State of the Graph Database

Sponsored by Neo4j
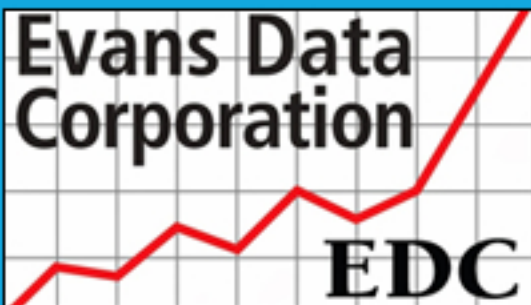
Highlights: 2016 State of the Graph Report

In the fall of 2015, Neo4j approached and commissioned Evans Data Corp to conduct a Q4 2015 survey of their graph database users in order to understand their behaviors, successes, and needs. Neo4j recruited and provided the respondents who were:

- individuals who downloaded O'Reilly's free Graph Databases book, courtesy of Neo4j;

- individuals who are familiar with Neo4j from attending GraphConnect 2015 in San Francisco;

- individuals who follow Neo4j on social networks or subscribe to Neo4j's newsletter.

This study of Neo4j users and familiars found that among those users, performance is the core consideration surrounding their graph database use. This and other findings are explored in this paper.

Performance Through Index-Free Adjacency

Native graph databases afford efficient querying through index-free adjacency, an architecture principle that requires connected nodes to physically point to each other. These adjacent nodes improve performance by localizing the search to include only those nodes that queries are designed to explore, allowing for quicker responses to queries. To this end, 57% of graph database developers in this survey typically use a native property graph model, wherein the data is already optimized for graph traversal.

Of course, graph databases are about more than efficient data processing; graph databases allow end users to get the precise data they need, and then explore the context surrounding their data queries.

"...the ability to model and query complex facts without the need to hire a team of SQL professionals,"

Data Scientist
Germany

Top Use Cases for Graph Databases

Graphs and graph databases are everywhere, from logistics applications that catalog inventories and sales, to applications that monitor financial activity in order to identify cases of identity theft and fraud. The three most common use cases for graph database users today are graph based search, social network applications, and master data analysis, but many use cases are poised to grow. We expect real-time recommendations and software analytics, in particular, to grow over the next two years.

Graph-based search is the most common activity that users perform with their graph databases, using search to quickly find accurate and relevant content without specific constraints of data type. This is a relatively new functionality that is made possible by flexible data structures, but it is already in use across industries that range from software development and IT to education and healthcare.

Graph-based search can be relevant in a variety of different departments, but it is particularly relevant within R&D departments. Graph-based search is a flexible use case that allows organizations to manage their intellectual property, and also allows consumers to find the content that they want and need. This flexibility, another

core consideration for graph database users and developers alike, allows organizations to quickly change their queries, and even their data structure, to fit their business needs. This survey found that 59% of Neo4j respondents currently perform graph-based search, and this functionality is prevalent in small and large organizations alike.

Graph-based search allows end users to quickly understand the commonalities between their customers, their assets, and their target market, but this is not the only function that graph databases allow. Social network applications allow end users to explore new interests and connect and collaborate with colleagues, and provide a better experience for based on social interactions.

"The data is a natural fit for a connect-ed graph so it will give us tremendous competitive advantage over the other vendors in the industry ."

