



O serviço de dados do iplanfor: um modelo incluindo Processos e tecnologia

Rodrigo Petry
Gerente de Sistema
de Informações
Geográficas

APRESENTAÇÃO

De acordo com a Lei Complementar N. 184 de 19 de dezembro de 2014, ficaram definidas diversas competências ao Instituto de Planejamento de Fortaleza – IPLANFOR relacionadas à gestão do conhecimento estratégico para a Prefeitura Municipal de Fortaleza, conforme a seguir:

Art 2º O Instituto de Planejamento de Fortaleza (IPLANFOR), [...] tem as seguintes atribuições:

[...]

VIII – produzir e difundir conhecimento para a melhoria da governança municipal;

[...]

XII – coletar, pesquisar, analisar, sistematizar, e divulgar informações sociais, econômicas, estatísticas, geográficas, cartográficas, infraestruturais, de mobilidade urbana, dentre outras informações relacionadas à Fortaleza.

As competências do IPLANFOR dispostas nos incisos VIII e XII do artigo 2º expostas acima são desempenhadas pela Diretoria do Observatório da Governança Municipal – DIOBS. Conforme afirmam Davenport e Prusak(1998), se faz necessário primeiramente a obtenção de dados e sua transformação em informações para, posteriormente, através de metodologias específicas, se realizar a produção de conhecimento. A DIOBS, portanto, possui a necessidade de obtenção de dados, não apenas de forma singular no tempo, mas também de forma contínua e incremental, pois o conhecimento deve ser produzido continuamente para fundamentar processos decisórios, políticas públicas e formulação de estratégias.

Portanto, para realizar o fornecimento contínuo de dados e informações de for-

ma a sanar as necessidades da DIOBS e demais órgãos e departamentos envolvidos em atividades de nível estratégico, esse artigo propõe um modelo de serviço de dados, definindo suas características tecnológicas e funcionais, levando-se em consideração a realidade organizacional da Prefeitura Municipal de Fortaleza.

JUSTIFICATIVA

No mundo atual vivemos na era da tecnologia, é cada vez mais fácil coletar e analisar grandes quantidades de dados, se tornando inclusive um desafio para as organizações analisar e obter significado através deles. Uma implicação importante do desenvolvimento dessa tecnologia é a capacidade de se tomar decisões baseadas na análise desses dados. É possível realizar regressões, criar simulações, modelos de otimização, análises preditivas, entre outros. (ALBRIGHT, WINSTON e ZAPPE, 2010).

Hammond, Keeney e Raiffa(2011, p.5) afirmam que “tomar decisões é o trabalho mais importante de um executivo. É também o mais difícil e arriscado”, decisões ruins podem ser catastróficas para a organização. Más decisões podem ser geradas a partir de informações e dados que não foram coletados. (HAMMOND, KEENEY e RAIFFA, 2011).

A função básica dos dados e informações é munir o gestor de conhecimento a respeito das operações de sua organização, como também, reduzir a incerteza a respeito do estado e do comportamento do macroambiente onde ela está inserida. Moritz e Pereira (2006, p. 13) afirmam que:

[...] o comportamento da organização é diretamente afetado, em termos de eficácia e eficiência, pela qualidade das decisões, as quais, por sua vez, são influenciadas pela qualidade das informações geradas, agindo como um processo integrado e sistêmico.

As atribuições do IPLANFOR relacionadas à coleta e fornecimento de dados e informações tem como principal objetivo abastecer a tomada de decisão na Prefeitura Municipal de Fortaleza, de forma a melhorar sua qualidade e eficiência. As referidas decisões são aquelas de nível estratégico, ou seja, decisões que tem relação com seus objetivos de longo prazo e afetam toda a prefeitura, e, em última instância, a própria sociedade. Em vista de suas importantes atribuições, esse artigo tem como objetivo delinear um modelo de serviço de dados que abasteça a produção de conhecimento desempenhada pela Diretoria do Observatório de Governança Municipal do IPLANFOR. Este modelo apresentará apenas características tecnológicas, arquiteturas, processuais e de recursos humanos do serviço, não é intuito do artigo tratar da governança dos dados que transitam pelo serviço.

