



GUIA DE

Instalação Dataverse

2017



Cariniana

GUIA DE INSTALAÇÃO DATAVERSE

Igor Dias Ferrer
Luiza Martins de Santana Araújo

Brasília, DF
Janeiro 2017

SUMÁRIO

1. Apresentação.....	4
2. Para quem	4
3. Preparação da Instalação	5
3.1.Instalação Avançada	5
3.2.Arquitetura e Componentes	6
3.3.Requisitos de sistema	7
3.3.1. Requisitos de hardware	7
3.3.2. Requisitos de software.....	7
3.4 Pré-requisitos de sistema	7
3.4.1. Instalando o Java.....	7
3.4.2. Instalando Glassfish.....	8
3.4.3. Instalando PostgreSQL.....	10
3.4.4. Instalação do Solr	12
3.4.5. Instalando o JQ	13
4. Instalação	13
4.1. Efetuar Login.....	15
4.1.1. Conta do super usuário.....	15
REFERÊNCIAS	17

1. Apresentação

A Rede Brasileira de Serviços de Preservação Digital – Cariniana, tem como objetivo garantir o acesso continuado a longo prazo dos conteúdos armazenados digitalmente. Em 2015, foi realizada uma aliança com a Harvard University, desenvolvedora do projeto Dataverse, permitindo o uso desta plataforma pela Rede Cariniana.

A plataforma Dataverse é uma arquitetura de software livre para a publicação, citação, análise e preservação de dados de projetos de pesquisa. No Dataverse, a preservação e acesso em longo prazo estão garantidos na identificação persistente, protegendo os documentos digitais com mecanismos que preveem a obsolescência dos dados.

Este guia consiste em uma tradução e adaptação do *Installation Guide*, desenvolvido pelo Dataverse Project, tendo por objetivo orientar os parceiros da Rede Cariniana em relação aos procedimentos para sua instalação.

2. Para quem

Este guia é direcionado aos técnicos de informática parceiros da Rede de Pesquisa em Preservação Digital - Cariniana.

3. Preparação da Instalação

Para instalação do Dataverse será utilizado um ambiente de desenvolvimento para testes.

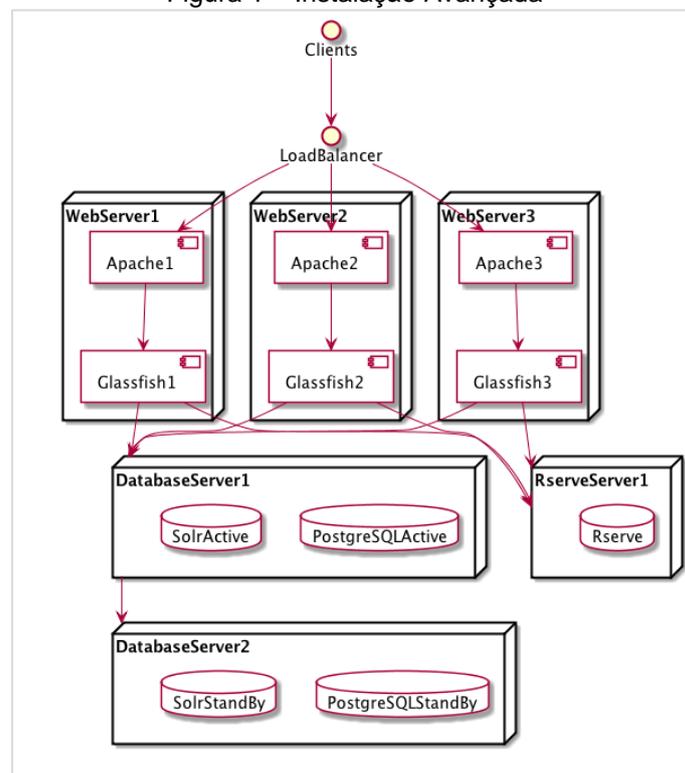
Instalar o Dataverse envolve a configuração do sistema seguido da execução de um script de instalação, que o guiará através do processo de instalação. Recomenda-se a leitura sobre a arquitetura e os componentes do Dataverse em primeiro lugar.

