



## SOBRE O SISTEMA

O LightBase Toolkit ("LightBase") é um sistema back-end completo para gestão de documentos, que possibilita o armazenamento e a indexação de grandes volumes de informação. Através do uso combinado de diversas tecnologias livres, permite estruturar e recuperar com extrema rapidez quaisquer informações salvaguardadas em seu banco de dados. É uma solução para controle administrativo de documentações diversas, com nível de detalhamento profundo e rapidez de indexação e busca, e que serve de base para o desenvolvimento de aplicações para diferentes nichos de mercado.

O conceito do LightBase reside na ideia de que a gestão de conjuntos densos de dados requer ferramentas que agilizem a busca por informações e reduzam a latência na interação cliente-servidor. Isto é alcançado por meio do uso de soluções de indexação e de bancos de dados consagradas que, em conjunto com modelos de dados multidimensionais especificamente projetados, permitem encontrar quaisquer informações dos documentos gerenciados rapidamente e com confiabilidade. Em síntese, o LightBase é definido como um banco de dados textual, de modelo documental mas que também guarda os dados de forma relacional.



## COMPOSIÇÃO

O ecossistema LightBase é composto por três módulos que, atuando em conjunto, proveem todas as suas funcionalidades:

- ⚙️ LightBase Converter (LBConverter) - Módulo de reconhecimento óptico de caracteres (OCR). Através dele, documentos digitalizados ou salvos em formato digital têm seu conteúdo extraído como texto puro que, então, pode ser indexado e armazenado para consulta e busca futuras.
- ⚙️ LightBase Index (LBIndex) - Módulo de indexação. Utilizando estruturas de dados de objetos multidimensionais, mantém indexação otimizada de todo o conteúdo dos documentos, permitindo pesquisas de resposta ágil completas em quaisquer termos existentes nos mesmos.
- ⚙️ LightBase Generator (LBGenerator) - Módulo de gestão dos documentos. É o responsável pelo armazenamento e controle da documentação, bem como pela interação com o front-end e mediação com os demais módulos através de serviços RESTful.

Os três componentes do ecossistema LightBase atuam em conjunto para fornecer os recursos de indexação e busca de conteúdos por meio de serviços. Juntos, funcionam do mesmo modo que um motor de um veículo: enquanto uma aplicação fornece as ferramentas para operação por usuários - assim como a carroceria e os instrumentos de um carro -, o LightBase opera a organização e a catalogação dos documentos, recebendo as solicitações da aplicação e entregando respostas adequadas.

## ECOSSISTEMA LIGHTBASE

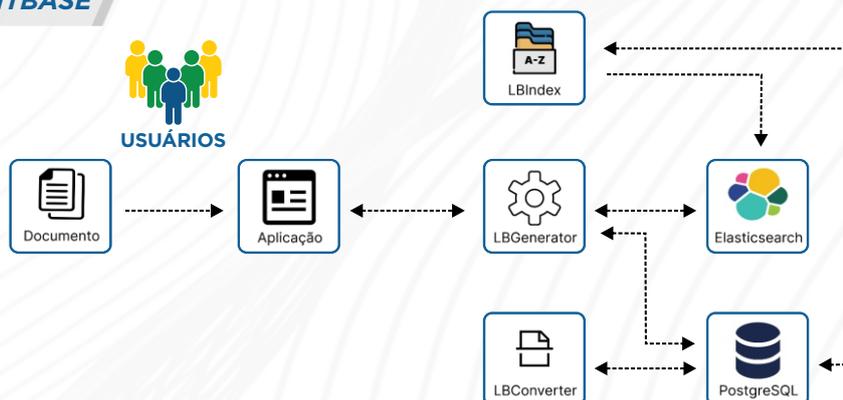


FIGURA 1  
ECOSSISTEMA LIGHTBASE





## ENTREGA DE VALOR



A perspectiva de utilização do LightBase está nas grandes bases de documentos que precisam ser pesquisados por seu conteúdo. Embora os bancos de dados tradicionais sejam as primeiras soluções pensadas para salvaguarda e gestão de dados de diferentes tipos, eles não atendem a critérios de rapidez de busca sobre dados textuais. Há diversas técnicas para se reduzir o custo computacional e de tempo para processar buscas sobre bancos de dados, como o particionamento de tabelas e a criação de índices, mas tais procedimentos não são viáveis para serem realizados sobre colunas de dados do tipo texto - e é exatamente para resolver este empecilho que entra o LightBase.

Os três módulos do LightBase provêm uma solução para a indexação textual, na medida em que combina o poder do Elasticsearch - desenvolvido especificamente para a finalidade de indexação e busca sobre quaisquer tipos de dados - com a capacidade de gestão e armazenamento de dados do PostgreSQL. Em síntese, o fluxo operacional é simples:

- 1** Um documento, seja ele um arquivo em formato PDF ou DOCX, é enviado ao LBGenerator, que por sua vez o encaminha para ter seu conteúdo extraído pelo LBConverter.
- 2** O texto extraído é enviado para o Elasticsearch para ser indexado. Ao mesmo tempo, o mesmo texto e o documento original são armazenados no PostgreSQL.
- 3** A partir deste momento, todo o conteúdo dos documentos está disponível para ser pesquisado com tempo de resposta extremamente reduzido. Quando uma busca é solicitada, o Elasticsearch a processa e entrega os dados dos documentos que contêm os termos buscados, bem como seus conteúdos.
- 4** Se o usuário assim desejar, pode baixar os documentos, cuja recuperação no PostgreSQL é, agora, realizada em tempo muito menor.
- 5** Durante todo o tempo, o LBIndex monitora alterações nos documentos já indexados, processando a atualização dos índices e do banco de dados.



