

Big Data no Setor Cultural: Uma Abordagem Inicial para a Gestão Pública no Ceará

Jeferson Kenedy M. Vieira¹, Davi Romero de Vasconcelos¹,
Marcos Dantas Ortiz¹, Alysson Diniz dos Santos²,
Georgia da Cruz Pereira², Amanda Lima Mendes³,
Natasha de Paula Farias³

¹Campus Quixadá – Universidade Federal do Ceará (UFC)
Quixadá, CE – Brasil

²Instituto UFC Virtual – Universidade Federal do Ceará (UFC)
Fortaleza, CE – Brasil.

³Instituto Mirante de Cultura e Arte
Fortaleza, CE – Brasil

{jefersonkenedy, daviromero, mdo}@ufc.br,
{alysson, georgia}@virtual.ufc.br,
{amanda.lima, natasha.depaula}@institutomirante.org

Abstract. This work presents an initiative for applying Big Data in cultural management in Ceará. The proposed solution aims to integrate and analyze data from various public sources to support cultural policy-making. The project has conducted a literature review and implemented a structured data collection and processing pipeline. Preliminary results include exploratory analyses and the development of an interactive dashboard for dynamic visualization.

Resumo. Este trabalho apresenta uma iniciativa para a aplicação de Big Data na gestão cultural do Ceará. A solução proposta busca integrar e analisar dados de diversas fontes públicas para apoiar a formulação de políticas culturais. O projeto já realizou um levantamento do estado da arte e implementou um pipeline de coleta e processamento de dados, estruturado em camadas. Os resultados preliminares incluem análises exploratórias e o desenvolvimento de um dashboard interativo para visualização dinâmica.

1. Contexto

O termo *Big Data* refere-se a conjuntos de dados tão extensos e complexos que não podem ser processados de maneira eficiente por ferramentas tradicionais de gerenciamento de dados [Wu et al. 2014]. Esses dados podem ser originados de diversas fontes, como redes sociais, websites, pesquisas, transações financeiras e registros de eventos culturais. Quando aplicado ao setor cultural, o *Big Data* possibilita a obtenção de *insights* em uma escala sem precedentes. A análise desses dados permite identificar padrões de consumo cultural, avaliar a popularidade de diferentes formas de arte e eventos, compreender tendências emergentes e mensurar os impactos econômicos das iniciativas culturais.

O investimento em cultura vai além do fomento às manifestações artísticas e da preservação do patrimônio cultural; trata-se de um fator estratégico para o desenvolvimento socioeconômico [Chen and Chen 2024]. O uso de *Big Data* no setor cultural pode fornecer aos tomadores de decisão, patrocinadores e artistas uma compreensão mais aprofundada do impacto de suas ações, permitindo decisões mais informadas e eficazes.

Essa iniciativa integra o Programa Cientista Chefe da Cultura (CC Cult), criado para articular as demandas da gestão pública com os conhecimentos desenvolvidos por pesquisadores de Instituições de Ensino Superior (IES). Desde sua criação em 2018, o programa tem se consolidado como uma política pública inovadora do Governo do Estado do Ceará, promovendo a colaboração entre universidades e o poder público estadual. Em seu segundo ciclo de execução (2023-2025), o CC Cult tem como foco o desenvolvimento de soluções tecnológicas voltadas à otimização da gestão cultural e ao aprimoramento da eficácia das políticas públicas na área [Almeida et al. 2024].

2. Processo adotado

Para embasar o desenvolvimento da solução de *Big Data* para Cultura, foi realizada uma *Rapid Literature Review* (RLR) [Ganann et al. 2010] com o objetivo de obter *insights* rápidos sobre o estado da arte na área. O estudo focou em identificar como o *Big Data* tem sido aplicado no ecossistema cultural e artístico, analisando seus impactos, principais tecnologias e ferramentas utilizadas, especialmente soluções *open source*, além das principais fontes de dados disponíveis. Foram consultadas bases renomadas, como Scopus, Science Direct, IEEE Xplore e ACM Digital Library, utilizando critérios como atualidade, disponibilidade de texto completo e relevância dos estudos. No total, foram analisados 27 artigos, que abordam aplicações de *Big Data* em áreas como gestão cultural, engajamento do público, inovação, impacto socioeconômico e preservação do patrimônio cultural.

