



New results on a comparative evaluation of software providing access to different relational databases exposed on the Grid



Authors:

Dr. DONVITO, Giacinto (INFN-Bari); Prof. MAGGI, Giorgio (INFN+Politecnico Bari); TAFFONI, Giuliano (INAF); VUERLI, Claudio (INAF); BARISANI, Andrea (INAF); PASIAN, Fabio (INAF); MANNA, F. (INAF); GIORGIO, Emidio (INFN); AIFTIMIEI, Cristina (INFN); JAIN, Atul (INFN+Politecnico Bari); CAROTA, Luciana (INFN); PIERRO, Antonio (INFN); VERLATO, Marco (INFN); FIORE, Sandro (SPACI Consortium and University of Salento); ALOISIO, Giovanni (SPACI Consortium and University of Salento); CAFARO, Massimo (SPACI Consortium and University of Salento); VADACCA, Salvatore (SPACI Consortium and University of Salento); NEGRO, Alessandro (SPACI Consortium and University of Salento); DEL FREO, Federico (EGG project)

Motivations

- To provide access in GRID to Bioinformatics data stored in relational DB's BioinfoGRID EU project (<http://www.bioinfoGRID.eu/>) LIBI Italian MIUR project (<http://www.libi.it/libi>)
- To provide access to multiple astronomical databases (archives, Astronomical Catalogues, etc.) for the Astrophysical community. The Virtual Observatory (<http://www.euro-vo.org/pub/> <http://www.ivoa.net>)
- To provide access to population data for porting "public administrations" applications to the GRID EGG Italian MIUR project
- Diffuse inside the Italian community the knowledge and the expertise to access to relational DB's from Grid.



Tools under evaluation

GRelC: Grid Relational Catalog
Developed by SPACI Consortium and University of Salento, Lecce
Giovanni Aloisio (giovanni.aloisio@unile.it), Supervisor
Sandro Fiore (sandro.fiore@unile.it), Project P.I
Massimo Cafaro (massimo.cafaro@unile.it), Team Member
Alessandro Negro (alssandro.negro@unile.it), Team Member
Salvatore Vadacca (salvatore.vadacca@unile.it), Team Member
Project site: <http://grelc.unile.it>



G-DSE (INAF + INFN)
Developed by INAF and INFN
Edgardo Ambrosi (ambrosi@cnae.infn.it)
Giuliano Taffoni (taffoni@oats.inaf.it)
Andrea Barisani (lcars@infis.units.it)
Project site: <http://www.oats.inaf.it/grid/G-DSE>



OGSA-DAI
Developed as part of the Open Middleware Infrastructure Institute UK (OMII-UK) project.
Project site: <http://www.ogsadai.org.uk/index.php>



