Exploitation of Multiple Data Sources to Meet the Future Needs of Extreme Weather Prediction

Dennis Hansen
Dr. Allan Weiner

Current System Deployment Trends

- Data sources for timely forecasting of extreme events are becoming more disaggregated (i.e., sensors distributed among multiple satellites)
  - Disaggregation of satellite capabilities will mean that observations from different sensor types will potentially not be temporally aligned
  - Systems will have to be designed to take into account:
    - Data being received from multiple sources over multiple communications channels
    - Potential incorporation of data from commercial satellites
    - Temporal alignment of the different sources
    - Making use of new science to account for potential loss of one or more data types
      - Potentially through real-time fusion of data to produce similar or equivalent products
    - Implementation of architectures that allow for rapid data fusion to create forecasts and other meteorological products
- Effects of sequestration and other budget pressures will intensify the trend of disaggregation strategies and potential commercialization while at the same time restricting resources necessary to upgrade processing systems to deal with the changes

Science Advancements in GOES-R ERA

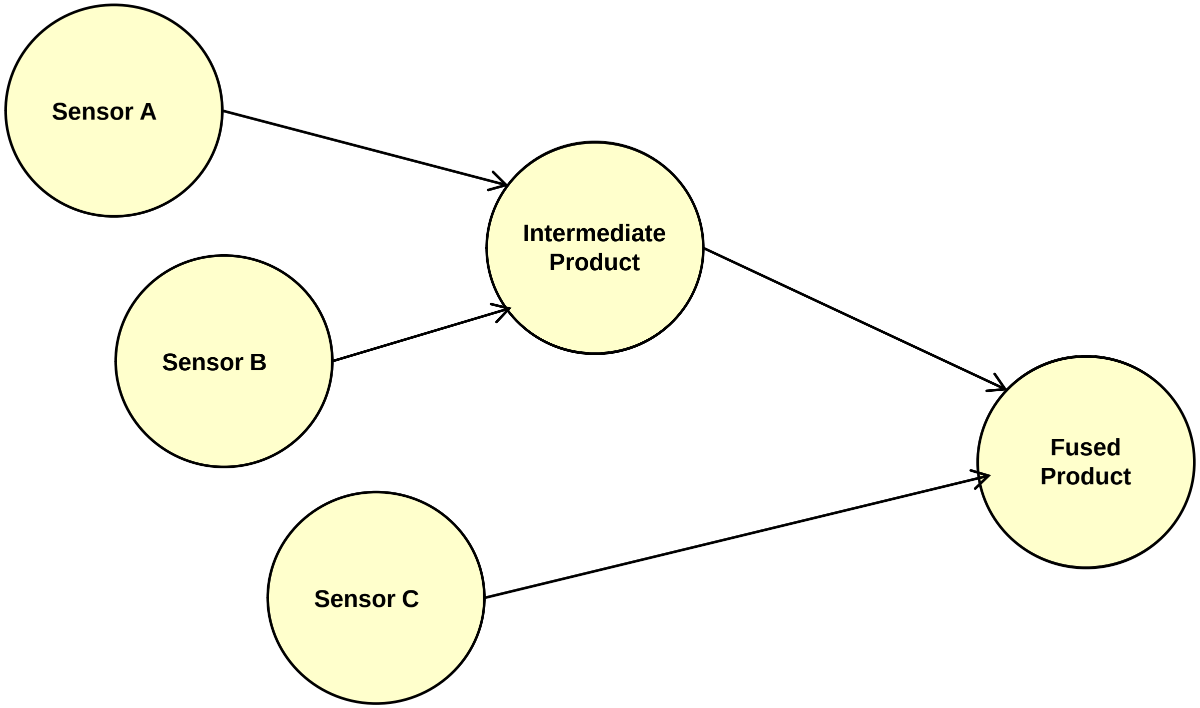
- Use of multiple data types to create new output products is a current reality
- Potential Examples include:
  - GOES-R Lightning Mapper (GLM) data fused with Advanced Baseline Imager (ABI) imagery and radar data may lead to longer Severe Weather Warning lead times.
  - High temporal resolution of the GOES-R ABI sounder combined with the high spectral resolution of the JPSS sounder and NWS forecasts may lead to refinements of the Severe Weather Watch box size.
- The primary advantage of current product generation is that the data is all retrieved from the same satellite within the same downlink data stream
  - Allows generation of intermediate products which are used to produce yet another end product
- Lessons learned from this form of data fusion will benefit future architectures aimed at processing data from disaggregated sources

