



NATIONAL RESEARCH
UNIVERSITY

Array DBMS: Problems, Technologies, and Perspectives

Ramon Antonio Rodrigues Zalipynis

<http://cv.gis.land>

Ramon Antonio Rodrigues Zalipynis

CHRONOSDB
ARRAY DBMS



ChronosDB is up to 75x
faster than SciDB

Michael
Stonebraker



[ACM Turing Award](#) (2014)

[PostgreSQL](#), [Vertica](#),
[Illustra](#), [VoltDB](#), [SciDB](#)



BITFUN

NEW TYPE OF
WORKLOADS

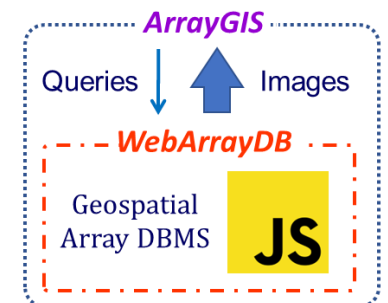


LECTURE
1,5 HOURS
VLDB 2021



NOVEL R&D
DIRECTION

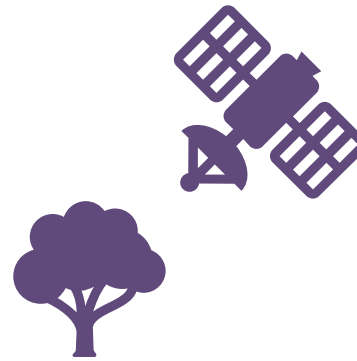
Specified in the paper



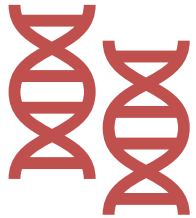
N -d arrays



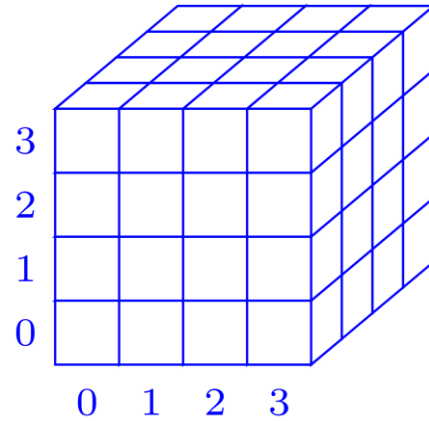
astronomy



geoinformatics



bioinformatics

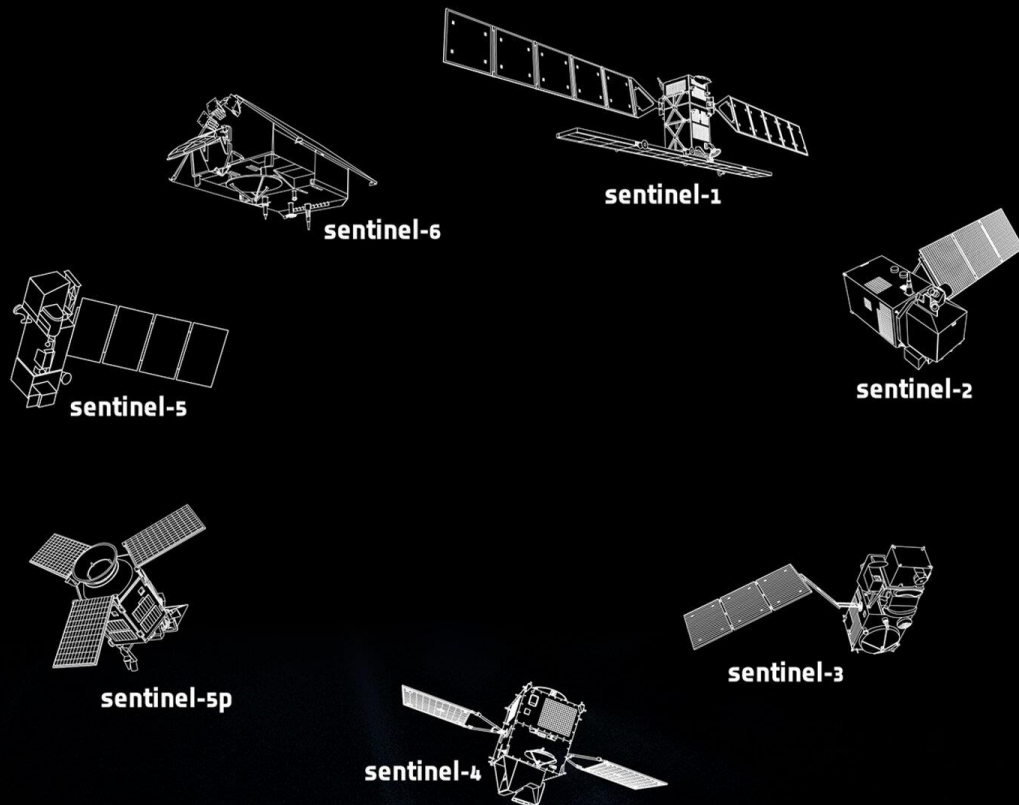


medicine



ecology

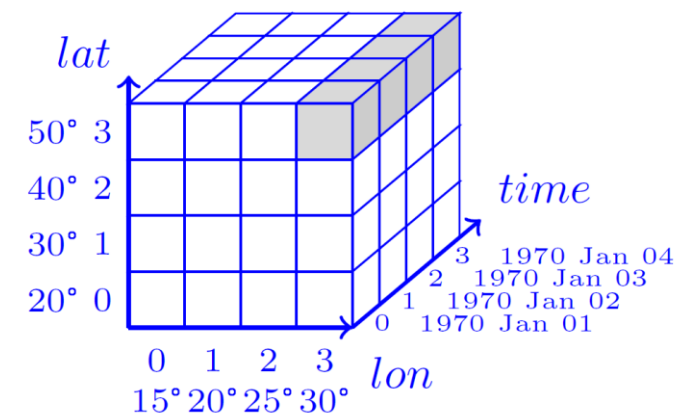
**Many core data
types are naturally
modeled by**



