LexPredict ContraxSuite Documentation

Installation and Configuration Guide
Release 1.0.9 - May 1, 2017

Getting Started

Licensing Information

ContraxSuite is available under a dual-license open-source model. Unless otherwise released by ContraxSuite, LLC, you can use and modify this software under the terms of the GNU AFFERO GENERAL PUBLIC LICENSE.

If you have question about licensing or would like to request a release from the AGPL license, please email license@contraxsuite.com.

ContraxSuite also relies on a software and data dependencies that are independently licensed. For more information about these dependencies and their respective licensing models, please refer to the Software and Data Dependency documentation in this release.

Support, Customization, Hosting, or Training

We can help! If you need assistance customizing, hosting, training, or supporting your ContraxSuite instance, please reach out to us to discuss options. You can always email us support@contraxsuite.com or create issue on Github.

Prerequisite Skills

This document assumes basic familiarity with the installation and configuration enterprise multi-tier applications. In particular, personnel with Linux, Java, and Python experience are best suited to successfully complete the installation process.

How to Get Support

For support or help in setting up the application, please contact support@contraxsuite.com.

Product Architecture

Application Design Principles and Components

The ContraxSuite application follows a service-oriented architecture. Each component supports resiliency and high-availability and can be scaled independently. The ContraxSuite application consists of the following core components:

#	Component	Purpose
1	Database	Store persistent structured data
2	Message Broker	Coordinate async, distributed activities
3	Enterprise Search Index	Provide full-text search capabilities
4	Distributed Task Engine	Execute async, distributed activities
5	Application Server	Manage application/logical layer
6	Web Server	Provide presentation layer for humans and API

Architecture Diagram

Single Server Development Environment

This section will be completed in Q3 2017, as detailed in the Public Roadmap.

Multi-Server Staging Environment

This section will be completed in Q3 2017, as detailed in the Public Roadmap.

Network Service and Port Diagram

The table below demonstrates a sample network service and port layout for the recommend configuration. This configuration varies greatly, based on whether the database type, message broker type, and whether a single-server or multi-server installation is performed.

#	Service	Purpose	Port	Protocol
	PostgreSQL	Database Access	5432	TCP/SSL
	ElasticSearch	Enterprise search	9200	TCP/SSL
	Redis	Message Broker	6379	TCP

	RabbitMQ (default)	Message Broker	5672	TCP/SSL
	uWSGI	Application Container	8001	TCP
	nginx	Web Application	443	HTTPS

N.B.: Some service configurations, e.g., redis, require tunnel or local socket file communication to maintain end-to-end network encryption.

Installation Guide for Linux (64-bit)

Docker Guide

Please see https://github.com/LexPredict/lexpredict-contraxsuite/blob/master/docker/QUICK_DEPLOY.md for docker quick start methods, which is our preferred setup method. We will be updating our documentation to reflect this change.

Quick Start Guide

Are you setting up a development or evaluation instance? If so, the single-server "Quick Start" installation guide might be a better choice for you. Please refer to the Quick Start Installation Guide for Linux in the documentation.

Server Requirements

While ContraxSuite can be run on a variety of Linux distributions and versions, we recommend the following configuration for development and testing:

- Operating System: Ubuntu 16.04 LTS 64-bit
- Virtualization Supported: Yes
- Minimum Requirements:
  - CPU / vCPU: 4 cores / 4 vCPU
  - RAM: 8GB
  - Disk: 40GB SSD or more
  - IP Addresses: 1 static

Pre-Installation Tasks

When installing ContraxSuite on one or more Linux servers, a number of pre-installation tasks must be performed in order to complete a successful installation. These pre-installation tasks ensure that key software dependencies and resources are available on the host to perform subsequent installation steps.

Storage Architecture

ContraxSuite utilizes three data stores - a traditional relational database, a message broker, and an enterprise search index. While the persistence of data is important for all three of these systems, the primary storage burden is placed on the relational database and enterprise search index.

As with any application, organizations need to first consider their requirements and preferences such as performance, security, recovery time objective (RTO), and recovery point objective (RPO). Based on these preferences and a budget, an organization can then properly architect to their needs.

In general, and especially for single server deployments, we strongly recommend that multiple storage devices are used to separate I/O paths in the application.

