Duncan asked "What's new?" in my thesis work. This will help determine the timeline for graduation.

This list is ordered with the most developed work on top. I think the Data Description and Loop Fusion are the most promising to work on for the next couple months.

Our work focuses on inferring and using parallelism in existing code. In contrast, most existing software for parallelism requires the user to write their code according to a specific programming model.

**Extensible Framework for parallelism** The makeParallel package analyzes and generates parallel code in a modular way. In contrast to alternative R implementations, we generate GNU R.

**Static Analysis** This is the first attempt that we're aware of to infer and represent task and data parallelism implicit in R code. This goes beyond existing work in R because it looks at an entire script holistically, which allows us to perform entirely different optimizations. Example optimizations include reading in subsets of the data, and running statements in a different order.

**Generating R Code to run in Hive** This model goes beyond existing R to SQL translators because it allows efficient user defined R functions that run inside a database. It uses existing interfaces, `stdin` on each Hadoop worker node.

**Forking Scheduling Algorithm** This is a new algorithm for task scheduling that picks the single best task to run in parallel at each step. It's useful because it works with some existing R parallel infrastructure.

**Data Description** We use a physical and logical description of the data to generate more efficient code. For example, each parallel worker can directly load a chunk of the data, rather than reading in everything at the manager and transferring. This is only novel in R- Hive metastore does this, as do protocol buffers.

**Loop Fusion** Combine multiple data parallel statements into one based on the code analysis. Compilers do this at a much lower level, but this is the first application to a high level data analysis language that we're aware of.

Cite something to make the bibliography appear [R Core Team, 2018].

# Bibliography

[R Core Team, 2018] R Core Team (2018). R language reference.