3.1. Instalação Avançada

Há alguns projetos de liderança da comunidade para usar ferramentas de gerenciamento de configuração para automatizar a instalação e configuração do Dataverse, porém o suporte para essas soluções é limitado ao que a comunidade Dataverse pode oferecer conforme descrito na página de cada projeto:

- <https://github.com/IQSS/dataverse-puppet>
- <https://github.com/IQSS/dataverse-ansible>

Figura 1 – Instalação Avançada



Fonte: <http://guides.dataverse.org/en/4.5/installation/index.html>

A equipe de desenvolvimento Dataverse auxilia a comunidade através da criação de um repositório no <https://github.com/IQSS>.

3.2. Arquitetura e Componentes

O Dataverse é um aplicativo da *Web Java Enterprise Edition* (EE) que é enviado como um arquivo war. Ao planejar sua instalação, você deve estar ciente dos seguintes componentes da arquitetura Dataverse:

- **Linux:** O CentOS é altamente recomendado, uma vez que todo o desenvolvimento e controle de qualidade ocorre nesta distribuição.
- **Glassfish:** é um servidor de aplicativos Java EE ao qual o aplicativo Dataverse (.war) deve ser implantado.
- **PostgreSQL:** é um banco de dados relacional.
- **Solr:** é um motor de busca; um esquema específico do Dataverse é fornecido.
- **Servidor SMTP:** para enviar e-mails para redefinições de senha e outras notificações.
- **Serviço de identificador persistente:** O suporte de DOI é fornecido. Uma assinatura EZID ou uma conta DataCite é necessária para uso em produção.
- **Integração do Dropbox:** para carregar arquivos da API do Dropbox.
- **Apache:** um servidor web que pode "inverter o proxy" Glassfish aplicações e reescrever o tráfego HTTP.
- **Shibboleth:** um sistema de autenticação descrito no Shibboleth . Seu uso com Dataverse requer Apache.
- **Geoconnect:** WorldMap: Dados Geoespaciais de Exploração descrevem o recurso; o código pode ser baixado diretamente do site <https://github.com/IQSS/geoconnect>.

3.3. Requisitos de sistema

3.3.1. Requisitos de hardware

Uma instalação básica do Dataverse funciona bem em hardware modesto. Por exemplo, como está descrito na instalação de teste em <http://phoenix.dataverse.org>, é apoiado por uma única máquina virtual com 2,8 GHz, 8 GB de RAM e 50 GB de disco.

Em contraste, a instalação de produção em <https://dataverse.harvard.edu> está atualmente apoiado por seis servidores com dois Intel Xeon 2,53 Ghz CPUs e 48 ou 64 GB de memória RAM. Os três servidores com 48 GB de execução de RAM são web frontends rodando Glassfish e Apache e são balanceados por um dispositivo de hardware. Os três servidores restantes com 64 GB de RAM são os servidores de banco de dados primário e de backup e um servidor. Vários TB de armazenamento são montados de uma SAN via NFS.

3.3.2. Requisitos de software

Antes de executar o script de instalação do Dataverse, você deve instalar e configurar os seguintes softwares listados abaixo, preferencialmente em uma distribuição do Linux, como por exemplo, CentOS.

3.4. Pré-requisitos de sistema

3.4.1. Instalando o Java

O Dataverse requer Java 8 (também conhecido como 1,8). O Dataverse deve funcionar corretamente apenas com o Java Runtime Environment (JRE) instalado, mas a instalação do Java Development Kit (JDK) é recomendada para que ferramentas úteis para solução de problemas de ambientes de produção estejam disponíveis. Recomenda-se o uso do Oracle JDK ou OpenJDK.

Pode-se baixar o Oracle JDK em: <http://www.oracle.com/technetwork/java/javase/downloads/index.html>.

