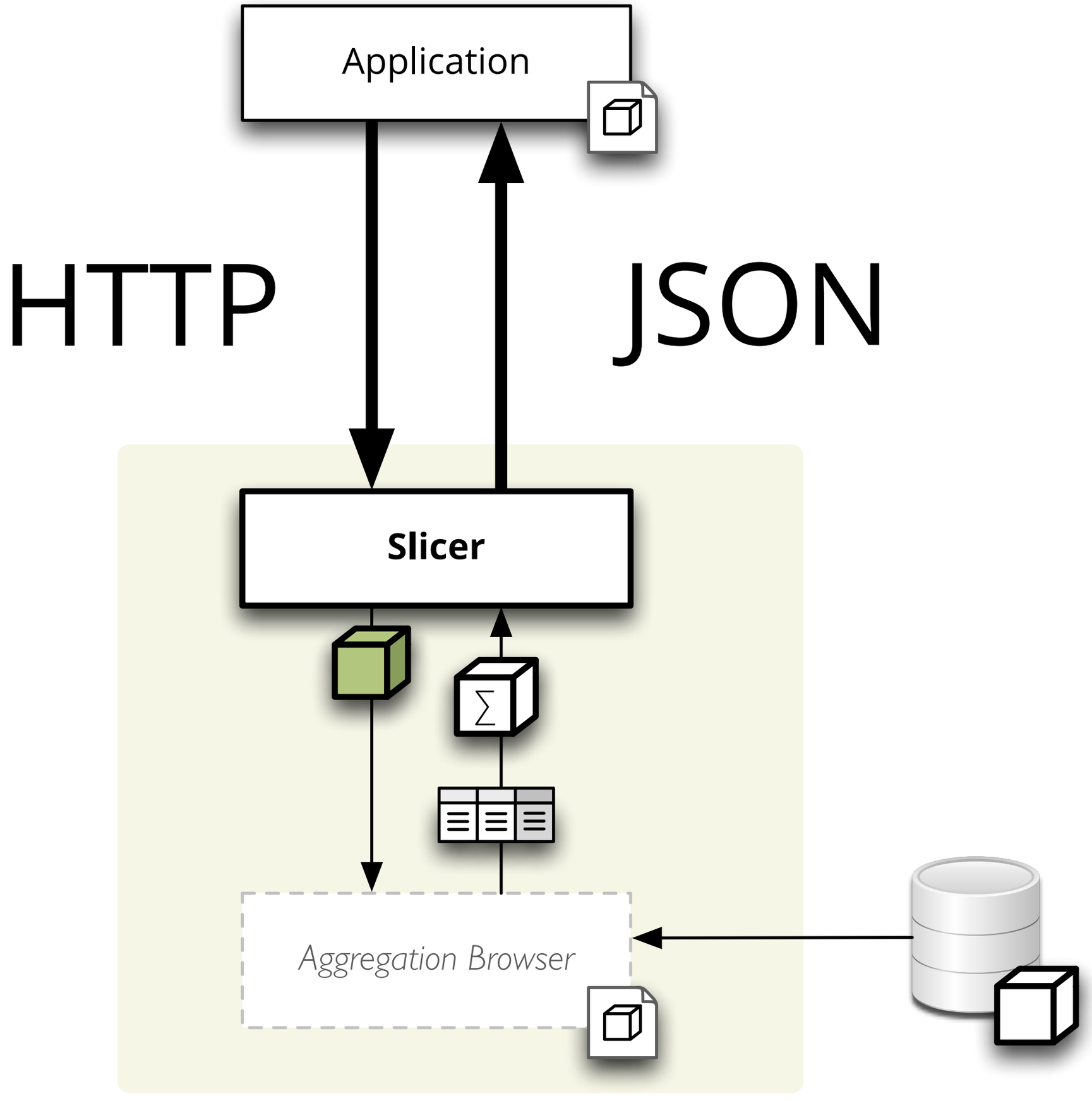
```
rows = ["item.category",
        "item.subcategory"]

columns = ["year"]

measures = ["amount_sum"]
```

```
table = result.cross_table(
                rows,
                columns,
                measures
            )
```

