

Webinar Recap: Distributed Computing & R Server

Colby Ford

Last week, our monthly webinar series covered Distributed Computing and What's New in R Server 9.0.1. If you missed the session, you can find the [recording here](#). We received a number of questions during the presentation and wanted to take the opportunity to provide some insightful answers for the audience.



If I am just getting started with R, how should I go about choosing between open-source and Microsoft R?

It really depends on your needs, but luckily both [open-source R](#) and [Microsoft R Open](#) are free to get started. Either solution will allow you to practice working in the environment, and once you get comfortable in R and you starting running into the limitations with R on large datasets (the ones that won't fit in your computer's memory), you can look into purchasing R Server, or spin up an R Server on Azure in HDInsight or on the Data Science Virtual Machine.

When thinking of SAS vs. R: as a SAS user how difficult or easy would it be for me to learn and start using R?

Both SAS and R do very similar things. However, SAS code is written in procedures where R is more of a script. Since you're already familiar with the algorithms/functions you use in SAS, learning the syntax of R really isn't that bad. Plus, the cost savings of R over SAS are tremendous! Check out the example below for a quick comparison.

	SAS	R	Microsoft R

Linear Regression	<code>proc reg data=mydata; model y=x1 x2 x3;</code>	<code>lm(y~x1+x2+x3, mydata)</code>	<code>rxLinMod(formula: y ~ x1 + x2 + x3, myxdfdata)</code>
-------------------	--	---	---

What use cases do you see being used with the R server technology?

I think the use cases depend on when you want to take your R practice to the next level:

- Larger, collaborative data science teams
- Larger amounts of data (tweets, click stream data, genomics, etc.) a.k.a. “Big Data”
- When operationalizing the models you make is crucial to providing business insight. Also, when you want to easily maintain, update, and rerun them.

Is there a benefit to using R if my dataset is in the range of two to four million rows of data?

In short, it depends. Two to four million rows might not be outside of the memory limit of normal R, but it is approaching it for many machines. The added value of R Server won't necessarily be in its power to handle more data, but in its ability to run computations on that data in a more reasonable timeframe. Millions of rows may take a long time to train the model or predict upon, but using R Server will definitely cut down on that time.

Do you have any virtual instructor-led R training?

While BlueGranite does not currently offer virtual training opportunities for R, our 3-day [Microsoft R Training](#) session takes place at your location and features hands-on, instructor-led lessons for up to 10 of your firm's attendees. A BlueGranite senior consultant will facilitate hands-on labs and provide material for your team on the fundamentals of R programming for data ingestion, exploratory data analysis, model building, evaluation, and operationalization. Attendees from your company will learn how to write effective R code that can be operationalized in production.

Additionally, there are many online resources that could help you get started. I would recommend looking into [free courses online](#) with edX as a beginning point and perhaps exploring an in-person training session in the future.

Azure seems to be the way everything is moving – would you agree?

I think so because of the flexibility and expandability of the Azure environment. Why would your organization want to pay thousands (or millions) on a big server system that will be out of date in a few years when you can just pay monthly only for what you use on Azure? Plus, the ease of setup is an added bonus. Instead of having to hire a crew to come in and setup Hadoop on your on-premise server, just spin it up on Azure in less than 30 minutes.

Thank you to everyone who joined us for the webinar! If you have any more questions or want to know more about R training opportunities, feel free to [reach out to us](#).