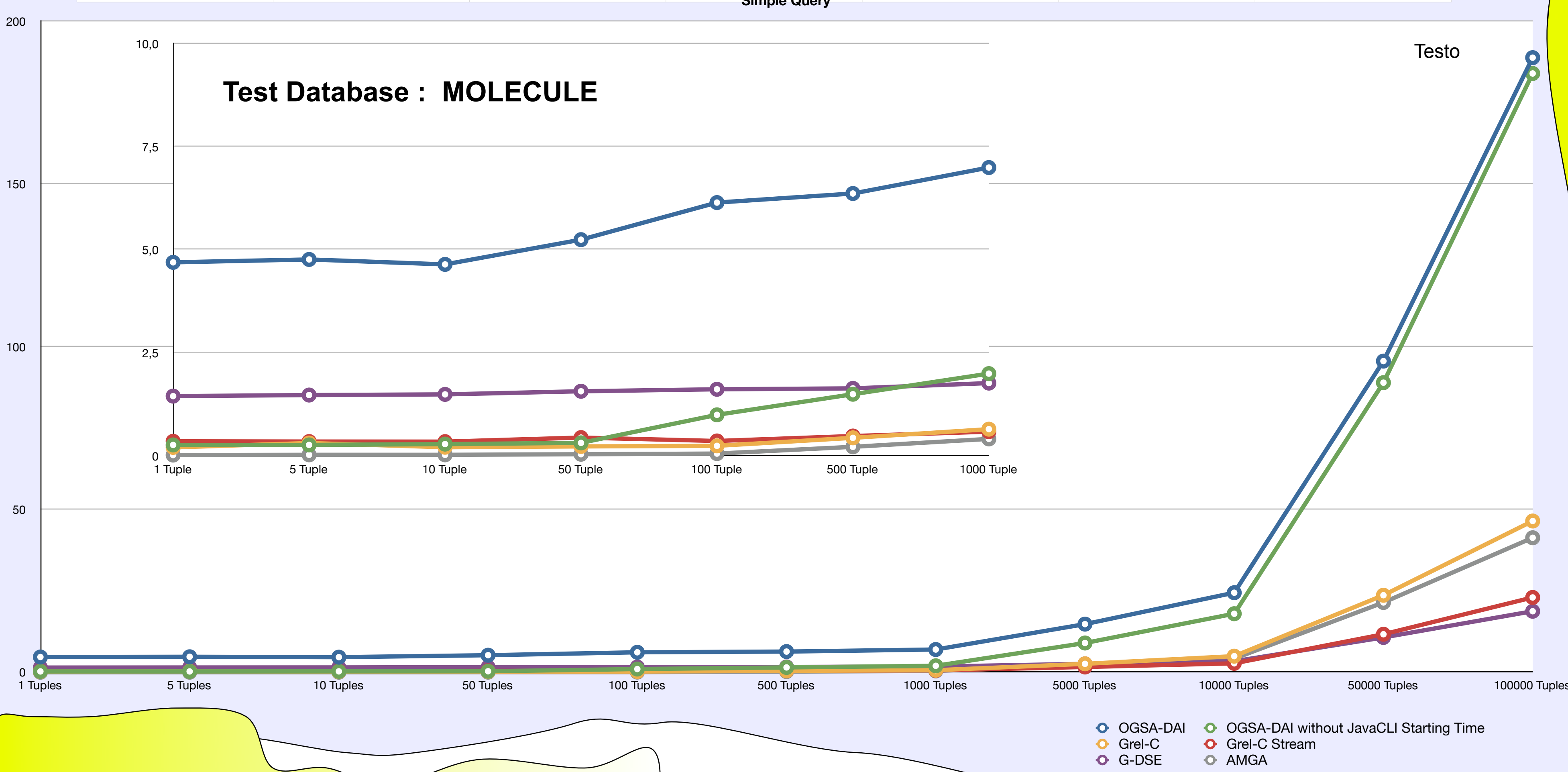
AMGA
Developed as part of gLite by:
Birger Koblitz: Initial design, project responsible.
Tony Calanducci: RPM building and testing. User support.
Salvatore Scifo: Java API maintainer
Claudio Cherubino: PHP client
Project site: <http://amga.web.cern.ch/amga/>



