IBM Research

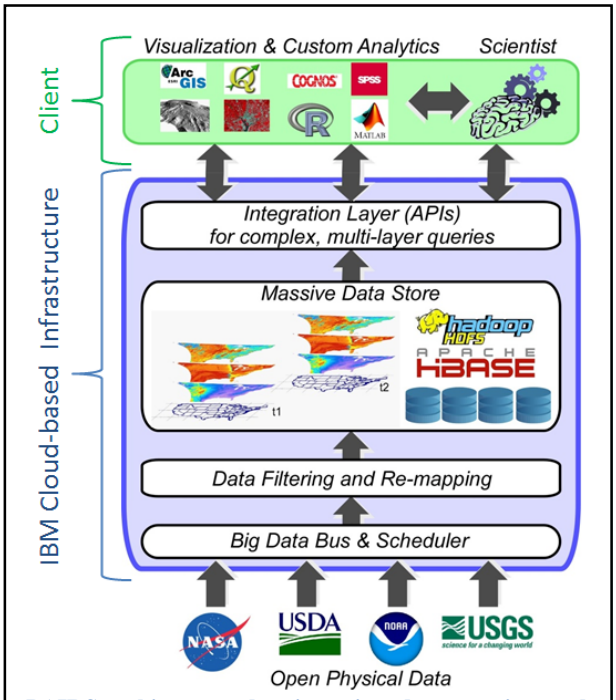
PAIRS Technology

Physical Analytics Integrated Data Repository and Services (PAIRS) is a Big Data analytics platform coupled with a massive data store of pre-processed satellite and weather data to run services such as weather forecasting and agriculture analytics. PAIRS offers an easy-to-use platform for assembling and evaluating geo-spatial Big Data sets, lowering time-to-discovery by reducing the data management burden.

The technology is based on the open source distributed data store Hbase/Hadoop to (cost) efficiently host and manage petabytes of data. PAIRS leverages efficient data layer indexing and data layer alignment, both for structured (e.g. satellite images, weather data) and unstructured (e.g. data from telco, social media, distributed sensor networks) data. As all data layers are aligned on global grids, any location on the globe can be queried in an intuitive and logical manner. The user accesses results in standard file formats such as GeoTiff, CSV, XML.

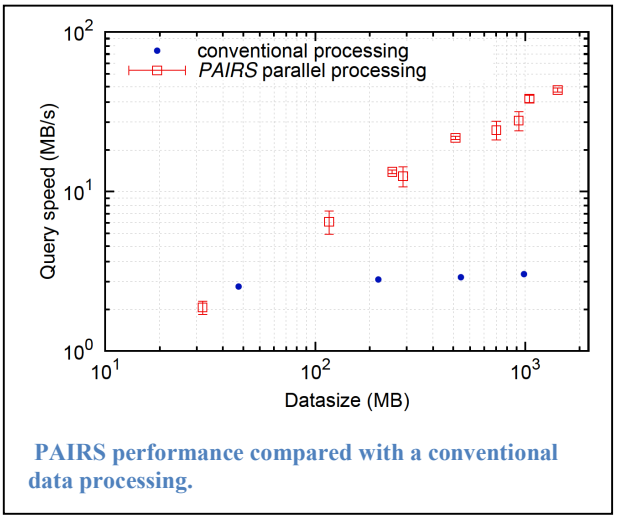
PAIRS processes and returns complex query results orders of magnitude faster than conventional data services. For example, a query such as:

"Show me all urban areas where is sunny for the next 10 days and where the population density is larger than 500 people per square mile and there are two coffee shops in 500 sq mile area"



PAIRS architecture; data ingestion, data curation, and analytics platform.

requires filtering and querying multiple layers across different spatial and temporal scales. PAIRS handles such multi-layer queries with speed and ease. Data discovery is possible across continental scale, potentially 10 to 100 times faster than conventional methods. Utilizing a distributed processing and storage, PAIRS is very efficient at querying data from the scale of single farm to the continental scale. With PAIRS, it is relatively easy to establish correlations and similarities between locations and events in space and time.



PAIRS performance compared with a conventional data processing.

PAIRS curates and updates satellite imagery, weather data, census, land use and business location data as they become available. PAIRS can integrate additional datasets including custom and proprietary data layers that will be automatically linked with existing data layers to run data discovery queries on them. Besides access to data layers, PAIRS is running three operational analytics - (1) significantly improved weather forecast data based on data blending (approx. 30% improvement compared with existing weather forecast), (2) solar radiation and wind forecasting for renewable energy production and (3) irrigation forecasts for precision agriculture.

PAIRS provides a cost effective, easy-to-use, scalable platform for managing geo-spatial data, ideally suited for solving complex business problems that require real time Big Data processing and advanced spatial-temporal analytics.