



IMPLEMENTATION OF NEW FEATURES IN I2B2

Paris N., Bréant S., Orlova N., Yao K., Nekooguyan N., Beauchet A. & Daniel C.

ASSISTANCE PUBLIQUE — HÔPITAUX DE PARIS (AP-HP), FRANCE

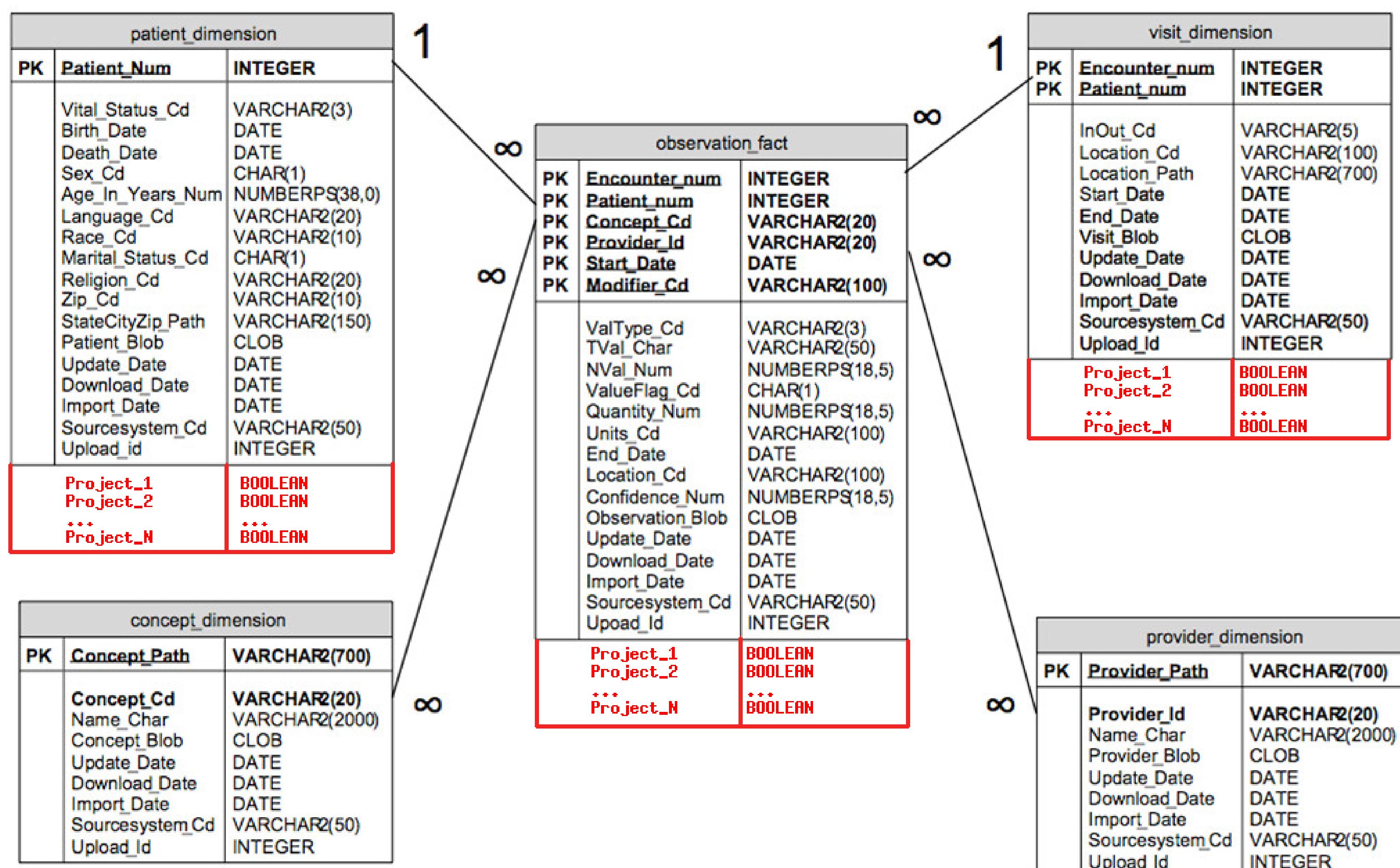
OBJECTIVES

Extend i2b2 to address specific needs of the AP-HP — a network of 37 hospitals in Paris:

1. building up as many projects as care-units (≈ 800) so that physicians can explore data of their patients
2. optimizing the pre-calculated number of patients for each concept of the ontologies
3. providing R statistical tool embedded in the webclient to end users

ISSUE 1: MULTIPLE PROJECTS AS LOGICAL VIEWS

Usually setting-up a new project in i2b2 requires a new physical datamart. Creation & maintenance of hundreds of projects, implies improvement in storage & performances design. This is carried out by creating virtual datamarts (logical views) based on only one physical database. The structure of 3 tables (figure below) of i2b2 is altered: one boolean field is added for each project. A logical view based on this field allows to filter data availables for a given project. Boolean fields are indexed to speed up queries.