O SERVIÇO DE DADOS

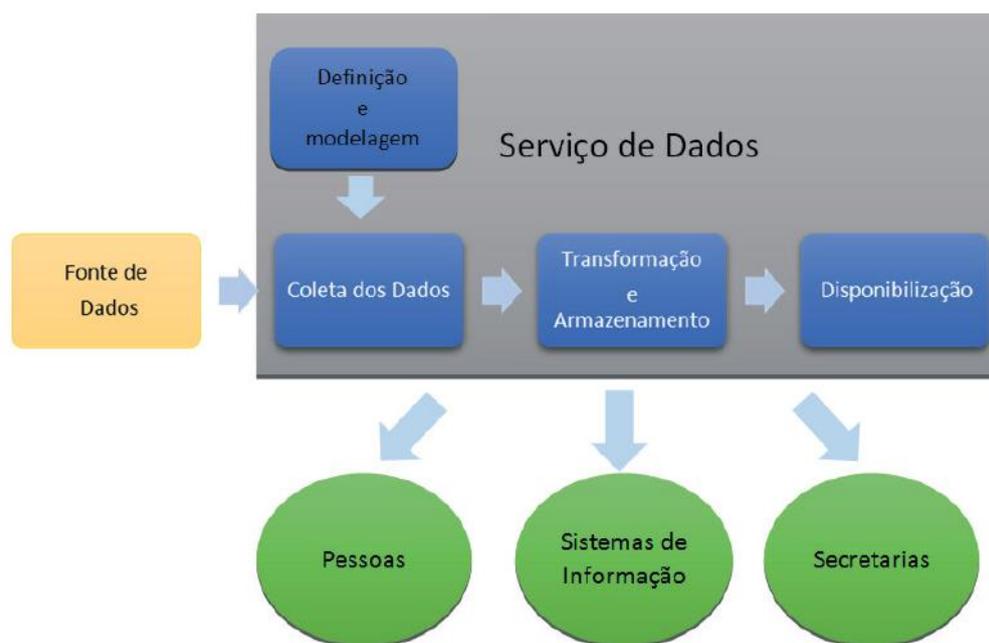
O serviço de dados consiste em um processo de definição e modelagem, coleta, transformação e disponibilização de dados às partes interessadas ou stakeholders. Essas partes interessadas podem ser sistemas de informação, departamentos do IPLANFOR,





colaboradores e secretarias da Prefeitura Municipal de Fortaleza. Veja o gráfico a seguir demonstrando o macroprocesso do serviço:

Cada fase do processo que constitui o serviço de dados utiliza tecnologias e



metodologias próprias para sua execução. O processo se inicia pela definição de quais dados são necessários e a sua estrutura de armazenamento (modelagem), em seguida é estabelecida a coleta dos dados nas respectivas fontes e estes então passam pelas transformações necessárias para que sejam disponibilizados aos seus usuários finais ou “clientes”.

De forma resumida, o serviço de dados é uma ferramenta de processamento de inputs e entrega de outputs de dados para as mais diversas finalidades dos seus usuários. Nos subcapítulos seguintes, cada etapa do macroprocesso será detalhado conjuntamente com a tecnologia utilizada para sua realização.

DEFINIÇÃO E MODELAGEM

A definição de quais dados serão coletados e sua respectiva estrutura de armazenagem e disponibilização final é o ponto inicial do serviço de dados. Nesta etapa os clientes do serviço demandam os conjuntos de dados de que necessitam e em qual formato e meio eletrônico devem recebê-los, são então revisadas as arquiteturas, entidades de dados e relacionamentos com dados já existentes no serviço.

FONTES DE DADOS

As fontes de dados do serviço podem ser internas ou externas à Prefeitura Municipal de Fortaleza, se caracterizam principalmente por serem heterogêneas, ou seja,

possuem ampla variação na sua formatação e meio de armazenamento. As fontes mais comuns de dados internos da Prefeitura Municipal de Fortaleza são as seguintes:

- Planilhas eletrônicas estruturadas, semiestruturadas e não estruturadas;
- Sistemas gerenciadores de bancos de dados relacionais (RDBMS);
- Arquivos CSV (Comma Separated Values);
- Arquivos de texto de coluna fixa;
- Arquivos DBF (Database File);
- XML (Extensible Markup Language);
- JSON (JavaScript Object Notation);
- Webservices:
 - SOAP (Simple Object Access Protocol);
 - REST (Representational State Transfer);
- Manuscritos.