Slicer

The HTTP OLAP Server



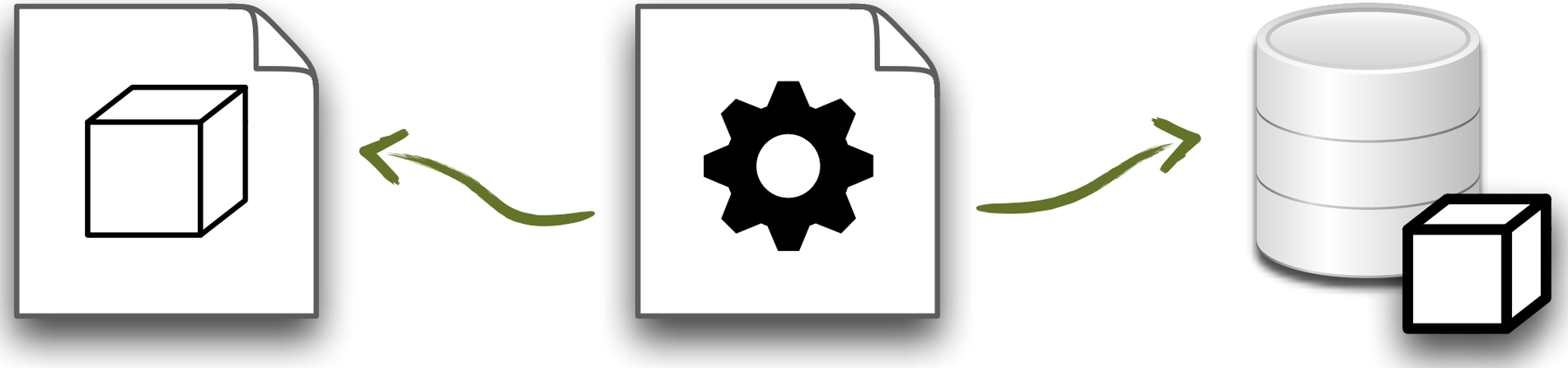

Application
HTTP
JSON
Slicer
Aggregation Browser

GET /model

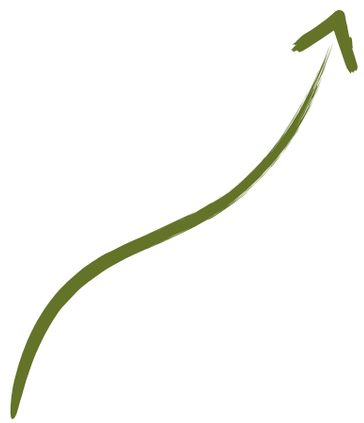
GET /aggregate

GET /values

GET /report

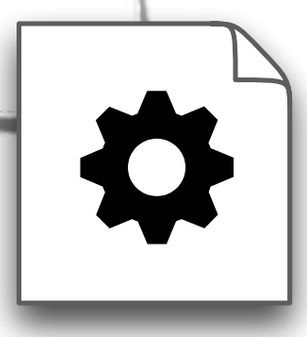


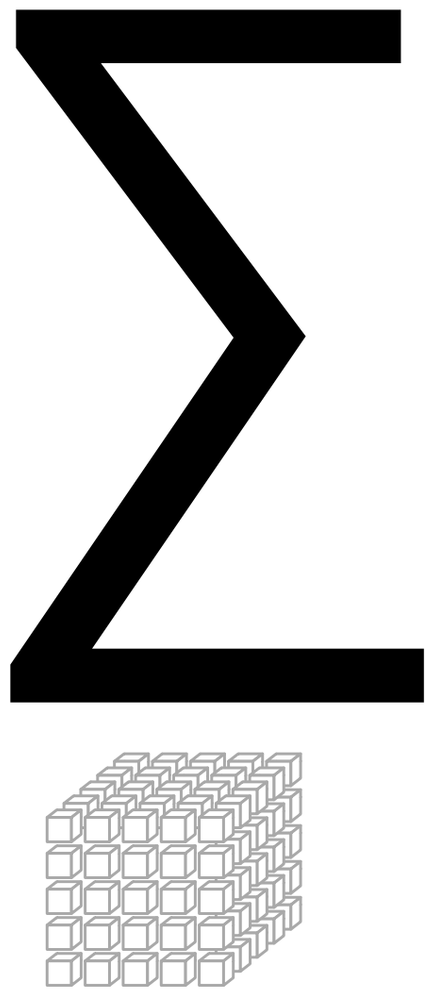
logical model configuration data



```
$ slicer serve slicer.ini
```

```
[server]
backend: sql
log_level: info

[model]
path: model.json
locales: en,sk

[workspace]
url: postgres://localhost/database
schema: datamart
fact_prefix: ft_
dimension_prefix: dm_
```