Data Scientist
Dublin, CA

Which of the following do you typically perform with your graph database?
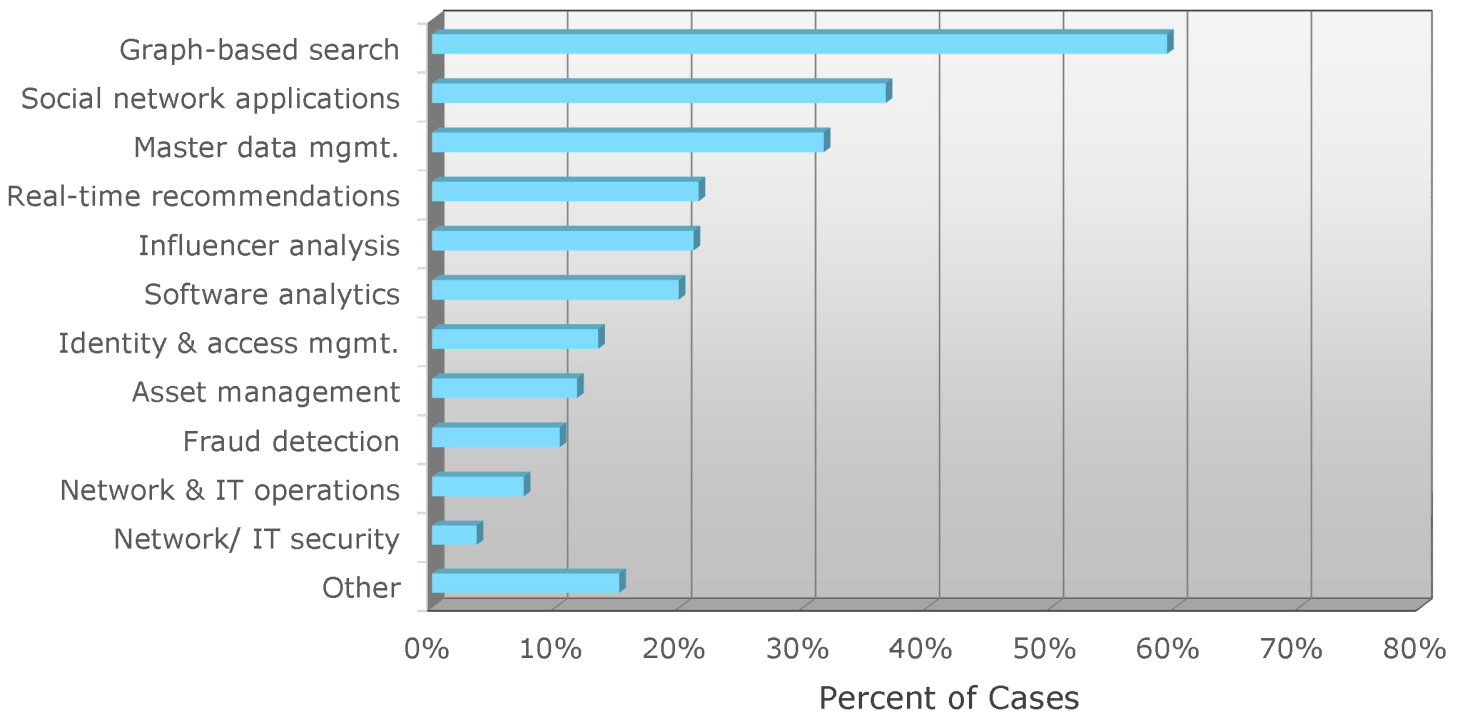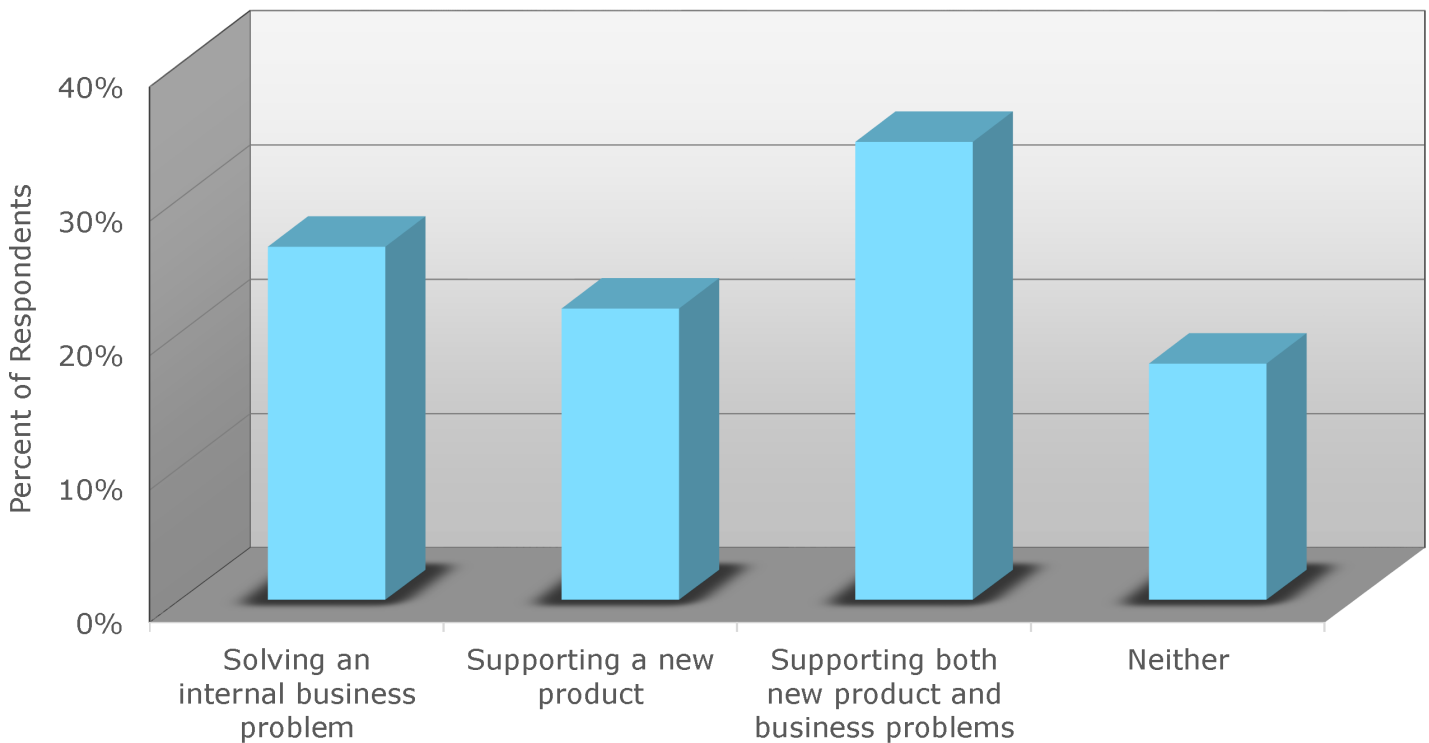

Neo4j State of the Graph Market Survey © 2016 Evans Data Corp.

