THE PETABYTE CHALLENGE:
2011 IOUG DATABASE GROWTH SURVEY

By Joseph McKendrick, Research Analyst
Produced by Unisphere Research, A Division of Information Today, Inc.
August 2011

Sponsored by **ORACLE**®

Produced by **UNISPHERE**® **RESEARCH**
Delivering Certainty
Thomas J. Wilson, President

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In this day and age, for many organizations, data is not only crossing into the hundreds of terabytes, but into the near-petabyte (PB) and multi-petabyte range.

Data is streaming into, out of, and through enterprises from a dizzying array of sources—transactions, remote devices, partner sites, websites, and non-stop user-generated content. Not only are the data stores resulting from this information driving enterprise data stores—both in core, mission-critical databases as well as other environments—to scale into the terabyte and petabyte range, but they occur in a multitude of formats, ranging from traditional structured, relational data to message documents, graphics, videos and audio files.

In a new survey conducted by Unisphere Research among members of the Independent Oracle Users Group (IOUG), close to one out of ten respondents report that the total amount of online (disk-resident) data they manage today—taking into account all clones, snapshots, replicas and backups—tops a petabyte.

The IOUG survey, conducted in partnership with Oracle Corporation, included input from 611 data managers and professionals. Respondents to the survey have a variety of job roles and represent a wide range of company types, sizes, and industry verticals. The greatest number of respondents have the title of database administrator, followed by director or manager. Close to one-third come from very large organizations with more than 10,000 employees. The largest industry segments in this survey are represented by software and tech companies, government agencies, financial services, healthcare, and utilities and telecommunications companies. (See Figures 50-52 at the end of this report.)

Key findings:
The following summarizes the survey results, which explore issues and solutions around managing fast-growing database environments. Key highlights and findings include:

- Almost all respondents report data growth over the past year and one-third of respondents report the amount of data within their enterprises grew by 25% or more in this time period. Almost one out of ten sites now has data stores in the petabyte range.

- A number of companies are compelled to preserve data for extended periods of time, e.g., to meet compliance requirements. As a result, more data is being kept online for longer periods of time—which increases storage costs. In fact, 12% of respondents say they simply now hang on to all data "forever."

- Many respondents report increasing issues in the performance of their applications as a result of data growth. However, many still look to hardware—additional server and storage systems—as the way to handle prolific, near-petabyte or multi-petabyte data.

- As data grows, the reflex reaction by most organizations is to buy and install more disk storage. Smart approaches are on the horizon, but still only prevalent among a minority of companies. Close to one-third now embrace tiered storage strategies, and only one out of five is putting information lifecycle strategies into place to better and more cost-effectively manage their data.

- More than one-third of respondents report they manage most of their company's information—including all information types, such as text, video, or audio—within core enterprise databases.

- Data managers in the survey are struggling with rapid data growth, but few have control over the storage technologies used to manage this growth. In many cases, those respondents "close to the ground" in data sites—such as DBAs—do not have a great awareness of accumulated or projected storage costs.

For many survey respondents, the surge of near-petabyte data environments is dramatically changing the information management landscape. As one respondent, a DBA with a large financial services firm, put it: "I would not say that 'Big Data' has made it more difficult, but we have to think and plan carefully before implementing any new strategy because the impact of any decision related to this volume of data will certainly be huge."

Part of this new reality includes a need for more comprehensive training and education. "We are trying to educate our employees about the correct way of writing, uploading, updating and/or indexing our data at our servers and database machines with the purpose of easing access to other employees," says another respondent, a development manager at a small high-tech firm.