

Creating work areas required by grotto

Introduction

Grotto requires a specific directory structure on the server. Also, the repository root in ergatis must have certain parameters. Before running any transcriptomic analysis on grotto, the following steps are required.

Procedure

1. Create the requisite directories.

To run transcriptomics pipelines, there are required directories. You can create them using the script `/local/projects/RNASEQ/script_repository/shell/create_rnaseq_analysis_dirs.generic.sh`.

For this example, we will imagine a project XYABC, you should use your own already existing project area.

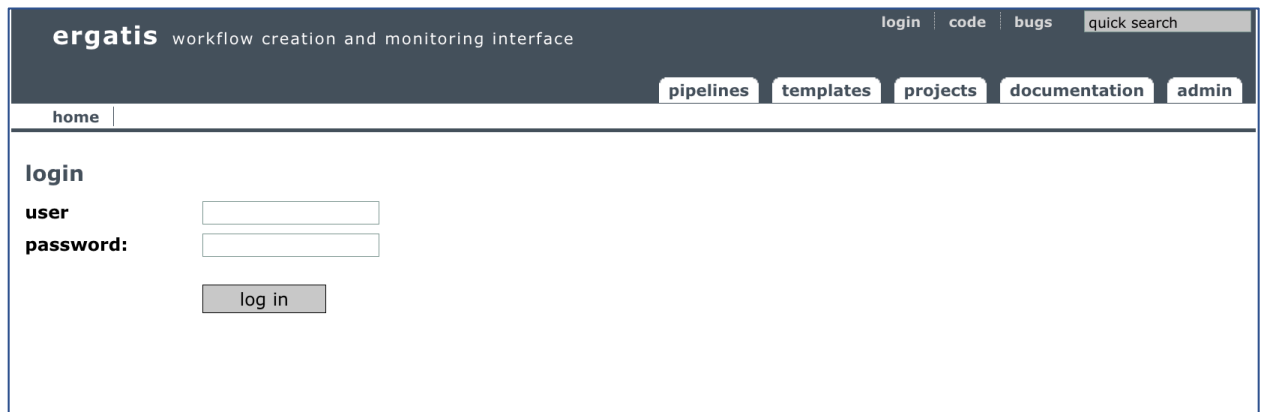
```
/local/projects/RNASEQ/script_repository/shell/create_rnaseq_analysis_dirs.generic.sh --p2 /local/projects-t2/XYABC --p3 /local/projects-t3/XYABC --t /local/scratch/XYABC
```

The script creates the necessary links between 'projects-t2' and 'projects-t3' to store large output files in 'projects-t3' instead of 'projects-t2'.

2. Create ergatis repository root using the ergatis GUI.

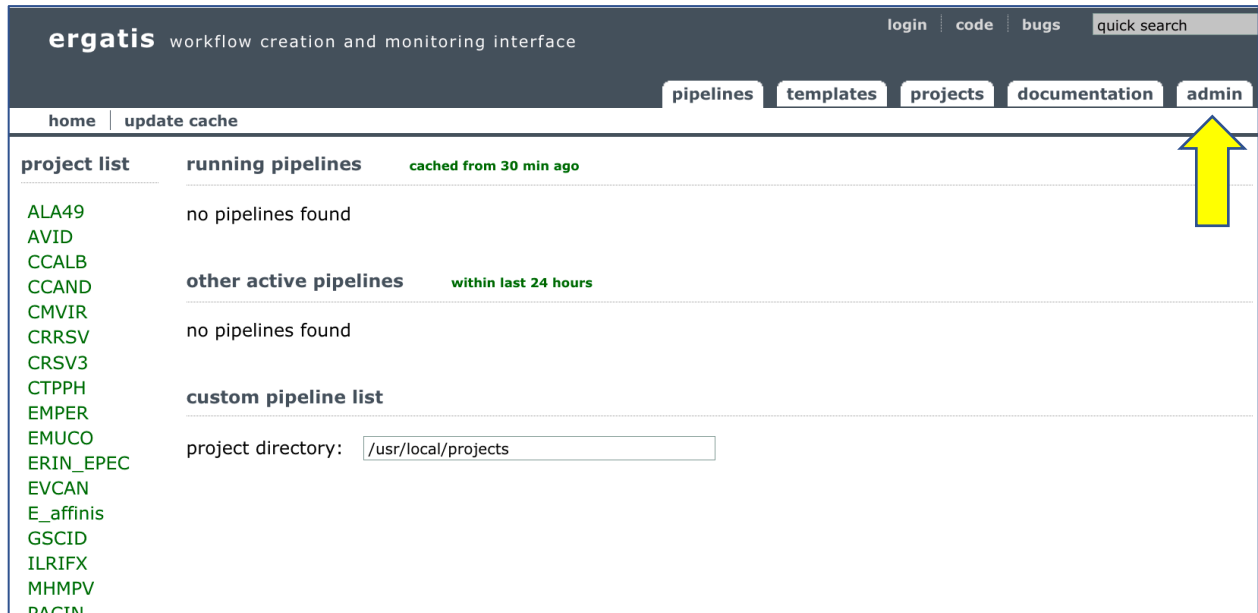
Open ergatis at <http://ergatis.igs.umaryland.edu/cgi/>.