The table below provides two recommended storage architectures for a single server deployment:

Recommend Deployment

Purpose	Suggested Mountpoint	Suggested Type
Operating System	/	
OCR Workspace	/tmp	SSD
RDBMS	/database	SSD
Enterprise Search	/search	SSD
Application	/opt	

Simple Deployment

Purpose	Suggested Mountpoint	Suggested Type
RDMBS	/database	SSD
OS, Application, etc.	/	SSD

From a sizing perspective, requirements will vary widely based on the type of document, number of documents, and ContraxSuite functionality applied. ContraxSuite and its software dependencies require approximately 2GB for complete installation However, the table below provides estimates for additional storage requirements per thousand documents:

Sizing Requirements per 1000 Documents (KiloDoc)

System	GB per KiloDoc
RDBMS	0.5
Enterprise Search	0.75

ContraxSuite can also be architected to provide encryption at rest through hardware, volume, or filesystem encryption, as well as RDBMS encryption. Organizations that require complete encryption at rest should consider which layers encryption should be enabled on, and incorporate these choices into performance analysis regarding CPU and IOPS.

Network Architecture

ContraxSuite can be deployed either on a single server or flexibly distributed across two or more servers. In single-server deployments, no additional network configuration is required. However, in situations where the application is deployed to multiple servers, care must be taken to properly configure network traffic between hosts.

In general, other than during document ingestion and during some distributed tasks such as clustering or classification updates, ContraxSuite is not a network-intensive application. However, some use cases and organizations may frequently ingest new documents or re-train and apply new machine learning systems, and these organizations should appropriately allocate high-throughput paths between the RDBMS, application, and distributed task servers.

Required System Software

The following software is required to execute the actual installation steps of the process below. Please refer to your underlying operating system support documentation or Google for instructions on installing these packages if not available.

- OpenSSH Client and Server
- zip/unzip
- tar/untar
- Git
- Python 3.x
- “sudo” or administrative privileges

Java Software

Java is a key component of the ContraxSuite application, powering the enterprise search engine as well as a number of natural language processing (NLP) libraries. ContraxSuite has been tested and designed to support the following Java versions and releases:

Version	Implementation	Status
8.x	Oracle Java SE	Supported, Recommended
7.x	Oracle Java SE	Supported
All	OpenJDK	Not Supported

Please note that OpenJDK is not currently supported due to incompatibilities with the Stanford NLP libraries.

More information regarding Oracle Java SE installation can be found at the references below:

- http://www.oracle.com/technetwork/java/javase/overview/index.html
- https://docs.oracle.com/javase/8/docs/technotes/guides/install/toc.html

Installing the Database

Currently, we strongly recommend that PostgreSQL is used as the relational database for ContraxSuite. ContraxSuite is tested, generally supported, and has some functionality that may only work with Postgres. The full list of compatible Linux databases can be found below:

Database	Version	Status
Postgres	9.x	Supported, Recommended
Postgres	8.x	Supported
MySQL	5.5	Supported
MySQL	5.6	Supported
Oracle	11g	Supported
Oracle	12c	Supported

Cloud deployments can be run on Database-as-a-Service offerings, including Amazon RDS.

More information and support for database installation can be found at the links below:

- https://www.postgresql.org/docs/9.6/static/tutorial-install.html
- https://www.postgresql.org/docs/9.5/static/tutorial-install.html
- https://www.postgresql.org/docs/9.4/static/tutorial-install.html

- https://dev.mysql.com/doc/refman/5.5/en/installing.html
- https://dev.mysql.com/doc/refman/5.6/en/installing.html
- https://docs.oracle.com/cd/E11882_01/nav/portal_11.htm
- https://docs.oracle.com/database/121/nav/portal_11.htm
- https://aws.amazon.com/documentation/rds/

Database architecture, much like storage architecture, should be designed around the organization's performance, RTO, and RPO requirements. For example, Postgres,

Installing the Message Broker

Currently, we strongly recommend that RabbitMQ or Redis are used as the message broker for ContraxSuite. The full list of compatible message brokers can be found below:

Broker	Version	Status
RabbitMQ	3.6+	Supported, Recommended
Redis	3.x	Supported
Redis	4.x	Supported
Amazon SQS		Supported

More information and support for message broker installation can be found at the links below:

- https://www.rabbitmq.com/admin-guide.html
- https://www.rabbitmq.com/install-debian.html
- https://www.rabbitmq.com/install-rpm.html
- https://redis.io/documentation
- https://redis.io/topics/quickstart
- https://aws.amazon.com/documentation/sqs/

Installing the Enterprise Search Index

ContraxSuite relies on an enterprise search index to provide full-text search and other customization capabilities. Currently, we strongly recommend that ElasticSearch is used to provide enterprise search. However, Solr can be used as an alternative in some contexts.

