1. What is the approximate depth of a Decision Tree trained (without restriction)on a training set with one million instances?

I guess around million? Because unconstrained decision tree severely overfits to the data

Depth = log_2(m), so log_2(10^6) ~= 20, or a little more.

2. Is a node's Gini impurity generally lower or greater than its parent's? Is it generally lower/greater, or always lower/greater?

Gini impurity = 1 - (a/a+b)^2 + (b/a+b)^2 if there are two subsets: a, b. I think it depends on whether you chose the metric as Gini impurity or not. If it is the case, the node should always have lower Gini impurity. Otherwise, why make the node?

Generally lower, due to CART's cost function. The Child's gini impurity may be greater than its parents so long as the other child's gini impurity compensates the cost.

3. If a decision tree is overfitting, is it a good idea to decrease max_depth?

Good idea. Higher max_depth usually results in overfitting model.

4. If a Decision Tree is underfitting, is it a good idea to try scaling the input features?

Probably not. Decision tree is already robust to the input features

5. If it takes one hour to train n=1M Decision Tree, how about n=10M?

Training complexity = O(n*mlog2(m)), so ten hours.

6. Will presort=True speed up training?

Dataset is large, so probably not.