I think that documentation should be organized better than a FAQ, but if information haven't found its final destination yet, better have some place for it. My goal is to keep the FAQ a staging area and not let it grow to *be* the documentation.

Exactly what does the RHadoop/rmr developer need to have installed on the hadoop nodes?

R and all the libraries needed, including rmr and its dependencies and any other library that the developer needs. This is far from perfect and we are aware that some people simply can't install anything on their cluster because of policies, let alone technical difficulties. There seems to be a way around it but it's not very easy to implement. In dev under tools/whirr there is a whirr script that will do this for you, tested on AWS EC2

What releases of R are supported?

See Compatibility

Is there any mechanism for installing R on the hadoop nodes?

As a demo, unsupported, you can look under tools/whirr.

What happens when there are R code failures?

If they are temporary, hadoop may retry a specific task a number of times. If that number is exceeded, the hadoop job will fail and the mapreduce call will fail. The stderr from R is your friend. In hadoop standalone mode, which is highly recommended for development, it simply shows up in console mixed with somewhat verbose hadoop output. In pseudo distributed and distributed modes it ends up in a file somewhere. This is a good resource about that.

What is debugging like on hadoop?

See [[Debugging rmr programs]]

Is there some minimal set of things that an R programmer needs to know about hadoop?

This isn't an easy question, but let me try. Understanding mapreduce is the first priority (the original google paper is still the reference point). Reading a variety of papers with different applications, probably the ones closer to one's problem domain. Cloudera's Hammerbacher has a collection on Mendeley and another one is on the atbrox blog. Somewhat off topic, I would also recommend people acquaint themselves with the parallel programming literature for architectures other than mapreduce.