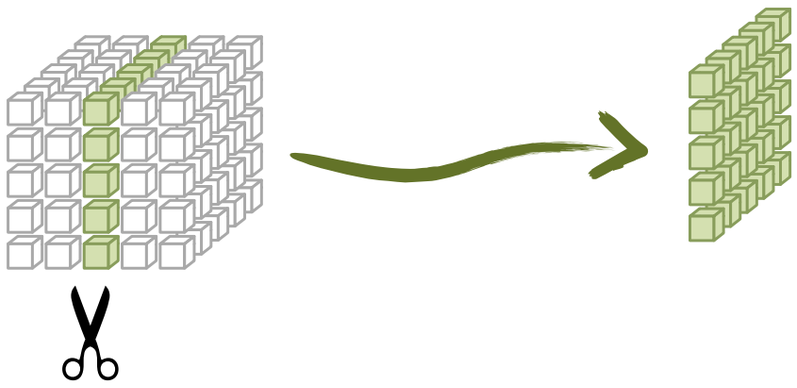
amount

GET /aggregate

GET aggregate


```
{
    "cell": [],
    "drilldown": [],
    "summary": {
        "record_count": 62,
        "amount_sum": 1116860
    }
}
```

$$\sum$$ amount



```
GET /aggregate?cut=date:2010
```

```
GET aggregate?cut=year:2010
```


```
{
    "cell": [
        {
            "path": ["2010"],
            "type": "point",
            "dimension": "year",
            "level_depth": 1
        }
    ],
    "drilldown": [],
    "summary": {
        "record_count": 31,
        "amount_sum": 566020
    }
}
```

```
GET aggregate?drilldown=year
```


```
{
    "cell": [],
    "total_cell_count": 2,
    "drilldown": [
        {
            "record_count": 31,
            "amount_sum": 550840,
            "year": 2009
        },
        {
            "record_count": 31,
            "amount_sum": 566020,
            "year": 2010
        }
    ],
    "summary": {
        "record_count": 62,
        "amount_sum": 1116860
    }
}
```

