SciDB

Philippe Cudre-Mauroux

Definition

SciDB is a distributed database management system for managing and processing multi-dimensional arrays in scientific applications. It was first designed and developed by a group of academics led by Michael Stonebraker before being productized by Paradigm4.

Overview

The first XLDB Workshop (1st Extremely Large Databases Workshop 2007) in 2007 brought together a group of scientists and industry members to discuss the capabilities of Data Base Management Systems (DBMSs) at managing non-relational data at extreme scales. A number of key shortcoming were identified, which were presented the following year at XLDB-2 (2nd Extremely Large Databases Workshop 2008). The lack of support for managing scientific data was in particular discussed. Consensus emerged that many Big Science projects presented unique challenges that could not be handled through generic DBMSs and would require a complete rewrite approach (Stonebraker et al 2007).

As a result, a group of scientists led by Micheal Stonebraker embarked on a new project to design and develop a new system for handling scientific arrays, as arrays represented a prominent datatype for many science users (including astronomers, oceanographers, seismologists or climate researchers). The resulting system, SciDB, was first presented at VLDB 2009 (Cudré-Mauroux et al 2009). This first version supported a number of important features for science users, including:

1. native storage for nested, multidimensional arrays;
2. scientific operators built as User-Defined Functions (UDFs) to

manipulate both the structure and the values of the arrays;
3. a shared-nothing design (Stonebraker 1986) allowing the system to scale out to many nodes and petabytes of data;
4. initial support for advanced features required by science users such as data versioning, provenance and uncertainty.

The team continued the development of SciDB as an open-source data management solution for Big Science projects. In 2011, Michael Stonebraker and Marilyn Matz co-founded Paradigm4 to further support the development of SciDB.

Architecture

SciDB is at its core a distributed array management system. The system takes as input large, multidimensional arrays potentially considering several values for each cell in the array. SciDB natively stores such arrays as a collection of *chunks*, each handling a portion of an array (e.g., a chunk or 10k x 10k x 5k array cells). Chunks are typically of equal size (e.g., 64 MB) and store attribute values vertically (i.e., a chunk only stores the values of a given attribute).

Chunks partition the array spatially, but can be overlapping to ease some operations like object detection, which would otherwise often involve stitching together multiple chunks. Chunks are written using various compression mechanisms on disk. The system adopts a *no-overwrite* storage strategy meaning that arrays cannot be updated in-place (they can however be appended or updated by creating a new version of the array). A system catalog keeps track of the attributes and the spatial position of each chunk.

SciDB distributes the chunks among a set of worker nodes. Queries consist of a tree of operators over one or multiple arrays. Operators can modify both the structure of an array (e.g., its size or dimensions) as well as its contents (attribute values). The system comes with a number of standard operators such as *Filter*, which filters cell values based on a threshold, or *SJoin*, which combines attributes from different cells. The system is extensible and allows the definition of both User-Defined Types (UDTs) and User-Defined Functions (UDFs).

Many queries, such as *filter* or *object detection*, can be processed fully in parallel on each worker node. Specific operations might however require to exchange, share or redistribute the chunks dynamically. SciDB defines a dedicated operator, called *ScatterGather*, to help support such cases. *Scatter/Gather* is a powerful operator allowing to dynamically redistribute the array over the nodes. Queries execution is orchestrated by a central node called the *coordinator*.

Academic Prototype

The first prototype of SciDB was developed as an open-source project by a group of researchers from MIT, Brow University, SLAC, NIISI RAS, the University of Washington, Portland State University, and Microsoft. The group was lead by Michael Stonebraker. The system was first demonstrated at VLDB in 2009 (Cudré-Mauroux et al 2009).

This academic prototype adopted the architecture described above, with a non-overwrite, native array storage system and distributed query execution. It also featured initial support for array versioning, data provenance and uncertainty, as well as a simple query optimizer.

The main use-case for this prototype was loosely modeled on an astronomy workload, and was later extended to a full benchmark for scientific data management systems (Cudre-Mauroux et al 2010). The use-case considered the end-to-end ingestion and processing of raw data from a sensor system. It included three types of operations: (i) manipulation of raw imagery, including processing pixels to extract geo-spatial observations; (ii) manipulation of observations, including spatial aggregation and grouping into related sets; and (iii) manipulation of groups, including a number of relatively complex geometric operations in several dimensions.

The academic prototype was the basis of several research projects in array data management, including:

- A new storage manager to efficiently encode and access versioned arrays (Seering et al 2012);
- A new storage manager for complex, parallel array processing implementing various partitioning and query execution strategies (Soroush et al 2011);
- A hybrid analytic system for array-structured data integrating R and SciDB (Leyshock et al 2013);
- A new structure to store and model multidimensional arrays supporting efficient inference processes (Ge and Zdonik 2010);
- Techniques to handle fine-grained lineage in scientific databases (Wu et al 2013);
- Techniques to incrementally adds nodes and to optimize data placement for n-dimensional array systems (Duggan and Stonebraker 2014).

