

Auf Basis der EMC Greenplum Database kann das Schweizer Staatssekretariat für Wirtschaft komplexe Auswertungen der Arbeitsmarktzahlen seiner heterogenen Nutzergemeinde rascher und flexibler zur Verfügung stellen.

Mit dem revidierten Arbeitslosenversicherungsgesetz (AVIG) von 1995 trat die Schweizerische Arbeitslosenversicherung in ein neues Zeitalter ein: Statt einfach passiv die Arbeitslosenzahlen zu verwalten, betreibt der Bund heute eine aktive Arbeitsmarktpolitik. Sie funktioniert nach dem Prinzip „Eingliederung vor Rente“ und hat die möglichst umfassende Reintegration der Stellensuchenden in den Arbeitsmarkt zum Ziel.

### Business Intelligence für den Arbeitsmarkt

Statistische Betrachtungen, die unter anderem als kurzfristiger Konjunkturindikator dienen, sowie die Steuerung des Vollzugs des AVIG benötigen aktuelle Informationen und flexible, der jeweiligen Fragestellung exakt angepasste Auswertungen. Zu diesem Zweck betreibt die Arbeitsmarktstatistik des Bundes, die dem Staatssekretariat für Wirtschaft (SECO) angegliedert ist, bereits seit 2003 ein Business-Intelligence-System auf Basis eines klassischen Data Warehouse. Schon mit dieser ersten BI-Generation konnte man die früher üblichen, mit allmonatlich 17'000 Seiten äusserst umfangreichen und entsprechend unpraktischen Statistiken in Papierform durch eine viel bequemere Browserabfrage ablösen.

Zu den rund 900 Nutzern gehören neben den zuständigen Bundesstellen Mitarbeitende der 120 regionalen Arbeitsvermittlungszentren (RAV) in den 26 Kantonen sowie der über 40 Arbeitslosenkassen – für die Arbeitsmarktstatistik ergibt sich eine komplexe und heterogene Nutzergemeinde mit unterschiedlichen Anspruchsgruppen, die dem föderalistischen System der Schweiz entspricht.

### Neue BI-Lösung für wachsende Ansprüche

Aufgrund steigender Anforderungen an die Arbeitsmarktstatistik startete das SECO im Jahr 2007 das Projekt LAMDA X (Labor Market Data Analysis). Die BI-Lösung wurde auf einem neuen Quellsystem aufgesetzt und das Frontend neu auf Basis von Microstrategy 9 entwickelt. Damit wurden flexiblere Auswertungen und künftig auch Abfragen von Mobilgeräten möglich.

Im Rahmen von LAMDA X wurden bis 2009 verschiedene BI-Anwendungen mit teils komplexen Berechnungen realisiert, darunter etwa die offizielle Arbeitsmarktstatistik, Auszahlungsstatistiken der Arbeitslosenversicherungen, Führungskennzahlen (KPI) für die RAV-Leiter sowie Arbeitsloseninformationen für die Öffentlichkeit, die unter [www.amstat.ch](http://www.amstat.ch) abgerufen werden können.

Komplexere Auswertungen rufen nach einer neuen Infrastruktur, wie Dr. Elmar Benelli, Leiter DataWarehouse LAMDA beim SECO, feststellt: „Es gab immer mehr Performanceprobleme. In der bestehenden Systemarchitektur wären diese nur mit ziemlichem Aufwand zu bewältigen gewesen. Einschränkende Faktoren waren die physische Trennung von Datenbanksystem und Daten sowie der Datentransport über das Netzwerk, das ja auch von anderen Teilnehmern benutzt wird.“

Zusammen mit dem für das Data Warehouse verantwortlichen Partner Saracus suchte das Team von Dr. Benelli ab dem Frühling 2011 bei verschiedenen Anbietern nach einer neuen Datenbankanlösung, die den Anforderungen von „Big Data“ genügt. Fündig wurde man bei EMC Greenplum: Im Dezember 2011 fiel der Entscheid für eine Proof-of-Concept-Installation, in der die EMC Greenplum Database das bisherige Datenbanksystem ersetzt. Die Greenplum-Datenbank arbeitet mit massiv-paralleler Verarbeitung (MPP). Ein Master-Server steuert die Verarbeitung auf beliebig vielen Segment-Servern – die Lösung ist flexibel skalierbar und lässt sich auf unterschiedlichen Hardwareplattformen betreiben. Der Proof of Concept der Arbeitsmarktstatistik umfasst einen Master-Server, ergänzt durch einen zweiten Server für Redundanz, sowie vier Segment-Server.