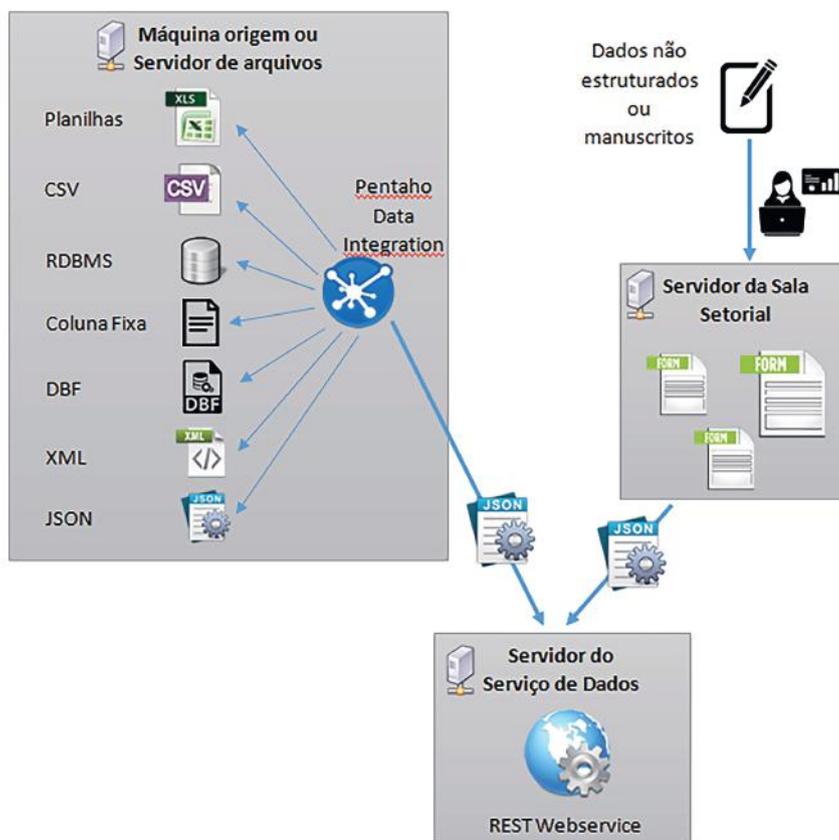
Há casos em que as fontes de dados não estão em formatos eletrônicos, isso não significa, no entanto, que não se possa realizar sua coleta por meio de sistematização. Para estes casos, a estratégia utilizada atualmente na DIOBS, e que será adotada para o serviço de dados, é a implantação do aplicativo da Sala Setorial, desenvolvido especialmente para coleta e sistematização de bancos de dados não informatizados ou sequer estruturados.

COLETA DOS DADOS

O serviço de dados oferecerá por padrão uma interface de comunicação via webservice para o recebimento de dados da fonte em formato JSON. Esta etapa consiste em estabelecer a transmissão permanente e automatizada de dados da fonte para o webservice de recebimento do serviço, com a finalidade de manter o conjunto de dados atualizados e de preferência sem a intervenção humana.

A estratégia de coleta varia conforme a característica de cada fonte de dados como enumerado no subcapítulo 3.2, e a infraestrutura de tecnologia de informação onde a fonte está instalada. A ferramenta principal de coleta utilizada pela DIOBS é o Pentaho Data Integration (PDI). O PDI é capaz de realizar a leitura de dados de diversas fontes e estruturas diferentes e transformá-los para o padrão JSON utilizado no input do serviço de dados. No caso de dados não sistematizados, não estruturados ou manuscritos, utiliza-se a Sala Setorial para sistematizá-los em um banco de dados eletrônico. Para mais informações sobre seu funcionamento é possível consultar o Caderno do Observatório da Governança Municipal de Fortaleza Volume II, que apresenta mais detalhes de sua implantação e funcionalidades.

Veja o esquema a seguir das estratégias para cada tipo de fonte de dados e infraestrutura de tecnologia da informação:



TRANSFORMAÇÃO

Após a coleta de dados da fonte, eles são transformados e armazenados em um banco apropriado para a sua estrutura. A transformação compreende os procedimentos realizados para higienizar, padronizar, harmonizar e relacionar os dados recebidos da fonte com os dados já existentes no serviço. No contexto da governança de dados, a etapa de transformação tem o objetivo de manter a acurácia, validade e harmonização dos dados do serviço. Ela visa garantir que o cliente receberá dados válidos, precisos e atualizados no output.

