

## **Potencializando o uso de dados em políticas públicas através do primeiro Datalake municipal no mundo no Rio de Janeiro.**

Ramon Pires Carnaval Barbosa

### **Resumo**

Este trabalho aborda a implementação do primeiro DataLake municipal do mundo, no Rio de Janeiro, destaca sua importância para o setor público, o potencial de uso para políticas públicas, levando em consideração aspectos como escalabilidade, qualidade dos dados, governança, infraestrutura e políticas públicas.

A implementação do DataLake do Rio de Janeiro permitiu a integração de diversos dados governamentais em um único repositório centralizado, possibilitando uma visão integrada e unificada dos dados. A solução apresenta uma escalabilidade que permite que os dados sejam processados e analisados em tempo real, contribuindo para aprimorar a gestão pública e o desenvolvimento de políticas públicas mais eficientes e efetivas.

O artigo destaca a importância da qualidade dos dados no DataLake, uma vez que é fundamental que os dados sejam precisos, completos e confiáveis para a tomada de decisão. Além de garantir a qualidade da informação, também se adota a governança de dados adequada, incluindo políticas claras de acesso, segurança, privacidade e conformidade regulatória.

Por fim, o estudo conclui que o DataLake do Rio de Janeiro representa uma iniciativa pioneira que pode ser replicada em outras cidades e países, contribuindo para aprimorar a gestão pública e o desenvolvimento de políticas públicas mais eficientes e efetivas.

## Resumo Expandido

O Datalake surge no contexto de rompimento de paradigmas tecnológicos e de cultura institucional condicionada a burocracia extensa e com excesso de rito e entraves que dificultavam o acesso e o desenvolvimento de trabalhos e políticas públicas orientados pelos dados. Outro fator relevante é a capacidade de armazenar e processar limitados pelos recursos tecnológicos.

O percurso necessário para acessar dados exigia envolver diversos atores operacionais, além de gestores responsáveis em autorizar o acesso. Isto significa que a pessoa motivada a elaborar estudos e análises ficavam mais distantes das bases. A cultura de dados precisava ser incentivada e disseminada com casos e exemplos que motivassem o engajamento dos interessados.

A infraestrutura envolve tecnologia da *Google Cloud Plataforma*, e a solução utilizada é a *BigQuery* na implementação do DataLake, permitindo a análise de grandes volumes de dados com rapidez e eficiência. A escolha da *BigQuery* foi baseada em sua capacidade de processamento escalável, além de permitir o gerenciamento de dados estruturados e não-estruturados.

Portanto, os objetivos do projeto consistem em reduzir o tempo de acesso aos dados, facilitar a localização e conhecimento da existência de determinadas bases de dados, aprimorar a qualidade dos dados, isto é, torná-los de fácil compreensão e prontos para o uso. Oferecer infraestrutura única, pois diversos há diversas bases de dados que estão em tecnologias diferentes, o que dificulta o acesso e sobretudo integração entre diferentes conjuntos de dados, logo, centralizar em única interface de acesso é crucial. Ainda é importante citar que engajar, motivar e disseminar a cultura de dados também se faz deveras importante.

O Fluxo de dados consiste na extração das fontes originais, carregamento para ambiente em nuvem onde ficará armazenado. Esta sequência retrata o processo de ingestão de dados no Datalake. Por outro lado, os conjuntos de dados são organizados em projetos, no sentido de entidades organizadoras, ou seja, um contêiner para armazenar os dados. Estes projetos podem ser públicos e não- públicos, o primeiro se refere ao projeto que podemos encontrar dados abertos, enquanto o segundo o acesso é restrito aos órgãos controladores ou com quem esteja compartilhado.

Cabe destacar que a participação social é temática importante para a discussão, é uma possibilidade de expandir a captação e a participação propriamente dita da sociedade civil e organizações sociais engajadas no setor dos dados e que necessitam do suporte do Estado para ampliar sua potência analítica e de

processamento. O diálogo entre essas esferas proporcionam a resolução de carências existentes para ambos e, sobretudo, fortalece a perenidade do projeto enquanto produto e serviço de Estado.