Preliminary Test Result

Select Query Results

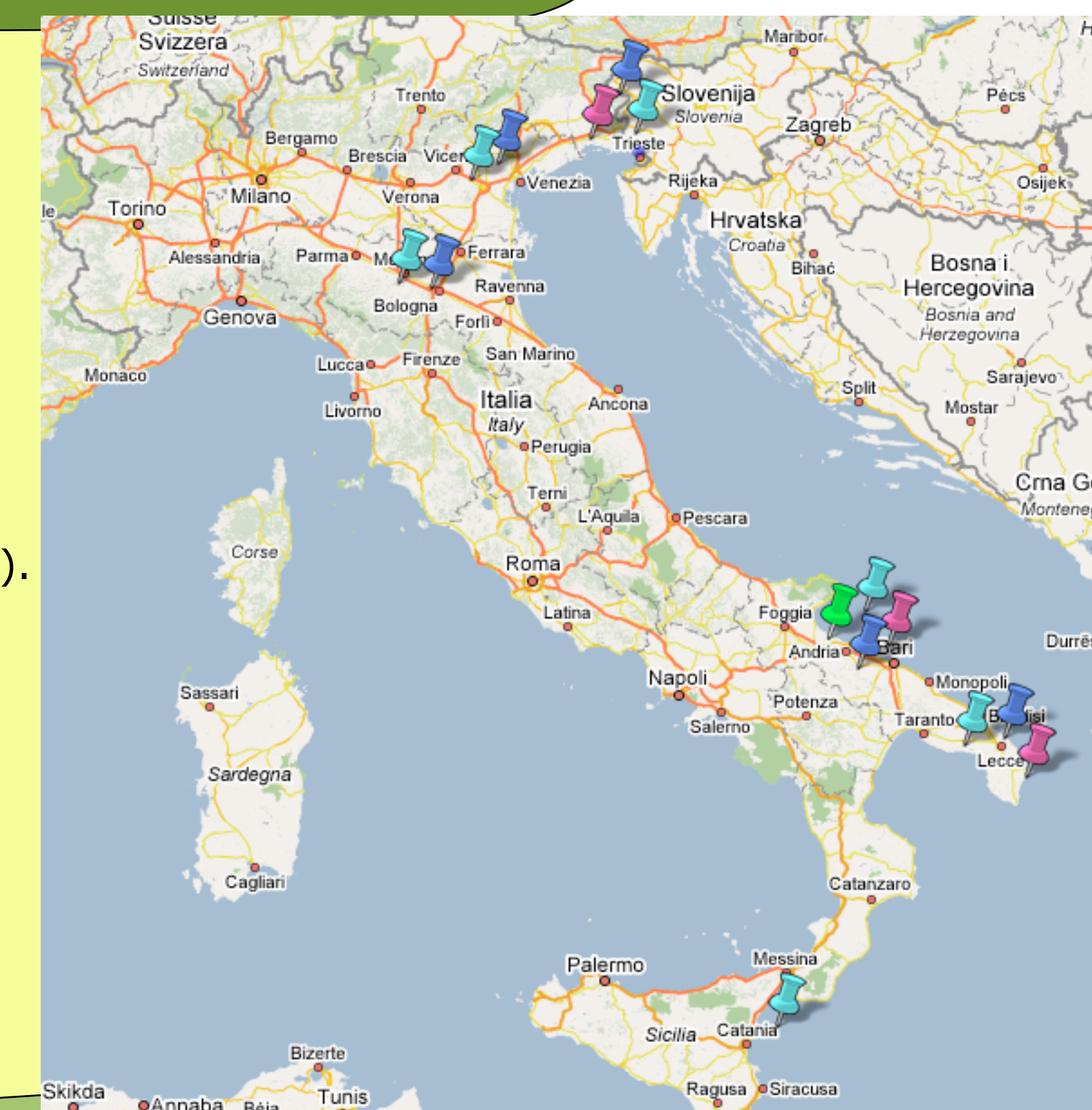
N. of Tuples	OGSA-DAI (s)	OGSA-DAI (s) Without Java Overhead	GRelC DIME (s)	GRelC STREAM (s)	G-DSE (s)	AMGA (s)
1	4,7	0,27	0,21	0,358	1,454	0,024
5	4,77	0,27	0,316	0,352	1,48	0,03
10	4,65	0,29	0,214	0,35	1,496	0,03
50	5,25	0,32	0,234	0,448	1,572	0,044
100	6,15	1	0,248	0,366	1,62	0,06
500	6,37	1,5	0,442	0,486	1,642	0,224
1000	7	2	0,652	0,592	1,77	0,416
5000	14,8	9	2,606	1,634	2,602	2,106
10000	24,46	18	4,962	2,754	3,398	4,208
50000	95,63	89	23,686	11,696	10,71	21,454
100000	188,86	184	46,486	23,002	18,764	41,336



TEST BED

- OGSA-DAI
- G-DSE
- AMGA
- GRelC-DAS

Test Database:
Bioinformatics database containing just a "molecule table" with about 500.000 tuples (350MB, PostgreSQL).
Other Databases:
Sakila (MySQL 23 tables)
World (MySQL 6 tables)
Dellstore (PostgreSQL 8 tables)
uniutrdb_test (MySQL 35 tables)
go_5_06 (MySQL 18 tables)
homo_sapiens (MySQL 74 tables)
Population db (PostgreSQL)



Tools Installation

- GRelC:**
 - Easy to install and manage via the grid data portal interface (GRelC Portal)
 - Postgresql and Mysql drivers currently available from the website. New drivers related to Oracle, DB2, etc. are in a preproduction phase.
- G-DSE:**
 - The installation and configuration procedure was not straightforward.
 - Installation based on yaim (glite tool)
 - Found problems in almost all the sites
- OGSA-DAI:**
 - The installation and configuration of the tool was not straightforward although a huge documentation is provided on the web site.
 - Greater variety of data sources accessed
- AMGA:**
 - The installation is quite easy and straightforward
 - Enabling authentication for a new user is not so easy
 - Providing more than one DB is not simple