Login with your IGS credentials.



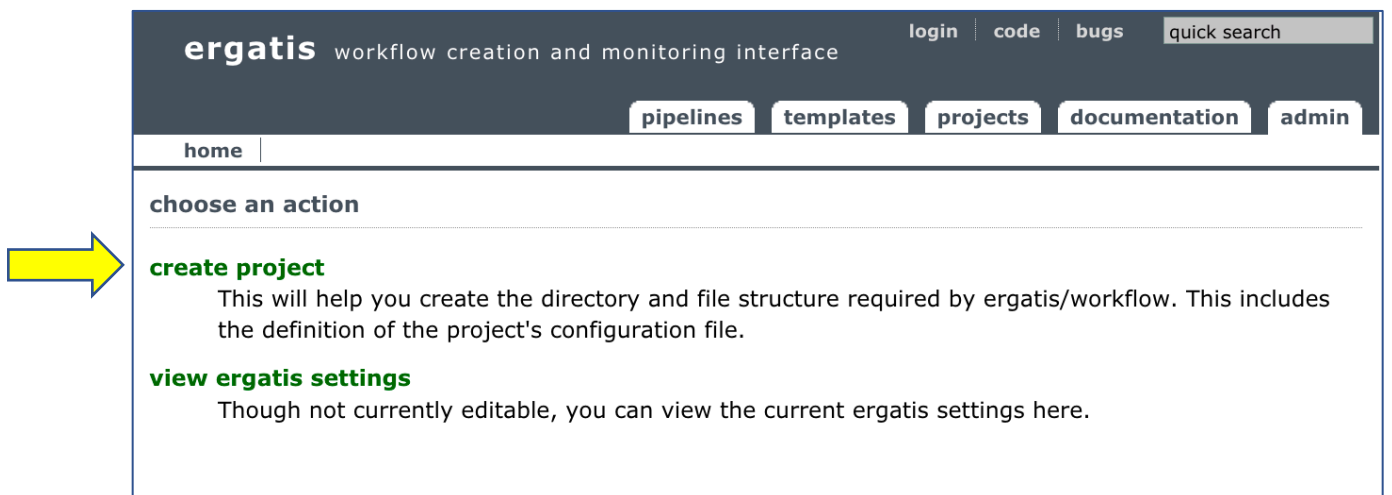
The screenshot shows the ergatis web interface. At the top, there's a dark header with the ergatis logo and the text "workflow creation and monitoring interface". To the right of the header are links for "login", "code", "bugs", and a "quick search" box. Below the header is a navigation bar with tabs for "home", "pipelines", "templates", "projects", "documentation", and "admin". The "home" tab is currently selected. The main content area displays a "login" form with labels for "user" and "password:", each followed by a text input field. Below these fields is a "log in" button.

Select the 'admin' tab.



The screenshot shows the 'ergatis' workflow creation and monitoring interface. The top navigation bar includes links for 'login', 'code', 'bugs', and a 'quick search' box. Below this, a secondary navigation bar contains tabs for 'pipelines', 'templates', 'projects', 'documentation', and 'admin'. The 'admin' tab is highlighted with a yellow arrow. The main content area is divided into two columns. The left column, titled 'project list', contains a list of project names: ALA49, AVID, CCALB, CCAND, CMVIR, CRRSV, CRSV3, CTPPH, EMPER, EMUCO, ERIN_EPEC, EVCAN, E_affinis, GSCID, ILRIFX, MHMPV, and PACTN. The right column, titled 'running pipelines', shows 'no pipelines found' and 'other active pipelines' (also showing 'no pipelines found'). Below this, there is a 'custom pipeline list' section with a 'project directory' field containing the path '/usr/local/projects'.

Select 'create project'.



The screenshot shows the 'ergatis' workflow creation and monitoring interface. The top navigation bar includes links for 'login', 'code', 'bugs', and a 'quick search' box. Below this, a secondary navigation bar contains tabs for 'pipelines', 'templates', 'projects', 'documentation', and 'admin'. The 'admin' tab is highlighted with a yellow arrow. The main content area is divided into two columns. The left column, titled 'project list', contains a list of project names: ALA49, AVID, CCALB, CCAND, CMVIR, CRRSV, CRSV3, CTPPH, EMPER, EMUCO, ERIN_EPEC, EVCAN, E_affinis, GSCID, ILRIFX, MHMPV, and PACTN. The right column, titled 'running pipelines', shows 'no pipelines found' and 'other active pipelines' (also showing 'no pipelines found'). Below this, there is a 'custom pipeline list' section with a 'project directory' field containing the path '/usr/local/projects'.

