

## Desafio Técnico Simbiose – Sincronização entre Elasticsearch e Cassandra

### Tecnologias Utilizadas:

- Cassandra (<http://cassandra.apache.org/>)
- Elasticsearch (<http://www.elasticsearch.org/>)
- Python (preferencialmente)

### Descrição do Desafio:

Criar um sistema que automaticamente sincronize dados entre o Cassandra e o Elasticsearch.

### Funcionamento Esperado:

Deve ser um programa que rode como "daemon" e pode ser escrito na linguagem de sua preferência (mas Python seria desejável, caso isso não seja um problema).

O programa deve verificar periodicamente (período deve ser parametrizável) se novos dados chegaram no Cassandra ou no Elasticsearch e sincroniza-los entre os dois “bancos” automaticamente.

Caso novos dados tenham chegado em ambos, ambos deverão ser sincronizados entre Cassandra e Elasticsearch.

Caso um mesmo dado tenha chegado tanto no Cassandra quanto no Elasticsearch (mesmo ID), deverá ser considerado o dado mais atual.

Exemplo: um dado com ID 45 chegou no Elasticsearch e também no Cassandra, mas o que chegou no Elasticsearch é mais atual, então o dado do Elasticsearch deve ser atualizado no Cassandra. Se o do Cassandra fosse o mais atual, deveria ser atualizado no Elasticsearch.

### Disponibilização do Código e Documentação:

É importante que seja criado um projeto no GitHub para hospedar esse fonte e que o mesmo esteja bem organizado, apresentável, tenha um bom README e seja intuitivo de se entender.

### Expectativa:

De acordo com nossa experiência, mesmo contendo tecnologias novas, acreditamos que esse projeto/desafio possa ser facilmente concluído em uma semana.

### Informações Adicionais:

- Você deve definir o modelo de dados no Cassandra e no Elasticsearch para que o requisito seja cumprido. Você pode definir quais colunas serão armazenadas no Cassandra ou no Elasticsearch. Apenas tenha em mente que teremos várias colunas de dados que serão sincronizadas entre os “bancos” (Cassandra e Elasticsearch), independentemente das colunas/campos que você criar no modelo.
- Os IDs serão UUIDs do tipo 4 ([http://en.wikipedia.org/wiki/Universally\\_unique\\_identifier](http://en.wikipedia.org/wiki/Universally_unique_identifier)), tanto no Cassandra quanto no Elasticsearch. O ID de um registro deve ser o mesmo no Cassandra e no Elasticsearch.
- Pode considerar que nunca são feitos deletes e que nunca é mudado o ID de um documento em um update. Contudo, lembre-se que em caso de conflito sempre deve valer a versão mais nova. Se você tiver conflito entre um insert e um update ou entre 2 updates, por exemplo, valerá o que foi feito por último.
- Não é permitido usar um banco de dados para sincronizar os dados entre os 2 bancos. Esteja também ciente que não conseguimos prever quais tipos de dado serão gravados, por exemplo, não sabemos se o que está sendo sincronizado é uma tabela de produtos, de clientes ou de logs.