Large Arrays

Array DBMS
N-d arrays

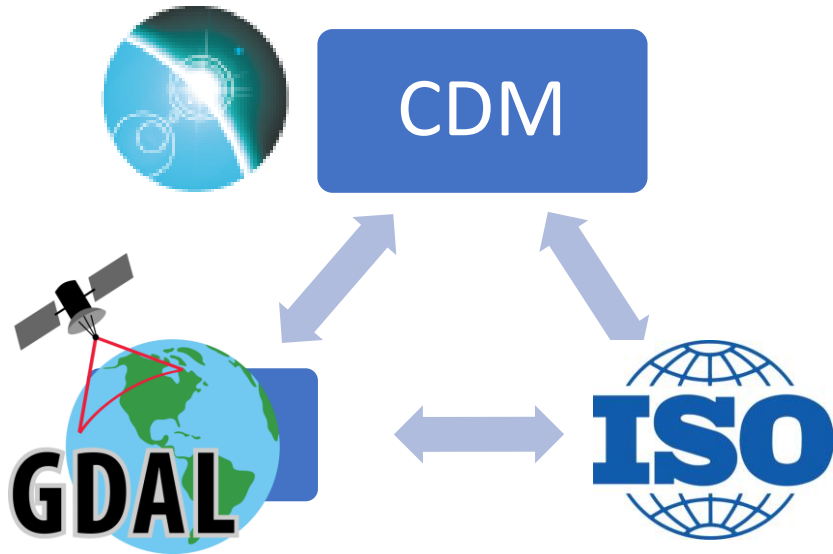
Sentinel Data
Access System
203
TiB/day



Data Models

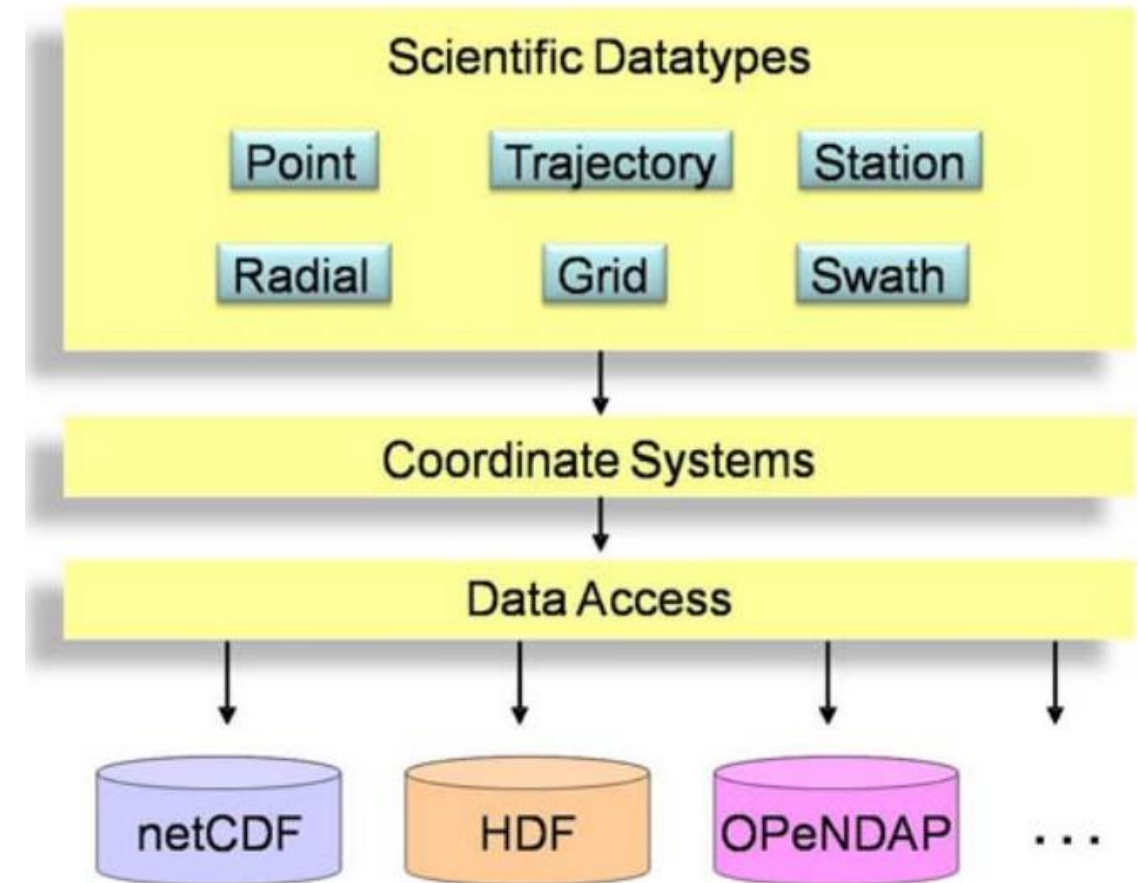
Industrial Data Models: CDM, GDAL, ISO 19123

The goal: provide a uniform access to a dataset that can be provided in different storage formats or service APIs.



decades of practical experience

Key drawback: the models work only with a single file, not with a set of files as a single array



[1] GDAL 3.1 implements the HDF model since 7 May 2020 <https://github.com/OSGeo/gdal/releases/tag/v3.1.0>