Population Test Result

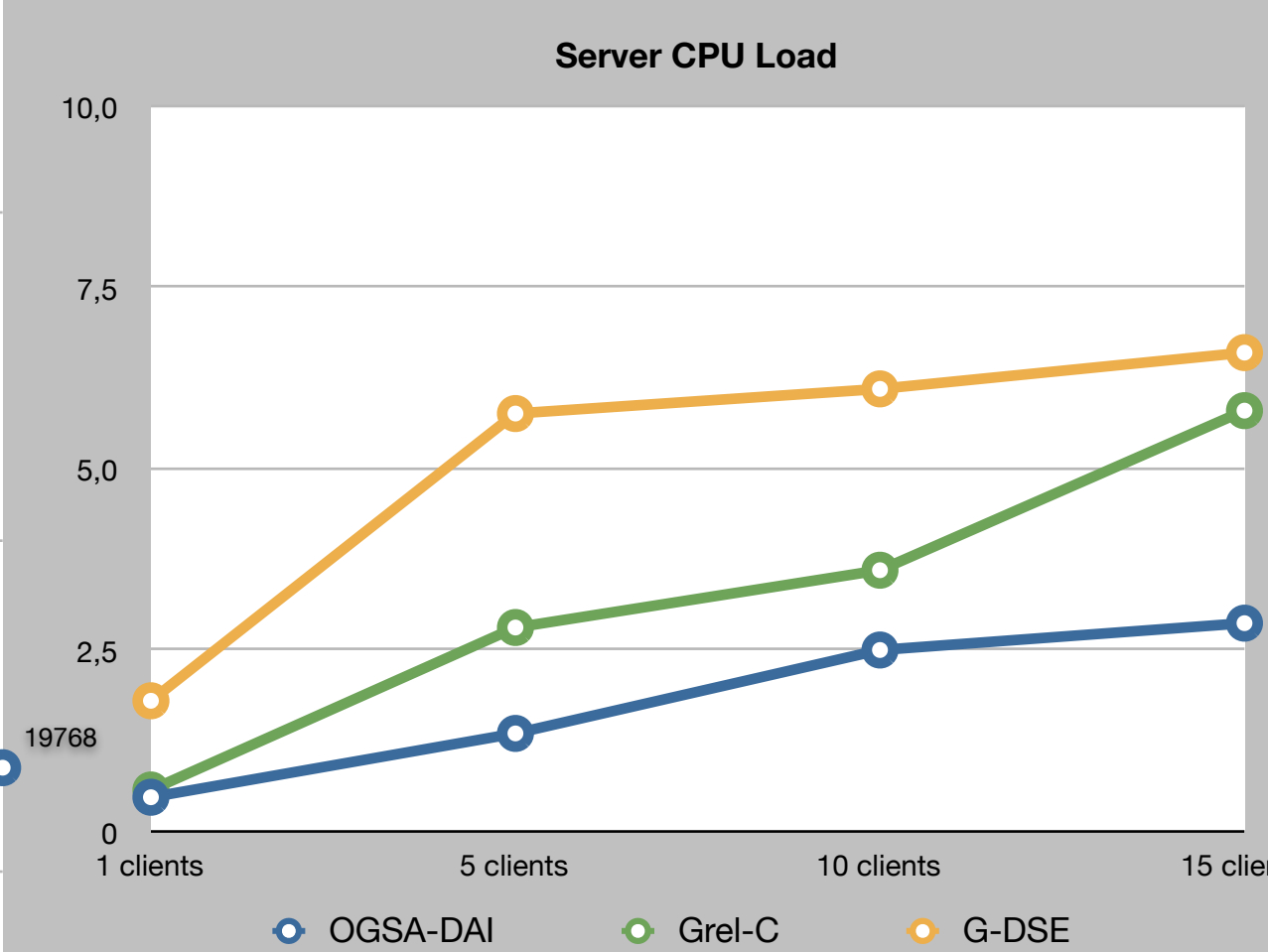
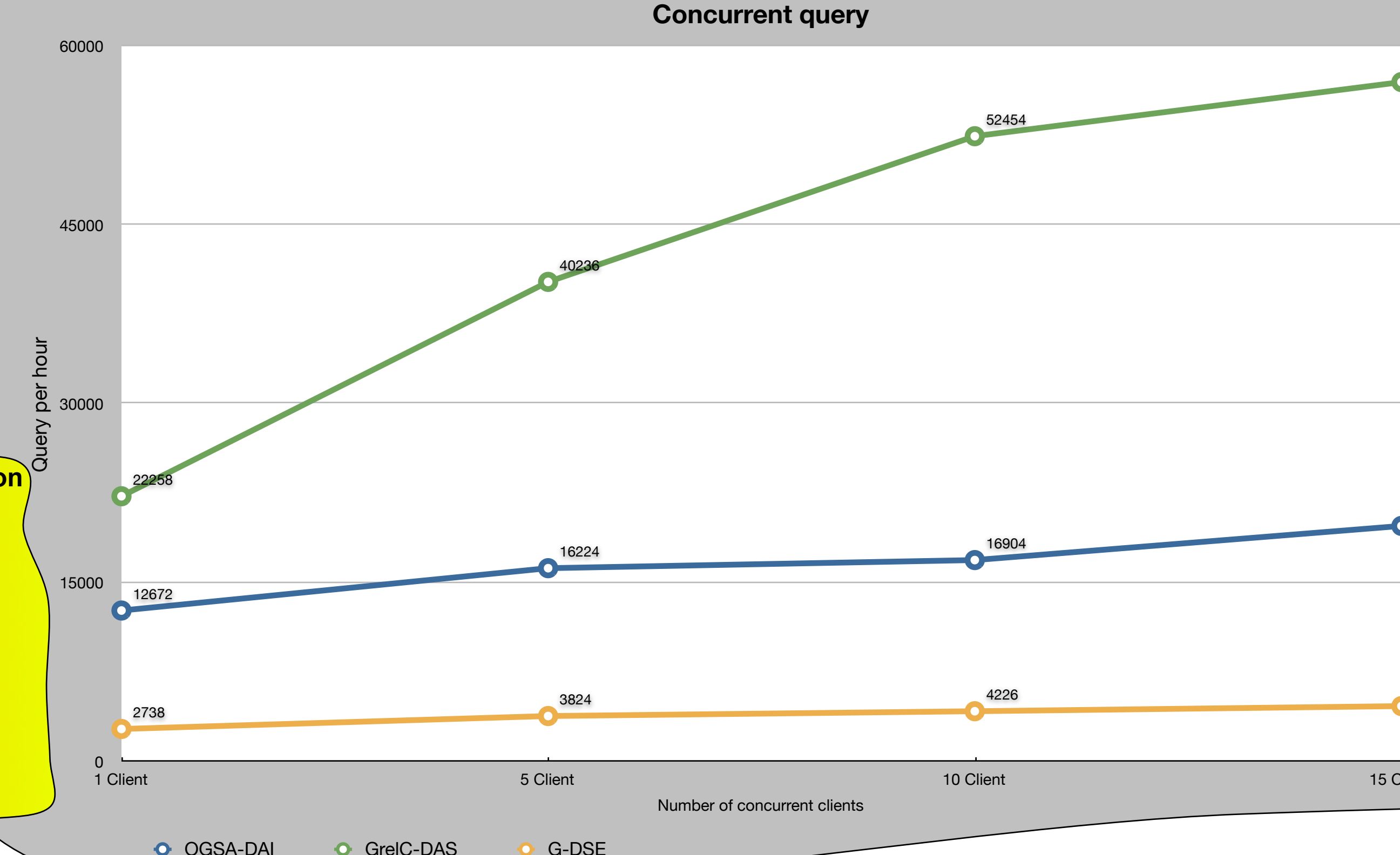
- Test has been performed on the population database using the real test case query.
- Real test query includes complex join query between two or more tables, inner join and outer join between tables.
- It has been found that the database itself is bottleneck in this case.
- No big difference found in the query response time between running query using database client and any of three tools.
- Current version of AMGA does not support such complex join query like inner join or outer join between tables and require a readjust of the DB.