GET report

list of cuts

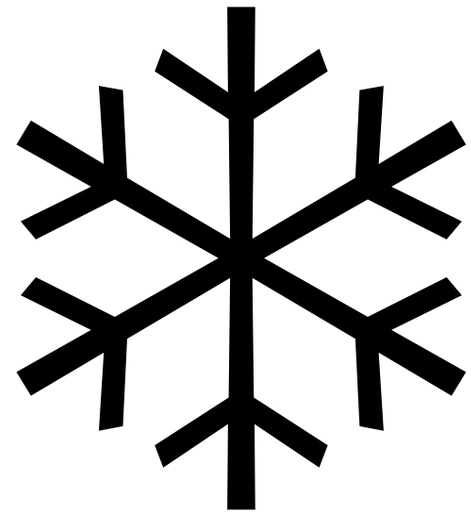
list of named queries

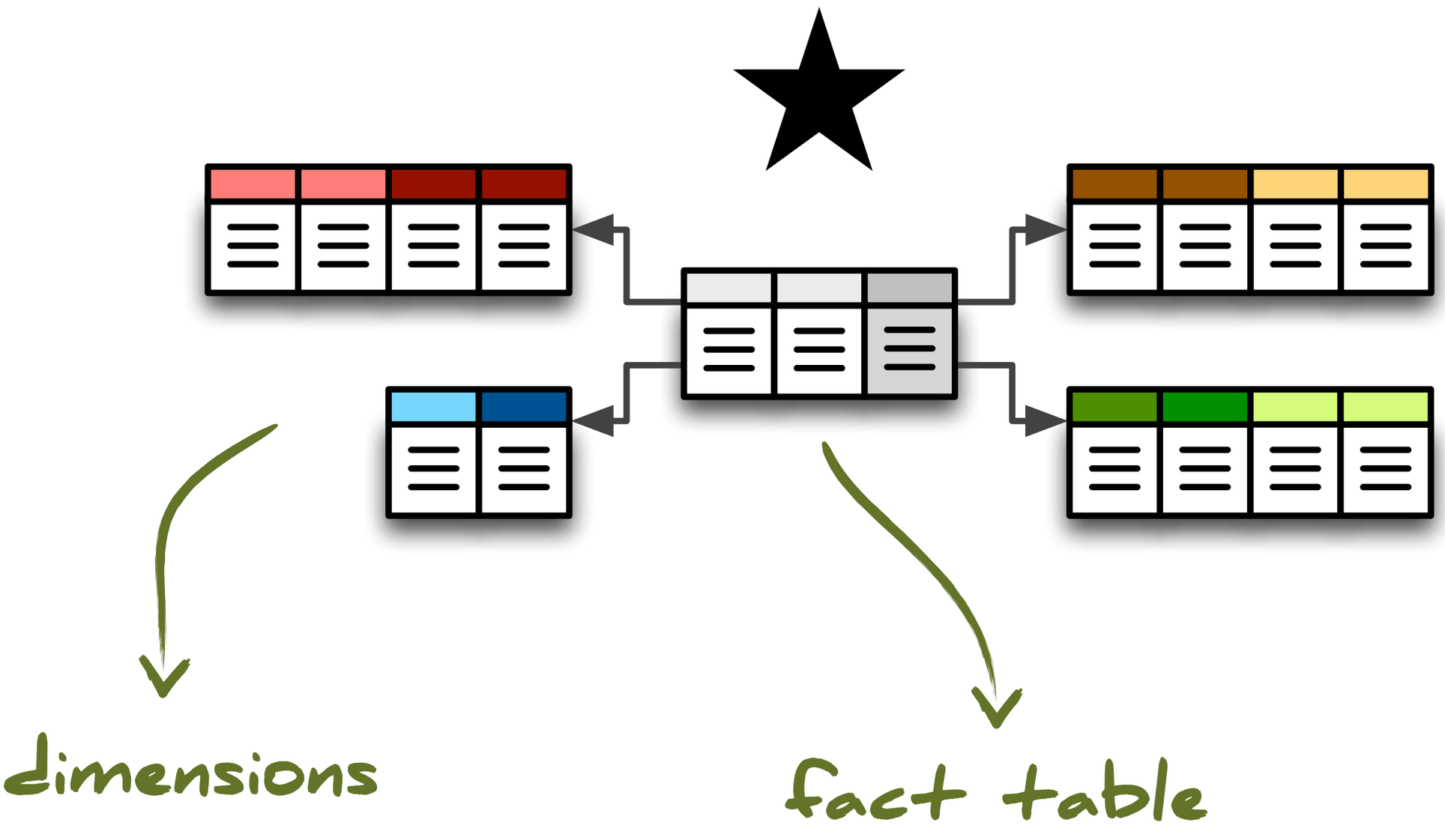
```
Content-Type: application/json
{
    "cell": [
        {
            "dimension": "date",
            "type": "range",
            "from": [2009],
            "to": [2011,6]
        }
    ],
    "queries": {
        "by_segment": {
            "query": "aggregate",
            "drilldown": ["segment"]
        },
        "by_year": {
            "query": "aggregate",
            "drilldown": {"date":"year"}
        }
    }
}
```