Paradigm4 Development

In 2011, Michael Stonebraker and Marilyn Matz co-founded Paradigm4 to further support the development of SciDB (Paradigm4 2011). The company hired Paul Brown to oversee the development of the system.

The system developed by Paradigm4 follows the architecture described above, but adds a number of key features and libraries to SciDB (Stonebraker et al 2011, 2013). It supports both a SQL-like, declarative language called *AQL* and a functional language (AFL). The system includes a full-fledged optimizer, and an executor able to distribute queries to thousands of nodes. Support was also added for handling variable-size chunks, various encoding schemes (e.g., delta encoding, run-length encoding or LZ encoding), uncertain data, and coarse-grained provenance.

A number of optimized operators come built-in with the system, including common linear algebra operators. The system also adds a number of dedicated operators for several vertical domain, e.g., for life-sciences (to support genomic analysis, digital biomarker discovery, or biomedical images) and finance (for multi-factor model generation or transaction cost

analysis). Further use-cases that have been explored include sensor analytics, insurance and E-commerce.

Cross-References

Big SQL, Big Spatial Data Management [Architectures].

References

Cudré-Mauroux P, Kimura H, Lim K, Rogers J, Simakov R, Soroush E, Velikhov P, Wang DL, Balazinska M, Becla J, DeWitt DJ, Heath B, Maier D, Madden S, Patel JM, Stonebraker M, Zdonik SB (2009) A demonstration of scidb: A science-oriented DBMS. PVLDB 2(2):1534–1537, URL http://www.vldb.org/pvldb/2/vldb09-76.pdf

Cudre-Mauroux P, Kimura H, Lim KT, Rogers J, Madden S, Stonebraker M, Zdonik SB (2010) Ss-db: A standard science dbms benchmark. URL http://www-conf.slac.stanford.edu/xldb10/docs/ssdb_benchmark.pdf

Duggan J, Stonebraker M (2014) Incremental elasticity for array databases. In: International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014, pp 409–420, DOI 10.1145/2588555.2588569, URL http://doi.acm.org/10.1145/2588555.2588569

1st Extremely Large Databases Workshop (2007) URL http://www-conf.slac.stanford.edu/xldb2007/

2nd Extremely Large Databases Workshop (2008) URL http://www-conf.slac.stanford.edu/xldb2008/

Ge T, Zdonik SB (2010) A*-tree: A structure for storage and modeling of uncertain multidimensional arrays. PVLDB 3(1):964–974, URL http://www.comp.nus.edu.sg/~vldb2010/proceedings/files/papers/R86.pdf

Leyshock P, Maier D, Tufte K (2013) Agrios: A hybrid approach to big array analytics. In: Proceedings of the 2013 IEEE International Conference on Big Data, 6-9 October 2013, Santa Clara, CA, USA, pp 85–93, DOI 10.1109/BigData.2013.6691558, URL https://doi.org/10.1109/BigData.2013.6691558

Paradigm4 (2011) URL https://www.paradigm4.com/

Seering A, Cudré-Mauroux P, Madden S, Stonebraker M (2012) Efficient versioning for scientific array databases. In: IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012, pp 1013–1024, DOI 10.1109/ICDE.2012.102, URL https://doi.org/10.1109/ICDE.2012.102

Soroush E, Balazinska M, Wang DL (2011) Arraystore: a storage manager for complex parallel array processing. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011, pp 253–264, DOI 10.1145/1989323.1989351, URL http://doi.acm.org/10.1145/1989323.1989351

Stonebraker M (1986) The case for shared nothing. IEEE Database Eng Bull 9(1):4–9, URL http://sites.computer.org/debull/86MAR-CD.pdf

Stonebraker M, Madden S, Abadi DJ, Harizopoulos S, Hachem N, Helland P (2007) The end of an architectural era (it's time for a complete rewrite). In: Proceedings of the 33rd International Conference on Very Large Data Bases, University of Vienna, Austria, September 23-27, 2007, pp 1150–1160, URL http://www.vldb.org/conf/2007/papers/industrial/p1150-stonebraker.pdf

Stonebraker M, Brown P, Poliakov A, Raman S (2011) The architecture of scidb. In: Scientific and Statistical Database Management - 23rd International Conference, SSDBM 2011, Portland, OR, USA, July 20-22, 2011. Proceedings, pp 1–16, DOI 10.1007/978-3-642-22351-8_1, URL https://doi.org/10.1007/978-3-642-22351-8_1

Stonebraker M, Brown P, Zhang D, Becla J (2013) Scidb: A database management system for applications with complex an-

alytics. Computing in Science and Engineering 15(3):54–62, DOI 10.1109/MCSE.2013.19, URL https://doi.org/10.1109/MCSE.2013.19

Wu E, Madden S, Stonebraker M (2013) Subzero: A fine-grained lineage system for scientific databases. In: 29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013, pp 865–876, DOI 10.1109/ICDE.2013.6544881, URL https://doi.org/10.1109/ICDE.2013.6544881