Mark

Monday, October 19, 2015 10:04 AM

There is a diference between information and data
Entropy in data defines the informmation?

Entropy = H(S)
N = items
S = stream

If you want to store N items of data from S requires atleast HN(S) bits of storage

High H(S)
    Gomez 2013 shows that you can only do 15%

Usual compression workflow
    Precondition
    De dedpulication
    Entrop coding

Lossy and lossless compression

Lossy is just too controversial
Mark will only talk about lossless

Already users do some sort of temporal reduction

Lots of complicated math that is in marks slides
But my summary is that
    Use an interpolator to predict a value and then keep the differences

    Ratio of compression is independent of many things

Improvement in compression ratio is about 80%

Blinkdb.org