## Flexibilität im Zentrum

„Die Flexibilität bei der Hardware war matchentscheidend“, hält Benelli fest. „Unser Data Warehouse ist mit rund 500 GB, davon 200 für die Data Marts, relativ klein. Wir benötigen also vor allem auch Skalierbarkeit nach unten. Ausserdem erlaubt uns EMC Greenplum den Betrieb auf preisgünstigen Industriestandard-Servern, die wir bei den üblichen Bundeslieferanten unkompliziert beschaffen können. Wir sind damit nicht von einem bestimmten Hardwareanbieter abhängig und bleiben auch in Zukunft flexibel. Bei den anderen Lösungen, die wir evaluierten, sind meist auch proprietäre Komponenten involviert“.

Die hohe Flexibilität bringt aber auch einen gewissen Mehraufwand mit sich, wie Elmar Benelli anmerkt: „Die EMC-Greenplum-Software liess sich einfach installieren. Auch die Migration des Frontend bot keine Probleme, sie war in einem halben Tag erledigt. Hardwareseitig war die Installation jedoch schwieriger als gedacht. Es war etwas aufwändig, die Server so zu konfigurieren, dass die EMC Greenplum Database optimal darauf läuft. Der Teufel steckt wie so oft im Detail.“ Mit einer vorkonfigurierten Data Computing Appliance, die von EMC Greenplum ebenfalls erhältlich ist, gäbe es solche Probleme nicht. „Dafür hätte der Betreiber keine freie Wahl der Hardware mehr, und dies war für uns besonders wichtig“, erläutert Benelli den Entscheid für die gewählte Software-Only-Variante.

## Leistung deutlich gesteigert

Flexibler Umgang ist für Elmar Benelli aber nicht nur bei der Hardwareauswahl wichtig. „Im BI-Umfeld muss man Flexibilität generell höher gewichten als Stabilität – die Anforderungen ändern sich naturgemäss ständig, anders als etwa in einem Transaktionssystem, wo die Nutzer grundsätzlich möglichst wenig Wechsel wünschen.“

Die mit dem Proof-of-Concept angezielte und auch erreichte Performance-Steigerung trägt zur erwünschten Flexibilität bei: „Im Frontend zeigt sich eine massiv bessere Leistung. Auch bei komplexen Abfragen erscheinen die Ergebnisse rasch“, schildert Elmar Benelli eine Erkenntnis aus der Testinstallation. Und: „Falls es bei einzelnen Auswertungen doch wieder zu Performance-Problemen kommen sollte, haben wir genügend Luft für Optimierungen, angefangen von der Partitionierung der Datenbank bis zum Einsatz von Distribution Key und Column Views. Wir haben bisher noch keine dieser Optimierungsmöglichkeiten genutzt.“ Der ETL-Teil, wo die Informatica-Plattform zum Einsatz kommt, biete ebenfalls bereits ansprechende Ladezeiten, müsse hingegen im Detail noch auf die neue Datenbank angepasst werden.

„Extrem schnell geht nun das Kopieren der Datenbank“, meint Elmar Benelli begeistert. „Wenn das Data-Warehouse-Team früher für Test- und Entwicklungszwecke oder für leistungshungrige Spezialauswertungen eine separate Datenbank aufbereiten musste, dauerte dies zwei Wochen. Mit der EMC Greenplum Database braucht es dazu nur einen Kopiervorgang, der in zwanzig Minuten erledigt ist.“

## Erfolgreiches Testprojekt mit Zukunft

„Das Team aus Mitarbeitern des SECO, der Saracus AG, des Bundesamts für Informatik und Telekommunikation (BIT) als Lösungsbetreiber sowie von Microstrategy und von EMC hat sehr gut zusammengearbeitet. Die Techniker haben sich schnell gefunden, waren immer gut erreichbar und haben Probleme jeweils zuverlässig gelöst.“ Laut Elmar Benelli konnte die Testphase Anfang Sommer 2012 erfolgreich abgeschlossen werden. Nun werden das bestehende System und die neue EMC-Greenplum-Lösung einige Monate parallel betrieben und dabei letzte Bereinigungen und Optimierungen eingebaut. Spätestens Ende Jahr soll die Schweizerische Arbeitsmarktstatistik dann definitiv über die neue Datenbank laufen.

## WEITERE INFORMATIONEN

Nutzen Sie die Möglichkeiten, die Big Data bietet - mit Produkten, Services und Lösungen von EMC Greenplum. Weitere Informationen finden Sie unter [www.greenplum.com](http://www.greenplum.com).