

LEAP(4BD): Lightweight Evaluation and Architecture Prototyping for Big Data



Make more informed decisions about Big Data system design

Ensure scalability as your data grows

Choose NoSQL databases that can support your future scalability needs

The exponential growth of data in the last decade has fueled a new specialization for software engineering—*big data* software systems. At the heart of big data systems is a collection of database technologies that are simpler and more lightweight, and provide higher scalability and availability than traditional relational databases.

These highly scalable “NoSQL” databases are typically designed to scale horizontally across clusters of low-cost, moderate-performance servers. They achieve high performance, elastic storage capacity, and availability by replicating and partitioning data sets across a cluster. Each database specifies its own proprietary data model and query language, as well as database-specific mechanisms for achieving distributed data consistency and availability.

This specification means that for software engineers building scalable, big data applications, there’s a dizzying range of potential databases that can be used as building blocks for a solution. The range of choices makes database selection a crucial software architecture decision, as selecting a database technology that can’t meet system requirements will be costly, reduce downstream productivity due to rework, and even lead to project cancelation. Therefore, architects must carefully compare candidate database technologies and features and select platforms that can satisfy application quality and cost requirements. In the inevitable absence of up-to-date, reliable technology evaluations, this comparison exercise is in practice a highly exploratory, unstructured task that uses an Internet search engine as the primary information gathering and assessment tool.

LEAP(4BD)

To meet the challenge of selecting a database for big data systems, the Software Engineering Institute (SEI) has developed the Lightweight Evaluation and Architecture Prototyping for Big Data (LEAP4BD) method. The LEAP(4BD) method provides a systematic approach for a project to select a NoSQL database that can satisfy its requirements.

A key feature of LEAP(4BD) is its NoSQL database feature-evaluation criteria. This ready-made set of criteria significantly speeds up a NoSQL database evaluation and acquisition effort. To this end, we have categorized the major characteristics of data management technologies based on the following areas:

- Query language
- Data distribution
- Consistency
- Performance
- Modifiability
- Data model
- Data replication
- Scalability
- Availability
- Administration and management

Within each category, we have detailed evaluation criteria that will differentiate big data technologies. These evaluation criteria are reusable across projects and pre-populated to capture the capabilities of most popular NoSQL technologies.

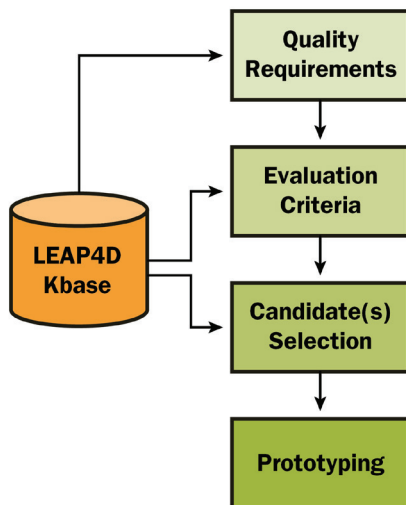
LEAP(4BD): Lightweight Evaluation and Architecture Prototyping for Big Data

How Does LEAP(4BD) Work?

In LEAP(4BD), we first work with the project team to identify features pertinent to the system under development. These features help identify a specific set of technologies that will best support the system. Then, we weight individual features according to system requirements and evaluate each candidate technology against these features.

The LEAP(4BD) method employs five steps to produce both qualitative and quantitative evaluation results:

- 1. Specify requirements.** First, we elicit the high-priority functional and quality requirements for the system. For big data applications, we focus on specific requirements for performance, scalability, availability, consistency, and security. We also define a use case that is representative of the customer's application domain. The use case defines both a data model and workload that will be used as the basis of performance and scalability assessment in Step 4.



- 2. Select candidate NoSQL databases.** Next, we select two to four candidate NoSQL databases for deeper evaluations. Selection criteria are both contextual (e.g., experience with a specific technology) and technical, and require evaluating database features against the requirements developed in Step 1.
- 3. Design a use case-specific data model.** Based on the use case defined for evaluation, we map the application's logical data model to the physical model supported by the candidate NoSQL databases. We also deploy the database and load the test data (synthetic or actual) into the database instances.
- 4. Execute performance and scalability tests.** We implement a test case driver that executes the specified workload on each database. Load is scaled by increasing the number of concurrent client requests to assess how each database reacts to increased workloads.

- 5. Report results.** The evaluation report includes both qualitative and quantitative results. The report details the performance and scalability results that we obtained from testing each NoSQL database in a consistent environment. It also describes how easily the logical data model maps to the specific NoSQL data models that were tested, and the specific features of each database that will influence the identified quality requirements for the application.

LEAP(4BD) is supported by a knowledge base that stores the results of our evaluations and comparisons of different NoSQL databases. We have pre-populated the LEAP(4BD) knowledge base with evaluations of specific technologies (e.g., MongoDB, Cassandra, and Riak) with which we have extensive experience. Each evaluation of a new technology adds to this knowledge base, making evaluations more streamlined as the knowledge base grows. Overall, LEAP(4BD) provides a systematic, quantitative, and highly transparent approach that quickly ranks the various candidate technologies according to project requirements.

Engage with Us

If you would like to use LEAP(4BD) in your next acquisition of a big data system, please contact SEI Customer Relations at info@sei.cmu.edu or 1-412-268-5800. We would like to work with you to select a NoSQL database that can support your future scalability needs.

Additional Resources

An Approach to Managing the Software Engineering Challenges of Big Data (podcast), <http://url.sei.cmu.edu/iq>

Related Web Sites

www.sei.cmu.edu/architecture

blog.sei.cmu.edu/archives.cfm/category/big-data

For Course Registration

www.sei.cmu.edu/go/big-data

For General Information

For information about the SEI and its products and services, contact Customer Relations
Phone: 412-268-5800
FAX: 412-268-6257
info@sei.cmu.edu
www.sei.cmu.edu