

## *Spatio -Textual Data Exploration at Scale*

Web: <https://www.ds.unipi.gr/spades/>



The [SPADES project](#) is a research project that is funded by the [Hellenic Foundation for Research and Innovation \(HFRI\)](#) and the General Secretariat for Research and Technology (GSRT), under grant agreement No [1667]. The instrument aims to support post-doctoral research, and the principal investigator of SPADES is [Akrivi Vlachou](#). The project is hosted at the [Department of Digital Systems](#) in the [University of Piraeus](#).

## Description of Work

Several research challenges are associated with the efficient support of spatio-textual processing at scale. Two main factors are critical for location-based services, thus determining their overall performance: the efficiency of query processing for spatio-textual queries and the quality of the retrieved points of interest. The efficiency of query processing directly influences the query throughput, which is very important in the context of scalable applications. Consequently, SPADES identifies a set of research and technological challenges that need to be effectively addressed, in order to support spatio-textual and spatio-temporal queries over massive data:

- 1. Support of advanced spatio-textual query types**
- 2. Novel distributed index structures for spatio-textual search**
- 3. Abstractions for parallel spatio-textual data processing**

## Objectives

- Novel rank-aware query types complying with the paradigm of spatial-keyword search that cover a wide variety of information needs, targeting at the mobile user.
- Provision of more expressive querying mechanisms for points of interest that combine spatial information and textual relevance with temporal information and user preferences.
- Advanced distributed indexing structures capable to support complex spatial-keyword queries effectively, by means of harnessing the merits of spatial data structures and text indexes.
- Novel partitioning mechanisms and load balancing techniques for spatio-textual queries, aiming at efficient parallel data processing.



## Impact

SPADES aims to address the limitations of spatio-textual data analysis and processing when applied in the context of Big Spatial Data, as witnessed by the lack of existing systems and techniques for this purpose. Achievement of this goal constitutes a substantial step forward in dealing with challenges emerging from management of Big Spatial Data. At a practical level, the research outcome will benefit applications such as spatio-textual search and retrieval, mining of spatio-textual data, next generation location-based services, and tourism-oriented applications to name a few.

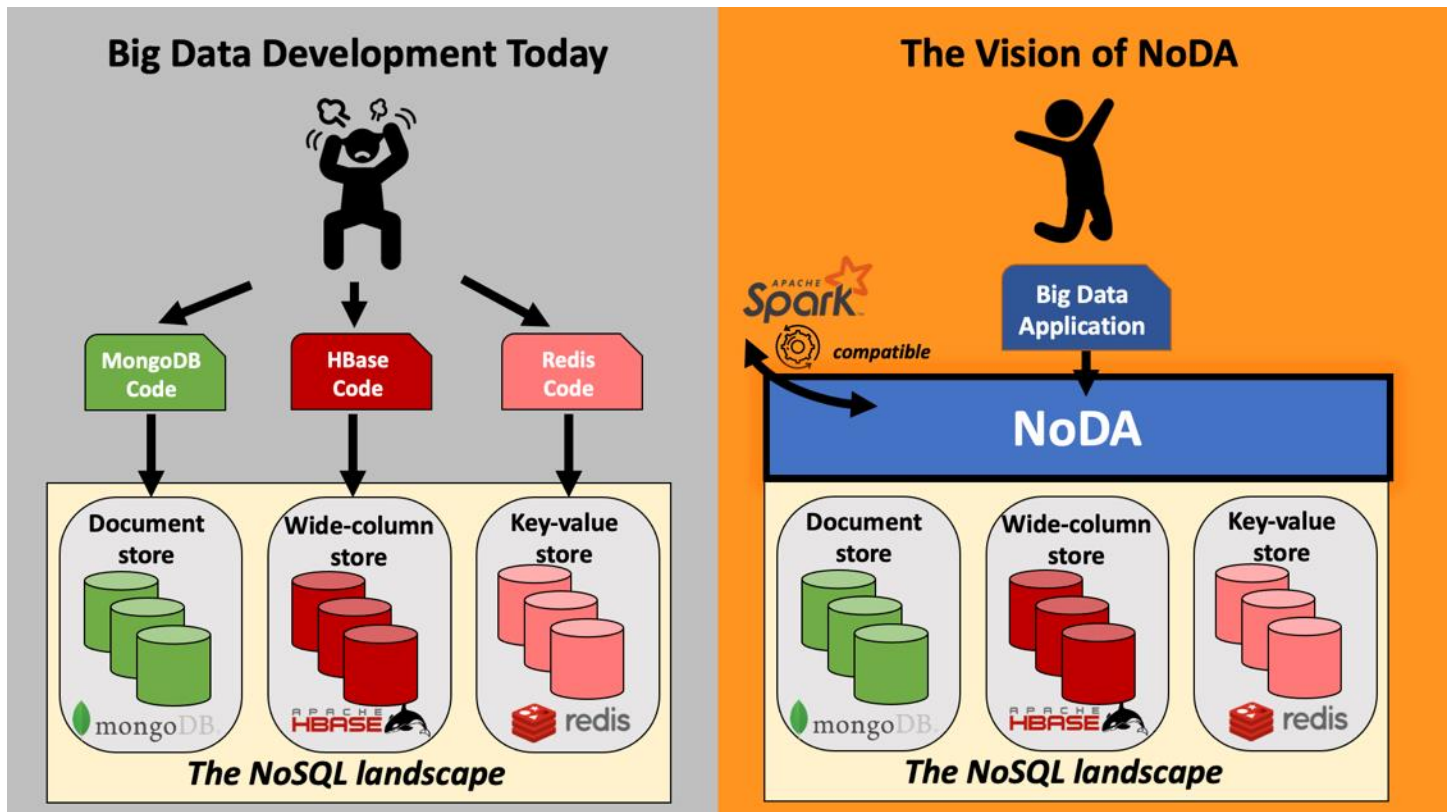
By exploiting SPADES the analysis of massive spatio-textual datasets (typically encountered in the aforementioned domains and especially in social networks) is going to be accelerated significantly. In consequence, applications will be able to query and analyze larger quantities of spatio-textual data in shorter time, thus speeding up the process of making new scientific discoveries.

