

NOAA's Big Data Project: Vision and Approach

Andy Bailey, BDP Technical Lead

NOAA Office of the Chief Information Officer



What is NOAA?

- United States Government Agency
- NOAA ~ National Oceanic and Atmospheric Administration
- Six line offices:
 - Marine and Aviation Operations (OMAO)
 - NOAA Marine Fisheries Service (NMFS)
 - National Ocean Service (NOS)
 - Office of Oceanic and Atmospheric Research (OAR)
 - National Weather Service (NWS)
 - National Environmental Satellite, Data and Information Services (NESDIS)

What does NOAA do?



OMAO



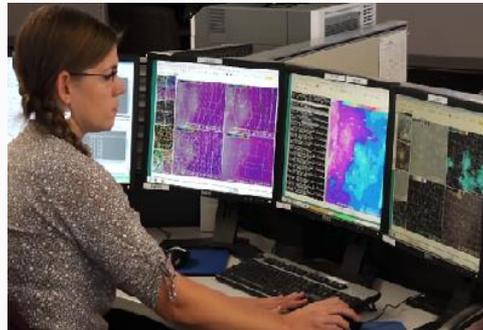
NMFS



NOS



OAR



NWS



NESDIS

NOAA
Data Expertise

CRADA Collaborators
Infrastructure Expertise



End User
Wider Consumer Community

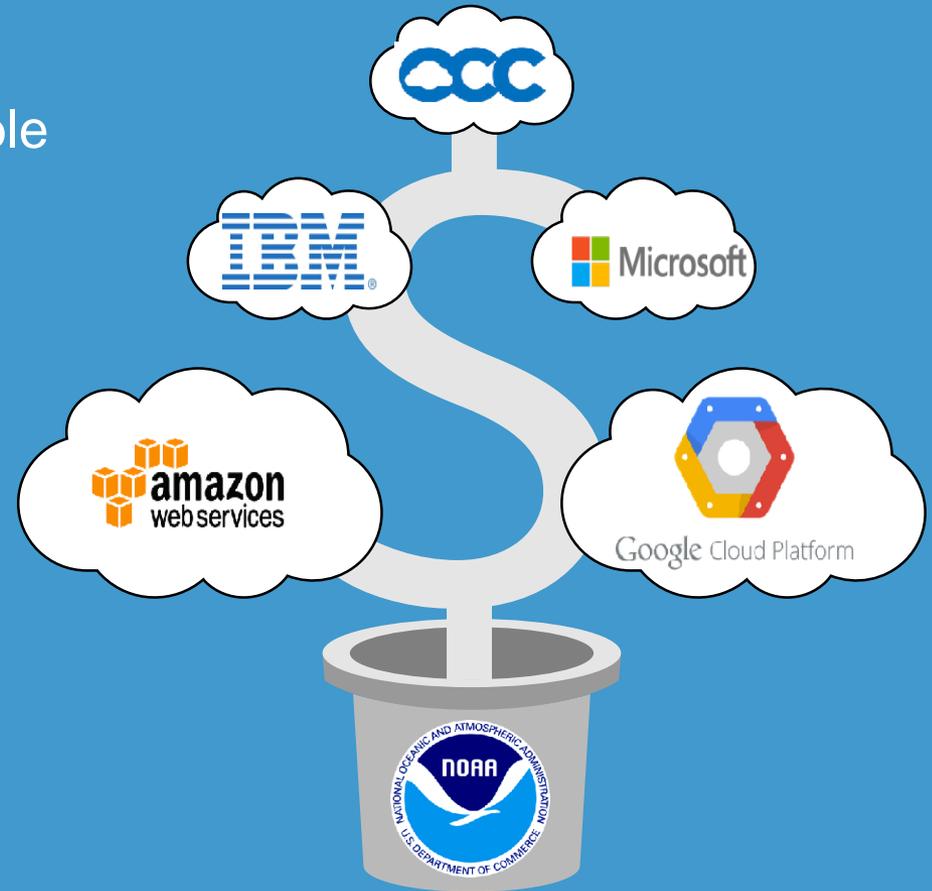
Third Party Partner
Value-Added Services

The Big Data Project

“Portals versus Platforms”

Keys

- NOAA’s open data - freely available
- NOAA’s subject matter expertise
- Industry’s infrastructure expertise
- Level playing field



Leverage the value of NOAA’s data to increase their utilization

Collaborative Research
and Development
Agreement

CRADA
Collaborators
Responded to
RFI

Value added
products
charged for

Collaborators can
generate revenue when
3rd parties process or
store the data. All can
charge for value added
products.

