



Big Data with R: Tutorial on the ff package

Ways to get started

1. Go to <http://cran.revolutionanalytics.com/>
2. Download precompiled binaries for either



3. Profit

LOOKS CLOUDY - BY FGURNEE



Stop! Not all data need apply



- R's RAM police are always on guard to Big Data
- R requires data to be stored locally into the computer's RAM
- But what if you have a project that requires you to give a unique name id to every star in the Milky Way (it could happen)

ff package saves the day

- Available from CRAN repositories
`install.packages("ff")`
- Stores data to disk but behaves like it's in RAM
- Transparently maps to the data only storing up in memory the associated link to the disk
- Provides methods to use data temporarily or archive for future
- Allows to work with Large Data, Standard HW, Minimal RAM, and Maximal Performance

R Data Types Supported

- Boolean, logical, quad, nibble, byte, ubyte, short, ushort, integer, single, double, raw
- Compounds
 - Factor, ordered, Date, POSIXct, POSIXlt

R Data Structures Supported

	example
vector	<code>ff(1:12)</code>
array	<code>ff(1:12, dim=c(2,2,3))</code>
matrix	<code>ff(1:12, dim=c(3,4))</code>
data.frame	<code>ffdf(gender=a, age=b)</code>
symmetric matrix with free diag	<code>ff(1:6, dim=c(3,3), symm=TRUE, fixeddiag=NULL)</code>
symmetric matrix with fixed diag	<code>ff(1:3, dim=c(3,3), symm=TRUE, fixdiag=0)</code>
distance matrix	
mixed type arrays	

Advanced ff features

- Parallel execution
 - Ex. snowfall package
- Bit filters
- Hybrid indexing

Enough talk... lets go to the
code!