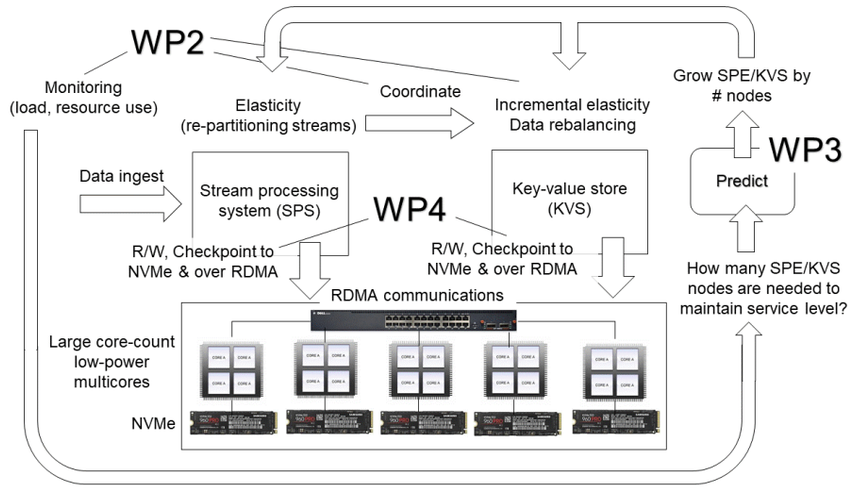
H.F.R.I.
Hellenic Foundation for Research & Innovation

Description of the funded research project

1st Call for H.F.R.I. Research Projects to Support Faculty Members & Researchers and Procure High-Value Research Equipment

STREAMSTORE



Elastic, Predictable NoSQL Data Storage for Scalable Streaming Applications

PI and Research Team

Principal Investigator: Prof. Kostas Magoutis, Univ. of Crete



Reader-friendly title: STREAMSTORE

Scientific Area: Mathematics and Information Sciences

Host Institution: FORTH, Institute of Computer Science

Project webpage: https://streamstore-project.eu

Budget: 169,422 Euro

Duration: 36 Months

H.F.R.I.
Hellenic Foundation for Research & Innovation

Research project synopsis

Processing high-volume streaming data is an important driver for creating business value in popular social networking and Internet services, giving rise to a new generation of stream-processing systems (SPSs). Stream processing often requires consulting and updating large persistent data structures to enrich or correlate with stored ground-truth data (social graphs, user profiles, etc.). Such persistent data structures are typically stored in distributed NoSQL key-value stores (KVSs). Advanced SPSs additionally require a high degree of elasticity, namely the ability to rapidly scale to adapt to higher load, often swiftly increasing stream-processing rate by 2x-3x. A challenge in such cases is that the KVS must also adapt at least as fast as the SPS to sustain the increased level of streaming throughput. KVS adaptation involves re-distribution of data across nodes, a process that takes time and impacts performance. In addition, elasticity actions at the KVS are not coordinated with (triggered by) elasticity actions at the SPS part of the system. The long term vision of STREAMSTORE is to fully support highly elastic scalable SPSs managing large operator states on KVSs. Rapid low-impact elasticity across all system layers become a first-class property for maintaining system response during periods of high workload variability. This vision requires the evolution of the elasticity, predictability, and availability of scalable data store technology beyond the state of the art. The targeted scientific breakthrough of STREAMSTORE is to deliver a highly elastic (rapidly scalable) KVS that is optimized for emerging server architectures and fully coordinated with state-of-the-art elastic scalable SPSs. Implementing it requires extending state-of-the-art software techniques (such as incremental elasticity, measurement-based performance prediction, and lightweight checkpointing) and aligning them with novel hardware capabilities of emerging server clusters, to enable highly elastic KVSs.

Project originality

STREAMSTORE advances the scientific state of the art by targeting the coordination of KVS and SPS actions during elasticity phases and by offering new techniques for the elastic management of distributed NoSQL KVSs. KVSs currently support reconfiguration (elasticity) actions but are typically not optimized for rapid, low-impact such actions when modifying the number of servers storing data. SPSs in production environments are known to experience drastic load surges often by 2x-3x. While SPSs can be rapidly reconfigured by allocating more tasks to process new streams, the associated KVSs take longer to reconfigure and have an impact on application performance while reconfiguring, raising the need for improved elasticity support for KVSs, as well as better integration and coordination in elasticity actions in the data store and stream-processing system. STREAMSTORE proposes the use of incremental elasticity technology (developed by the PI's Team) for improving the elasticity of scalable stream-processing systems, an application domain in which the impact of KVS limitations has not been considered at depth. STREAMSTORE will exploit untapped research potential in leveraging novel hardware features of multicore server architectures (flexible allocation of a large number of cores, fast low-overhead communication and storage devices), which fit very well with the hardware requirements of incremental elasticity. STREAMSTORE will further advance the state-of-the-art in performance prediction in scalable SPS and KVSs, focusing on the use of measurement-based ML techniques such as support vector regression (SVR), artificial neural network (ANN) regression, or multivariate adaptive regression splines (MARS), building upon recent work by the PI's Team on performance modeling of distributed KVSs. STREAMSTORE will showcase the use of state-of-the-art networking techniques (RDMA) and novel storage technologies to improve service performance and availability of the combined SPS/KVS platform.

Expected results & research project impact

STREAMSTORE research advances the scientific state of the art by addressing the alignment of the elasticity actions between SPSs and the KVSs they depend on for joining streaming data with data at-rest. The main STREAMSTORE research result will be a coordinated, scalable, elastic SPS and KVS platform leveraging new networking and storage technologies at a targeted TRL of 4. Economically and socially, STREAMSTORE research is expected to have strong impact on important areas such as social networking (evidenced by the recent interest of Internet service giants such as Google, Twitter, and LinkedIn), surveillance operations for counter-terrorism activities, high-speed financial trading, and other stream processing applications where high-volume processing of events, large amounts of stored operator states, and periods of high variability in incoming stream load are anticipated and must be handled appropriately to ensure that user experience is not affected during such periods. Thus it directly improves the ability to respond to surges of societal activity, and rapid detection of trends and high-value events at low latency. High engagement of citizens into social media (requiring higher value/quality and more timely information) as well as the adoption of such systems in e-governmental services, will increase the citizens' trust in the dependability of critical infrastructure and directly contribute to further digitization of society. STREAMSTORE is also in line with emerging technology trends and contributes to a technology vision where infrastructure-level software (SPS and KVS platforms) must take advantage of new (commercially available) server hardware for the high level of parallelism, and high-throughput networking and storage devices. The research output of STREAMSTORE will validate these emerging platforms, highlighting their benefits in the stream-processing and data-serving domains. STREAMSTORE technology is also expected to impact application domains beyond stream processing, such as implementing service-level agreements (SLA) for scalable persistent data.

Importance of this funding

H.F.R.I. funding of Blue Sky research through funding schemes for faculty members gave this PI and his Team a significant boost in their freedom to focus on cutting-edge research over emerging distributed systems and technologies, allowing them to advance an ambitious research-oriented agenda. By being investigator-driven and bottom-up in its approach, H.F.R.I. research funding allowed this PI and his Team to put a strong focus and effort on pushing the envelope in a high-impact research area, with a greater degree of efficiency and flexibility than previously possible through national funding schemes.

H.F.R.I.
Hellenic Foundation for Research & Innovation

COMMUNICATION

185 Syggrou Ave. & 2 Sardeon St. 2
171 21, N. Smyrni, Greece
+30 210 64 12 410, 420
communication@elidek.gr
www.elidek.gr