

The paper offers 5 key novel contributions beyond the state of the art:

- a reverse-engineering tool for big data applications;
- anti-patterns and heuristics analysis of recovered big data topologies/architectures;
- transparent formal verification of recovered topologies;
- support for Hadoop Map Reduce Applications (Batch) as well as Apache Storm applications (streaming);
- industrial evaluation using case-study research;