Atualmente, os dados utilizados no projeto estão sendo analisados separadamente, conforme são extraídos de diferentes fontes estruturadas e abertas. Entre as principais bases exploradas estão o Mapa Cultural do Ceará, que detalha o ecossistema cultural do estado; o Ceará Transparente, com informações sobre contratos e investimentos da Secretaria de Cultura do Estado do Ceará (SECULT-CE); além de bases nacionais como as disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e a Pesquisa de Informações Básicas Municipais (MUNIC), que oferecem um panorama da gestão pública e suas relações com políticas culturais. No entanto, a perspectiva futura do projeto é desenvolver uma solução capaz de integrar e cruzar esses dados, permitindo análises mais amplas e estratégicas. Além disso, a incorporação de dados de redes sociais e indicadores qualitativos ampliará o alcance da ferramenta.

O processo de Extração, Transformação e Carga (ETL) adotado no projeto foi estruturado para garantir a coleta, processamento e integração eficiente de grandes volumes de dados provenientes de diferentes fontes. Inicialmente, os dados são extraídos de bases estruturadas, como o Mapa Cultural do Ceará¹, Ceará Transparente², MUNIC³, Sistema

¹<https://mapacultural.secult.ce.gov.br/>

²<https://cearatransparente.ce.gov.br/>

³<https://www.ibge.gov.br/estatisticas/sociais/educacao/10586-pesquisa-de-informacoes-basicas-municipais.html>

de Informações e Indicadores Culturais (SIIC)⁴ e Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua⁵, utilizando APIs e downloads de arquivos. Em seguida, esses dados passam por uma etapa de pré-processamento, que inclui a remoção de informações duplicadas, padronização de formatos e tratamento de valores nulos. Para garantir um fluxo automatizado e escalável, a orquestração desse pipeline é realizada pelo Apache Airflow, permitindo o agendamento e monitoramento das tarefas.

O processamento dos dados ocorre em múltiplos estágios, organizados em três camadas: Bronze, Silver e Gold. No nível Bronze, os dados são armazenados no Delta Lake sem alterações, preservando sua forma bruta. No nível Silver, os dados passam por transformações estruturais, como agregações, remoção de inconsistências e normalização. No nível Gold, as informações são refinadas para análises mais avançadas e integração com outras bases. Para processar esses dados em larga escala, é utilizado o Apache Spark, que permite a execução distribuída de operações complexas. Além disso, a análise exploratória e a visualização dos resultados são feitas com Python, utilizando bibliotecas como Pandas para manipulação de dados e Matplotlib para criação de gráficos informativos.

Uma versão de um *dashboard* interativo está sendo desenvolvida, à princípio apenas com dados coletados a partir do Mapa Cultural do Ceará, com dados do PROSIEC (Programa de Fortalecimento do Sistema Estadual da Cultura do Ceará). O *front-end* do projeto é desenvolvido com Next.js e React, utilizando Ant Design para a interface. A visualização de dados é feita com bibliotecas como Chart.js, Highcharts e D3.js, incluindo suporte para geovisualização. O código é estruturado em TypeScript e segue boas práticas de formatação e padronização.

Atualmente, a equipe do projeto é composta por cinco professores-pesquisadores e 9 bolsistas, incluindo graduados e graduandos. Os bolsistas estão divididos em três grupos de trabalho: o time de design, o time de *front-end* e o time de *back-end*.

3. Solução

O projeto *Big Data* da Cultura foi concebido para melhorar a gestão e a análise de dados do setor cultural, possibilitando um acompanhamento mais preciso de iniciativas, investimentos e impactos socioeconômicos. A solução desenvolvida até o momento concentra-se na coleta e organização de dados provenientes de diversas bases públicas e governamentais. Uma proposta de arquitetura é apresentada na Figura 1.

Os resultados preliminares indicam que a estruturação dessas bases já permite a extração de *insights* relevantes para a tomada de decisão. Foram realizadas análises exploratórias sobre participação cultural, distribuição de investimentos e impacto das políticas públicas, além de estudos comparativos entre diferentes regiões do estado. A Figura 2 ilustra uma análise feita a partir dos dados extraídos do Mapa Cultural.

