After Watson has completed the initial two steps of question analysis and hypothesis generation, it's time to move on to step three, where each of the hypotheses are scored.

In this step, Watson computes confidence levels for each possible answer or hypothesis.

This is necessary to accurately estimate the probability of a proposed answer being correct.

Watson will only buzz and to answer a question if a confidence level for one of the hypotheses is above a threshold.

To compute these confidence levels, Watson combines a large number of different methods.

First, Watson starts with a lightweight scoring algorithms to prune down the large set of hypotheses.

Recall that in step two, about 200 different hypotheses were generated.

An example of a lightweight scoring algorithm is computing the likelihood that a candidate answer is actually an instance of the LAT.

For the Mozart symphony question where the LAT is "this planet," a candidate answer like "Earth" would have a very high score, but a candidate answer like, "the moon" would have a lower score.

If the likelihood is not very high, Watson throws away the hypothesis.

They candidate answers that pass this step proceed to the next stage of the scoring algorithms.

Watson lets about 100 candidate answers pass on to the next stage.

Then Watson goes into more advanced scoring analytics.

Watson needs to gather supporting evidence for each candidate answer.

One way of doing this is through a method called passage search, where passages are retrieved that contain the hypothesis text.

To simulate this, let's see what happens when we search for two of our hypotheses on Google.

Our first hypothesis is "Mozart's last and perhaps most powerful symphony shares its name with Mercury." And our second hypothesis is "Mozart's last and perhaps most powerful symphony shares its name with Jupiter." On Google, if we search for Mozart, symphony, and Mercury, we get about 900,000 results.

And we get some good results.

They definitely mention the three words we searched for, but Mercury is only next to symphony once.

And there's no mention about this being his last symphony.

Now, if we search for Mozart, symphony, and Jupiter, we get about 1.5 million results.

And they look much more promising.

We see the phrase "last symphony" a couple times and "Jupiter symphony" more than once.

Therefore, the hypothesis with Jupiter seems to be more supported than the hypothesis with Mercury.

Now, the scoring analytics determine the degree of certainty that the evidence supports the candidate answers.

More than 50 different scoring components are used.

One example is analyzing temporal relationships.

Consider the Jeopardy question-- "In 1594, he took a job as a tax collector in Andalusia." Two candidate answers are Thoreau and Cervantes.

However, this algorithm would determine that Thoreau was not born until 1817.

So it would give a higher score to Cervantes.

Once all of the scoring algorithms are run, Watson needs to select the single best supported hypothesis.

Before this can be done, similar answers need to be merged, since multiple candidate answers may be equivalent.

As an example, the candidate answers "Abraham Lincoln" and "Honest Abe" refer to the same person.

So the scores for these two candidate answers need to be combined.

Watson should not be viewing similar answers as competing choices.

Now, Watson is ready to rank the hypotheses and estimate an overall confidence for each.

To do this, predictive analytics are used.

To compute an overall confidence level for each candidate answer, Watson uses logistic regression.

The training data is a set of historical jeopardy questions and all of the candidate answers.

Each of the scoring algorithms is used as an independent variable.

Then, logistic regression is used to predict whether or not a candidate answer is correct using the scores.

This gives each score a weight and computes an overall profitability or confidence that a candidate answer is correct.

If the highest confidence level for one of the candidate answers for a question is high enough, Watson buzzes in to answer the question.

In total, the Watson system is composed of eight refrigerator-sized cabinets and has high-speed local storage for all information.

It originally took over two hours to answer one question.

And the team had to reduce this to two to six seconds.

In the next video, we'll see how Watson progressed in the six years between starting and playing on Jeopardy, what happened during the game, and what the Watson team is working on now.