

Exploring the role of Bioconductor in AnVIL

2020-08-18

Attending: Martin Morgan,

:00 - :10 Context: vision and challenges

Original vision

- Enable cloud computing for the large and impactful Bioconductor community
- Naively
 - Launch cloud-based RStudio
 - Gain access to Bioconductor packages and resources
 - Better experience than on the desktop
 - System configuration challenges satisfied (docker images)
 - Fast package management
 - Scalable computing
- Good things with the current situation
 - Access to large and protected data
 - Account management
 - Cloud computing infrastructure offered by Leo
 - 'Custom environments' have allowed us to move development forward (but not to promote our efforts to users)

Current status / challenges

- Jupyter notebooks supporting R are not a good enough solution
- RStudio is not supported, and the roadmap to support has become less rather than more clear.
- R and Bioconductor requiring ongoing maintenance (e.g., releases of R and semi-annual releases of Bioconductor); is sufficiently on the radar of our Broad counterparts?
- Technical call meetings are the primary means of learning about recent developments, but it is hard to know how much confidence to place in future directions.

:10 - :40 Exploring paths to first-class exposure to RStudio

- Roadmap to RStudio in Leonardo?
- Roadmap to RStudio / Bioconductor via kubernetes?
- Other approaches or options we should be considering

:40 - :60 Interfacing with Terra and other AnVIL components

- Are REST APIs the best bet for continued programmatic access?

- An alternative -- rely on python clients invoked via R?
- Options for improving transparency of developments. E.g.,
 - Leonardo API deprecations (we were notified about this [good]) -- it would be good to know what other deprecations are on the horizon, for all AnVIL APIs.
 - Improved practices, e.g., semantic versioning of the APIs
 - Published road maps?