



Archive Management of NASA Earth Observation Data to Support Cloud Analysis

Christopher Lynnes, Kathleen Baynes, Mark McInerney
NASA Goddard Space Flight Center



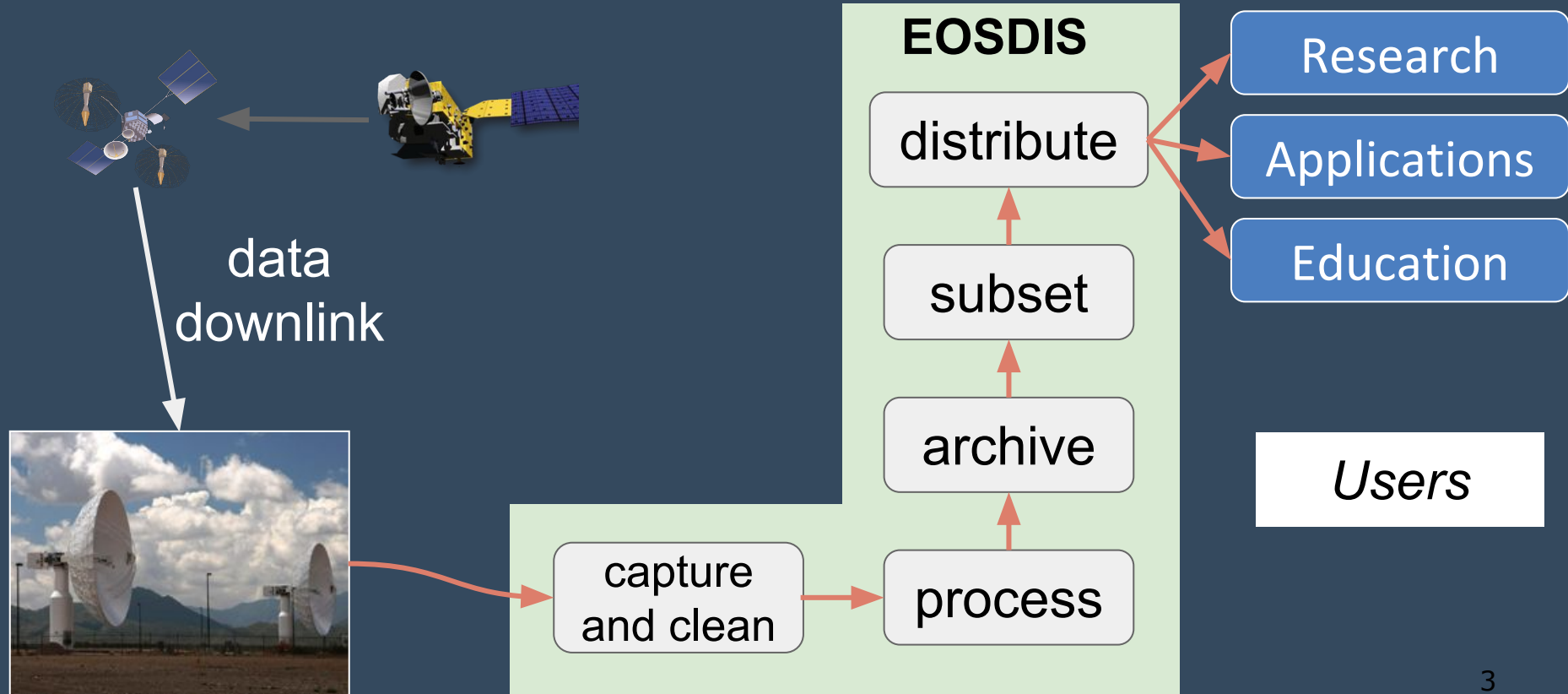
NASA flies a variety of Earth Observing satellites

- Hydrosphere
 - Physical Oceanography
 - Hydrology
- Atmosphere
 - Composition
 - Dynamics
- Biosphere
 - Land
 - Oceanic
- Cryosphere
 - Ice
 - Snow
- Lithosphere
- Anthroposphere



