

Feasibility Study for Applying Cloud Infrastructure to Data Analysis Problems

PI: Brian Schlining

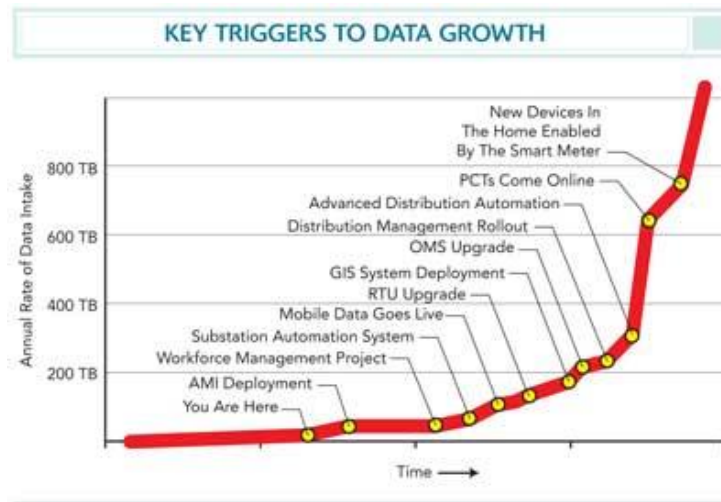
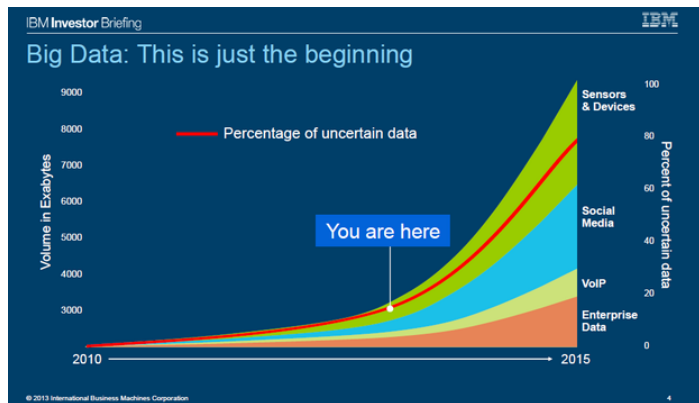
Lead Scientist: Steve Haddock

FSACIDAP

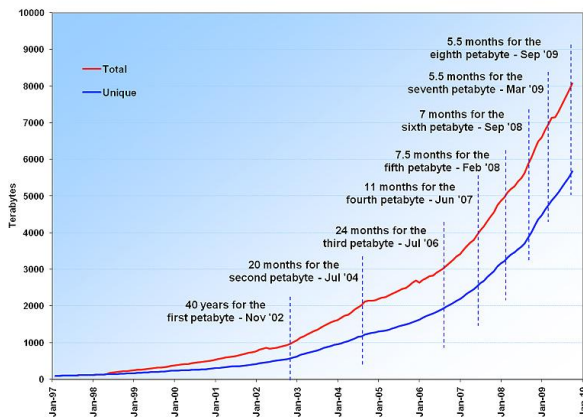
PI: Brian Schlining

Lead Scientist: Steve Haddock

Problem



NCAR MSS - Total Data in Archive



MBARI is continually collecting larger data volumes

If our data volume follows trends in other areas of computing, it will outstrip our ability to analyze it efficiently on a single compute node.

Solution - Parallelization

Break a problem down into tasks that can be executed semi-independently of each other



1 computer => 1 or few tasks at a time



many computers => many simultaneous tasks

Cost-effective Solution? - The Cloud



Also known as: “Someone else’s computers”

Cost-effective Solution? - The Cloud



“Rent” time as-needed on external hosted cloud services.

Google

 vmware

ORACLE®

amazon.com®

 Microsoft


rackspace
the open cloud company

Scale applications up and out on-demand.



Case Study 1: Deep-sea Guide

Create 478,352 data products for the Deep-sea Guide.

Runs on a virtual machine on IS's infrastructure.

- 9 GB RAM
- Larger than normal share of CPU power.

Case Study 1: Deep-sea Guide

- The analysis runs for 24 hours on 4 cores
- Constrained by:
 - RAM limit on shared VM appliance.
 - Unable to run more complicated analysis.
 - Graphics generation is single threaded. Graphics output runs for hours after the numerical analysis has been completed
 - Only 4 CPU cores; limits ability to run parallel computations.

Case Study 1: Deep-sea Guide

- Wants
 - Scale up RAM as needed
 - Scale out CPUs as needed

Remove limits on computing resources

Other Data Sets:

- AVED
 - LISST-Holo
 - Acoustic Data
 - Seismic Data (including real-time stream analysis)
 - Genomic analysis
- } Or any large image or video data set

What is Proposed ...

- **Evaluate the state-of-the-art** in relevant areas of cloud computing.

- Distributed Analysis Frameworks



- Infrastructure support tools



What is Proposed ...

- **Cost analysis** of feasible architectures deployed on different vendors clouds.
- **Develop a prototype application.**
- Create a best-practices and/or how-to guide for other engineers to **provide guidance in developing and deploying applications.**
- **Create in-house expertise.**