Fair and Level

NOAA must offer
equal access to
the data for all
collaborators.

BDP
Specifics

No Net Cost to
Tax Payers

As part of the
CRADA, NOAA
may recover
costs to get data
to collaborators

Data remains
free and open

Original data can be
downloaded for free.
Collaborators can recover costs
associated with data acquisition.

Augmentation
not
Replacement

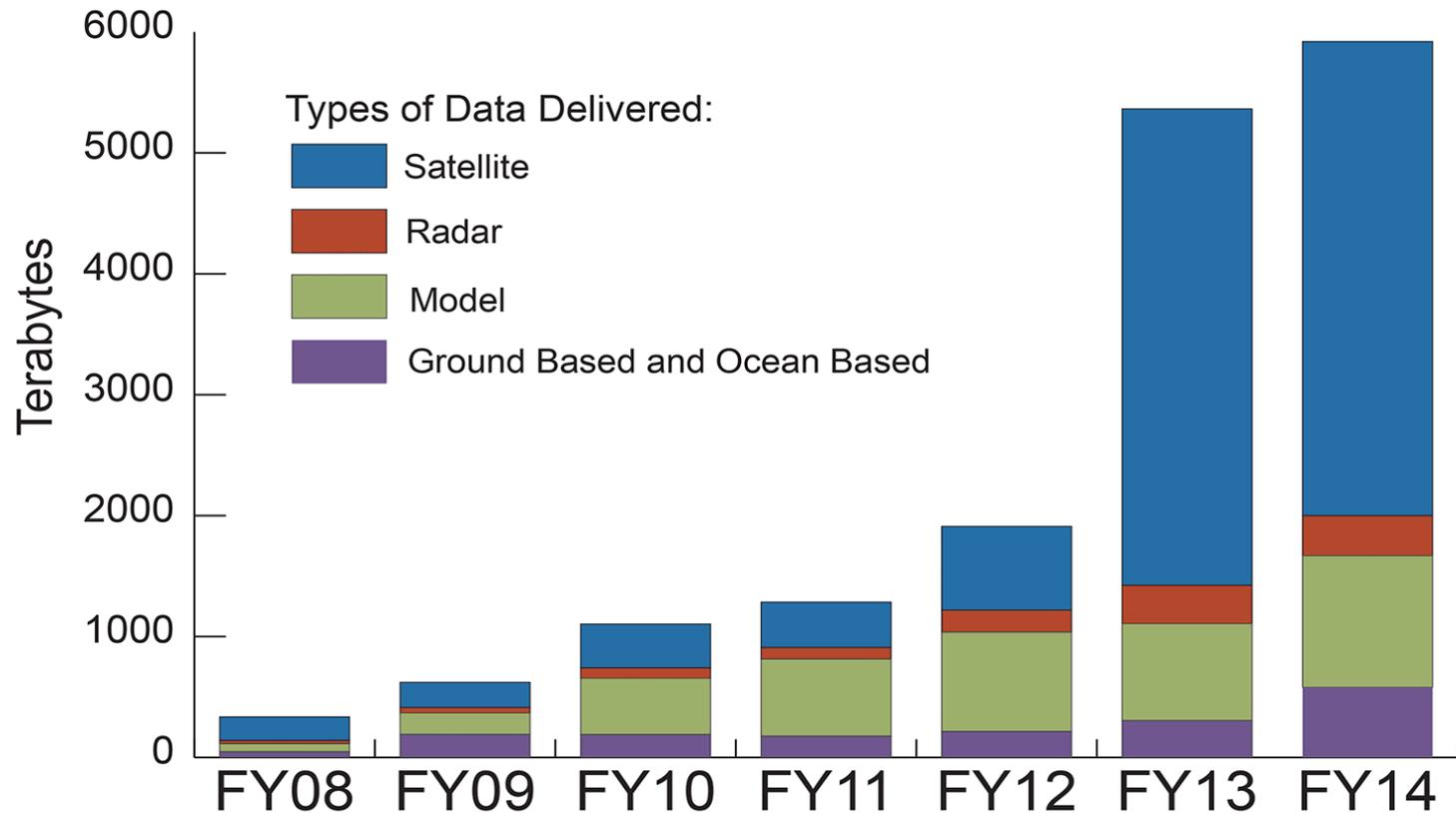
All existing NOAA service
outlets remain. BDP offers
alternatives and
advantages