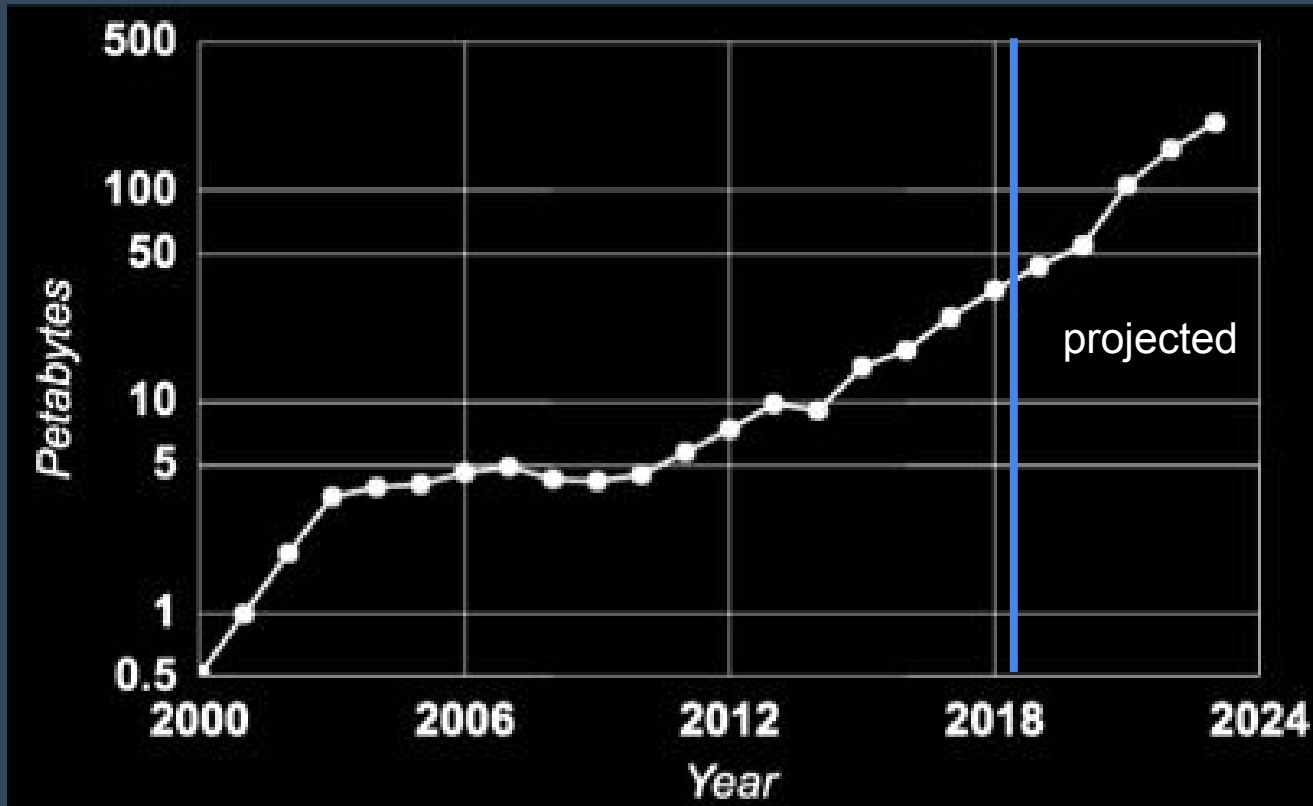


Earth Observing System Data and Information System (EOSDIS)



