

Rc²: R Collaboration in the Cloud

UseR! 2011

E. James Harner and Mark Lilback

Department of Statistics
West Virginia University

August 16, 2011

Outline

Introduction

Architecture

Use Cases

Course Management

Overview

Rc² (R cloud computing) is a web 2.0 front-end to R which is:

- cloud based (via online storage and Software-as-a-Service);
- highly scalable;
- collaborative (via shared sessions, workspaces, and files);
- cross platform (via WebKit browsers);
- touch optimized (via client specific style sheets and scripting).

This allows researchers to collaborate over the Internet without concern for data becoming out of sync.

Users can start long-running computations and Rc2 will notify the user(s) when the process is complete.

Architecture

Rc² has a 4-tier architecture:

client WebKit-based browser (iPad native client planned)

application Jetty webserver and appserver, with applications written in Ruby (Sinatra) and Java (servlets)

database PostgreSQL for meta-data, user profiles, courseware, etc.; Apache Cassandra (NoSQL—key-value) for data files, R code, R objects, etc.

compute engine R with Rserve connectivity

The three backend tiers can be clustered across multiple servers.

Client

Client interfaces (HTML 5 and JavaScript) are comparable in speed and functionality to desktop R interfaces and include:

- a multi-tab text editor for .R and .Rnw files;
- a command line;
- R text output;
- a workspace display;
- an interactive graphics display supporting multiple plots;
- voice chat capability.

Clients communicate asynchronously with the server using websockets. Websockets allow persistent client connections and are fully supported by Jetty and WebKit-based browsers.

Workspaces

A workspace is a superset of an R workspace. It has a list of associated files (no directories) along with all objects that would be stored in an `.Rdata` file.

Workspaces can be:

- shared with another user for collaboration;
- cloned to a list of users as a copy (useful for teaching);
- flagged as complete if cloned.

Cloned files can be read-only references or copied. By default, `.R` and `.Rnw` files are copied, while data files are referenced to save space.

File Hierarchy

Users can upload/manage a full file structure of directories and files. Files can be:

- uploaded directly or created during an R session;
- added to a workspace from the file hierarchy as either a copy or a read-only reference.

Files saved in an R session are placed in the file system in the same path as the workspace hierarchy, with a folder created with the workspace name.

Files vs. Workspaces

Most users will never interact with their file hierarchy. Rather, most will use either files they created within an R session or files that were cloned into their workspace, e.g., by a collaborator or teacher. Cloning greatly reduces setup and complexity for new users (e.g., students).

Power-users can create directories with commonly used libraries and packages, and reference them in multiple workspaces.

Graphics

A custom R device package is used to capture graphics. When a script is executed, a graphics device is created and if anything is drawn in it, it is returned to the browser.

Each graphic generated is returned as a separate image; each acts as a different device. If you generate 3 plots in one script, they will be displayed. If you run a different script in another tab with 2 plots, these 2 plots will be displayed, but they are tied to their tab.

Graphics in the browser can be closed, zoomed out, and walked through as a slide-show.

Security

A 3-value token is used for autologin, which:

- disables an account if someone attempts to hijack a session;
- limits usage to one web browser (but multiple tabs) at a time;
- logs all activity for reports and security auditing,

Rc² has a fine-grained permission system so a student in one class can be a GTA in another.

Students

Students

- log in and see a hierarchical list of classes, assignments, quizzes, examples, etc.;
- select a class and assignment and begin an R session;
- edit work in a `.R` or `.Rnw` file and then execute and debug until complete;
- submit the assignment, quiz, etc.

Generally, students submit a `.Rnw` or `.R` file, but they can also directly submit the pdf file.

Teachers

Teachers

- log in and create one or more workspaces, generally one for each assignment, quiz, etc.;
- upload .R, .Rnw, .txt files, etc.;
- clone workspaces to all the students in the class;
- load each student's .Rnw and .R files into the master workspace for execution and grading.

Data files for students are read-only, whereas .Rnw and .R files are unique copies for each student.

Collaborators

User1

- logs in and creates a workspace job1;
- starts his/her R session;
- invites User2 to log in and join the job1 workspace/session.

User1 can:

- assign control to User2;
- take back control and make a quick edit, then return control to User2.

If User1 logs off, User2 can:

- manage the session and keep working;
- complete his/her work and log out.

When User2 logs off, all files and R objects are automatically saved to job1. If User1 logs back in and loads job1, the most recent additions by User2 are restored.

Classroom

Instructor

- logs in and starts an R session;
- notifies students via their preferred method (SMS, Twitter, Facebook, or email) that a session is starting in X minutes.

Students

- log in and join the session.

Instructors can:

- walk through examples, switch files, make edits, which all sync with student clients;
- communicate by voice with students as they demonstrate statistical principles in R;
- pass control to a student;
- take back control and correct mistakes.

Course Management

A full course management system is planned for multiple courses, with each course potentially containing multiple sections, with each section potentially linked to multiple labs.

Administrators, content creators, instructors, graduate assistants, and students all have different permission levels.

Rc² will support assignments, quizzes, and exams. Instructors can select questions from a large pool of questions and certain types of questions can be automatically graded by Rc².

Rc² will support student interfaces to assignments, etc. and has an extensive reporting system.