



GEORGETOWN UNIVERSITY

**MASSIVE DATA
INSTITUTE**

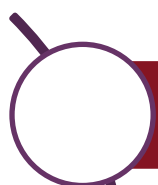
McCourt School of Public Policy

MDI Infrastructure Services

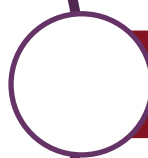
The Massive Data Institute (MDI) has available infrastructure for faculty to use to support research related to public policy. We use a combination of Google and Microsoft Cloud resources to meet the needs of our faculty and researchers. To date it has already been used by faculty to investigate fraud in presidential elections, understand online movements, process legislative documents and predict forced migration. Below is a list of compute resources that are available. If you are interested in using these resources, please fill out the following form:

<https://forms.gle/gscXnW4xFCZ1zTGQ6>

MDI



Rstudio for Large Simulations



Linux Server with a Database



Spark Cluster

Rstudio for Large Simulations

MDI has created a high-performance R server that has access to 74 vCPUs and 144GB memory (RAM). We can also setup smaller versions as needed (16 core and 64 GB memory is a typical configuration). These servers can be accessed via the web through R or RStudio, allowing researchers to run R code remotely. These machines are shutdown when they are not in use.

Cost:

- Free for 4 weeks of use per year
- For smaller machines - \$1000 per year
- For larger machines - \$2500 for 6 months

Linux Server with a Database

MDI will setup a server for faculty to use to store data and to run smaller jobs on. These servers typically have 16 GB memory, 128 GB storage, and 4vCPU, but they can be configured differently if needed. We will install the database and necessary software and add accounts for faculty and students using the server. It is assumed that faculty using this option are familiar with a Linux environment and will use ssh to connect to the server. Servers are not shutdown when they are not in use and data is backed up.

Cost:

- Free for 6 months
- For smaller machine - \$500 per year
- For larger machine - \$2500 per year

Spark Cluster

MDI has Spark clusters of different sizes available for running parallel compute jobs. Typically, our clusters have between 4 and 25 nodes.

Cost:

- Free for 3 months if number of nodes is 5 or less.

- Free for 1 month if more than 5 nodes.
- For smaller cluster - \$500 per year (*configured with 5 small nodes or less*)
- For medium cluster - \$2000 per year (*configured with 12 medium nodes or less*)
- For large cluster - \$5000 per year (*configured with 25 nodes or less*)

If you need more than a single server or a single resource, we have flat package rates.

- Small tier - \$2500 (A small number of servers)
- Medium tier - \$5000 (A mix of servers, Spark and RStudio usage)
- Large tier - \$10,000 (A larger mix)
- Platinum tier - \$15,000 or higher