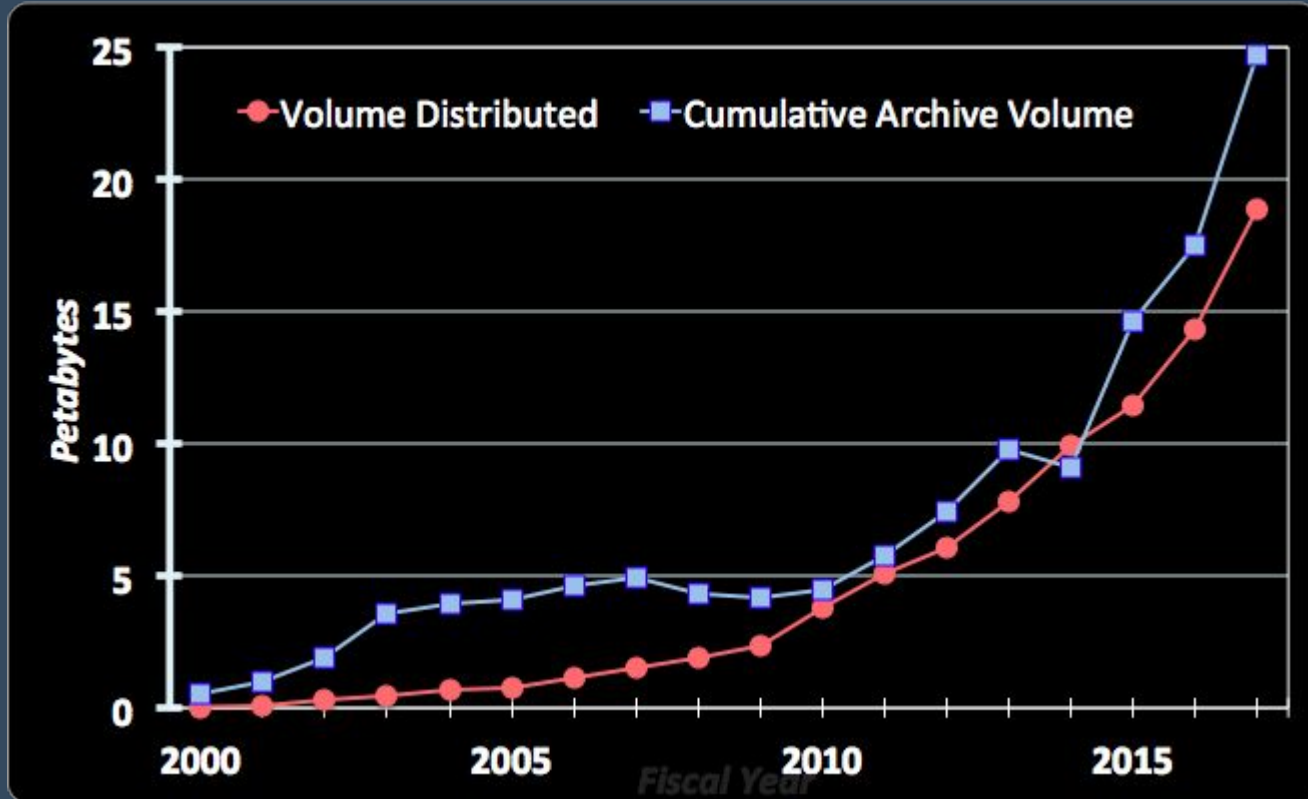


Over time, EOSDIS archive volumes increase exponentially





Distribution increases similarly to cumulative volume





Three things that will become more difficult and tedious:

1. Provisioning big data storage and supporting utilities
 - floorspace, cooling, power, network
2. Moving big data to local storage and managing them
3. Analyzing (really) big datasets on local compute

