

Prólogo

La persistencia de la información es un punto clave en cualquier sistema de información. En las últimas décadas, el modelo de persistencia más utilizado en cualquier proyecto informático ha sido el modelo relacional. Se trata de un modelo que ofrece un gran número de ventajas tales como estabilidad puesto que ha sido utilizado de manera intensa en las últimas décadas, ofrece un lenguaje de consultas estandarizado para poder manipular cualquier base de datos relacional, y se basa en un modelo formal matemático denominado el álgebra relacional. Además, el propio modelo es simple e intuitivo dado que se basa en el concepto de manipular tuplas de datos simples. Estas características han influido en su expansión y su utilización en una gran variedad de proyectos de naturaleza muy distinta. El modelo se concibió en un contexto concreto con respecto a la cantidad de datos y tipos de procesamiento que eran necesarios realizar. Sin embargo, la llegada del denominado fenómeno *big data*, ha cambiado radicalmente las necesidades de procesamiento, de manera que el modelo relacional no se adapta adecuadamente a este nuevo contexto. Es por esta razón que en los últimos años han ido surgiendo un conjunto de nuevos modelos de persistencia diferentes al relacional, concebidos para dar respuesta a estas nuevas necesidades propias del *big data*. Estos modelos reciben el nombre genérico de bases de datos NoSQL. Este libro se centra en uno de estos modelos, el denominado orientado a familias de columnas. Para explicar los conceptos fundamentales se usa Apache Cassandra. Esta base de datos presenta como una de sus ventajas, el mantener cierto parecido conceptual con las bases de datos relacionales (por ejemplo el lenguaje de consultas es muy parecido al lenguaje SQL). Es por ello que resulta muy atractivo a to-

das aquellas personas que proceden del mundo relacional dado que les resulta un entorno familiar.

El libro se estructura de la siguiente manera. En el primer capítulo se realiza una introducción al fenómeno *big data*. A continuación, en el capítulo 2 se presentan las limitaciones de las bases de datos relacionales. En el capítulo 3 se realiza una introducción a las bases de datos NoSQL. En el capítulo 4 se presentan las características principales de la base de datos Apache Cassandra. El capítulo 5 se dedica a describir el modelo de datos de Apache Cassandra. A continuación en el capítulo 6 se presenta como manipular Apache Cassandra desde Python. Por último en el capítulo 7, se presenta algunos problemas de aplicación resueltos.

Espero que el libro sea ameno al lector, y que cumpla su objetivo de servir de medio de acercamiento a este nuevo ámbito de las bases de datos NoSQL y en particular a Apache Cassandra.

Madrid, 11 de julio de 2018
Antonio Sarasa Cabezuelo