

# **Diseño de una base de datos orientada a grafos para la gestión de datos en iDatos**

**Jesús Alberto González Yanes**

Instituto Canario de Estadística, ISTAC

[jgonyanp@gobiernodecanarias.org](mailto:jgonyanp@gobiernodecanarias.org)

**Rafael Betancor Villalba**

Instituto Canario de Estadística, ISTAC

[rbetvilb@gobiernodecanarias.org](mailto:rbetvilb@gobiernodecanarias.org)

**Luz Marina Moreno de Antonio**

Departamento de Ingeniería Informática y de Sistemas, Universidad de La Laguna

[lmmoreno@ull.edu.es](mailto:lmmoreno@ull.edu.es)

**Isabel Sánchez Berriel**

Departamento de Ingeniería Informática y de Sistemas, Universidad de La Laguna

[isanchez@ull.edu.es](mailto:isanchez@ull.edu.es)

## **PALABRAS CLAVE**

BB.DD. orientada a Grafos, iDatos, estadísticas multifuente

## **RESUMEN**

En la actualidad, en el marco del Plan Estadístico 2018-2022 se pretende impulsar el Sistema de Datos Integrados (iDatos) con el fin de producir estadísticas multifuente, apoyándose en una gestión eficiente de datos maestros compartidos en múltiples registros, de forma que faciliten el enlazamiento de los diferentes orígenes de datos. Dentro de este plan se han marcado objetivos que potencien el uso de registros administrativos y diversas fuentes de datos. El gran volumen de datos manejado y su continuo crecimiento exige el uso de tecnologías Big Data que garanticen la eficacia y el rendimiento de la solución que se proponga.

En este contexto encontramos que a una misma unidad de análisis le corresponderá datos que están dispersos en distintas fuentes ya sean administrativas u otras fuentes complementarias Big Data. El uso de datos maestros permite determinar un único elemento de referencia y facilita la construcción de las tablas que registran las relaciones entre los diferentes registros. El conjunto de datos maestros que se contemplan son: direcciones, edificios, viviendas y locales, población y hogares, y por último empresas y establecimientos. La solución actual almacena las relaciones en una base de datos relacional, sin embargo, en los últimos años han proliferado en diferentes contextos las bases de datos orientadas a grafos en las que las relaciones constituyen el elemento crucial en el modelo de datos. El beneficio del almacenamiento nativo de grafos viene dado por la infraestructura de distribución de los datos que se diseña y construye especialmente para tener un buen rendimiento y una alta escalabilidad en el tratamiento de los modelos de grafos, idóneos para la representación de las relaciones. Frente a las bases de datos relacionales y otras soluciones NoSQL, cuando se pretende explotar las relaciones entre datos masivos relacionados hay un aumento evidente de rendimiento.

Este trabajo se ha ocupado del análisis del problema y diseño del esquema de una Base de Datos Orientada a Grafos que de soporte al sistema iDatos. Para obtener conclusiones respecto a la viabilidad e idoneidad del uso de esta tecnología Big Data se ha trabajado con un conjunto de datos de prueba que permite implementar el grafo resultante y diseñar consultas para comparar el rendimiento en una base de datos relacional similar a la utilizada actualmente por el ISTAC y la solución propuesta.