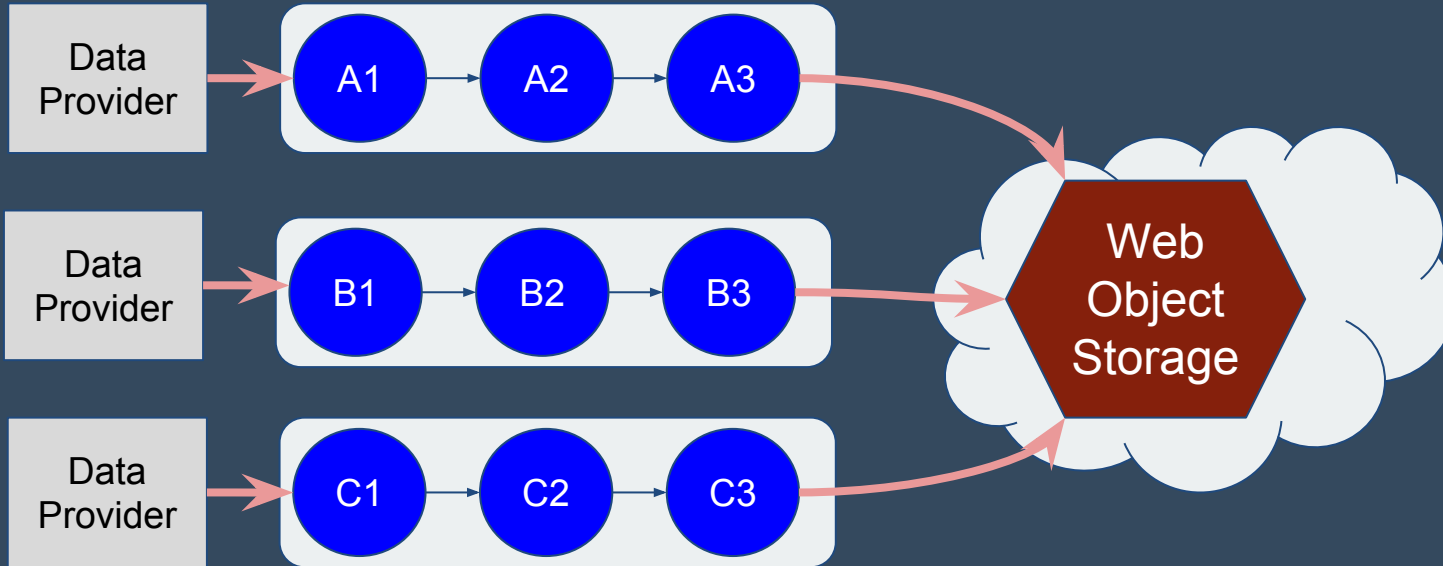


Cloud Computing Benefits

1. Infinite data storage available
2. Infinite computing power available
3. Computing power “next to” data storage
4. Rapidly evolving ecosystem of high-performance analytics tools