Simple Example of Fused Product Flow

Sensor A
1 sample per second

Sensor B
1 sample per minute

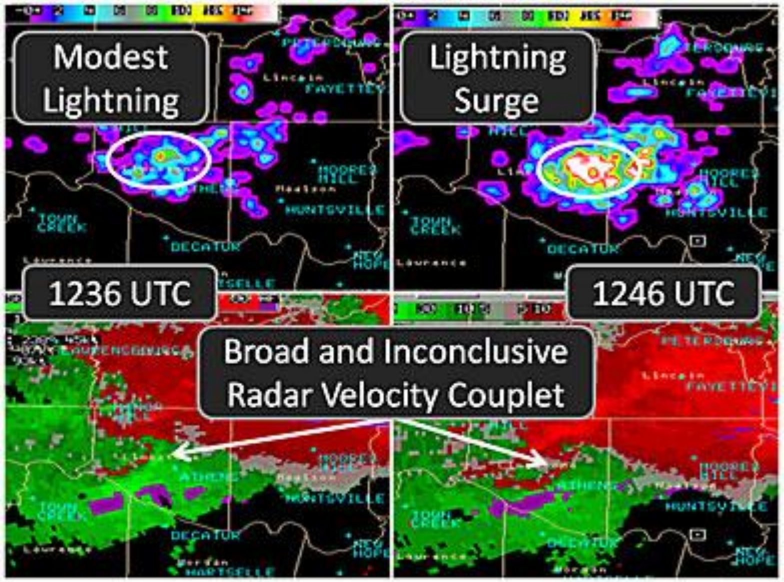
Sensor C
1 sample per 30 minutes



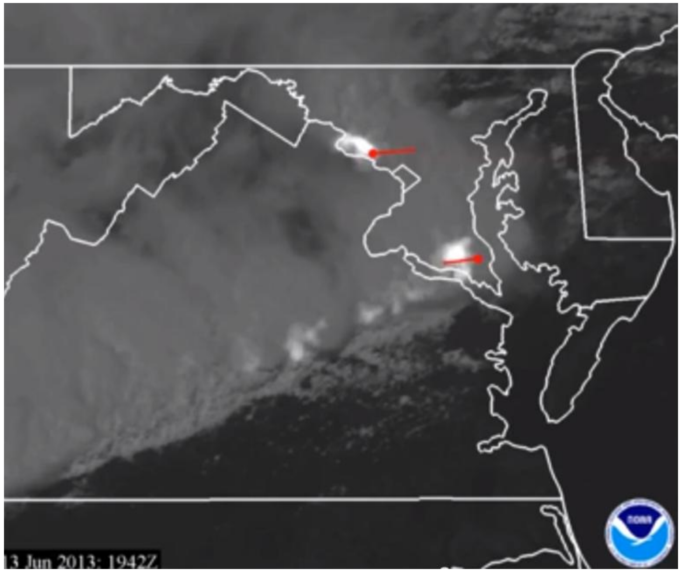
- Sensor A processing becomes the Base for process timing
- Intermediate product processing begins upon notification of the arrival of Sensor A data
- The most current Sensor B data is used in the data fusion
- Fusion of the final product started on notification of intermediate product availability and then processes using the most current value of Sensor C

- NOTE: Sensors A, B, & C may on one satellite or disaggregated among multiple satellites or even ground based observations

Example of Merged Product



http://www.goes-r.gov/spacesegment/glm-lightning-detect.html
Image Courtesy of Geoffrey Stano and SPoRT
Total lightning (Upper) from the North Alabama LMA coincident with NEXRAD radar-derived storm relative velocity (Lower) at 1236 (Left) and 1246 (Right) UTC on 6 May 2003. The lightning surge of over 200% occurs 14 minutes prior to a confirmed tornado touchdown.



http://www.youtube.com/watch?v=9vVh5V2h_sg
Lightning trail during DC Maryland tonado outbreak of 13 Jun 2013

- GOES-R GLM data (similar to the top of the left insert) can be combined with radar and visual imagery to aid forecasters in the detection and prediction of severe weather events

