

4.5. Sobrevivência

Os registros de dados devidamente criticados e validados, após terem sido padronizados e possivelmente enriquecidos, são comparados contra a base de dados alvo e devidamente agrupados com a visão corporativa.

No entanto, a gravação na base única é feita usando o registro conhecido como Golden Record (Melhor registro). O Golden Record é composto de informações de quaisquer das instâncias desta entidade, podendo conter dados de diversos registros cuja informação tenha sido definida por meio de uma regra de sobrevivência estabelecida.

As regras de sobrevivência de registros visam estabelecer critérios para que o registro resultante do processo de unificação contemple as melhores informações possíveis (mais adequadas ao negócio), de forma automatizada;



Um diagrama de sobrevivência de registros. À esquerda, uma seta curva azul aponta para a direita, indicando a seleção de um registro vencedor. À direita, há duas tabelas. A primeira tabela contém quatro registros com diferentes combinações de nome, sobrenome, data de nascimento e ID. A segunda tabela, localizada abaixo da primeira, contém um único registro que é o vencedor, destacado em negrito.

Prof	Antonio J Fernandes	19450815	20030514
Dr	Antonio José Fernandes		20021220
Prof	Antonio José Fernandes	19450815	20080630
Prof	Antonio José Fernandes	19540815	20080425

Prof	Antonio José Fernandes	19450815	20080630
-------------	-------------------------------	-----------------	-----------------

Como cada empresa possui sua particularidade, a solução permite a customização das regras para a definição/escolha da melhor informação. Essa escolha pode levar em consideração diversos critérios como:

- **Menor / Maior (quantidade de caracteres);**
- **Mais frequente / Mais frequente não nulo;**
- **Igual a / Diferente de;**
- **Maior que / Menor que;**
- **Pelo menos um;**
- **Nomeação de atributo por meio de verificação dos sistemas de origem;**

A partir da aplicação das regras é possível verificar os resultados e, caso necessário, alterá-las para sanar qualquer inconsistência verificada nos golden records.

Além dessas regras mais simples, é possível customizar regras mais complexas que levarão em consideração mais de uma informação ao mesmo tempo como por exemplo a regra abaixo:

Valor **não nulo**, conforme priorização abaixo. Para desempate levar em consideração o registro **mais atual**

1º - **SISTEMA_RH**

2º - **SISTEMA_MARKETING**

3º - **SISTEMA_FINANCEIRO**

Vamos entender a regra acima:

- Primeiro é priorizado a informação **não nula**, independente da origem;
- Com o grupo de registros que possuem a informação preenchida é priorizado por exemplo a origem da informação. O usuário de negócio com o conhecimento corporativo, definiu por exemplo que o nome da pessoa no **SISTEMA_RH** tem uma qualidade superior aos outros sistemas pois são validados por exemplo no E-Social.
- Por fim, caso esse sistema possua registros duplicados para desempate, o analista por exemplo definiu que o registro **mais atual** deve prevalecer.

Outro ponto importante e destacado na solução é que essas regras são customizadas por grupo de informação. Para a escolha do **NOME** é possível criar algo semelhante ao exemplo acima, porém para a escolha do melhor **E-MAIL** é possível definir uma regra diferente. Nesse caso pode ser que o sistema de marketing tenha uma informação mais atual e mais confiável. E é essa composição que formará o **Golden Record**.

Revision #17

Created 22 July 2022 12:29:29

Updated 13 December 2022 10:43:06