Outro aspecto importante a ser considerado consiste na modulação da governança estabelecida para engajar, envolver e desenvolver órgãos e projetos estratégicos e setoriais. Obter a parceria dos órgãos diante das necessidades e oportunidades em implementar uma solução que alcance a população é a estratégia estabelecida para que todo trabalho seja perene. Os dados possuem poder de guiar e direcionar as tomadas de decisões a partir de *insights* originados em análise exploratória. No entanto, os indicadores estruturam rotinas estratégicas e pragmáticas que criam informações sobre a cidade. (Kitchin et all, 2014)

Dada escalabilidade, volumetria e potencial de processamento da solução, diversos projetos podem rodar integrando inúmeras bases de dados, alcançamos a ideia de *think tank*. Pois, elaborar um projeto de dados envolvendo um órgão ao menos com problemas complexos, nos conduz à explosão de ideias que os dados nos permitem alçar, logo, traduzimos uma infinidade de possibilidades.

Entretanto, pensar uma solução não se reduz aos dados, envolve um sistema de produção intelectual. As fases do diagnóstico *latu sensu*, observação das áreas de negócio envolvidas, compreensão do público alvo, isto é, decodificar as dores do cidadão (persona), estabelecimento de ideação, prototipagem, validação, implementação e melhoria contínua da solução e concepção são elementos cruciais. Portanto, resumimos à lógica do *Design Thinking*.

A construção das ideias também reflete modelos externos pesquisados por outro método utilizado por diversos setores. O Benchmarking é ferramenta necessária para (re)orientar o caminho adotado nas soluções. Trabalhar inovação em governos é sempre um desafio porque ele se distingue do usual da lógica de mercado, pelo menos deveria. O Escritório de Dados já estabeleceu comunicação com outras cidades e governos como Fortaleza, Espírito Santo, Recife, Cidade do México. Ora como referência outra em busca de ratificação de suas estratégias. Neste sentido, conhece outras plataformas, soluções, modelos de negócios etc., agregam valor na formulação e concepção do produto desenhado e implementado.

A experiência adquirida, a infraestrutura de dados construída e utilizada nos conduziu para a busca de mais dados. Futuramente buscamos implementar o MAPRio (Monitoramento de Áreas Públicas). O projeto consiste na instalação de equipamento de câmeras na frota de ônibus e carros institucionais da Prefeitura com tecnologia embarcada para identificar locais e ativos públicos que requeiram manutenção e conversação, a partir da leitura por inteligência artificial que registrará no banco de

dados. Com isso, será possível distribuir as demandas aos órgãos públicos competentes. Com isso, seria possível estabelecer uma amplitude de medições e parâmetros para obter melhores resultados e satisfação da população. Todavia, ainda requer maturação sobre o modo de relação entre tecnologia, dados, serviços, métricas e população.

Hodiernamente, o Datalake reúne mais de cinquenta por cento dos órgãos municipais. Há aproximadamente trinta tabelas de dados públicos, enquanto de modo não-público temos 329 tabelas também estruturadas e operacionais. No âmbito de projetos, em nosso portfólio, temos o número total de 46, tendo entregado mais de 50%. A realização disso tudo envolve uma equipe apenas de 9 pessoas, considerando que há demandas administrativas e burocráticas que requer atenção e urgência.

Por fim, à luz da reflexão do modo de captação dos dados, os impactos que podem ocorrer em uso indiscriminado, a ausência de instrumentos claros de acesso, a desregulação institucional ou a manipulação política são elementos iniciais que devem ser observados diante da responsabilidade do Estado enquanto responsável por manter a segurança íntegra. Sobretudo, não romper os limites do uso para políticas públicas eficientes e eficazes em razão do controle social, esse é um grande desafio e debate que carece de vozes.

#### **Referências:**

1. KITCHIN, Rob; LAURIAULT, Tracey P.; MCARDLE, Gavin. Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards, *Regional Studies. Regional Science*, 2014.
2. D, Edwards; J. C., Thomas. Developing a municipal performance-measurement system: Reflections on the Atlanta dashboard. *Public Administration Review*, 65 (3), 369–376.10.1111/puar. 2005
3. Dashboard Gerencial da Cidade de Londres. Disponível em: <<https://citydashboard.org/london/>> Acesso em 25 de maio de 2023.
4. Portal de dados abertos da Cidade do México. Disponível em: <<https://datos.cdmx.gob.mx/>> Acesso em 24 de maio de 2023.
5. CASSINO, João; SOUZA, Joyce; SILVEIRA, SÉRGIO. Colonialismo de Dados: Como opera a trincheira algorítmica na guerra neoliberal. São Paulo. SP: Autonomia Literária, 2021.