High Level View of Approach

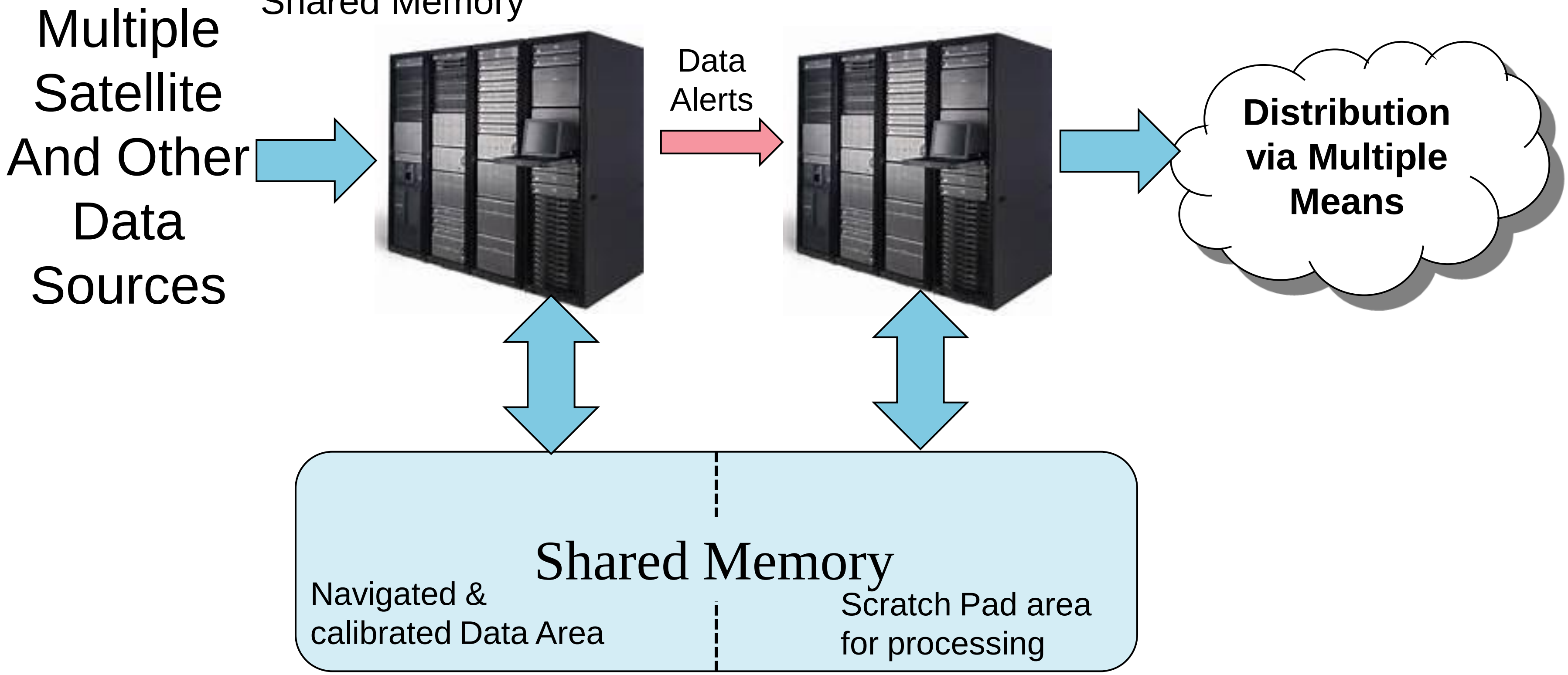
HARRIS

Data Ingest

- Protocol processing
- Identification of data
- Navigation and Calibration
- Organization of data in Shared Memory

Data Fusion

- Selection of data with closest temporal fit
- Preliminary processing of each data type
- Data fusion algorithm execution
- Format for distribution to product users



The Challenge/Solution

- Fused data production in real-time presents memory management and process threading challenges
  - Locating the data to minimize access time relative to the processes
  - Minimizing data movement within memory to avoid process delays
  - Notification methods to trigger processing of the next step in the process flow
- A services based approach can meet the needs of a process flow of independent and interdependent steps
  - Allows for independent action of each processing step
  - Can rapidly respond to availability of new data or completion of a previous processing step
- The means to establish the processing precedence must be made simple to allow the scientist to concentrate on the science of the process rather than programming the interaction