Para as próximas etapas, pretende-se avançar na consolidação de um *dashboard* interativo, que permitirá a visualização dinâmica dos dados e facilitará a tomada de decisão por gestores públicos e demais interessados no setor cultural. A Figura 3 apresenta a tela inicial da proposta de *dashboard* interativo que está sendo desenvolvida. Com a

⁴<https://www.ibge.gov.br/estatisticas/multidominio/cultura-recreacao-e-esporte/9388-indicadores-culturais.html>

⁵<https://servicodados.ibge.gov.br/api/docs/agregados?versao=3>

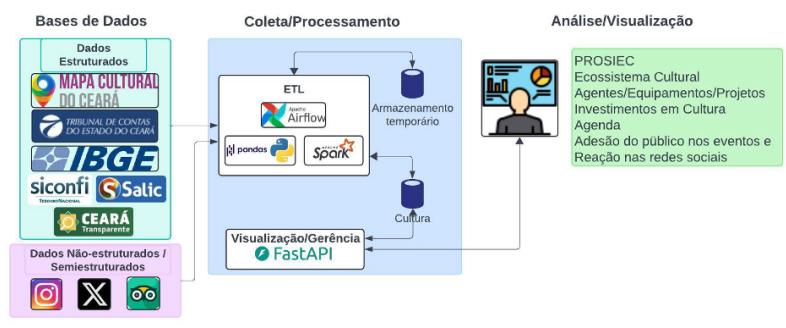


Figura 1. Proposta de Arquitetura

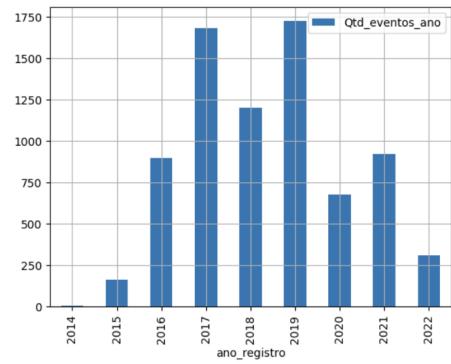


Figura 2. Quantidade de eventos por ano, com dados extraídos do Mapa Cultural do Ceará

integração progressiva das bases e a evolução da plataforma, espera-se fornecer um sistema robusto para monitoramento e formulação de políticas culturais baseadas em dados.

O projeto Big Data da Cultura vai além da aplicação de técnicas de Big Data, trazendo inovação ao setor cultural por meio da integração e análise de dados públicos. Essa iniciativa fortalece a transparência, embasa a formulação de políticas culturais e democratiza o acesso a informações, gerando impactos sociais relevantes e contribuindo para a valorização da cultura por meio da tecnologia.

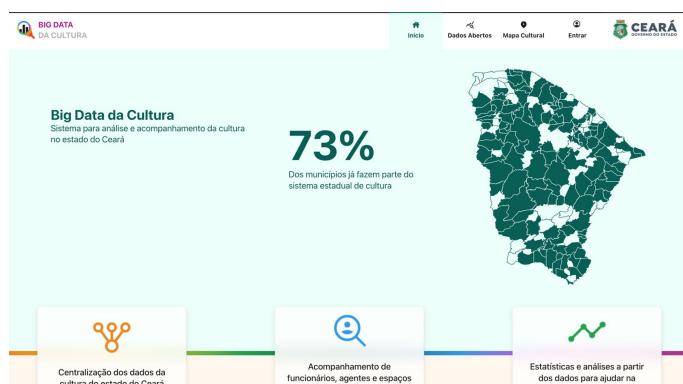


Figura 3. Tela inicial do dashboard interativo

Referências

- Almeida, C. L. S. d., de Azevedo Júnior, I. L., and Rodrigues, L. C. S. (2024). Cientista chefe da cultura: um programa do ceará feito para o brasil. *O Públíco e o Privado*, 22(46):113–139.
- Chen, X. and Chen, C.-K. (2024). An initial exploration of market dynamics and financial investment strategies in the cultural industries. *Journal of Statistics and Management Systems*, 27(6):1169–1184.
- Ganann, R., Ciliska, D., and Thomas, H. (2010). Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*, 5:1–10.
- Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1):97–107.