Figura 2 – Instalação do Java

```
login as: igor
igor@192.168.0.133's password:

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Mon Feb 13 16:49:15 2017 from 10.0.10.42
igor@Lab-Cariniana:~$ install java-1.8.0-openjdk-devel
```

Fonte: Autoria própria

Insira o código:

```
# yum install java-1.8.0-openjdk-devel
```

Se você tiver várias versões do Java instalado, Java 8 deve ser o padrão quando java é invocado a partir da linha de comando. Você pode testar isso executando `.java-version`.

No CentOS você pode fazer a instalação do Java 8 com o comando `alternatives`, pedindo para selecionar a versão do Java a partir de uma lista

```
# alternatives --config java
```

Você pode também realizar o seguinte comando:

```
# alternatives --set java /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/bin/java
```

3.4.2. Instalando Glassfish

O Glassfish na versão 4.1 é necessário. Deve ser evitado a versão 4.1.1 do Glassfish, pois podem ocorrer problemas que ainda não foram solucionados.

Importante: uma vez que o Glassfish é instalado, uma nova versão da biblioteca Weld (v2.2.10.SP1) deve ser baixado e instalado. Isso corrige um

problema sério na biblioteca fornecida com Glassfish 4.1 (veja <https://github.com/IQSS/dataverse/issues/647> para detalhes).

Baixe e instale o Glassfish (instalado em `/usr/local/glassfish4`). Por meio do comando:

```
# wget http://dlc-cdn.sun.com/glassfish/4.1/release/glassfish-4.1.zip
# unzip glassfish-4.1.zip
# mv glassfish4 /usr/local
```

A tela abaixo será exibida no terminal:

Figura 3 – Instalação Glassfish

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Mon Feb 13 16:49:15 2017 from 10.0.10.42
igor@Lab-Cariniana:~$ wget http://dlc-cdn.sun.com/glassfish/4.1/release/glassfis
h-4.1.zip
```

Fonte: Autoria própria

Baixe o Weld v2.2.10.SP1 e instale-o na pasta de módulos:

```
# cd /usr/local/glassfish4/glassfish/modules
# rm weld-osgi-bundle.jar
# wget http://central.maven.org/maven2/org/jboss/weld/weld-osgi-bundle/2.2.10.SP1/
weld-osgi-bundle-2.2.10.SP1-glassfish4.jar
# /usr/local/glassfish4/bin/asadmin start-domain
```

Figura 4 – Instalação Weld

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Mon Feb 13 16:49:15 2017 from 10.0.10.42
igor@Lab-Cariniana:~$ wget http://central.maven.org/maven2/org/jboss/weld/weld-o
sgi-bundle/2.2.10.SP1/weld-osgi-bundle-2.2.10.SP1-glassfish4.jar
```

Fonte: Autoria própria

Verifique a versão do Weld:

```
# /usr/local/glassfish4/bin/asadmin osgi lb | grep 'Weld OSGi Bundle'
```

Parar Glassfish e mudança de cliente para servidor sob `<jvm-options>-client</jvm-options>`:

```
# /usr/local/glassfish4/bin/asadmin stop-domain  
# vim /usr/local/glassfish4/glassfish/domains/domain1/config/domain.xml
```

3.4.3. Instalando PostgreSQL

A versão 9.x é necessária. As versões anteriores não foram testadas. Para iniciar a sua instalação insira:

```
# yum install postgresql-server  
# service postgresql initdb  
# service postgresql start
```

O script de inicialização padrão deve funcionar bem. Ative-o com este comando:

```
# chkconfig postgresql on
```

Configurando o Acesso ao Banco de Dados para o Aplicativo Dataverse

O aplicativo e o script de instalação se conectarão ao PostgreSQL sobre TCP / IP, usando autenticação de senha. Nesta seção, explicamos como configurar o PostgreSQL para aceitar essas conexões.