Why is NOAA interested in this?

- NOAA's data are increasingly popular and valuable.
- Under the current scheme, NOAA struggles to keep up with public demand – budgets for capacity and security aren't keeping pace with data access costs.
- NOAA wants to learn about solutions while we also promote use, democratize data access, facilitate research and enable new economic opportunities for partners.

Why is NOAA interested in this?

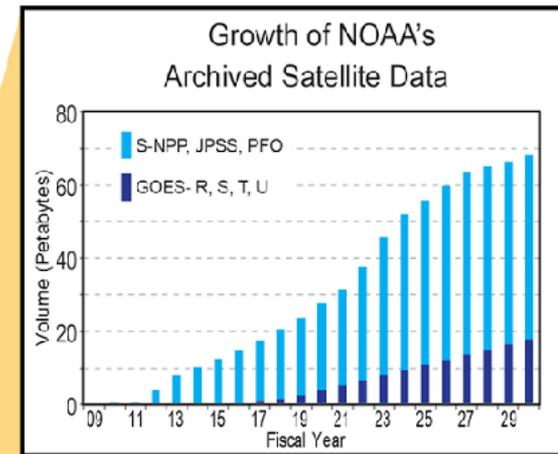
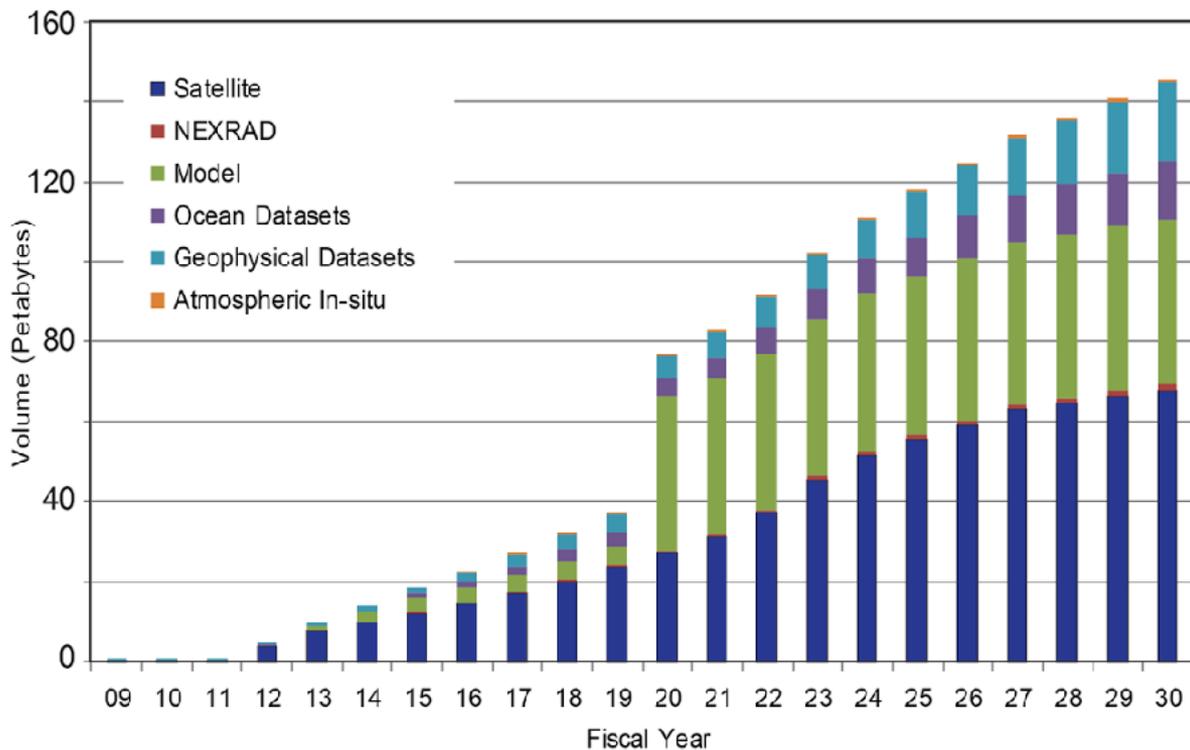
NOAA Server Load



