**Assegurando a Qualidade dos Dados**

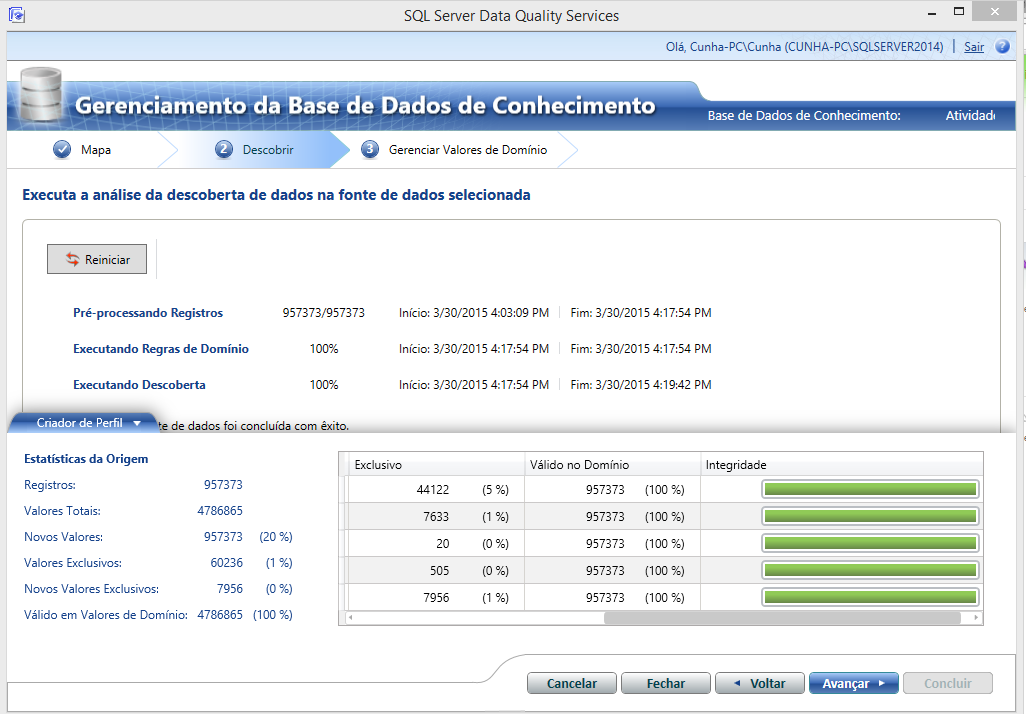
Em um processo de *data warehouse*, é de fundamental importância, que seja assegurada a qualidade dos dados. Para isto, temos de garantir que as informações fornecidas ao *data warehouse,* estejam corretas. Portanto, deve-se gerenciar as informações em uma empresa, envolve a integração dos dados, a limpeza dos dados, a correspondência dos dados para remover qualquer duplicata, a padronização dos dados, o enriquecimento dos dados, a adequação dos dados com os requisitos legais e assim por diante, bem como o armazenamento dos dados em um local centralizado com todas as configurações de segurança necessárias.

Neste projeto, optei por realizar a tarefa de assegurar a qualidade dos dados, usando a ferramenta *Data Quakity Services* (**DQS**), pois a mesma oferece todos os recursos necessários para permite a limpeza, a correspondência, a padronização e o enriquecimento dos dados; portanto, é possível, através dessa ferramenta, enviar informações confiáveis para business intelligence e data warehouse.

**Assegurando a qualidade dos dados da tabela dimAlunosRepDiscCurso**

A tabela dimAlunosRepDiscCurso, é uma das muitas tabelas que servirão de entrada para o data warehouse, ela contém informações dos alunos, em função de sua situações (Aprovado, reprovado, cursando, etc), por disciplinas dentro dos cursos, em cada ano e período letivo. Vamos criar uma base de dados de conhecimento do DQS denominada **AlunosReprovadosPorDisciplinasECurso** com o conhecimento (metadados) sobre dados do alunos reprovados por disciplinas e cursos. Nesta base de dados de conhecimento serão executadas as atividades de limpeza e correspondência nos dados. A atividade de limpeza identifica dados incorretos ou inválidos, corrige os dados incorretos, propõe correções/sugestões, padroniza os dados e enriquece os dados com mais informações. A atividade de correspondência compara dados e identifica registros semelhantes (mas um pouco diferentes) nos dados que o ajudam a remover dados duplicatas.