[2] S. Nativi et al. Unidata's Common Data Model mapping to the ISO 19123 Data Model. Earth Sci. Inform. 1 (2008), 59–78.

Evolution: ChronosDB Array Model

ChronosDB data model

- (1) **formalized** industrial experience
- (2) treats arrays in **multiple files**
arbitrarily **distributed** over cluster
nodes as a single array

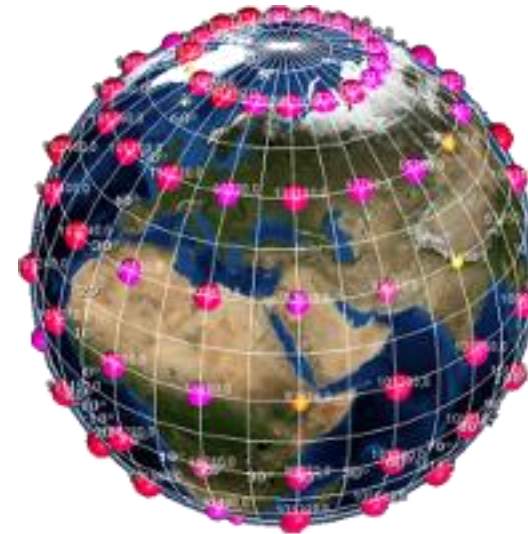
ChronosDB Data Model

mappable



ChronosDB data model

- (3) a rich set of data types
(Gaussian, irregular grids, etc.)