Why is NOAA interested in this?

Archive Projections for NOAA data

Growth of NOAA's Archive



Big Data Project Methodology

01

Business Discovery

CRADA Collaborators & any Third-Party Partners work together to identify datasets of interest & develop business cases

02

Initial Technical Discussion

Develop a strategy for data delivery from NOAA to BDP Collaborators

03

In-Depth Data Discussions

Engage NOAA SMEs, BDP Collaborators for technical interchanges

04

Product Development

Collaborators and their Partners create services

- ◆ Develop markets & financial opportunities based on NOAA data
- ◆ Generate revenue and profits

05

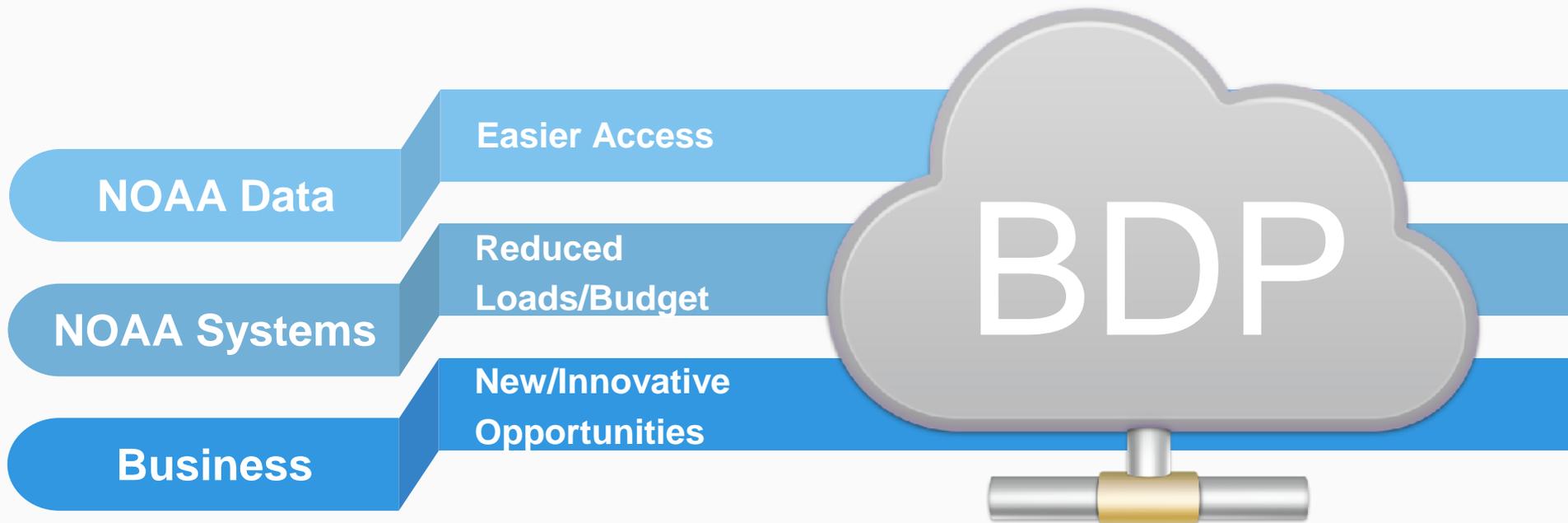
Augmented NOAA Services

NOAA continues all of it's existing data services

- No interruption of existing services to customers, but new options
- BDP activities are an augmentation of existing services