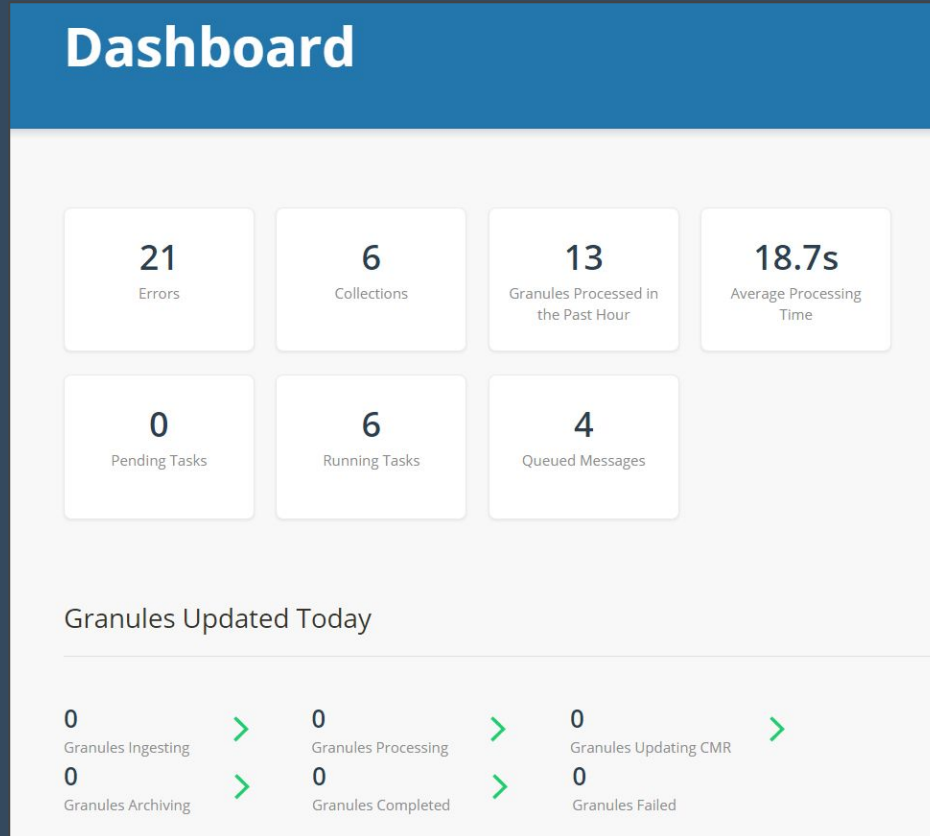
EOSDIS “Cumulus” Science Data Archive



Orchestrated by
Step Function service

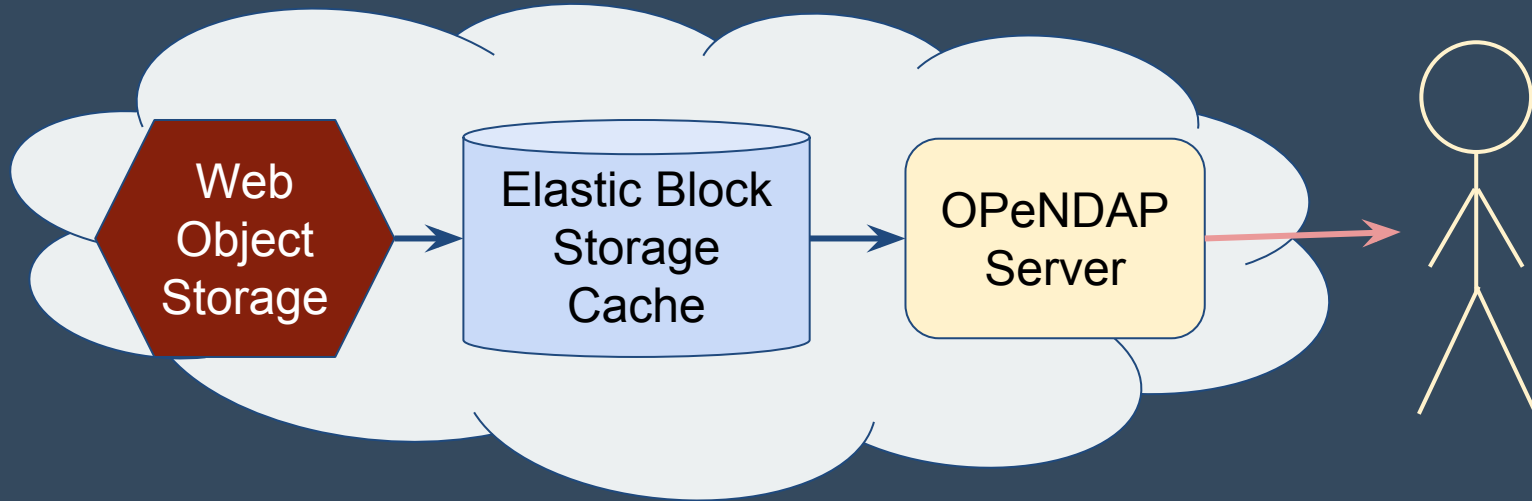


Cumulus Dashboard



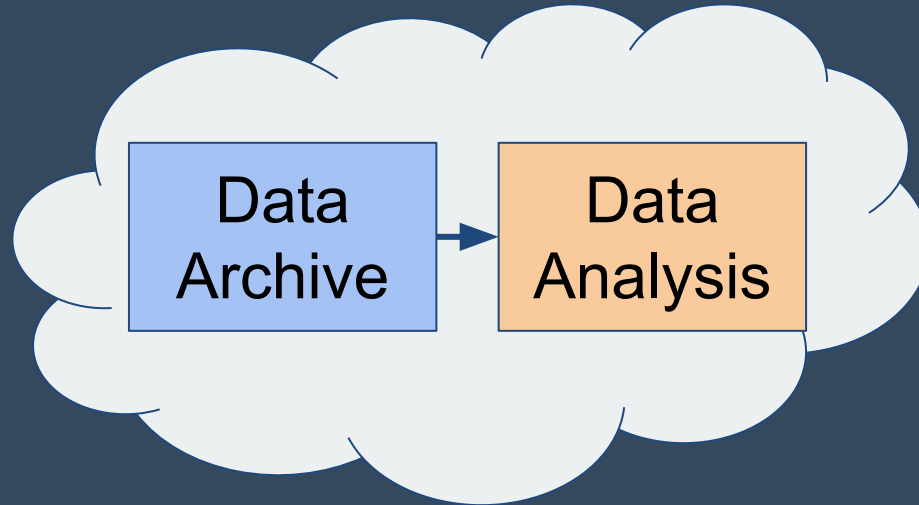


Services for Data In Web Object Storage



