1   Publishable summary

LeanBigData (http://leanbigdata.eu/) aims at:

- Architecting and developing three **resource-efficient** Big Data management systems typically involved in Big Data processing: a novel transactional NoSQL key-value data store, a distributed complex event processing (CEP) system, and a distributed SQL query engine. The efficiency of these systems is one of the main innovations of the project.

- Providing **an integrated big data platform** with these three main technologies used for big data, NoSQL, SQL, and Streaming/CEP that will improve response time for unified analytics over multiple sources of data avoiding the inefficiencies and delays introduced by existing ETL-type approaches.

- Supporting an **end-to-end big data analytics solution** enhancing the lifecycle of data analytics by: 1) automated discovery of anomalies and root cause analysis that will provide end-users with a starting point at time 0; 2) Supporting data scientists to manipulate the result set of analytical queries in an agile way by means of a visual and interactive interface to discover insights by enabling an easy declarative manipulation of the results sets.

During the second year of LeanBigData the following results have been achieved:

1) A startup, LeanXcale, has been created and incorporated in the project; 2) the first version of the platform has been produced. It integrates the OLAP query engine, the enhanced transactional system, the new CEP, the data collection framework, data sensors, and initial version of the use cases; 3) design and implementation of the new LeanBigData monitoring system; 4) the frontend of the key-value data store has been fully implemented; design and initial implementation of the protocols for elasticity and fault tolerance 5) the tiered fast storage layer (TFL) prototype has been implemented and integrated with the front-end key value store via the shared block device abstraction; 6) redesign and implementation of the transactional processing to increase its efficiency; 7) the CEP final prototype is completed; 8) The parallel-distributed implementation of the CEP operators has been completed. Table operators have been designed and implemented to integrate the CEP with the ultra-scalable database and the key-value data store. 9) The data collection framework is fully implemented; 10) Different data sensors (CPU stats, storage and as well as memory stats and energy consumption) are integrated; 11) the Distributed Query Engine for analytic processing provides now intra-query and intra-operator parallelism; 12) a vision based gesture recognition system has been implemented; 13) the anomaly detection algorithms have been implemented, as well as a preliminary version of the distributed clustering; 14) the dashboard is being implemented based on Ganglia as monitoring system; 15) initial versions of all the use cases have been produced and integrated with the platform.