Tangible BDP Benefits



Example BDP Success Story

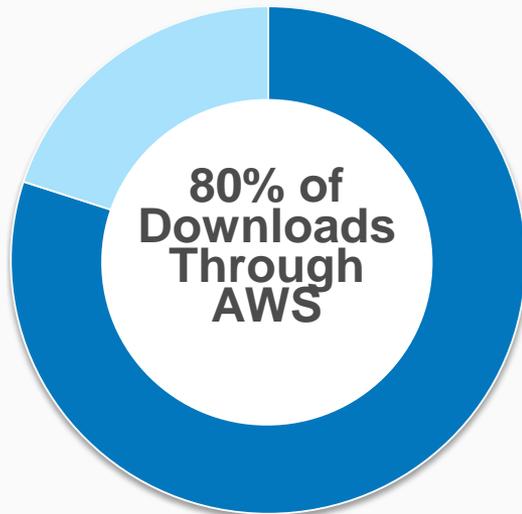
WSR88D level 2 radar data

- **Entire** 88D Archive transferred to AWS and OCC 2015
(as well as two others who haven't made their services public)
- **Options: NOAA** Redirects to BDP Collaborators' services
- Single access point for **archived** and **real-time data**
- 3rd parties - **Climate Corp and Unidata**- were key to success

Example BDP Success Story

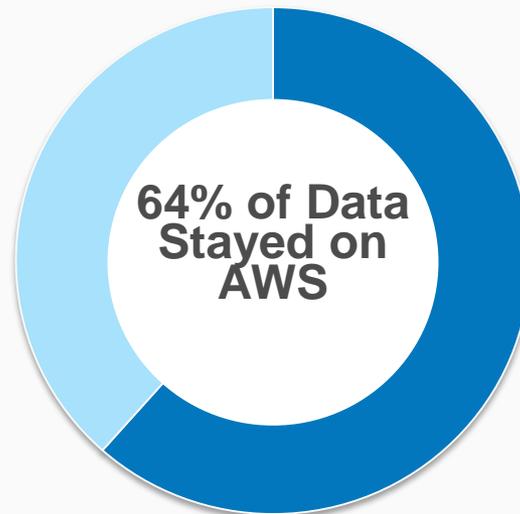
WSR88D level 2 radar data Win - Win - Win

