

Efficient Recommendation of De-identification Policies using MapReduce

ABSTRACT

Many data owners are required to release the data in a variety of real world application, since it is of vital importance to discovery valuable information stay behind the data.

EXISTING SYSTEM

Big data is important to exchange and share data among different parties. However, publishing those data containing sensitive information could violate individual's privacy. In order to get sufficient protection while maintain high data utility, privacy-preserving data publication (PPDP) .In existing re-identification attacks on the AOL and ADULTS datasets have shown that publish such data directly.

DRAWBACKS

- It causes a tremendous threat to the individual data privacy.
- No balance between data utility and privacy.

PROPOSED SYSTEM

In proposed, De-identification policies is one of the models that can be used to achieve such requirements, however, the number of de-identification policies is exponentially large due to the broad domain of quasi-identifier attributes. To better control the trade off between data utility and data privacy, skyline computation can be used to select such policies, but it is yet challenging for efficient skyline processing over large number of policies. In this paper, we propose one parallel algorithm called SKY-FILTER-MR, which is based on MapReduce to overcome this challenge by computing skylines over large scale de-identification policies that is represented by bit-strings. To further improve the performance, a novel approximate skyline computation scheme was proposed to prune unqualified policies using the approximately domination relationship.

ADVANTAGES

- The power of filtering in the policy space generation stage was greatly strengthened to effectively decrease the cost of skyline computation over alternative policies.
- Privacy protection is provided

SYSTEM REQUIREMENTS

H/W System Configuration:-

- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

S/W System Configuration:-

- Operating System : Linux
- Application Server : Tomcat5.0/6.X
- Backend coding : Java
- Tool : Virtual Box
- Environment : Ubuntu
- Technology : Hadoop