Considere, por exemplo, os seguintes dados advindos de uma fonte qualquer:

Dados da fonte			
Pessoas			
Código	Nome	Cód. Bairro	Bairro
1	Jorge	1	Aldeota
2	Paula		Varjota
3	Luis		Varjota
4	Graça	3	

- Podemos enumerar alguns problemas com os dados acima:
- O nome do bairro de Paula está incompleto e não possui código;
 - O bairro de Luis não possui código;
 - O bairro de Graça possui código mas o nome está nulo;

Veja abaixo o esquema que exemplifica uma transformação de dados que higieniza e padroniza os dados das pessoas, conformando com dados existentes no serviço, considerando que queremos incluir os dados de pessoas na banco de dados do serviço.



Dados da fonte			
Pessoas			
Código	Nome	Cód. Bairro	Bairro
1	Jorge	1	Aldeota
2	Paula		Varjota
3	Luis		Varjota
4	Graça	3	

Dados do serviço	
Bairros	
Código	Nome
1	Aldeota
2	Varjota
3	Montese



Dados do Serviço			
Pessoas			
Código	Nome	Cód. Bairro	Bairro
1	Jorge	1	Aldeota
2	Paula	2	Varjota
3	Luis	2	Varjota
4	Graça	3	Montese

As seguintes operações foram realizadas para determinar os bairros de cada pessoa e inserir dados precisos, limpos e harmonizados com os existentes do serviço:

- O registro de Jorge foi inserido diretamente na tabela pessoas, pois seu bairro estava exatamente igual ao da tabela de bairros;
- O bairro de Paula foi determinado utilizando um algoritmo de Fuzzy Match, ou seja, foi determinado pelo bairro com nome de maior similaridade escrita que consta na tabela de bairros;
- O bairro de Luis não possuía código, porém o seu nome "Varjota" é idêntico ao bairro encontrado na tabela de bairros, podendo assim determinar que seu código é "3";
- Por fim, o bairro de Graça que estava nulo foi determinado pela coluna "Cód. Bairro", assim sendo resultando que ela é do bairro "Montese".

As transformações podem realizar tanto a inclusão de novos dados nas tabelas existentes do serviço de dados, quanto atualizar dados antigos, corrigir dados incorretos e enriquecer dados incompletos.

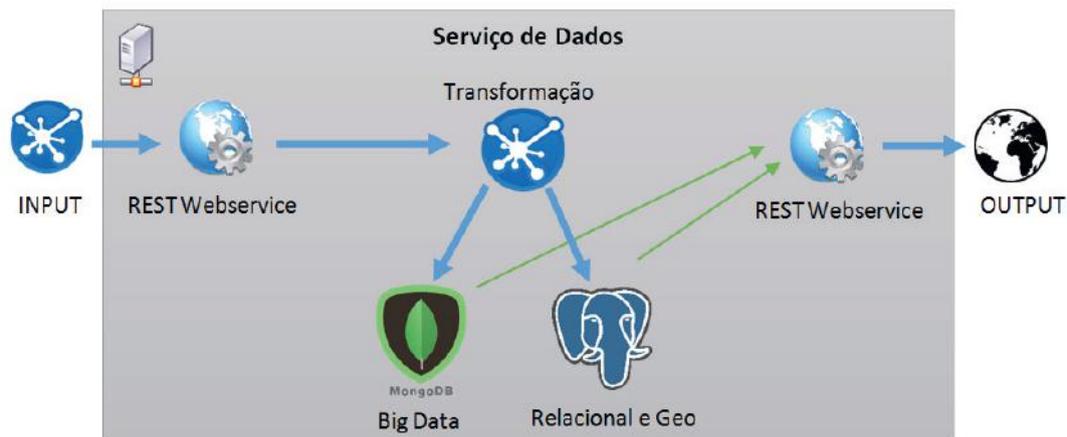
ARMAZENAMENTO

Após os dados serem coletados e transformados, segue-se o seu armazenamento em sistemas gerenciadores de bancos de dados (SGBD) para que se aproveite as capacidades de pesquisas complexas, indexação e estruturação que eles oferecem.