Check the data
collection here:

<http://wikience.org>

- (4) supports multiple file
formats, but is independent
from them

NetCDF

HDF4

Grib2

HDF5

GeoTIFF

...



Array Schema

Illustrative dataset Landsat 8, GeoTIFF

- paths 191–198,
rows 24–27
- 01–15 July 2015

gdalinfo is the name of
both a ChronosDB
command and a well-
known command line tool

```
gdalinfo Landsat8.Level_1.SurfaceReflectance.Band4
Driver: GTiff/GeoTIFF
Size is 38673, 24937
Coordinate System is:
PROJCS["WGS 84 / UTM zone 32N", <skipped> AUTHORITY["EPSG","32632"]]
Origin = (-53110.000000000000000,5878570.000000000000000)
Pixel Size = (30.000000000000000,-30.000000000000000)
Metadata:
  AREA_OR_POINT=Area
Corner Coordinates:
Upper Left  ( -53110.000, 5878570.000) ( 0d47'37.30"E, 52d46'20.14"N)
Lower Left  ( -53110.000, 5130460.000) ( 1d50'34.16"E, 46d 6'11.15"N)
Upper Right ( 1107080.000, 5878570.000) ( 17d59'48.38"E, 52d42'52.21"N)
Lower Right ( 1107080.000, 5130460.000) ( 16d50'57.11"E, 46d 3'26.67"N)
Center      ( 526985.000, 5504515.000) ( 9d22'26.95"E, 49d41'33.18"N)
Subarray=2048x2048 Block=2048x1 Type=UInt16, ColorInterp=Gray
NoData Value=0
```



<https://vldb.org/pvldb/vol11/p1247-zalipynis.pdf>

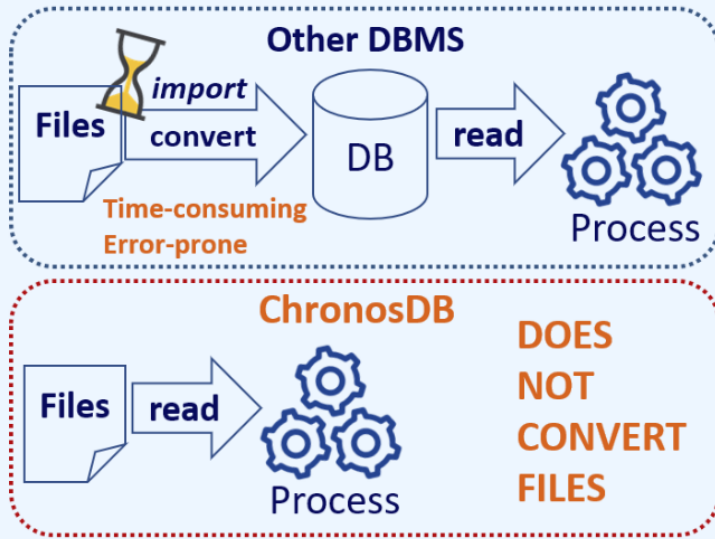
- users perceive an array as a single object
- output format & meaning = well-known **gdalinfo** tool designed to handle **one file at a time**

Users avoid learning a new
array schema notation and
inspect ChronosDB arrays in a
way they are accustomed to

<https://gdal.org/programs/gdalinfo.html>

Array Storage

ChronosDB Advantages: <http://chronosdb.gis.land/>



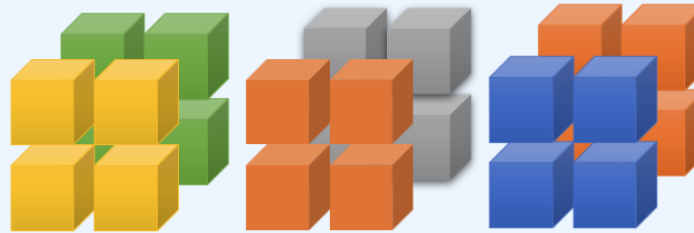
AVOID DATA IMPORT

ChronosDB directly operates on files: no time-consuming and error-prone data ingestion into an internal DBMS format.

