lgtm enterprise

LGTM Enterprise System Architecture

Release 1.19, December 2018

Semmle™

Semmle Inc.
44 Montgomery Street
3rd Floor
San Francisco, CA 94104

LGTM Enterprise release 1.19
Document published December 13, 2018

Contents

Introduction

About this document

This document is an excerpt from the LGTM Enterprise administrator help. It provides a series of basic architecture diagrams intended to provide a high-level overview of the main components of LGTM Enterprise and how they are connected. For more detailed information about the services mentioned in this document, see the LGTM Enterprise administrator help.

Related documentation

- *LGTM Enterprise System Requirements* (PDF)
- *LGTM Enterprise Installation Guide* (PDF)
- LGTM Enterprise administrator help

  To access this, click **Admin help** at the top of the administration pages in LGTM Enterprise, or browse to help.semmle.com/lgtm-enterprise/admin.

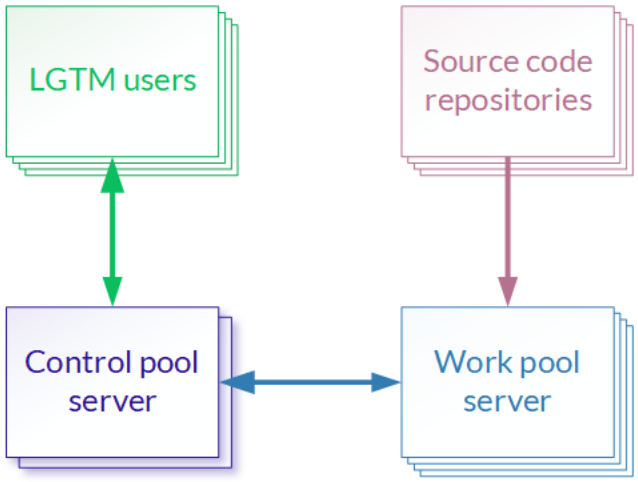
Server overview

The servers used in a deployment of LGTM Enterprise can be classified into two types:

- **Work pool** servers
  The servers in the work pool host one or more LGTM workers (processes that build code and generate analysis data). For more information, see "Work pool" on page 7.
- **Control pool** servers
  The servers in the control pool:
  - Store the persistent state of the LGTM cluster
  - Coordinate the work of the work pool, process and store its results
  - Host the web interface to the cluster

  For more information, see "Control pool" on the next page.

Overview of data flow:



For a simple deployment you might have just one control pool server and perhaps five work pool servers.

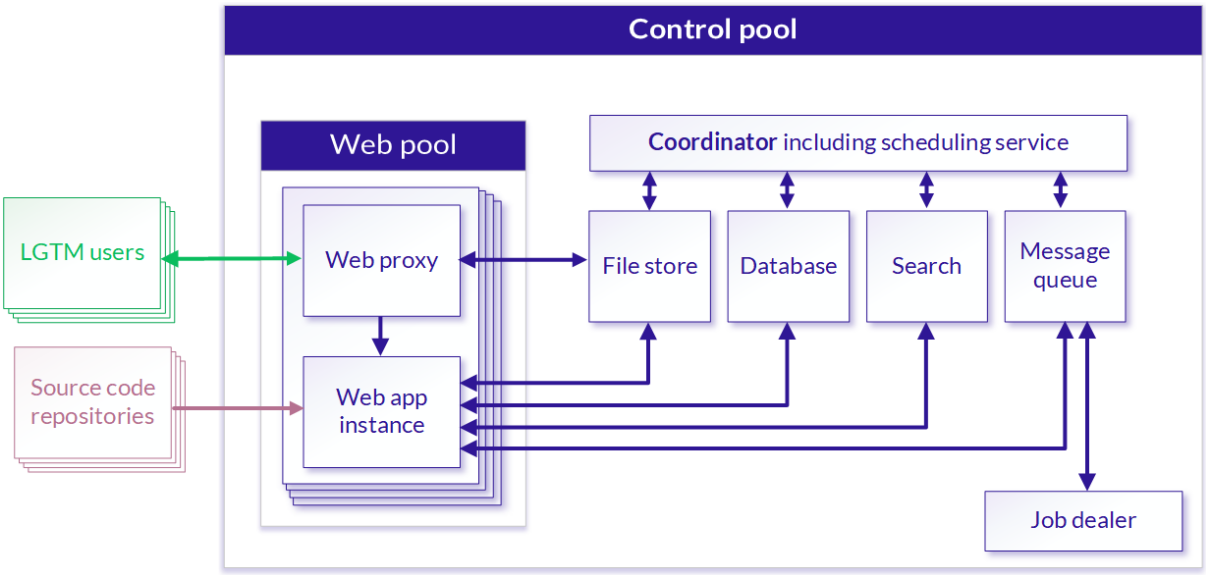
> **Note**
> Except where otherwise stated, the arrow directions on diagrams give an indication of data flow between system components.