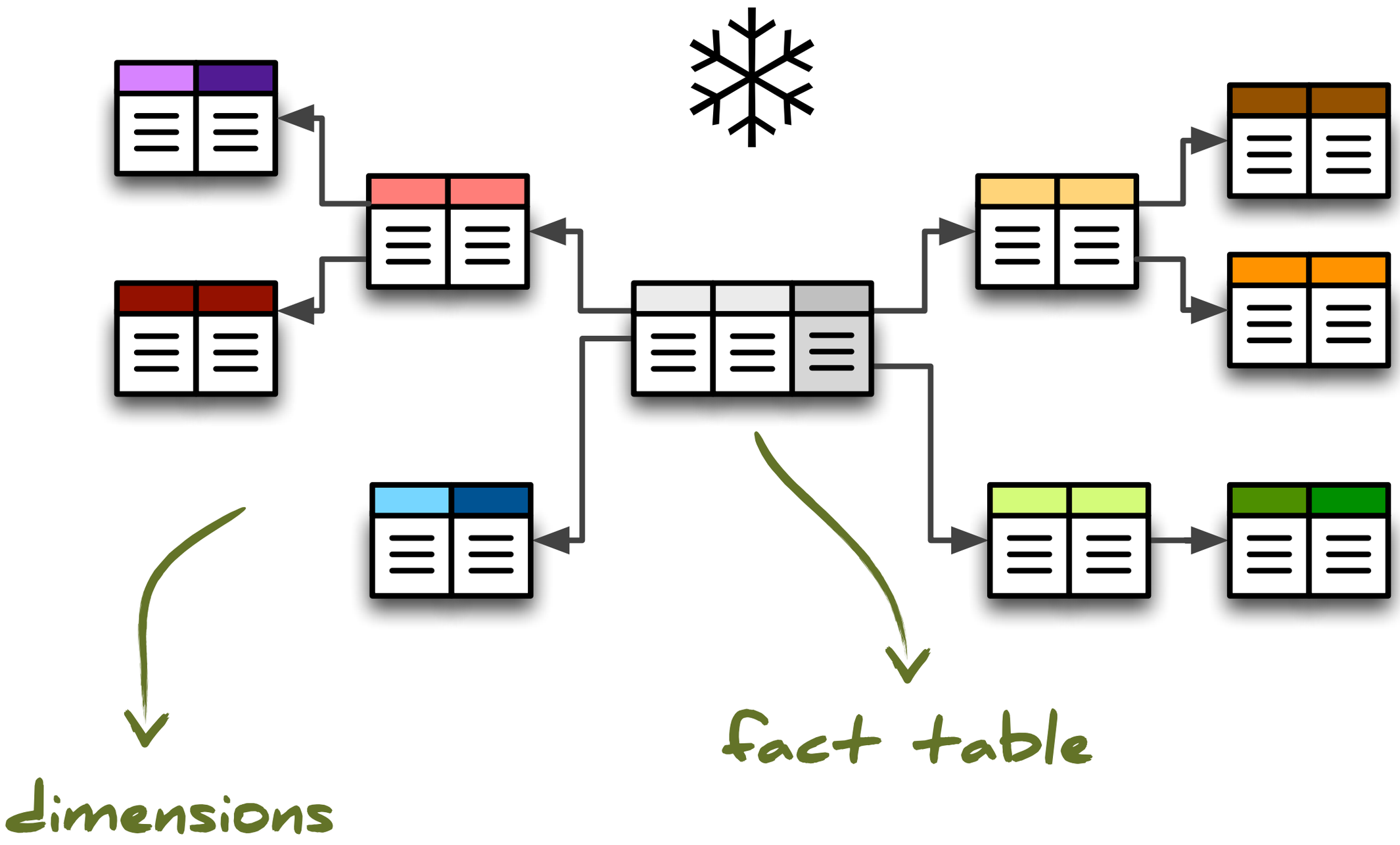
SQL Backend

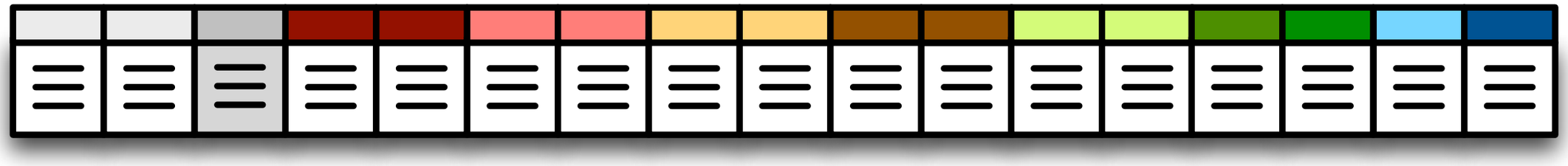
What data it works with?

