An Innovative Infrastructure with a Universal Geo-spatiotemporal Data Representation
Supporting Cost-effective Integration of Diverse Earth Science Data

Michael Lee Rilee\textsuperscript{1,3} and Kwo-Sen Kuo\textsuperscript{1,3}
\textsuperscript{1}Rilee Systems Technologies LLC; \textsuperscript{2}Bayesics LLC; \textsuperscript{3}NASA Goddard Space Flight Center

Abstract

\begin{itemize}
  \item The SpatioTemporal Adaptive Resolution Encoding (STARE) is a unifying schema encoding geospatial and temporal information for organizing data on scalable computing/storage resources, minimizing expensive data transfers.
  \item STARE provides a compact representation that turns set-logic functions into integer operations, e.g. conditional subsetting, taking into account represented spatiotemporal resolutions of the data in the datasets.
  \item STARE geo-spatiotemporally aligns data placements of diverse data on massive parallel resources to maximize performance.
  \item Automating important scientific functions (e.g. regridding) and computational functions (e.g. data placement) allows scientists to focus on domain-specific questions instead of expanding their efforts and expertise on data processing.
  \item With STARE-enabled automation, SciDB+STARE provides a database interface, reducing costly data preparation, increasing the volume and variety of interoperable data, and easing result sharing.
  \item Using SciDB+STARE as part of an integrated analysis infrastructure dramatically eases combining diachronically different datasets.
\end{itemize}

Why SciDB?

Resource Consumption Advantages
\begin{itemize}
  \item Minimize download and local data management
  \item Free end-user resources for research and science
\end{itemize}

Performance Advantages
\begin{itemize}
  \item Array data model is better suited for scientific data than relational databases.
  \item Tightly coupled analysis and storage layers allows better optimization than Spark.
\end{itemize}

SciDB Query

To spatiotemporally "join" the two datasets with STARE indexing (5-min at level 7, i.e. "78-km resolution):  
```
join (mmg, precip, tmms_2B31) by (MMG).join (mmg, precip, tmms_2B31) by (MMG)
```
To subset temporally for visualization:
```
select * from mmg_mmms_09120303
where tempIndex x 11 = 0 (8 time periods)
```

Example of STARE to represent Hurricane Irma forecast (vertical = time)

This figure illustrates the use of STARE to represent Hurricane Irma forecast data. The forecast data is spatiotemporally processed, allowing for efficient data access and analysis. The visualization highlights the temporal evolution of the forecast, showing how STARE enables seamless integration and analysis of complex Earth science data.

Summary

SciDB on STARE
\begin{itemize}
  \item Provides a unifying scheme for comparing and combining diverse data sets
  \item Naturally supports data placement alignment for efficient use of SciDB
  \item Set and logic operations are efficient, straightforward to code
  \item Transparent use of high-end parallel/distributed compute & storage
  \item Scientists can work with data via high-level queries
  \item Growing set of functions for representation, regidding in SciDB enabled by STARE
\end{itemize}

Related Presentations:
STARE in Visualization: IN23F-07, IN33C-0141; in the SciDB array database: IN41B-0035 (this work), IN33E-04, and the path to enabling machine learning: IN11E-07.

Acknowledgement
Funding for this research is provided by NASA Earth Science Technology Office (ESTO) through the Advanced Information Systems Technologies (AIST) program and NASA's Advancing Collaborations for Earth System Science (ACCESS) program, for which we are greatly grateful.