Control pool

The control pool contains one or more servers. These host:

- Scheduling service—runs periodic tasks at set intervals.
- Database service—provides permanent storage of LGTM data.
- File storage service—provides storage of frequently accessed data. This gives immediate access to repositories in LGTM format, recent build/analysis log files, unprocessed job results, and the latest snapshot for each project (used by the query console).
- Message queue service—manages work. The associated job dealer ensures that build/analysis jobs are processed by a suitable worker.
- Search service— provides help search functionality.
- Web pool which runs:
  - Web application—handles requests to the LGTM web interface and also processes the results of tasks and jobs.
  - Web proxy service—acts as a proxy and SSL terminator for the LGTM web interface.

For more information about these services, see the "Services" topic in the LGTM Enterprise administrator help.



The server that hosts the scheduling service is referred to as the coordinator. If the control pool components are distributed across multiple servers, it is important to know which server is the coordinator because LGTM command-line actions must be run from this server.

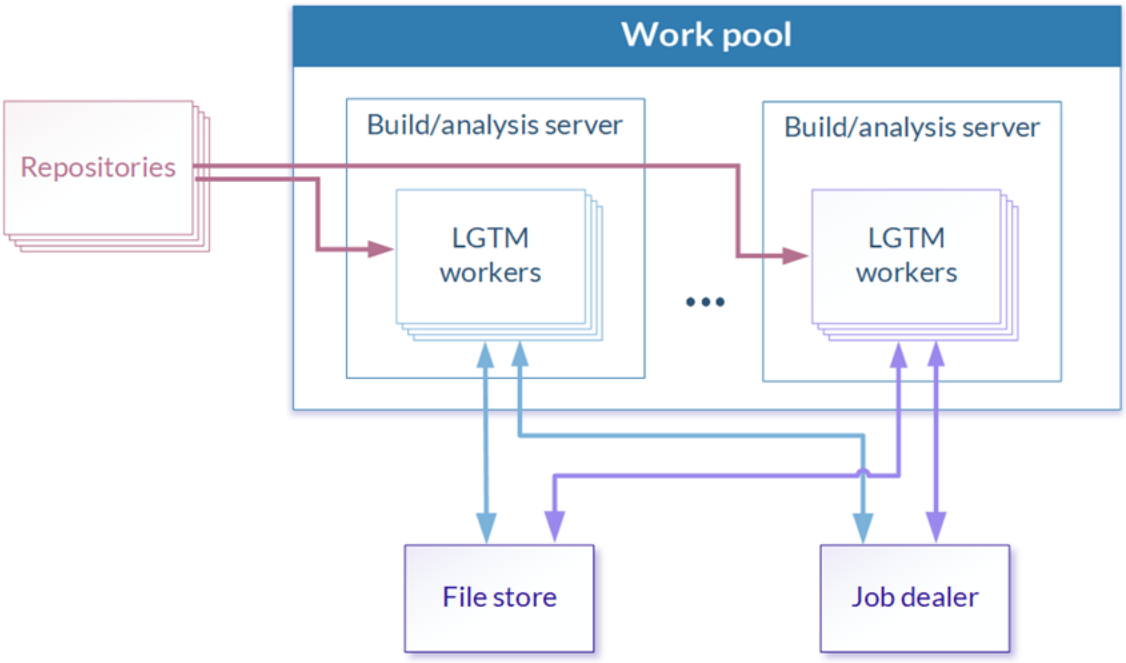
Work pool

The work pool consists of one or more build/analysis servers. In the simplest demo setup the work pool and the control pool may be hosted on the same server and you may only have the minimum of two workers (one for users' custom queries, one for all other tasks).

For deployments, the work pool will consist of multiple servers, each of which may host multiple LGTM workers.

Each worker can access three types of resource:

- **Repositories**
  Workers can download source code from a repository.
- **File store**
  Workers download files from the file store in the control pool. Files include Semmle Core files and project configuration files. After analysis, workers upload data to the file store—for example, project snapshots containing query results and a source archive.
- **Job dealer**
  Workers poll the job dealer for the next job to work on. Some workers are configured only to retrieve jobs that run user-defined queries from the Query console, to ensure these are processed quickly. If a worker has one or more labels assigned to it, it retrieves jobs with matching labels in preference to unlabeled jobs, and ignores any jobs with other labels. On completion of a job, the worker informs the job dealer.