## REQUISITOS DE SISTEMA

O ecossistema LightBase funciona com seus três componentes atuando em conjunto, cada um sendo um serviço independente. Além dos três componentes, o ecossistema também faz uso do sistema gerenciador de bancos de dados PostgreSQL e do sistema de indexação Elasticsearch. Todos os componentes podem ser instalados nos mesmos servidores ou cada um em servidores dedicados.

O ecossistema LightBase funciona com seus três componentes atuando em conjunto, cada um sendo um serviço independente. Além dos três componentes, o ecossistema também faz uso do sistema gerenciador de bancos de dados PostgreSQL e do sistema de indexação Elasticsearch. Todos os componentes podem ser instalados nos mesmos servidores ou cada um em servidores dedicados.

### SISTEMAS OPERACIONAIS SUPORTADOS

- Distribuições Linux baseadas em Debian (10+), RedHat (9+) ou SuSE. Sugere-se a utilização do CentOS Stream versão 9.

### CONFIGURAÇÕES MÍNIMAS DE SERVIDOR

- Processadores 64 bits dual-core, com frequência de 2 GHz, 8 GB de memória RAM, armazenamento variável a depender da quantidade de documentos prevista.

### CONFIGURAÇÕES RECOMENDADAS DE SERVIDOR

- Processadores 64 bits quad-core ou superiores, 16 GB de memória RAM, armazenamento variável a depender da quantidade de documentos prevista.

### PACOTES NECESSÁRIOS

- Todas as dependências são instaladas com o setup dos módulos, não sendo necessário nenhuma instalação ou configuração à parte.





## DESEMPENHO OPERACIONAL



O LightBase é capaz de operar em níveis ótimos de desempenho para diferentes perfis de acesso. Seus serviços são executados em múltiplos processos, o que garante plena disponibilidade e baixa latência mesmo em situações de estresse e múltiplas conexões. Além disso, graças ao fato de que as buscas textuais não são processadas no banco de dados, a resposta às requisições são fornecidas em pouquíssimos segundos\* - e por vezes milissegundos -, o que dá ao LightBase excelente capacidade de gerenciamento de recursos computacionais.

\* a depender do correto dimensionamento dos servidores



## SUPORTE

Como já relatado, o LightBase é um motor de indexação e busca de documentos. Para que suas potencialidades possam ser exploradas plenamente, torna-se imprescindível o emprego de uma aplicação, que por vezes é desenvolvida de acordo com a demanda de utilização do cliente.

A LightBase Consultoria em Software Público Ltda. possui expertise em desenvolvimento de aplicações e uma equipe extremamente qualificada de desenvolvedores e analistas de sistemas. Deste modo, é capaz de desenvolver aplicações sob demanda e prestar todo o suporte necessário, desde a implantação até o treinamento.



## DATASHEET

### ECOSSISTEMA LIGHTBASE



⚙️ **Versão corrente - 1.0**

⚙️ **Linguagem de programação - Python, versão 3.10**

⚙️ **Dependências Python** - alembic 1.4.3, argparse 1.4.0, bcrypt 4.0.1, Beaker 1.12.1, beautifulsoup4 4.12.2, certifi 2023.5.7, Chameleon 4.0.0, configparser 5.3.0, decorator 5.1.1, elasticsearch 8.9.0, jsonpath-rw 1.4.0, linecache2 1.0.0, Mako 1.2.4, meld3 2.0.1, ordereddict 1.1, PasteDeploy 3.0.1, pbr 5.11.1, pdfminer 20191125, ply 3.11, pycopg2-binary 2.9.6, Pympler 1.0.1, pyramid 1.10.8, pyramid-beaker 0.8, pyramid-chameleon 0.3, pyramid-restler 0.1a5\*\*, pyramid 1.10.8, python-dateutil 2.8.2, repoze.lru 0.7, repoze.profile 2.3, requests 2.30.0, simplejson 3.19.1, six 1.16.0, SQLAlchemy 1.2.19, superlance 2.0.0, supervisor 4.2.5, traceback2 1.4.0, translationstring 1.4, unittest2 1.1.0, urllib3 1.26.16, uWSGI 2.0.21, venusian 3.0.0, voluptuous 0.13.1, waitress 2.1.2, WebOb 1.8.7a\*\*, xlrd 2.0.1, zope.deprecation 5.0, zope.interface 6.0.

⚙️ **Dependências externas** - antiword 0.37, catdoc 0.95, leptonica 1.83.1, tesseract 5.3.1, unrtf 0.21.10, xpdf 4.04.

⚙️ **Método de instalação** - Bash Script.

⚙️ **Licença** - GNU General Public License versão 3.0.

\*\* Dependências não originais, modificadas pela LightBase para compatibilidade

### ELASTICSEARCH



⚙️ **Versão corrente - 8.9**

⚙️ **Método de instalação** - Script de instalação do LBGenerator; manual (orientações no site do desenvolvedor).

⚙️ **Licença** - Elastic License v2.

**POSTGRESQL**



⚙️ **Versão corrente - 15**

⚙️ **Método de instalação - Script de instalação do LBGenerator; manual (orientações no site do desenvolvedor).**

⚙️ **Licença - PostgreSQL License.**