Zillow's Big Data and Real-time Services in R

Zillow is the leader in providing data and analysis for consumers and professionals in the real estate market place. The flagship product is the Zestimate™, an estimate of the value for over 90 million homes in the US. Millions of models are behind the Zestimate and the models are constantly being retrained. The model fitting and scoring infrastructure rests on top of R, allowing rapid prototyping and deployment to production servers. In order to more fully integrate the Zestimate infrastructure with the Zillow production environment, we have developed ZillowRServe, a client/server interface based on the RServe package.   ZillowRServe allows the Zestimate to operate as a service as well as providing basic monitoring and controlling of the Zestimate application. ZillowRServe provides a simple generic application that can automatically spawn and manage a number of parallel R processes. It also provides utilities specific to Zillow for dealing with real-estate data. For false tolerance and load balancing, we deploy multiple redundant ZillowRServe on multiple machines. We also embedded a ZillowRServe client in stored procedures for use within SQL server to handle real-time data import.