

suggestions for reducing data redundancy

Tony Damiani, Jeff Christen and Sarah Christen met on 8/19/2011 to discuss options for reducing data redundancy. Notes for this meeting below.

after STARS go live there was a pilot project to allow campus developers to create 'shared functions'. These functions were developed in a development environment, migrated to and tested in test then voted on by a campus developer group. Those that were deemed useful to all were migrated to production for others to use and are still being used today. This pilot also allowed for local data to be brought into a shared environment. there were two ways for this to happen. 1. If data was something that all campus groups could use, useful mapping tables or other commonly used data, these were brought in the production datamart as SVAT (Student Value Added Tables). 2. If the data was only used by one area it was often done in a private (personal) schema but only used in the test environment, which is a nightly copy of production. This met the needs for the pilot project. Along with this project there was a group of people from CIT, Central Student office and campus developers who met on a regular basis to facilitate this pilot project.

Action: Sarah will follow up with Stephanie Herrick and Deb Fyler who lead this Pilot for any notes or suggestions on how this could be done again and what improvements might be possible.

We discussed trying to recreate a similar project now for KFS (as well as KC and workday right around the corner) data and to create an environment which encourages reduction of data redundancy. We also talked about standard use of tools but decided to focus first on reducing redundancy and creating solutions to facilitate this process.

Today the only option we offer campus for joining local data with central administrative data is a Direct Connect id which is used to pull the data out of the central system. Campus developers can then create a new database, supported and maintained by the unit which includes copies of large portions of the central datamart as well as their local data. Allowing campus to import their local data into the central data stores, to be joined with central administrative data would result in cost savings in support, maintenance and database space.

Goals of this pilot would need to include:

low maintenance for central DBAs. While we want to realize the campus cost savings that could be gained by this type of change, we also need to ensure that processes are efficient and don't result in huge efforts on a team which already carrying a heavy load.

should be a funded pilot/project so that campus does not need to 'pay' for database space

need to ensure good QA in test prior to migration to prod to avoid bad data joins in a production environments which could result in bringing down production reporting for campus users.

Each IT service center has a separate database schema within the production warehouse. This allows campus developers to control who has access, and what data gets pulled into their individual schema. Because it is in the same database, joins across schemas are possible

