Big Data, Cloud and the NOAA CRADA at the Climate Corporation

THE CLIMATE CORPORATION

The Climate Corporation (TCC) provides decision-making tools for farmers

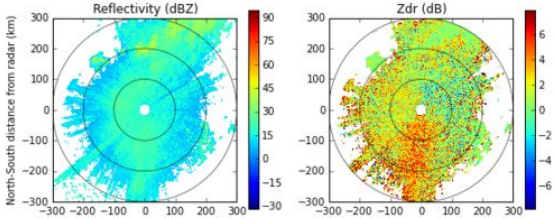
Weather is both a *feature* that we provide to our users, and an *input* to our agronomic models

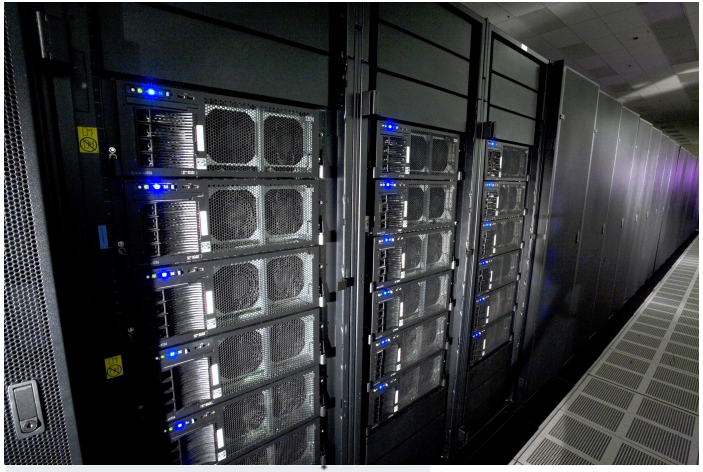
© 2015 The Climate Corporation All Rights Reserved

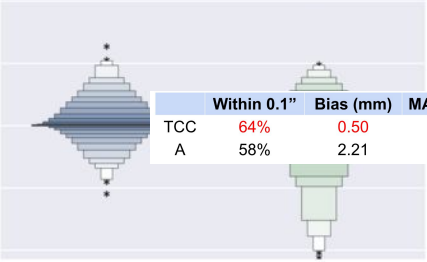
Big Data refers to computational limitations of the analysis that you are performing

```
def plot_radar_images(plots):
    ncols = 2
    nrows = len(plots)/2
    for plotno, plot in enumerate(plots, start=1):
        ax = fig.add_subplot(nrows, ncols, plotno)
        display.plot(plot[0], plot[2], ax=ax, title=plot[1],
             colorbar_label='',
             axislabels=('East-West distance from radar (km)' if plot
                         'North-South distance from radar (km)' if pl
        display.set_limits((-300, 300), (-300, 300), ax=ax)
        display.set_aspect_ratio('equal', ax=ax)
        display.plot_range_rings(range(100, 350, 100), lw=0.5, col='k
    plt.show()

plot_radar_images(plots)
```







	Within 0.1”	Bias (mm)	MAE (mm)
TCC	64%	0.50	2.73
A	58%	2.21	4.20

http://goo.gl/NB5hDO

© 2015 The Climate Corporation All Rights Reserved

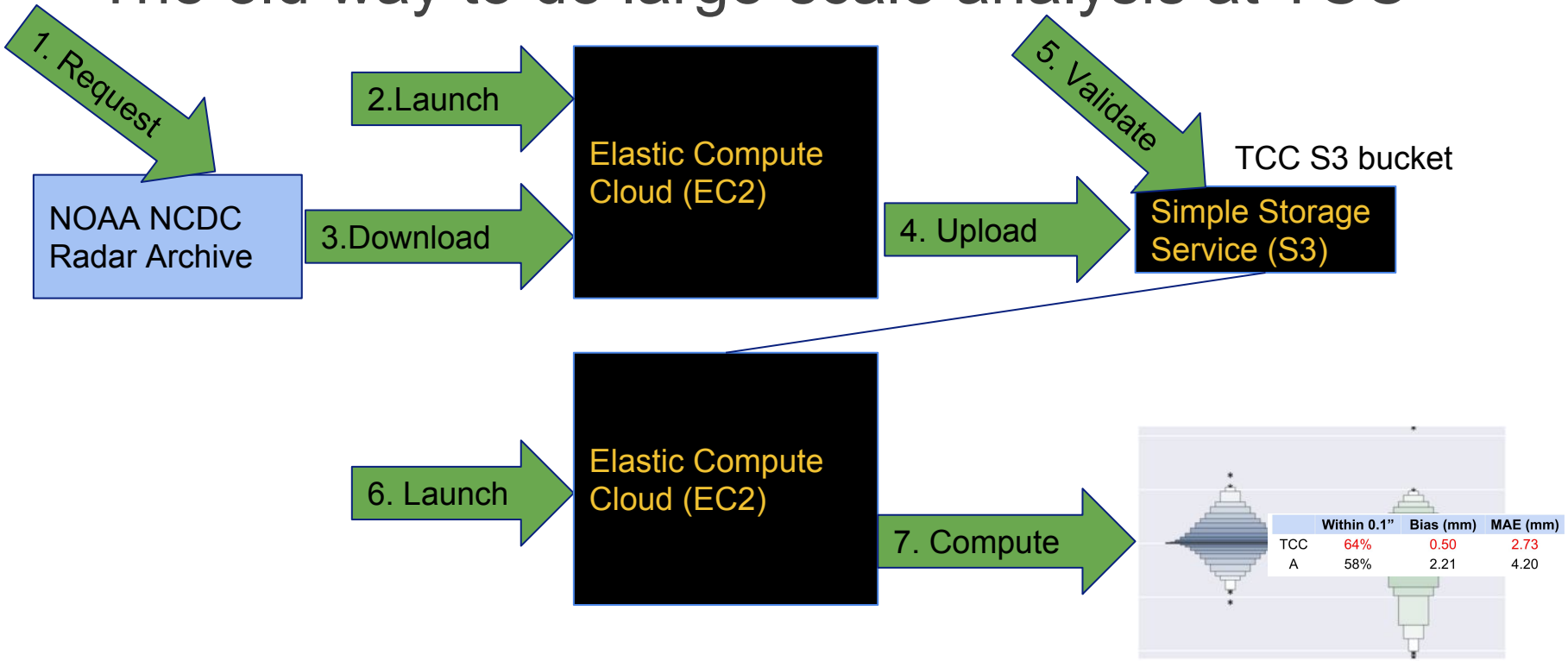
The Cloud is remote computational infrastructure, typically something you rent