Stress Test Result

To test the performance of each tool under load:

- We use a fixed number of clients that run constantly the same query against the server.
- We measure the number of queries that each service is able to provide for a fixed time-window (30 min).
- This gives an idea of the performance that each software is able to reach under high load
- We strictly keep under control the load on both client and server side

- OGSA-DAI clients are loading much more the client machine instead of loading the server
- For each execution 4 seconds are spent to start the JVM and only 0.2 seconds in order to execute the query.
- We solved the problem building a program that runs an infinite loop instead of running many times the CLI
- In this way the client is not overloaded and a good scalability is achieved



CONCLUSIONS

- OGSA-DAI:**
 - has lot of advanced features
 - Widely distributed
 - Good Performance if used with API
- GRelC:**
 - Very fast for small/medium/large datasets.
 - Fast DML query submission (very low GSI latency).
 - Easy to install. Already ported on gLite
 - Easy to manage via the grid data portal interface (GRelC Portal)
- GDSE**
 - Very efficient for huge query
 - Easy to manage
 - Good integration with gLite environment (BDII and VOMS enabled)
- AMGA:**
 - Very quick for small query
 - Fast DML query submission (very low GSI latency).
 - A lot of advanced functionality
 - Good performance for large datasets
 - It should be standard part of the next release of gLite

SECURITY

Tools	GSI	VOMS Authentication	Transport Layer Security	Data Encryption
OGSA-DAI	Yes	No	Yes	Yes
GRelC-DAS	Yes	Yes	Yes	Yes
G-DSE	Yes	Yes	Yes	No
AMGA	Yes	Yes	Yes	Yes