Byte,
Int,
Float,
...

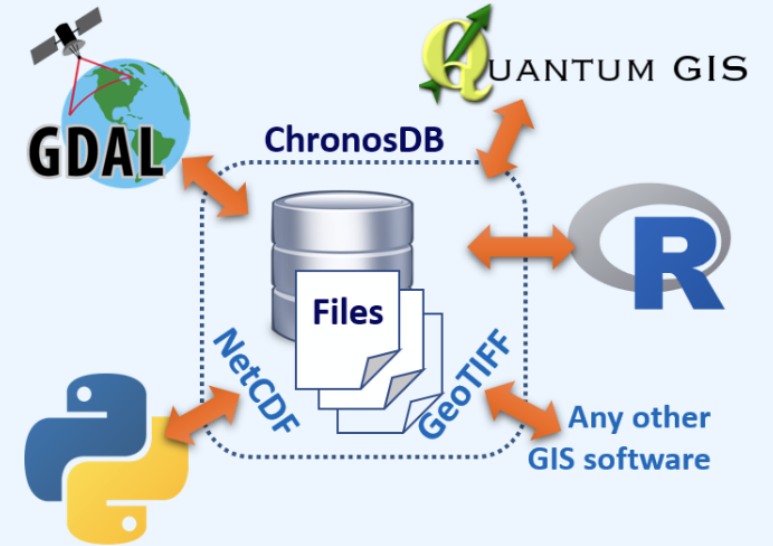
Missing
values
support

JPEG, LZW
PACKBITS
DEFLATE,
...



POWERFUL STORAGE

ChronosDB currently works with GeoTIFF & NetCDF files. They support chunking, compression, multiple data types and more.



SHARE DATA READILY

ChronosDB stores data in ordinary files in standardized formats. You can directly access the data from any other software.

Array Indexing

BitFun: Fast Answers to Queries with Tunable Functions in Geospatial Array DBMS

<http://bitfun.gis.land/vldb2020/>

Ramon Antonio Rodrigues Zalipynis
rodrigues@gis.land

Novel R&D Directions



Xian, Shaanxi, China
SIGMOD/PODS 2021

Convergence of Array DBMS and Cellular Automata: A Road Traffic Simulation Case

<http://sigmod2021.gis.gg/>



VLDB 2022

SimDB in Action: Road Traffic Simulations Completely Inside Array DBMS

<https://wikience.github.io/simdb2022/>

Web GISs and WebArrayDB



R.A. Rodrigues Zalipynis, N. Terlych. WebArrayDB: A Geospatial Array DBMS in Your Web Browser, VLDB 2022

[Paper homepage
wikience.github.io/webdb2022](http://wikience.github.io/webdb2022)

**ArrayGIS is freely available at
<http://webdb.gis.gg>**

If you would like to know more

- Lecture (tutorial) at VLDB 2021 (CORE A*)
- 1.5 hours (90 minutes)
- only 8 tutorials were accepted to VLDB 2021
- main program, main conference day
- separate section, chair: Ramon Antonio

VIDEO

<http://vldb2021.gis.gg>

Day 3: Wednesday, Aug 18								
Local Time	Copenhagen Time	Blomstersalen	Congress Hall	Ballonsalen	Columbine	Vandsalen	Pjerrot	Harlekin
10:00	9:00	Industry 2: Databases for Data warehouse and Invited Talk (3 talks)	Research Session 16: NLP for Databases (5 talks)	Research Session 17: Machine Learning for Databases (5 talks)	Research Session 18: Query Optimization (5 talks)	Research Session 19: Data Streams I (5 talks)	Research Session 20: Graph Management I (5 talks)	Tutorial 3: Array DBMS: Past, Present, and (Near) Future
10:15	9:15							
10:30	9:30							
10:45	9:45							
11:00	10:00							
11:15	10:15							



NATIONAL RESEARCH
UNIVERSITY

Thank you for your attention!

<http://cv.gis.land>