

Ficou curioso?

Nessa aula falamos um pouco mais sobre os perfis da engenharia de dados. Gostaria de ressaltar que a alocação entre esses perfis ou o generalista não vem em nenhuma receita. Equipes pequenas usualmente começarão em perfis generalistas e poderão se especializar de acordo com o necessário. Mas nada impede que se aloquem de novo como generalistas em caso de necessidade.

Já sobre o caminho que o dado faz, é importante que ele seja pensado como um todo, para que as comunicações sejam bem pensadas. A forma como ele é armazenado por exemplo vai refletir no resto do processo. Por exemplo, caso ele seja enviado a um banco de dados a transferência será diferente em relação a um armazenamento em arquivos de planilha.

Porém, mesmo que as etapas sejam pensadas como um todo, o melhor é que a implementação seja distribuída, ou *serverless* como já comentamos. Ou seja, um serviço para coleta, um serviço para transformação e um outro para armazenamento. Por que isso é útil? imagina que você possui tabelas de três equipamentos sendo carregadas pra nuvem e elas são trabalhadas de forma independente em cada uma das etapas, que também são independentes.

O problema em apenas uma dessas tabelas não atrapalhará o processo das outras, essa é a vantagem de processamento independente. Agora, caso o problema ocorra na etapa de armazenamento, é possível que façamos uma estrutura em que a transformação continue ocorrendo normalmente, e quando o armazenamento voltar a operar, ele carregue os dados transformados, sendo essa a vantagem de serviços modularizados.

Outra vantagem de modularizar serviços é, por exemplo, se você encontrou uma ferramenta que te atende melhor, se você mantiver as entradas e saídas, pode-se fazer uma troca de forma mais fácil da antiga pra nova.

Ferramentas AWS que podem ser úteis para você se informar:

- AWS Cloud9: Instâncias para desenvolvimento, podendo ser feito por mais de uma pessoa online.
- AWS Codecommit: Armazenamento e versionamento de código.