Para que haja flexibilidade no quesito estruturação, tipo e arquitetura dos dados que comporão o serviço, Os SGBD selecionados para este modelo de Serviço de Dados são os seguintes:

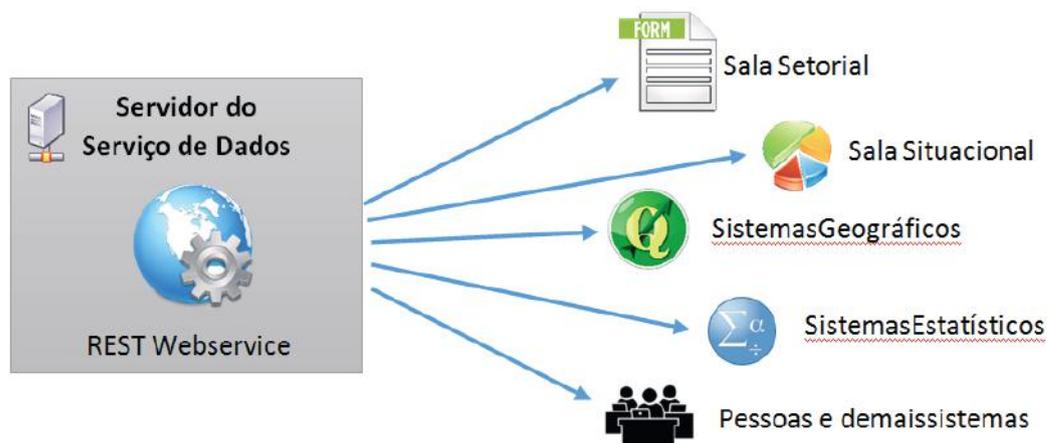
- PostgreSQL: Banco de dados gratuito e open-source, utilizado para armazenamento de dados estruturados e relacionais, conta com a poderosa extensão PostGIS para armazenamento de dados georreferenciados;
- MongoDB: Banco de dados gratuito e open-source, utilizado para armazenamento de dados não relacionais, estruturados e não estruturados e Big Data.

O esquema abaixo demonstra a etapa de armazenamento que ocorre após a transformação dos dados, direcionando-os para o SGBD mais adequado à sua arquitetura:



DISPONIBILIZAÇÃO

No fim do macroprocesso do Serviço de Dados, está a disponibilização dos dados coletados e transformados. A disponibilização aos stakeholders se dará de forma ininterrupta e acessível através de um webservice RESTful instalado no mesmo servidor que recebe os dados oriundos da coleta. A tecnologia RESTful foi selecionada pois grande parte dos sistemas de informações atuais já possuem compatibilidade de leitura dos dados disponibilizados por ela.



CONSIDERAÇÕES FINAIS

A Prefeitura Municipal de Fortaleza é uma organização complexa e possui objetivos desafiadores. Assim como outras organizações, seus recursos são limitados e seu uso deve ser otimizado, soma-se a isso um mundo que está em constante mudança exigindo que as decisões sejam tomadas em tempos cada vez mais curtos.

É essencial para atingir a missão da Prefeitura que o processo decisório global seja eficiente e abastecido de conhecimento produzido a partir de dados confiáveis e atualizados. O modelo de serviço de dados apresentado neste artigo visa justamente esse objetivo, porém, não tem a pretensão de ser a solução tecnológica completa de governança de dados.

A Prefeitura Municipal lida com problemas crônicos de fragmentariedade de seus sistemas e bancos de dados, dificultando uma visão única e integrada de informações, algo de suma importância para a redução das incertezas na tomada de decisão. Portanto, o modelo aqui proposto possui um caráter integrador e distribuidor de dados para tentar amenizar essa situação de silos desconexos de informação, podendo assim representar um estimulador para discussão e construção de uma política de gestão de dados mais complexa, integrada e completa.

REFERÊNCIAS

ALBRIGHT, Christian; WINSTON, Wayne; ZAPPE, Christopher. Data analysis and decision making. 4 ed. Mason: South-Western Cengage Learning, 2010.

DAVENPORT, H.; PRUSAK, L. Working Knowledge: How organizations manage what they know. Boston: Harvard Business School Press, 1998.

FORTALEZA. Lei Complementar n. 0184, de 19 de dezembro de 2014. Dispõe sobre a competência, estrutura e organização do Instituto de Planejamento de Fortaleza (IPLANFOR), criado pela Lei Complementar n. 0106, de 18 de abril de 2012, e dá outras providências. Câmara Municipal de Fortaleza, 2014. Disponível em: <http://216.59.16.201:8080/sapl/sapl_documentos/norma_juridica/4252_texto_integral> Acesso em: 10 ago 2016.

HAMMOND, John; KEENEY, Ralph; RAIFFA, Howard. The hidden traps in decision making. In: HARVARD UNIVERSITY. Harvard Business Review on making smart decisions. Boston: Harvard Business Review Press, 2011. p. 5-25.

MORITZ, Gilberto; PEREIRA, Maurício F. Processo decisório. Florianópolis: SEAD/UFSC, 2006.