Services Management Based on GOES-R

HARRIS

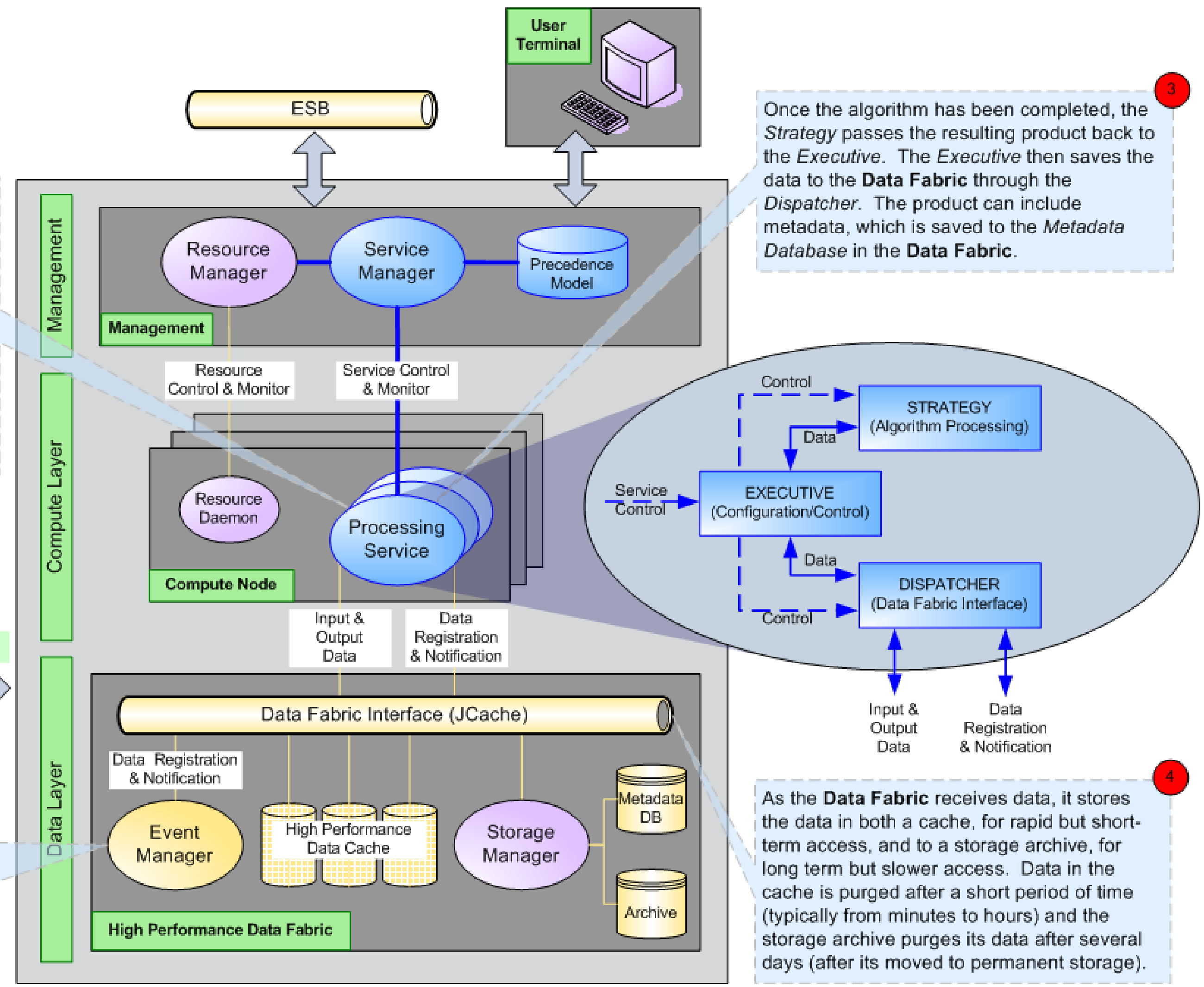
SBA
Custom
COTS

2 The **Data Fabric** (through the **Event Manager**) notifies the services requiring for L0 Data that it is now available. The *Executive* receives this notification through the *Dispatcher* interface. The *Executive* determines that there is enough data to begin processing and retrieves the data from the **Data Fabric** through the *Dispatcher* interface. The *Executive* then executes the service's algorithm with that data through the *Strategy*.

1 MM delivers L1B data into the Data Fabric.

MM
L1B Data

5 The **Event Manager** notifies all interested services that new data is available. When the service determines it has enough data to begin processing, the new data is downloaded and the previous three steps are repeated. Since new data is now written into the **Data Fabric**, this step is repeated as well until data entered into the **Data Fabric** is no longer required by any service.



3 Once the algorithm has been completed, the *Strategy* passes the resulting product back to the *Executive*. The *Executive* then saves the data to the **Data Fabric** through the *Dispatcher*. The product can include metadata, which is saved to the *Metadata Database* in the **Data Fabric**.

4 As the **Data Fabric** receives data, it stores the data in both a cache, for rapid but short-term access, and to a storage archive, for long term but slower access. Data in the cache is purged after a short period of time (typically from minutes to hours) and the storage archive purges its data after several days (after its moved to permanent storage).

Summary

- Disaggregation of sensor resources will force a change in how we meet the needs of forecasters to provide timely extreme weather warnings
  - New algorithms based upon data fusion in addition to new sensors
- The concept of data fusion is not new and when implemented has resulted in new and useful products
- In the future, data fusion may be needed to fill gaps caused by sensor failures or the lack of desired sensors
- Data fusion in real-time presents processing architecture challenges
- Systems such as GOES-R have shown that service based architecture concepts can be successfully employed to meet product latency needs
- Concepts from GOES-R can reduce operational complexity