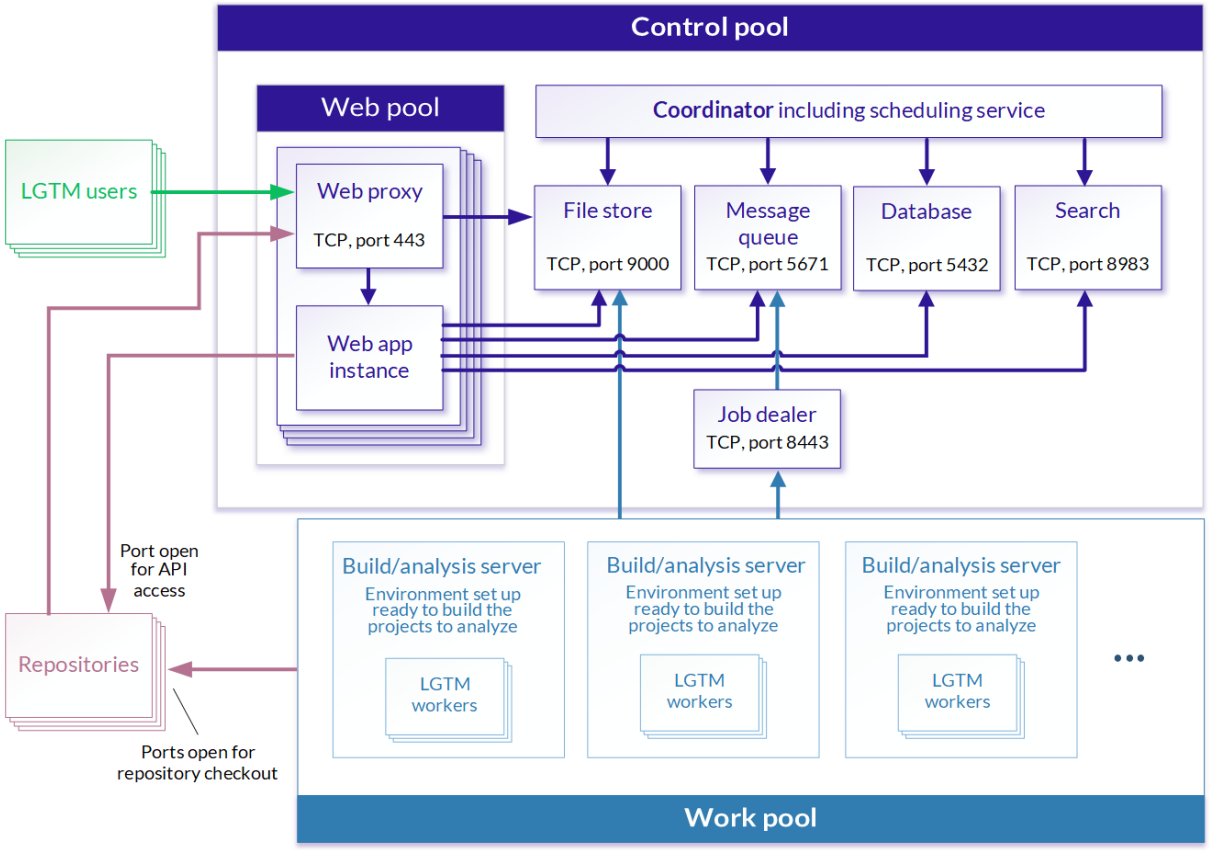


Note
The arrow directions on the diagram give an indication of data flow between system components.

Network connections

For a live site deployment, the system is scaled by being installed on a server cluster containing control pool and work pool servers. When the web pool contains more than one machine, you can use an external load balancer, or round-robin DNS, to share the load between the web pool machines.

Note
In the following diagram, the arrow directions indicate connection requests from one system component to the component to which the arrow points. Where an arrow points from A to B, A must be able to see B, and B must accept requests from A on the specified port.



Communication security

All LGTM components use secure channels, protected by SSL certificates, to communicate with each other. LGTM also uses secure connections to fetch source code your repositories, provided that you configure the repository host and LGTM to use secure connections.

Worker processes

The architecture of LGTM workers is similar to that of continuous integration tools like Jenkins and Atlassian Bamboo. The build and analysis processes run using a single operating system user name for multiple builds. LGTM does not enforce any barriers between builds running on the same machine, either concurrently or consecutively.

Individuals with write access to the LGTM build and analysis configuration for a project can specify exactly which commands are run during the analysis. LGTM configuration files are stored in the repository with the source code, consequently any changes to them are subject to your organization's code review process.

If a malicious developer wanted to introduce an LGTM configuration that accessed the source code of other projects, they would need to submit it as a normal code change. Depending on the configuration of the version control system, this may be apparent in commits, limiting the severity and likelihood of such an activity.

If you are concerned about securing one or more code bases from potential attacks by your own developers, you are advised to run a separate installation of LGTM Enterprise for such projects with workers run on dedicated machines.