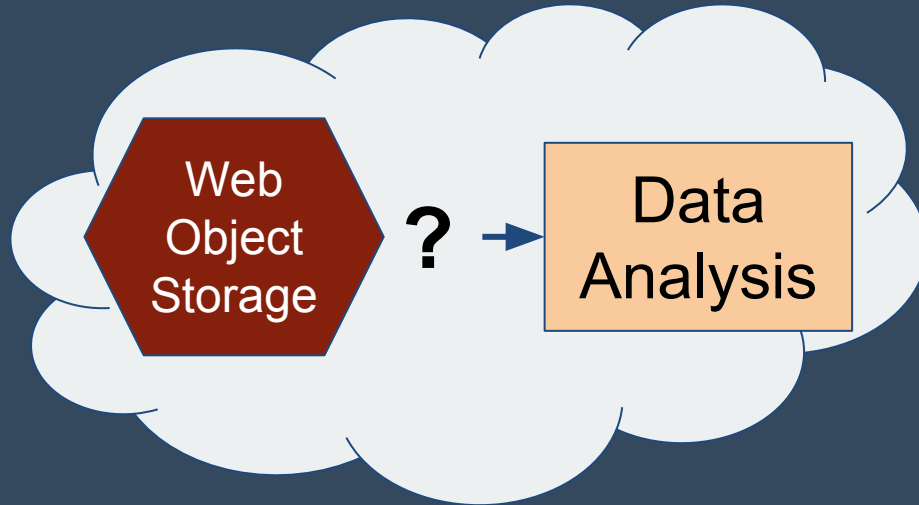


The Big Win: Data-proximal Analysis



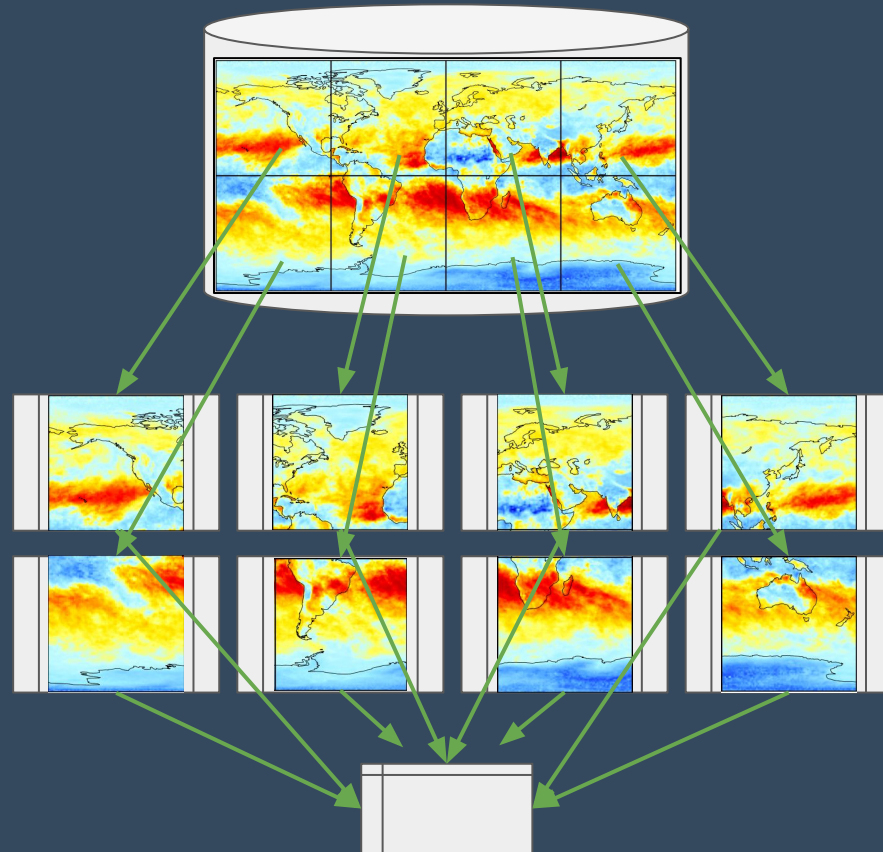


Catch #1: Web Object Storage



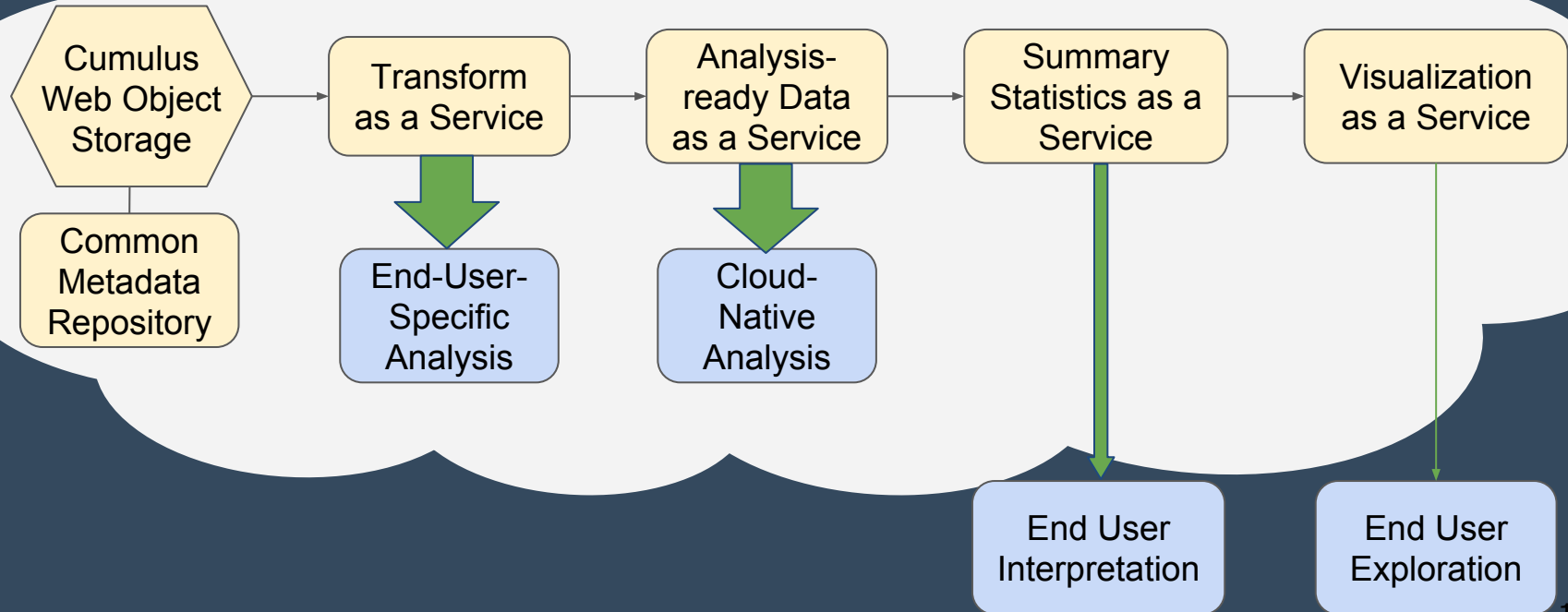


Catch #2: Data Processing Paradigm



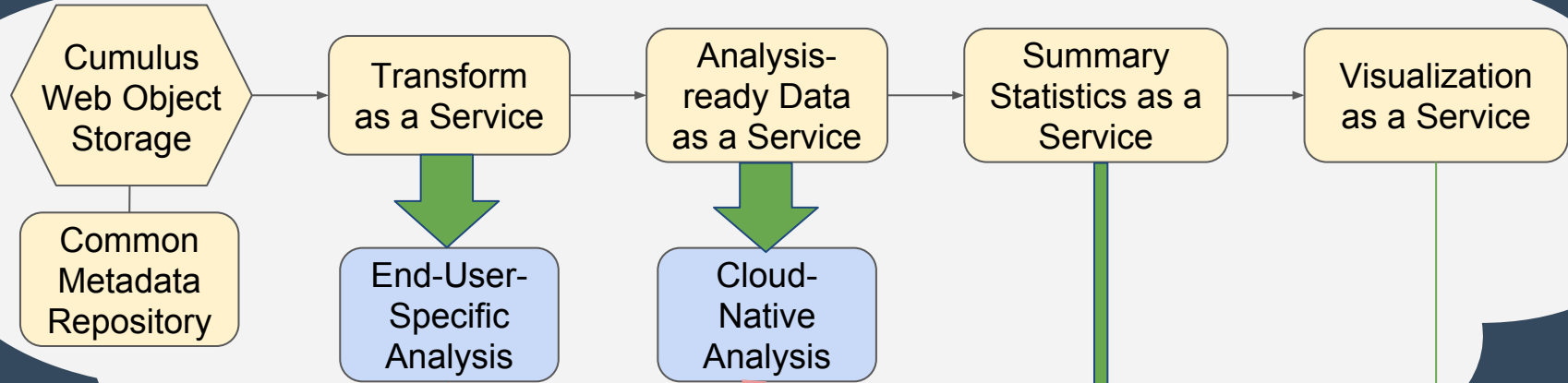


