

ELK Stack - Análise de Dados Governamentais

Daniel Vitor Ferreira Silva¹, Rafael Souza Oliveira²

¹Instituto Federal de Goiás (IFG)
Anápolis – GO – Brazil

Resumo. Desde 2012, com o início da vigência da Lei nº 12.527, os dados de gastos dos governos devem ser disponibilizados publicamente. Desde então, o problema deixou de ser o acesso a tais dados, mas sim como tratar, visualizar e interpretar esses dados de forma mais eficiente. Por isso, esse relatório faz uma apresentação educacional do chamado Elastic-Stack enquanto busca construir uma forma alternativa de fornecer visualização dos dados de gastos disponibilizados pelo governo. Além disso, também é objetivo deste ensinar ao leitor como reproduzir a construção de tal visualização.

1. Introdução

A Lei nº 12.527, sancionada em 18 de novembro de 2011, regulamenta o direito constitucional de acesso dos cidadãos às informações públicas e é aplicável aos três poderes da União, dos estados, do Distrito Federal e dos municípios. Esta Lei representou um importante passo para a consolidação do regime democrático brasileiro e para o fortalecimento das políticas de transparência pública.

Pensando nisso, este trabalho foi desenvolvido com o intuito de usar as ferramentas proporcionadas pelo chamado Elastic-Stack para criar uma visualização desses dados garantidos pela Lei de Acesso à Informação. Além disso, o principal objetivo aqui é apresentar as ferramentas e fornecer as informações necessárias para que qualquer pessoa interessada possa entender e replicar o processo de casa.

Por sua vez, o Elastic-Stack nada mais é que um conjunto de softwares que trabalham em conjunto, para em resumo: Obter dados em qualquer formato, fazer buscas e análises nesses dados e por fim, criar visualizações.

2. As Ferramentas

Nessa seção, discutiremos um pouco mais sobre os 3 pilares principais que compõem o Elastic-Stack ou ELK-Stack, sendo eles: Elasticsearch, Logstash, Kibana.

2.1. Elasticsearch

Trata-se do elemento central do nosso Stack. Ele permite armazenar, pesquisar e analisar grandes volumes de dados com rapidez e quase em tempo real, além de fornecer respostas em milissegundos. Ainda, o Elasticsearch é capaz de obter respostas de pesquisas de forma muito rápida porque, em vez de pesquisar o texto diretamente, ele pesquisa sobre um índice. Ele usa uma estrutura baseada em documentos em vez de tabelas e esquemas, como bancos de dados tradicionais.

A estrutura básica do Elasticsearch é composta por: Documentos, Índices, Nodes, Clusters e Shards. De forma resumida, Documentos são a unidade básica de informação

que pode ser indexada no Elasticsearch. Já os Índices são conjuntos de documentos que têm relação entre si. Um Node é uma instância do Elasticsearch, que armazena dados. Por sua vez, um Cluster é uma coleção de nós que se conectam juntos. Quando um nó é criado, por consequência, um cluster é criado. Por fim, o elasticsearch fornece a capacidade de subdividir o índice em várias partes; e essas partes são chamadas de Shards

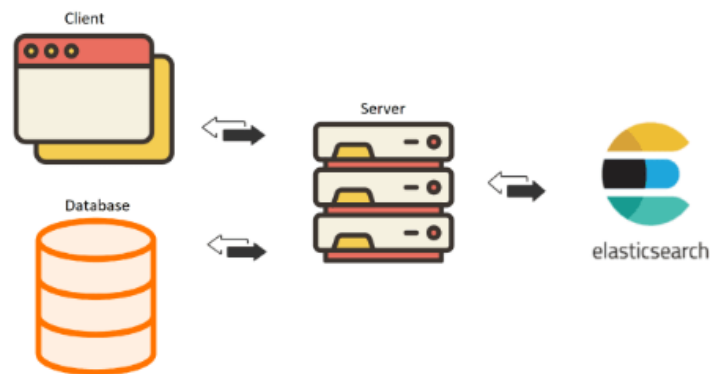


Figura 1. Simplificação de como o elasticsearch funciona.