RESULTS:

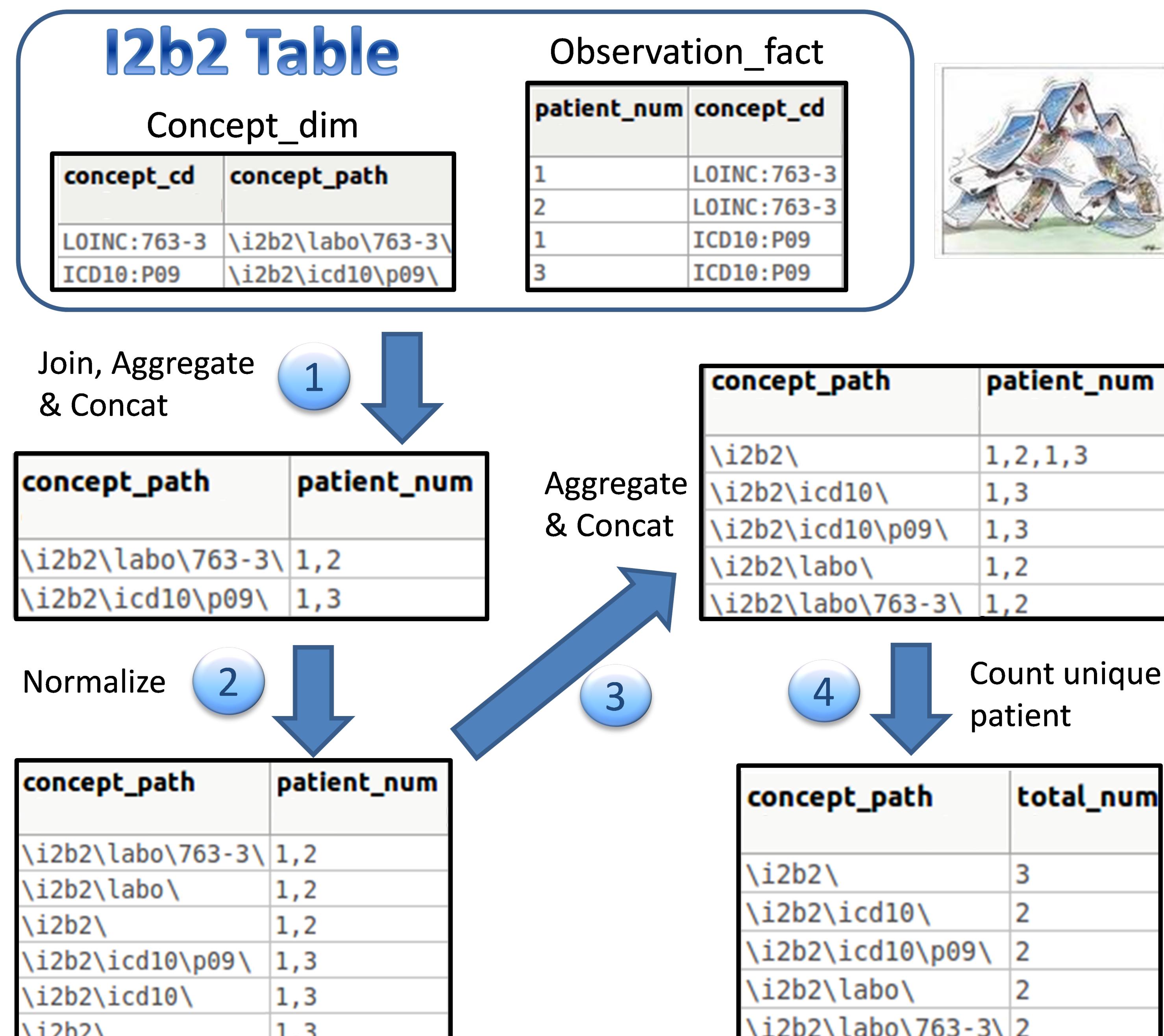
- STORAGE:
 - slight increase of database's size for logical views and boolean fields
 - only one version of data & i2b2 indexes
- PERFORMANCES:
 - queries time consuming is equivalent in both virtual & physical project design
 - data & indexes are shared by all projects (loaded in database cache)
- MAINTENANCE: one version of physical data simplifies & speeds up ETL processes

CONTACT

Nicolas PARIS, nicolas.paris@aphp.fr

ISSUE 2: PATIENT NUMBER PRE-CALCULATION ALGORITHM

In order to enhance the i2b2 SQL query generation, patients numbers are pre-calculated for all levels of ontologies with the “Castel of cards algorithm” (implemented in R & data.table package):



RESULTS: 50M facts spread over 300k ontologies are computed in 15 min on a 1.6GHz server, using a maximum of 2GB of RAM.

ISSUE 3: WEBCLIENT OPENCPU PLUGIN

OpenCPU = HTTP API for data analysis based on R.

- provides SSL security, cache-control, gzip compression
- configuration of memory allocation & time out
- ready to use JavaScript client library
- separation of concerns (R not embedded in java)
- fast & easy to develop & maintain