https://cloud.google.com/

https://aws.amazon.com/

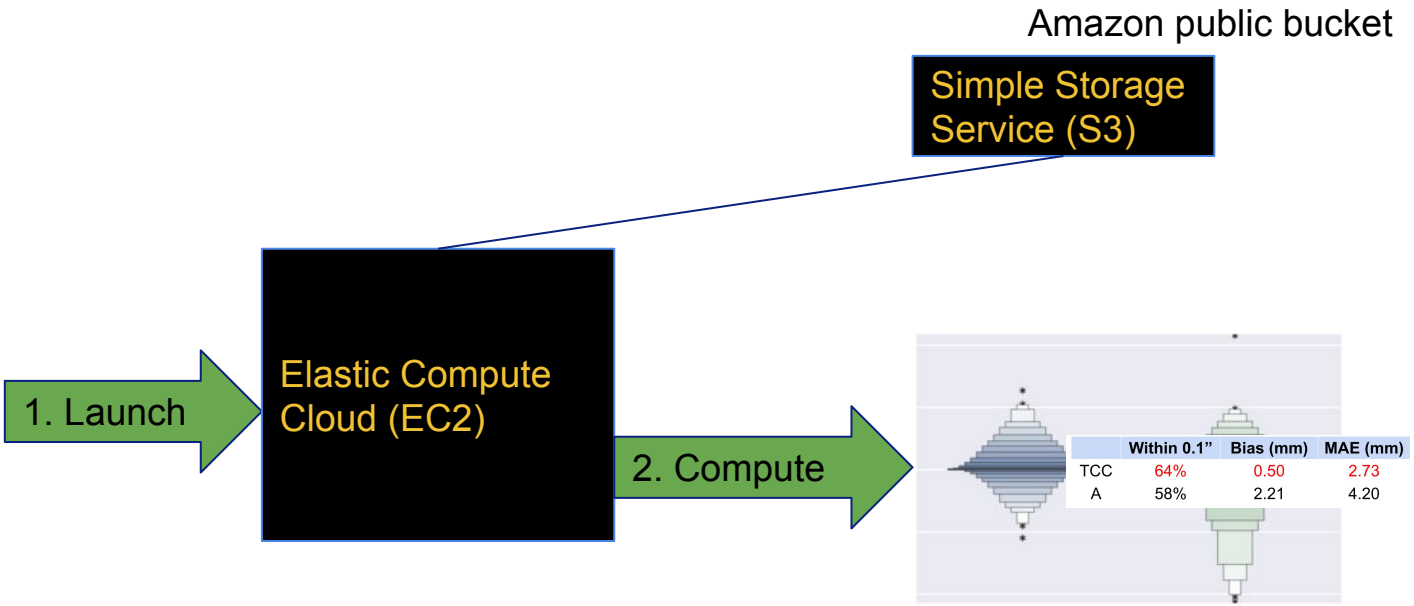
© 2015 The Climate Corporation All Rights Reserved

The old way to do large-scale analysis at TCC



© 2015 The Climate Corporation All Rights Reserved

The NOAA CRADA puts NEXRAD data on S3

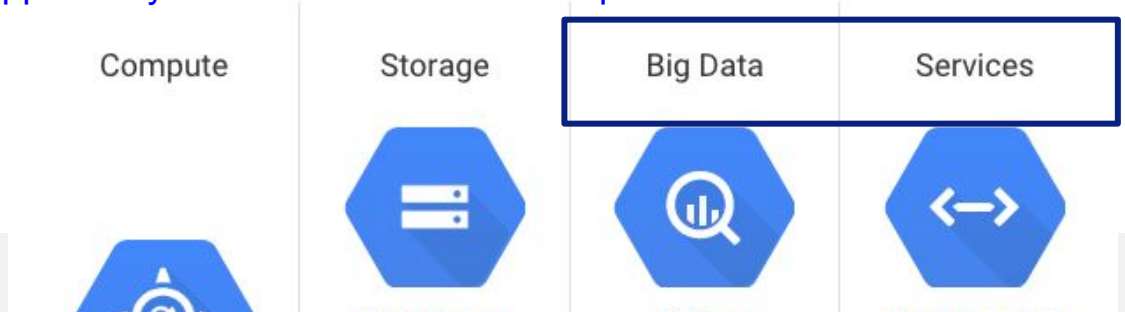


© 2015 The Climate Corporation All Rights Reserved

Everyone wins

TCC projects are several weeks shorter
TCC evaluations of new methods happen on larger datasets

We don’t pay Amazon for the S3 bucket to store NEXRAD data
Instead, we pay Amazon for the EC2 instances to process the larger dataset

NOAA data is used more widely, but without overwhelming NCDC
TCC/AWS found long-standing problem in NOAA archive, improving data quality

Opportunity around the other two components of the cloud