## Results

The results of the project are expected to be exploited by the local tourism business that could benefit from the provision of innovative location-based services over vast quantities of data to tourists. Such services entail queries that are processing-intensive and typically run for minutes rather than seconds, and often produce results that are not truly useful to the end user. SPADES promises to facilitate the access to location-based information, a task of particular importance to tourists. In this respect, the research results of SPADES are expected to benefit also society at large.

# Research Highlight: Uniform Access to NoSQL Stores

Even though NoSQL stores (MongoDB, CouchDB, HBase, Cassandra, Redis, etc.) comprise the state-of-the-art technology for Big Data management, they still rely on different languages and programming APIs, thus hindering application development.



In practice, as illustrated in the figure, a big data developer that wishes to implement a MongoDB application must first learn the API of MongoDB (similarly for HBase, Redis, etc.). Unfortunately, these APIs differ significantly, as they (may) support different data models. Consequently, the code of the same application is completely different from one NoSQL system to the other, a fact that complicates the transition between NoSQL systems.

In the context of SPADES, a novel method is proposed for unifying different NoSQL stores by means of a common programming interface. This method is called NoDA (**No**SQL **D**ata **A**ccess Operators) [3] and bears similarity to other interfaces (ODBC, JDBC) that have been used successfully in the relational database field for decades.

In plain terms, NoDA is a programming API (see the blue layer at the right part of the figure) that consists of basic data access operators (such as filter, project, aggregate, sort), which are implemented for each NoSQL system separately, thus offering a simple and familiar language to the big data developer in order to implement applications, with the following advantages:

- It is simple to use and easy to learn, as it hides the peculiarities of each NoSQL system.
- It is unified, thus allowing code portability from one NoSQL system to the other, in the same spirit as in relational databases.
- It offers additionally a declarative, SQL-like interface, which makes it more user-friendly both for big data developers and data analysts.

# Publications

## Journals

1. Panagiotis Nikitopoulos, Georgios A. Sfyris, Akrivi Vlachou, Christos Doulkeridis, Orestis Telelis: [Pruning Techniques for Parallel Processing of Reverse Top-k Queries](#). In Distributed and Parallel Databases Journal, Springer, 2020

---

## Conferences & Workshops

1. Akrivi Vlachou, Christos Doulkeridis, Nikolaos Koutroumanis, Dimitrios Pouloupoulos, Kjetil Norvag. [The SPADES Framework for Scalable Management of Spatio-textual Data](#). In Proceedings of 24th Pan-Hellenic Conference on Informatics (PCI'20), Athens, Greece, November 2020
2. Stella Maropaki, Sean Chester, Christos Doulkeridis, Kjetil Norvag: [Diversifying Top-k Point-of-Interest Queries via Collective Social Reach](#) In Proceedings of 29th ACM Conference on Information and Knowledge Management (CIKM'20), October 19-23, 2020.
3. Nikolaos Koutroumanis, Panagiotis Nikitopoulos, Akrivi Vlachou, Christos Doulkeridis: [NoDA: Unified NoSQL Data Access Operators for Mobility Data](#). In Proceedings of the 16th International Symposium on Spatial and Temporal Databases (SSTD'19), Vienna, Austria, August 2019.
4. Georgios M. Santipantakis, Apostolos Glenis, Christos Doulkeridis, Akrivi Vlachou, George A. Vouros: [stLD: Towards a Spatio-temporal Link Discovery Framework](#). In Proceedings of the International Workshop on Semantic Big Data (SBD'19) (workshop held in conjunction with SIGMOD'19), Amsterdam, The Netherlands, July 2019, pp.4:1-4:6.
5. Panagiotis Nikitopoulos, Georgios A. Sfyris, Akrivi Vlachou, Christos Doulkeridis, Orestis Telelis: [Parallel and Distributed Processing of Reverse Top-k Queries](#). In Proceedings of the 35th IEEE International Conference on Data Engineering (ICDE'19), Macau SAR, China, April 2019, pp.1586-1589
6. Akrivi Vlachou, Christos Doulkeridis, Apostolos Glenis, Georgios M. Santipantakis, George A. Vouros: [Efficient Spatio-temporal RDF Query Processing in Large Dynamic Knowledge Bases](#). In Proceedings of the 34th ACM/SIGAPP Symposium On Applied Computing (SAC'19), Limassol, Cyprus, April 2019, pp.439-447.

---

## Technical Reports

1. Akrivi Vlachou, Christos Doulkeridis, Nikolaos Koutroumanis, Dimitrios Pouloupoulos, Kjetil Norvag. The SPADES Framework for Scalable Management of Spatio-textual Data. Technical Report (2020).

---

## Contact Us

For more information please contact Akrivi Vlachou ([avlachou@unipi.gr](mailto:avlachou@unipi.gr)).

For more information about the research group and the department, please visit the respective home pages:  
[Department of Digital Systems](#)