A Cloud-Native Generalized Analytics Platform





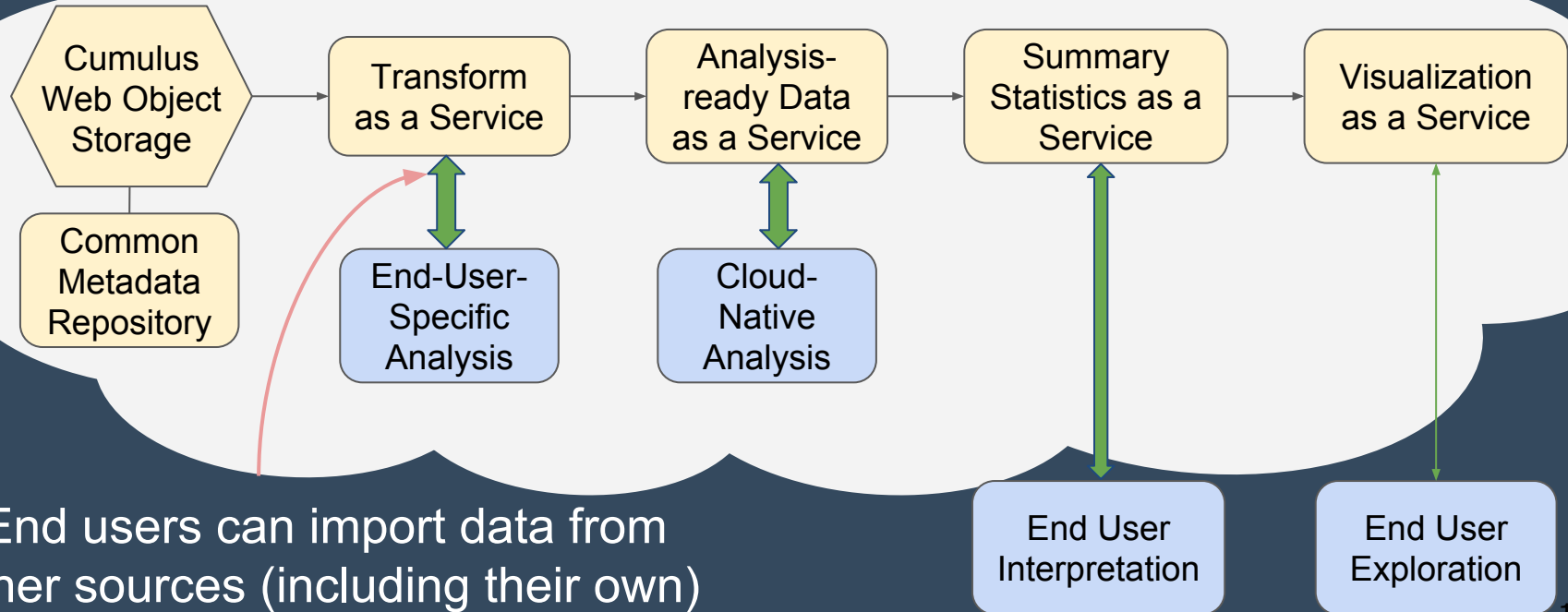
A Cloud-Native Generalized Analytics Platform



Different end users get data from different parts in the value chain based on specific needs and capabilities