2.2. Logstash

É uma ferramenta de coleta e tratamento de dados com recursos em tempo real, funcionando como uma espécie de pipeline de dados. Porém, para este artigo usamos apenas uma entrada de dados, a fim de facilitar a compreensão e reprodução do processo. O Logstash é dividido em Input, Filter e Output.

No input são passadas as informações de onde os nossos dados serão buscados por ele. No nosso caso, um único arquivo CSV. O filter é a parte responsável pelo tratamento dos dados, seja para fins de normalização ou de manipulação para conseguir apenas os dados que realmente interessam. Já o output é quem manda esses dados filtrados para o elastic e cria, no mesmo, um índice para eles.

2.3. Kibana

Por fim, temos o Kibana: Uma ferramenta de visualização e exploração de dados. Ele oferece recursos poderosos e fáceis de usar, como histogramas, gráficos de linhas, gráficos de pizza, mapas de calor e suporte geoespacial integrado. Além disso, ele, também, fornece uma estreita integração com o Elasticsearch, o que torna o Kibana a opção padrão para visualizar dados armazenados no mesmo.

A interface do Kibana permite que os usuários consultem dados nos índices do Elasticsearch e visualizem os resultados por meio de opções de gráficos padrão ou apps integrados como o Lens, o Canvas e o Maps. Ainda, os usuários podem escolher entre diferentes tipos de gráficos, alterar as agregações de números e filtrar para ver segmentos específicos de dados. Dessa forma, no trabalho proposto esta será a ferramenta usada para a criação de um Dashboard, composto de diversos desses elementos visuais que permitirão uma compreensão e análise mais profunda dos dados do governo.

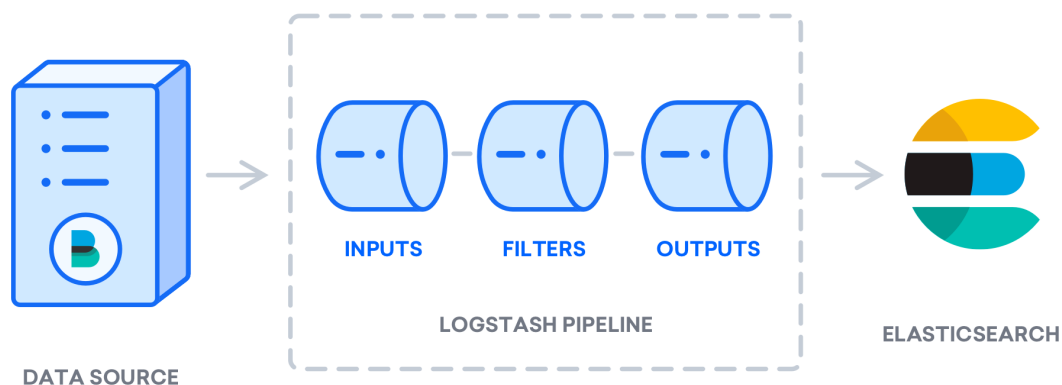


Figura 2. Funcionamento do Logstash.

3. Os Dados

Uma implicação direta da Lei nº 12.527 foi a criação dos chamados Portais de Transparência. Que nada mais são do que os sites criados para disponibilizar para a população como o dinheiro público é gasto. Ao se explorar tais Portais, é possível encontrar os mais diversos tipos de dados que abrangem o escopo de gastos governamentais. Como por exemplo, Receitas, Despesas, Transferências, Folhas de Pagamento e entre outros.

Para esse projeto, foram utilizados como objeto para aplicação das ferramentas citadas até aqui, dados retirados do Portal da Transparência de Anápolis - GO. Mais especificamente, da folha de pagamentos dos funcionários municipais, referente a Maio de 2022. Tal folha de pagamento é disponibilizada pelo site como um extenso arquivo CSV, contendo os seguintes campos: Nome, Referência, Matrícula, Tipo de Folha, Proventos, Descontos, Líquido, Cargo, Vínculo, Data Admissão e Lotação. Ou seja, todos os dados necessários para identificar um servidor público, sua função, salário, entre outros. Porém, seja no formato de tabela ou de CSV, é difícil extrair algum conhecimento desses dados em seu estado bruto. Portanto, o objetivo buscado aqui foi de usar os softwares, e seus conceitos, introduzidos anteriormente, para se tratar e melhorar a visualização de tais dados. Assim, facilitando o processo de se adquirir informações relevantes através dos mesmos.