Engine	Version	Status
ElasticSearch	2.x	Supported, Recommend

ElasticSearch	5.x	Supported
Solr	5.x	Partial

Installing the Distributed Task Engine

ContraxSuite relies on a distributed task engine to provide asynchronous and scalable capabilities. Currently, ContraxSuite is implemented using the Celery distributed task queue engine.

Engine	**Version**	**Status**
Celery	4.0	Supported, Recommend
Celery	4.1	Supported

Celery document is available at docs.celeryproject.org. In particular, the following resources are useful for the setup, configuration, and maintenance of the daemon and workers:

- http://docs.celeryproject.org/en/latest/getting-started/index.html
- http://docs.celeryproject.org/en/latest/userguide/daemonizing.html
- http://docs.celeryproject.org/en/latest/userguide/workers.html

Installing the Application Server

ContraxSuite relies on an application server or container provider to execute its core functionality. This application server must support the Web Server Gateway Interface (WSGI) standard, which allows for Python applications to interact with presentation-layer services.

Currently, we recommend that uWSGI is used to run ContraxSuite. We develop, test, and host ContraxSuite using uWSGI. However, a number of other application servers or containers are available and can be used:

Application Server/Container	**Version**	**Status**
uWSGI	2.x	Supported, Recommended
Gunicorn	All	Supported
Werkzeug	All	Supported

More information related to the installation, configuration, and maintenance of these containers is available here:

- https://uwsgi-docs.readthedocs.io/en/latest/index.html

- https://uwsgi-docs.readthedocs.io/en/latest/Install.html
- http://gunicorn.org/
- http://docs.gunicorn.org/en/latest/deploy.html
- http://werkzeug.pocoo.org/docs/0.12/

Installing the Web Server

ContraxSuite relies on a web server to interact with end-users and expose internal and external APIs. Any web server that supports WSGI standards can be used to run the application, including IIS for Windows.

Web Server	Version	Status
nginx	1.9	Supported, Recommended
nginx	1.10	Supported
Apache	2.4	Supported
IIS	8.5	Supported
IIS	10	Supported

ContraxSuite can be run behind a hardware or software load balancer, such as an F5 or Amazon Elastic Load Balancer (ELB) service.

As of Release 1.0.4, Nginx is also required for data exchange in some cases for the distributed task engine.

ContraxSuite is designed to support full HTTPS/SSL encryption end-to-end. Organizations without access to an existing wildcard certificate or Certificate Authority can utilize the Let's Encrypt project to obtain an SSL certificate without cost: https://letsencrypt.org/.

Installing ContraxSuite and LexPredict Knowledge Sets

ContraxSuite can be deployed from the public Github repository at:
https://github.com/LexPredict/lexpredict-contraxsuite

Users can select the desired release branch, download the release ZIP or clone the branch, and deploy as a standard Django application.

Developers and basic installations can also rely on the deployment automation available from the deployment repository below, described in the Quick Start Installation Guide for Linux:
https://github.com/LexPredict/lexpredict-contraxsuite-deploy

LexPredict knowledge sets are described in the Knowledge Set documentation and can also be retrieved from Github at: https://github.com/LexPredict/lexpredict-legal-dictionary

N.B.: Beginning with ContraxSuite Release 1.0.1, knowledge sets can be automatically retrieved from Github.

More information about customizing and configuring these knowledge sets is available in the Developer Guide.

Default ContraxSuite UI Setup

Most ContraxSuite implementations involve customized user interface and user experience development. However, to facilitate the testing and evaluation of ContraxSuite, a default UI can be configured for ContraxSuite out-of-the-box. This default UI uses two packages detailed in the list in the Software and Data Dependency documentation. Once the Third Party Theme and Software is licensed, the default theme can be installed using the automation described in the Quick Start Installation Guide in the Local Machine Installation or Remote Machine Installation sections.

Testing ContraxSuite

ContraxSuite ships with over 2,000 agreements and plans, allowing users to quickly load documents, test functionality, and learn to customize and develop their own tools.

In addition, LexNLP, the core library for ContraxSuite parsing, includes over 200 types of unit tests for more than 500 examples of English language from real-world public contracts. Additional unit tests in English and other languages are scheduled for inclusion in Q3 and Q4 2017.

ContraxSuite sample data can be retrieved from the public Github repository below:
https://github.com/LexPredict/lexpredict-contraxsuite-samples

These samples and unit tests include:

- Construction agreements
- Credit agreements
- Employment agreements
- Severance agreements
- Software license agreements
- Retirement plans

Installation Guide for Windows

Installation instructions for Windows are currently scheduled for completion in Release 1.0.5, January 1, 2018. Please see the ContraxSuite Public Roadmap for more information about timing for Windows-related support.