Collaborative provenance for workflow-driven science and engineering

Altıntaş, İ.
List of Tables

3.1 Summary of the characteristics of different scientific workflow provenance approaches. .................................................. 39

5.1 Example queries across workflow executions and collaborations .......... 60

6.1 Relation instances of the provenance schema corresponding to the example in Figure 6.2. .................................................. 68

6.2 Basic QLP constructs and functions, where $e_n$ is a node expression comprised of either a data artifact identifier, a run identifier, a data artifact type (denoting the set of artifacts having that type), or a workflow (denoting the set of runs of the workflow). We use $p$ to denote a QLP path query, and $r_i$ to denote a run. $a$ is attribute for user collaboration graph, which can be either of these $\{\phi, N, W, S, NW, NS, WS, NWS\}$. .......................................................... 75

6.3 Example queries expressed using the dependency functions defined for QLP. 76

6.4 Mapping of basic Open Provenance Model nodes and edges to the collaborative entities and relationships. .......................... 78

7.1 Relation instances of the provenance schema corresponding to the example in Figure 7.4. .................................................. 89

7.2 Example CAMERA queries 1 through 6 expressed in Datalog and QLP extensions. ............................................................. 93

7.3 Example CAMERA query 7 expressed in Datalog and QLP extensions. 94

8.1 Collaborative provenance views for data dependency, run dependency and user collaboration expressed in PostgreSQL. .................. 102

8.2 Example CAMERA queries Q1 through Q6 expressed in PostgreSQL. ...... 103

8.3 Example CAMERA query Q7 expressed in PostgreSQL. .................... 104

8.4 The size of database (in number of tuples) for different datasets. ........... 105

8.5 The query execution time (in ms) for collaborative provenance views over datasets of 5, 10, 25 and 50 run dependencies. ................. 105
8.6 The query execution time (in ms) for evaluation queries 1 through 7 over datasets of 5, 10, 25 and 50 run dependencies.

9.1 The publish and run observables in interoperable PC3 scenario. The contents of the table shall be read as follows: e.g., the second row refers to run1 performed by u2 using wf_preload published by u1. In run1, u2 used d_{062941} as an input and the run produced d_{062941−1} as its output.