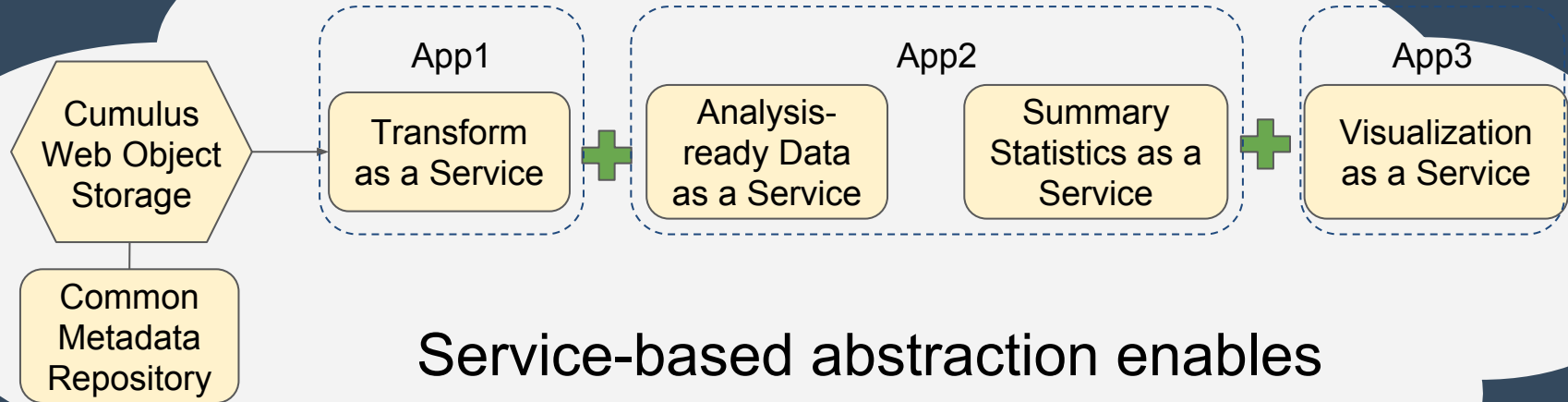
A Cloud-Native Generalized Analytics Platform



End users can import data from other sources (including their own)



A Cloud-Native Generalized Analytics Platform



Service-based abstraction enables mix-and-match reuse vs. monolithic apps