4. Replicando

Para se replicar o processo são necessários os seguintes passos: 1) Baixar o repositório; 2) Seguir o processo de instalação como descrito no arquivo README.

Em seguida, é necessário executar o Logstash para buscar e carregar os dados do arquivo CSV. Para isso, entre na pasta 'logstash' e coloque o arquivo de dados dentro da pasta 'csv'. Caso queira fazer com um arquivo de dados diferente do usado neste trabalho, abra a pasta 'pipeline' dentro de 'logstash' e mude o nome caminho do arquivo

para o nome do qual estiver usando; em seguida, ajuste o campo ‘filter’ com os campos correspondentes aos do novo CSV. Por fim, mude o nome do index na seção ‘output’.

5. Visualização e Resultados

Caso obtenha sucesso seguindo os passos descritos na seção anterior, terá obtido um resultado parecido com o mostrado na figura 3, abaixo.



Figura 3. Dashboard.

A figura 3 apresenta um dashboard com vários elementos visuais exibidos de forma simultânea. Onde, todos passam a mesma informação: Quantidades de funcionários por órgão, porém cada elemento agrega essa informação com seu próprio atributo visual extra.

Além disso, a interface do Kibana permite adicionar filtros à visualização, alterando os gráficos do dashboard em tempo real. Na figura 4, podemos ver o mesmo dashboard apresentado na figura 3, porém agora com filtros para mostrar somente os servidores da Educação, com o cargo de Auxiliar de Serviços de higiene e com período de alimentação de 40h mensais.

6. Conclusão

Contudo, pôde se observar que o Elastic-Stack se apresenta de fato como uma poderosa ferramenta de monitoramento e observação de dados. Especialmente quando comparado aos Portais de Transparência, proporcionando algumas vantagens com relação a esse, como a liberdade para escolha do intervalo de tempo para análise e a possibilidade de adição de filtros em tempo real. As diferentes visualizações disponibilizadas em conjunto, na forma de dashboard, permitem a gestores e usuários leigos a identificação rápida de informações de maior relevância, proporcionando efetivamente uma melhoria na transparência pública.

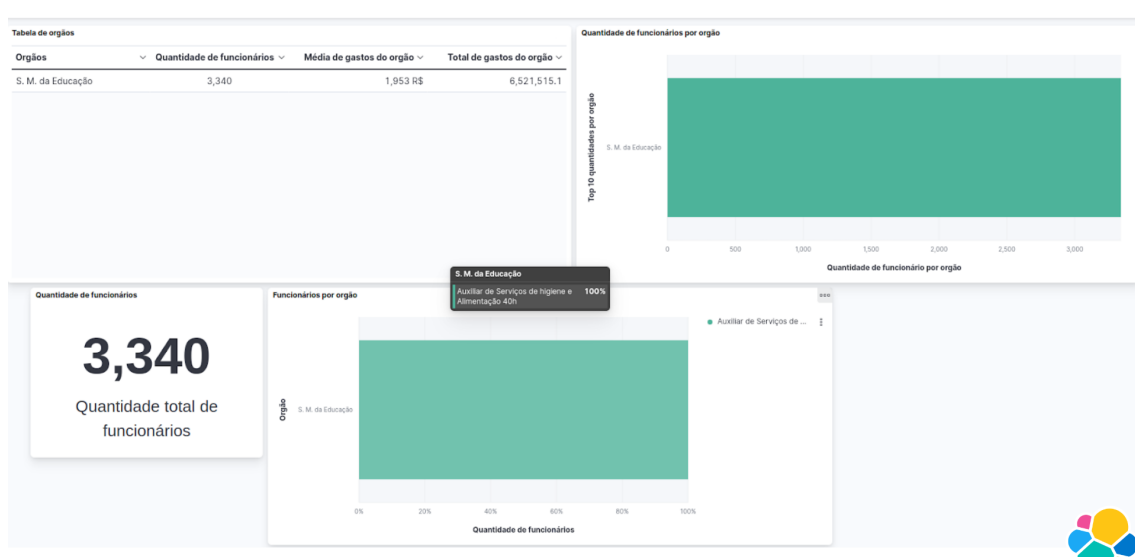


Figura 4. Dashboard com filtros.