Para esta base de conhecimento, optamos em analisar os domínios: matricula, cod\_disciplina, cod\_curso, situação\_matricula e coeficiente\_redimento. A figura 1 exibe a base de conhecimento, após ter sido feita uma análise de conhecimento na tabela.



**Figura 1.** Análise da descoberta de dados na fonte de dados dimAlunosRepDiscCurso.

**Fonte**: autor

Na figura 1 pode-se ver a integridade, todas com as barras verdes, indicando que não há ausência de dados, em nenhum dos domínios analisados. A figura q1 também mostra o total de registros analisados (957373).

A figura 2 mostre que todos os dados estão corretos, para todos os domínios analisados. Não existe dados errados ou inválidos.

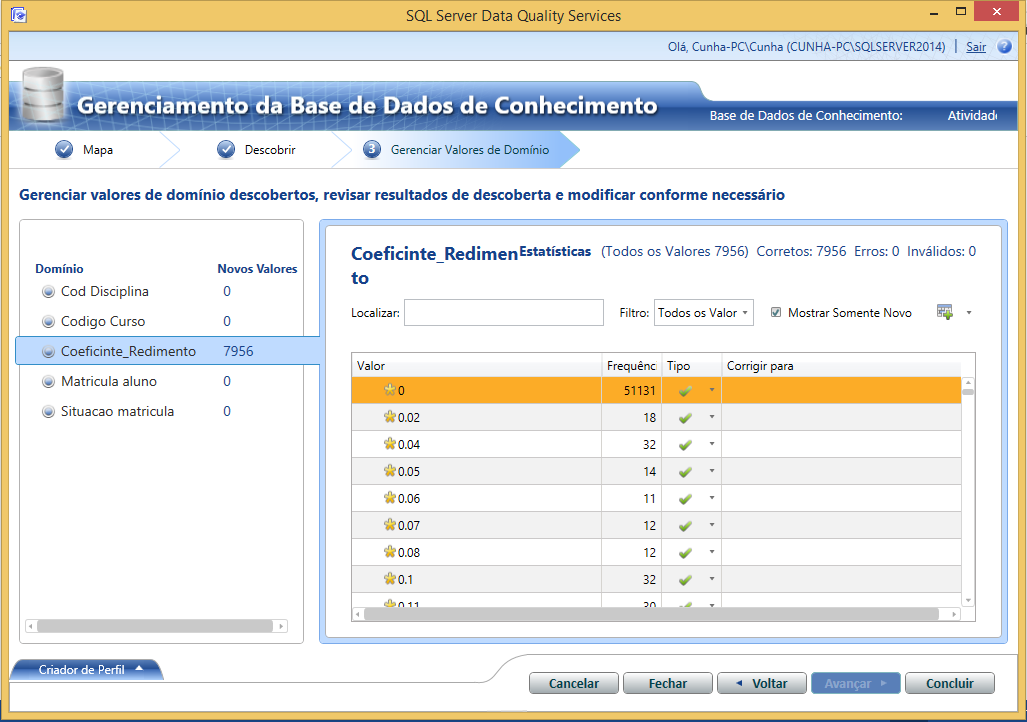


Figura 2. Tela de Gerenciamento da Base de dados de Conhecimento.

Através desta tela, pode-se ver para cada domínio se existe dados incorretos ou inválidos e dá a opção de corrigi-los, adicionando novo valor ao domínio. A figura 1 mostra que para o domínio Coeficiente\_Redimento, todos os valores estão corretos.