Se o PostgreSQL está sendo executado no mesmo servidor que o Glassfish, localize a entrada local host (127.0.0.1) que já está no `pg_hba.config` deve modificá-lo para se parecer com isto:

```
host all all 127.0.0.1/32 md5
```

Uma vez que você realizou os pré-requisitos e executou o script de instalação, ele irá pedir que você digite o endereço do servidor Postgres. Basta aceitar o valor padrão `127.0.0.1`.

O script do instalador Dataverse vai precisar para se conectar ao PostgreSQL como o usuário administrador, a fim de criar e configurar o banco de dados que o Dataverse estará usando. Se, por qualquer motivo, não estiver se conectando (por exemplo, se você não souber ou não lembrar qual é a sua senha de administrador do Postgres), você pode optar por desabilitar temporariamente todas as restrições de acesso em conexões localhost, alterando a linha acima para:

```
host all all 127.0.0.1/32 trust
```

Note que esta regra abre o acesso ao servidor de banco de dados via local host única. Ainda assim, em um ambiente de produção, isso pode constituir um risco de segurança.

Se a aplicação Dataverse está sendo executado em um servidor diferente, você vai precisar adicionar uma nova entrada para a `pg_hba.config` o acesso de seu endereço de rede:

```
host all all [ADDRESS] 255.255.255.255 md5
```

“Address” é o endereço IP numérico do servidor Glassfish. Insira este endereço quando o instalador solicitar o endereço do servidor PostgreSQL. Em algumas distribuições, o PostgreSQL está pré-configurado para que ele não aceite conexões de rede. Verifique se a `listen_address` linha no arquivo de configuração `postgresql.conf` não é comentada e se parece com isso:

```
listen_addresses='*'
```

O arquivo `postgresql.conf` estará localizado no mesmo diretório como o `pg_hba.conf` acima.

Atenção: **PostgreSQL deve ser reiniciado** para que as alterações de configuração tenham efeito:

```
# service postgresql restart
```

3.4.4. Instalação do Solr

Faça o download e instale o Solr com estes comandos:

```
# wget https://archive.apache.org/dist/lucene/solr/4.6.0/solr-4.6.0.tgz
# tar xvzf solr-4.6.0.tgz
# rsync -auv solr-4.6.0 /usr/local/
# cd /usr/local/solr-4.6.0/example/solr/collection1/conf/
# cp -a schema.xml schema.xml.orig
```

A razão para fazer o backup do schema.xml é porque o Dataverse requer um esquema Solr personalizado para operar. Este schema.xml está contido no zip "dvininstall" fornecido em cada versão Dataverse, disponível em <https://github.com/IQSS/dataverse/releases>. Baixe este arquivo zip, extraia schema.xml a partir do seguinte comando:

```
# cp /tmp/schema.xml schema.xml
```

Com o esquema específico de Dataverse no lugar, agora você pode iniciar o Solr:

```
# java -jar start.jar
```

Solr Init Script

O comando acima irá iniciar Solr no primeiro plano, que é bom para uma verificação rápida para saber se o Solr aceitou o arquivo de esquema, mas é recomendável iniciar Solr com um script de inicialização.

Você pode tentar ajustar esse script de inicialização Solr para as suas necessidades. O Solr deve estar em execução antes do script de instalação ser executado.

Protegendo Solr

Solr deve ser firewall de todos os hosts, exceto o servidor executando o Dataverse. Caso contrário, qualquer host que possa alcançar a porta Solr (8983 por padrão) pode adicionar ou excluir dados, pesquisar dados não publicados e até reconfigurar o Solr.

3.4.5. Instalando o JQ

JQ é uma ferramenta de linha de comando para analisar a saída JSON que é usada pelo script de instalação Dataverse. O site <https://stedolan.github.io/jq> explica várias formas de instalação, mas um método relativamente simples é descrito abaixo. Atenção: você deve transferir a versão de 64 ou 32 bits com base na sua arquitetura. No exemplo abaixo, a versão de 64 bits está instalada.

```
# cd /usr/bin
# wget http://stedolan.github.io/jq/download/linux64/jq
# chmod +x jq
# jq --version
```

4. Instalação

Agora que os pré-requisitos estão no lugar, é possível executar o script de instalação do Dataverse e verificar se a instalação foi bem-sucedida, fazendo login com uma conta de "super usuário".

