The AsterixDB Big Data Management System (BDMS) is the result of about 3.5 years of R&D involving researchers at UC Irvine, UC Riverside, and UC San Diego. The AsterixDB code base consists of roughly 250K lines of Java code that has been co-developed at UC Irvine and UC Riverside.

Initiated in 2009, the NSF-sponsored ASTERIX project has been developing new technologies for ingesting, storing, managing, indexing, querying, and analyzing vast quantities of semi-structured data. The project has been combining ideas from three distinct areas --- semi-structured data management, parallel databases, and data-intensive computing (a.k.a. today's Big Data platforms) --- in order to create a next-generation, open-source software platform that scales by running on large, shared-nothing commodity computing clusters. The ASTERIX effort has been targeting a wide range of semi-structured information, ranging from "data" use cases---where information is well-typed and highly regular---to "content" use cases---where data tends to be irregular, much of each datum may be textual, and the ultimate schema for the various data types involved may be hard to anticipate up front. The ASTERIX project has been addressing technical issues including highly scalable data storage and indexing, semi-structured query processing on very large clusters, and merging time-tested parallel database techniques with modern data-intensive computing techniques to support performant yet declarative solutions to the problem of storing and analyzing semi-structured information effectively. The first fruits of this labor have been captured in the AsterixDB system that is being released in Beta release form on June 6th, 2013. We are hoping that the arrival of AsterixDB will mark the beginning of the "BDMS era", and we hope that both the Big Data community and the database community will find the AsterixDB system to be interesting and useful for a much broader class of problems than can be addressed with any one of today's current Big Data platforms and related technologies (e.g., Hadoop, Pig, Hive, HBase, MongoDB, and so on). One of our project mottos has been "one size fits a bunch"---at least that has been our aim. For more information about the research effort that led to the birth of AsterixDB, please refer to our NSF project website: http://asterix.ics.uci.edu/.

The project itself is hosted on Google code at: http://code.google.com/p/asterixdb/.

In a nutshell, AsterixDB is a full-function BDMS with a rich feature set that distinguishes it from pretty much any other Big Data platform that's available today. We believe that its feature set makes it well-suited to modern needs such as web data warehousing and social data storage and analysis. It has:

- A semistructured (NoSQL style) data model based on extending JSON with object database ideas;
- An expressive and declarative query language that supports a broad range of queries and analysis over semistructured data;
- A parallel runtime execution engine that has been scale-tested up to 1000+ cores and 500+ disks;
- Partitioned LSM-based data storage and indexing to support efficient data ingestion and searching;
- Support for access to externally stored data (e.g., in HDFS) as well as to data stored in AsterixDB;
- A rich set of primitive data types, including spatial, temporal, and textual data;
- Indexing options that include B+ trees, R trees, and inverted keyword (exact and fuzzy) indexes;
- Support for fuzzy and spatial queries as well as for traditional parametric queries;
- Basic transactional (concurrency and recovery) capabilities akin to those of a NoSQL store.