A model is an optimized statistical algorithm that consists of one or more of the following:

* Input variables
* Outcome variables
* Code representing a specific statistical analysis in either R or python or H20 (Version2)
* Mappings from generic code to specific data source field names (Version2)
* Pointers to 3rd party data (Version2)
* Data transform rules (Version2)

The purpose of this model management solution is to:

* Create models consisting of transformation and analytics code
* Assign model meta-data
* Search for models
* Executing models in Spark (Version2)
* Creating training and test data sets (Version2)
* Monitor model processing
* View results and alter models (Version2)
* Save models upon change (Version2)
* Organize model for model ensembling using model templates
* View model ensemble
* Select models for inclusion in model ensemble
* Store model ensemble

Preconditions:

1. An existing model ensemble is being executed
2. Execution is occurring within Spark
   1. Data is being retrieved from DB and loaded into Spark
   2. Data is being processed
   3. Scores are being obtained
   4. Scores are being stored in DB
3. A model directory is in place
4. The model directory tracks model ensemble membership

Steps:

1. Data is loaded into data store
2. One or more business outcomes are selected for testing
3. A random sample of data of a specific size is retrieved for the purposes of training and loaded to Spark
4. A random sample of data of a specific size is retrieved for testing and loaded to Spark
5. A model is executed for training
6. At various points in the training process and at the completion of training the model is executed against a test data set
7. The results are displayed comparing new and old models based on business outcomes
8. Model ensembling process occurs adding new model to existing ensemble and testing.
9. Ensemble directory changes