NOAA Wins



■ AWS ■ NCEI

AWS Wins



■ AWS ■ Downloaded

End User Wins



■ AWS job time ~days
■ Through NCEI ~Years

Example BDP Success Story

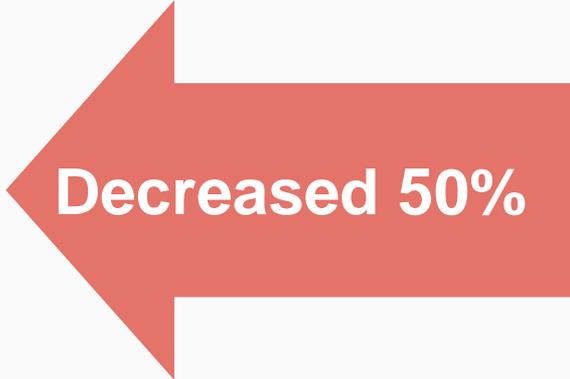
WSR88D level 2 radar data

Data Usage

Increased 2.3X

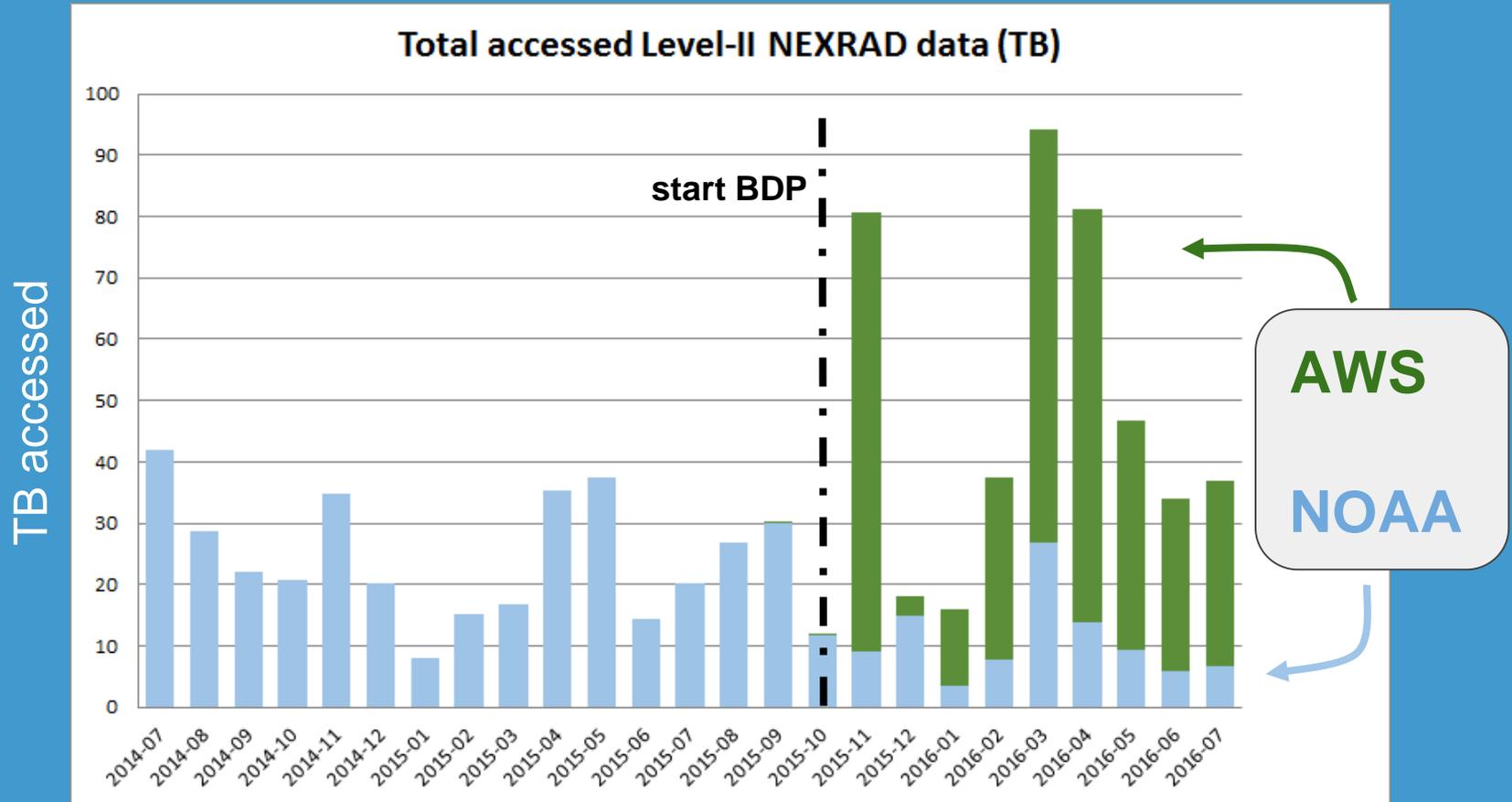


Decreased 50%



NCEI Server Load

NEXRAD Weather Radar Data



AWS: Oct '15 <https://s3.amazonaws.com/noaa-nexrad-level2> (1991+)

OCC: Jun '16 <http://occ-data.org/NOAANEXRAD/> (2015+) (S. Ansari et al, 2016)

Challenges

- Chicken and egg conundrum
 - Importance of 3rd party
- How to transfer massive datasets in real time
 - e.g. GOES16
- What happens when CRADA expires
 - Reluctance to play as April 2018 nears?
- Overcoming internal NOAA angst
 - Dissemination workers, 'That's my job'
 - Researchers, 'The data isn't ready yet'

Questions to Ponder/Discuss

Could NOAA go all in with public dissemination via the cloud?

Can NOAA use the same (free) public stores of data to also accomplish "mission" stuff like processing and science?

Is this an opportunity for NOAA to use new analytics, database, visualization and AI tools on the cloud?

Will the cloud change the time-scale of data-intensive research for everyone?

Is the cloud a good fit for big community projects like CMIP6?

Discussion

andy.bailey@noaa.gov

<http://www.noaa.gov/big-data-project>