Type project path as '/local/projects/XYABC/rnseq/ergatis', where XYABC is your project area.
Click the 'create project' button.

cmccracken [logout]codebugsquick search


ergatis workflow creation and monitoring interface

home | pipelines | templates | projects | documentation | admin

create a project

This form is used to create a new project area where you can run new pipelines. Enter the path where you want the project area to be created. The directory you enter should exist already and all necessary files and directories necessary will be created inside it.

- ☐ enter a project root
- ☐ check project root existence
- ☐ check project root permissions
- ☐ create workflow directory
- ☐ create runtime directory
- ☐ create pipeline directory
- ☐ create lock files directory
- ☐ create project id repository
- ☐ validate project id repository
- ☐ create output directory
- ☐ create default project configuration
- ☐ create project templates directory



Click the 'configure project config file' button.

ergatis workflow creation and monitoring interface cmccracken [logout] | code | bugs | quick search

[home](#) | [pipelines](#) | [templates](#) | [projects](#) | [documentation](#) | [admin](#)

create a project

This form is used to create a new project area where you can run new pipelines. Enter the path where you want the project area to be created. The directory you enter should exist already and all necessary files and directories necessary will be created inside it.

[create project](#)

- ✓ enter a project root
- ✓ check project root existence
- ✓ check project root permissions
- ✓ create workflow directory
- ✓ create runtime directory
- ✓ create pipeline directory
- ✓ create lock files directory
- ✓ create project id repository
- ✓ validate project id repository
- ✓ create output directory
- ✓ create default project configuration
- ✓ create project templates directory

Almost finished. You now need to configure the project config file:

[configure project config file](#)

Edit the project config file.

- Enter your project and project code. The project code is specific to your project and is used for billing.
- Change “package-latest” to “package-rnaseq” for ergatis dir, lib dir, bin dir, docs dir, and software config parameters.
- Click the save configuration button.

ergatis

workflow creation and monitoring interface

cmccracken [logout]

code

bugs

quick search

pipelines

templates

projects

documentation

admin

home

edit config file

/local/projects/RNASEQ/XYABC/rnaseq/ergatis/workflow/project.config

You are now editing the configuration file for a new project. Most of the settings have been populated from the ergatis.ini file within your installation and, if correct, you shouldn't have to change any of these values except the first (project abbreviation).

project

Usually the name or abbreviation for a project (no spaces). It will serve as the prefix of many IDs created, so you should keep it to 3-6 characters.

projectXYABC

repository root

The root of your project directory, under which should exist the 'workflow' directory

/local/projects/RNASEQ/XYABC/rnaseq/ergatis

This area will be used as temporary space for the project. It should be cleaned out regularly, usually by some cron job.

tmp dir

/local/scratch/\$_PROJECT\$/\$_COMPONENT_NAME\$/\$_PIPELINEID\$/\$_OUTPUT_TOKEN\$

SunGrid Engine supports grid usage monitoring by project codes, allowing admins to keep track of grid time across different groups. This will cause Workflow Engine to utilize SGE's -P directive. Required: if you don't know what to put here, just use 'general'

project code

xyabc-proj-code

All variables below must be full paths - no variable references or tokens.

This should be the path to the id repository under the 'workflow' directory for your project.

project id repository

/local/projects/RNASEQ/XYABC/rnaseq/ergatis/workflow/project_id_repository

The ergatis scripts and libraries should be found here.

ergatis dir

/usr/local/projects/ergatis/package-rnaseq

Path to the ergatis lib directory, usually under the ERGATIS_DIR

lib dir

/usr/local/projects/ergatis/package-rnaseq/lib

Path to the ergatis bin directory, usually under the ERGATIS_DIR

bin dir

/usr/local/projects/ergatis/package-rnaseq/bin

Path to the ergatis docs directory, usually under the ERGATIS_DIR

docs dir

/usr/local/projects/ergatis/package-rnaseq/docs

Turn on pipeline email notification by default. If this flag is set any user creating a pipeline under this project will always receive email notifications regarding the pipelines status upon completion or failure

email notification

0

Default debug setting

debug

0

include

Path to the software config file. It contains the path to the binary for each component as well as common data paths.

software config/usr/local/projects/ergatis/package-rnaseq/software.config

save configuration

Once this is complete, there should be a message on the screen that you have been successful. You are now ready to launch an analysis pipeline on grotto.