High Performance and Cost Effectiveness for the Labor Market Statistic in Switzerland

EMC²

GREENPLUM.

Based on the EMC Greenplum Database, the Swiss State Secretariat for Economic Affairs is *now* able to provide *analyses* to the user community *faster and* with greater flexibility.

In 1995, a new era started for the unemployment insurance business in Switzerland, marked by the amended Unemployment Insurance Act (AVIG). Instead of passively managing unemployment statistics, the Swiss government now is executing a proactive labor market policy. The new principle of 'Reintegration before retirement' aims at a lasting reintegration of jobseekers into the labor market.

Business Intelligence for the Swiss Labor Market

Up-to-date information and flexible analysis options are required for statistical observations which are used for various purposes, e.g. as short-term economic indicators or for controlling the correct execution of the Unemployment Insurance Act. At the same time, it must be possible to fine-tune analyses to the relevant question in every case. For this purpose, the Labor Market Statistics - a department of the Swiss State Secretariat for Economic Affairs - has been operating a business intelligence system since 2003, based on a traditional data warehouse. The business intelligence system of the first generation already saw the replacement of the paper printouts of monthly statistics (consisting of 17,000 pages each and every month) by a much more convenient browser query option.

Among the approximately 900 users are responsible federal offices, the employees of 120 regional offices (RAV) of the national employment service in the 26 Swiss cantons and more than 40 unemployment insurance funds. The result: A complex and heterogeneous user community with various requirements, reflecting the federal system of Switzerland.

A new Business Intelligence Solution for growing demands

Due to increasing requirements regarding the labor market statistics, the State Secretariat for Economic Affairs (SECO) initiated the LAMDA X (Labor Market Data Analysis) project. The new business intelligence solution was based on a new source system, and the front end was newly developed on the basis of Microstrategy 9. This enabled a more flexible analysis as well as mobile access.

Until 2009 a number of business intelligence applications for several purposes - e.g. the official labor market statistics, the statistics of payments executed by unemployment insurers, the key performance index for regional office executives and an application for public information regarding unemployment, accessible online at www.amstat.ch - were implemented. Some of these applications required complex calculations.

A more complex analysis requires a new infrastructure, as Dr. Elmar Benelli, Data Warehouse LAMDA Manager at the SECO, explains, "We experienced increasing performance problems. With the existing infrastructure such problems would have been manageable only with substantial effort and expenditures. The limiting factors were the physical separation of the database system and the data as well as the data transport across the network which is also used by other parties."

In cooperation with Saracus, the partner responsible for the data warehouse, Dr. Benelli's team began searching for a new database solution to cover Big Data requirements in spring 2011. This solution - EMC Greenplum - was discovered in December 2011. It was decided to implement a proof-of-concept installation, replacing the former database system with the EMC Greenplum database. The Greenplum database operates with massive parallel processing (MPP). A master server controls the computing

ESSENTIALS

Industry
Government

Business Challenge
- Support of complex calculations
- Hardware flexibility – no proprietary lock-in
- Downward scalability for small data marts

Solutions
- EMC Greenplum database on industry standard servers
- Massive parallel processing

Results
- Significant performance increases
- Support of constantly changing requirements
- Significant gain of time in preparing database

EMC²

executed by an unlimited number of segment servers. The solution offers flexible scalability and can be operated with various hardware platforms. The proof-of-concept installation for the Labor Market Statistic comprises one master server, one additional server for redundancy and four segment servers.

Focus in Flexibility

"Hardware flexibility was decisive," Benelli points out. "Our data warehouse with 500 GB - 200 GB thereof for the data marts (subsets of the data warehouse) - is relatively small. Consequently, we mostly needed downward scalability. Furthermore, EMC Greenplum enables an operation on low-cost industry-standard servers which can be easily acquired from the usual government suppliers. So we remain flexible in the future as well, and are independent of a specific hardware provider. Other solutions we evaluated included proprietary components that would not allow of such flexibility." On the other hand, such flexibility requires some additional effort. Benelli explains, "The EMC Greenplum software was easy to install, and the migration of the front end was executed without any problems and within half a working day. However, the hardware installation was more complex than expected. The configuration of the servers for an optimum operation of the EMC Greenplum database required some effort - often the devil is in the details, which was also the case here. "

Such problems could have been avoided by implementing a preconfigured data computing appliance which is also offered by EMC Greenplum. Benelli explains why the software-only option was chosen by the customer: „By using a preconfigured appliance we would have lost our freedom of choice regarding hardware, but this freedom of choice was extremely important to us."

Significant performance increases

Flexibility is very important to Benelli - not only regarding hardware choices. As he explains, "In general, flexibility must be rated higher than stability in a business intelligence environment. As opposed to a transaction system, where users basically do not want much change, one of the characteristics of a business intelligence environment are constantly changing requirements."

The desired performance increases were achieved with the proof-of-concept installation. This also supports the required flexibility. "We experienced massive performance increases at the front end. Even in case of complex queries the results are displayed quickly," Benelli explained. This insight was gained during operation of the test installation. And: "Should performance problems be experienced during the computing of specific analyses, we have room for improvement - from database partitioning to the use of distribution key and column views. We have not applied any of these improvement procedures yet." The ETL part with the Informatica platform provides acceptable loading times already, but still must be fine-tuned to conform to the new database.

Benelli pointed out: "Copying of the database now goes extremely fast. Previously, the preparation of a separate database for test and development purposes or special analyses with high performance requirements took the data warehouse team two weeks. Now, with the EMC Greenplum database, all it takes is copying, which can be done in 20 minutes."

Successful test project with a bright future ahead

"The cooperation of the team, comprising SECO employees, the Saracus AG, the solution operator Bundesamt für Informatik und Telekommunikation (BIT) and Microstrategy as well as EMC was excellent. Engineers stayed in touch constantly, were always available and reliably solved all problems that arose." According to Benelli, the test phase was successfully completed in early summer 2012. Now the existing system and the new EMC Greenplum solution will operate in parallel for a few months to allow of final improvements and fine-tuning. The new database will completely take over the Swiss labour market statistics by the end of the year at the latest.

EMC²