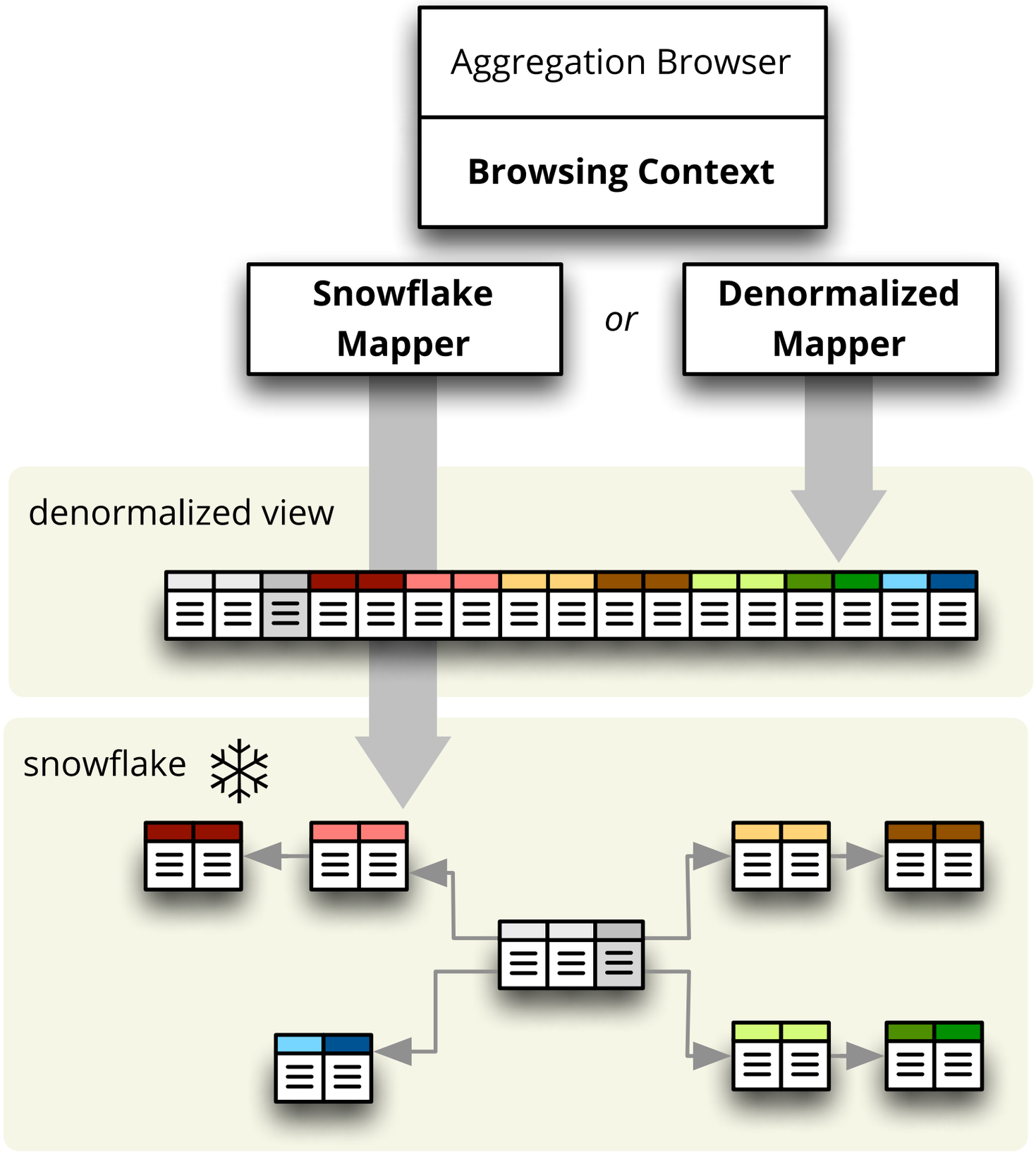
or




dimensions
fact table


dimensions
fact table

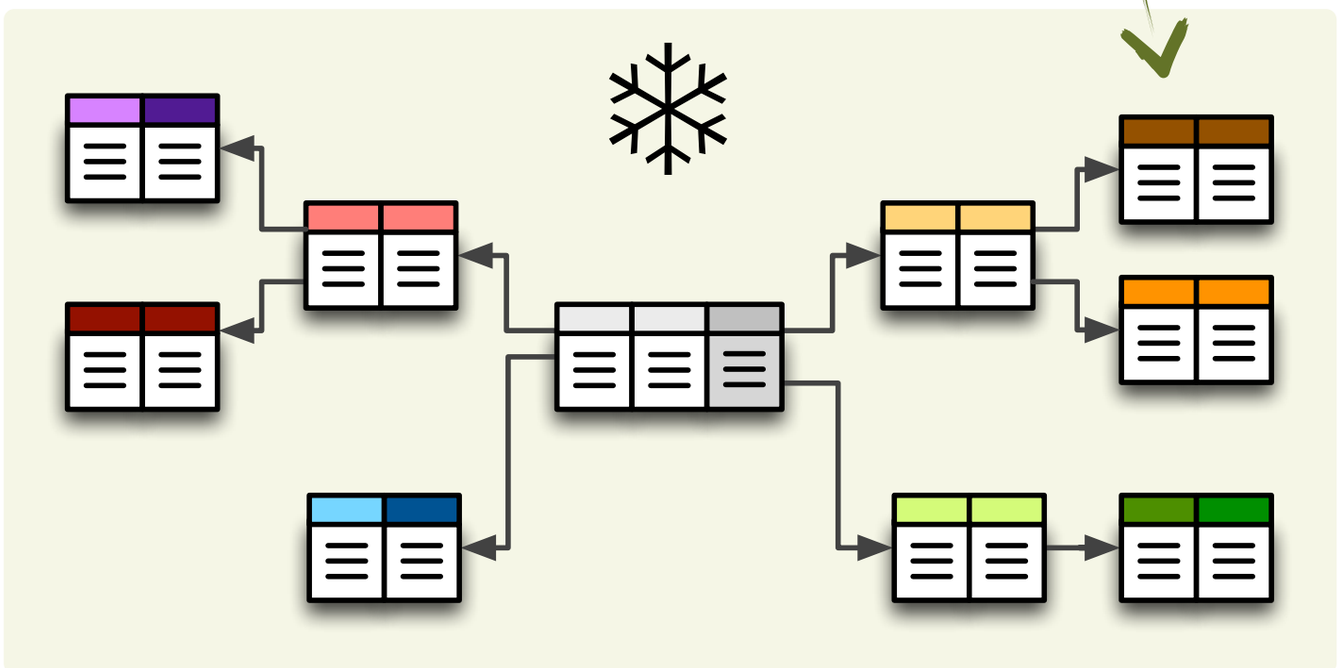



Aggregation Browser
Browsing Context
Snowflake Mapper
or
Denormalized Mapper
denormalized view
snowflake

logical

Name	Amount	Share
open procedure	6 078 066 210 €	28.12 %
unknown	5 479 519 613 €	25.35 %
competitive dialogue	4 223 187 949 €	19.54 %
restricted procedure	3 297 236 367 €	15.25 %
negotiated procedure without publishing	2 537 073 324 €	11.74 %

physical



SQL Features

- does not require DB write access
- denormalisation
  - *denormalised browsing, indexing*
- simple date datatype dimension
  - *extraction of date parts during mapping*
- multiple schema support

Slicer

command-line tool

- model validation

```
slicer model validate model.json
```

- model translation

```
slicer model translate model.json translation.json