É fornecido um instalador interativo com script. Este script irá configurar o seu ambiente Glassfish, criar o banco de dados, definir algumas opções necessárias e iniciar o aplicativo. Algumas tarefas de configuração ainda serão necessárias após a execução do instalador.

Você deve baixar o instalador em <https://github.com/IQSS/dataverse/releases> antes de configurar e iniciar o Solr.

Descompacte o arquivo zip e após a descompactação crie o diretório `dvinstall`. Execute o script de instalação:

```
# cd dvinstall
# ./install
```

Alterações a partir do Dataverse 4.3:

Não é necessário estar como root para executar o instalador. Basta certificar-se que o usuário que executa o instalador tem a permissão de gravação no diretório Glassfish. Por exemplo, se o seu diretório Glassfish é propriedade do root e tentar executar o instalador como um usuário comum, ele não vai funcionar. O diretório do Glassfish deve ser propriedade do mesmo usuário que o estará executando e você provavelmente não precisará executá-lo como root. Execute o aplicativo nas portas padrão HTTPs 80 e 443, em vez de 8080 e 8181. No entanto, outra maneira de conseguir isso seria manter Glassfish rodando em uma porta alta e escondê-lo atrás um Proxy Apache, via AJP, correndo na porta 80.

O script solicitará alguns valores de configuração. Se esta é uma instalação de teste, é possível aceitar os valores padrões fornecidos para a maioria das configurações:

- Endereço de Internet do seu host: *localhost*
- Diretório Glassfish: */usr/local/glassfish4*
- Servidor SMTP (correio) para retransmitir mensagens de notificação: *localhost*
- Endereço do servidor Postgres: *[127.0.0.1]*
- Porta do servidor Postgres: *5432*
- Nome do banco de dados Postgres: *dvndb*
- Serviço de indexação remota do Solr: *LOCAL*
- Este Dataverse usará o aplicativo TwoRavens: *NÃO INSTALADO*
- Servidor Rserve: *localhost*
- Porta do servidor Rserve: *6311*

Importante: Note que o instalador irá configurar o Dataverse para deixar o acesso sem restrições às APIs de administração do localhost. Considere as implicações de segurança deste arranjo, visto que qualquer pessoa com acesso shell para o servidor pode potencialmente fazer alterações no seu Dataverse. Uma solução alternativa seria bloquear o acesso aberto; e para permitir apenas solicitações que forneçam um "token de desbloqueio" pré-definido (senha). Se preferir como uma solução, consulte o script fornecido [post-install-api-block.sh](#) para obter exemplos sobre como configurá-lo.

4.1. Efetuar Login

O Glassfish é executado na porta 8080 e 8181, em vez de 80 e 443, respectivamente, ao acessar o link <http://localhost:8080> (substituindo o nome do host) será exibida uma página de login.

4.1.1. Conta do super usuário

Use a conta de “super usuário” criada pelo instalador para se certificar de que o login está ativo no Dataverse.

O script [dvininstall/setup-all.sh](#) executado pelo instalador, define a senha para de cadastro do super usuário, enquanto que o login e endereço de e-mail são fornecidos por um arquivo referenciado em [dvininstall/data/user-admin.json](#). Utilize as seguintes credenciais para iniciar sessão:

URL: `http://localhost:8080`

Username: `dataverseAdmin`

Senha: `admin`

REFERÊNCIAS

DATAVERSE PROJECT. **Installation Guide**: version 4.5. [S.l]: The Institute for Quantitative Social Science, 2016. Disponível em: <<http://guides.dataverse.org/en/4.5/installation/index.html>>. Acesso em: 17 jan. 2016.