```

- workspace testing

```
slicer test config.ini
```

- denormalization

```
slicer denormalize --materialize --index config.ini
```

Future

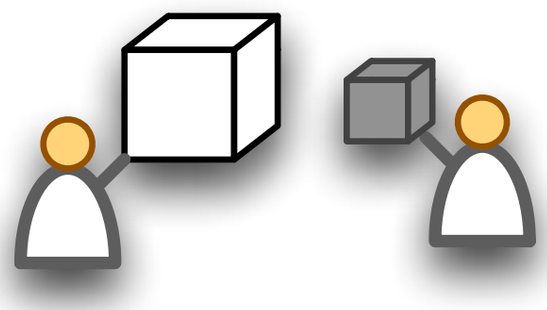
- formatters for visualisation libraries
- JavaScript library* → help needed
- backends
- derived measures

*http://github.com/Stiivi/cubes-js

Open Data

- shared repository of models
- shared repository of dimensions
- public cubes

  open Slicer HTTP APIs



http://github.com/Stiivi/cubes/wiki

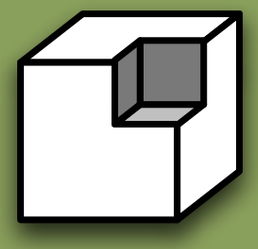
stay light

Nutrition Facts

Serving Size 1 cube

Amount Per Serving

	% Daily Value
Total Fat 0g	**0%**
Saturated Fat 0g	
Trans Fat 0g	



Thank You

source:

github.com/Stiivi/cubes

documentation:

packages.python.org/cubes/

examples:

github.com/Stiivi/cubes-examples

this presentation:

bit.ly/cubes-ep2012