Moreover, graph databases allow users to make connections in ways that they had never been able to before. Are you looking for a Doug Smith who went to school with Janet Moore at A&M? Graph databases make this possible, and allow developers to create applications that make recommendations based on a user's circle of friends. "Graph databases are a natural fit for a social network" an application developer with a Bay Area social media startup states; this developer, along with 37% of Neo4j's sample, uses their graph for developing, targeting, or pulling data from social networks. Social network applications, including collaborative tools and recommendation engines, will also remain a key use case for graph database users moving forward, but real-time recommendations will also be a significant implementation.

From finding new candidates through employment sites to creating the most compelling recommendations for users, recommendation engines allow organizations to leverage the relationships found in their data to provide optimal results for a variety of projects. "The flexibility of solutions developed using graph database will be a huge advantage in terms of delivering value to clients," a Colorado-based IT consultant states. Indeed, through the use of a recommendation engine, a service can pair clients with the end users that are the most likely to turn into return customers. It can also empower customers to select the right products based on their needs and habits. Furthermore, graph databases can also help vendors stock their stores with the most effective inventory. Forty-nine percent of survey respondents anticipate taking on real-time recommendations through graph databases in the next two years, and some specifically peg recommendations as being a big help to their work in that time. A Madrid-based researcher anticipates graph databases as "helping [me] create better recommender systems for flexible data models." The many uses of graph databases can augment all aspects of data management, from modeling through analysis.

Was the need for evaluating or moving to graph databases primarily:

Neo4j State of the Graph Market Survey © 2016 Evans Data Corp.

Graph database users most frequently turn to graph databases to solve a business problem, and four out of five survey respondents perceive new technical capabilities gained from graph databases to be the greatest benefit that they have gathered from graph database use. However, there are many compelling reasons for moving to graph databases, and most have done so for greater ease of modeling relationships. As a Bay Area-based data scientist states, "graphs are the ultimate modeling tool for relationship-based real-world scenarios," which makes graph databases an effective tool for ingesting data and turning it into insight.

Graph databases extend beyond high tech and IT domains, into industries such as education, medicine, and finance/insurance. These and other industries can use graph based search functionality to cast a spotlight on such activities as fraud detection and data security protocols. Fraud detection programs can take it a step further, by observing behavioral patterns, the types of vehicles that claimants drive, the location, and the time of day, or the amount of time between a policy being opened and the incident.   Medical uses of graph databases range from detecting patterns for individual patients to using graphs to understand the spread of contagious diseases. Graph databases can assist in epidemiological efforts through examining commonalities in data yielded from case studies and outbreak investigations alike.

> "Graphs are the ultimate modeling tool for relationship-based real-world scenarios,"
>
> *Data Scientist*
> *Bay Area, CA*

Resources for Graph Database Users

From the performance of native graph processing to the query power of graph-based search and recommendations engines, those on the outside looking in might assume that there's a steep learning curve to graph databases. After all, perceiving data in terms of graph databases when your entire organization has grown up with relational databases requires a change in perspective; however, graph database developers and users alike tend to cite the ease of modeling relationships among their primary reasons for moving to graph databases, and this ease also extends to the languages that they use to query their graph databases.

A German consultant working in the academic field reports that the greatest benefit to graph database use is "the ability to model and query complex facts without the need to hire a team of SQL professionals," which brings graph databases one step closer to the analysts and data scientists who need the data. In order to aid this process, Neo4j provides informational and instructional media, as well as easy to use, intuitive tools, such as the Neo4j database and the Cypher query language.

> "This is the first tool I can [use to] talk [about] this stuff with management, and they get it "
>
> *Data Scientist*
> *Springfield, VA*

The Languages of Graph Development

Virtually any application that provides information from a database requires users to query those databases. However, a lot happens behind the scenes, from data architecting to data ingress, and IT professionals must consider all of the machine-readable language that goes into queries. Do these databases use SQL? SQL has limitations that make some tasks very difficult.

This survey asked Neo4j's respondent set about the ease of use that Cypher, Neo4j's declarative language for its graph database, provides vs SQL. Over half of the respondents rated Cypher as easier than SQL, with an other quarter rating its ease of use as on par with SQL.

Cypher Ease of Use Compared to SQL (App Devs, IT Pros, Architects, Current Cypher Users)
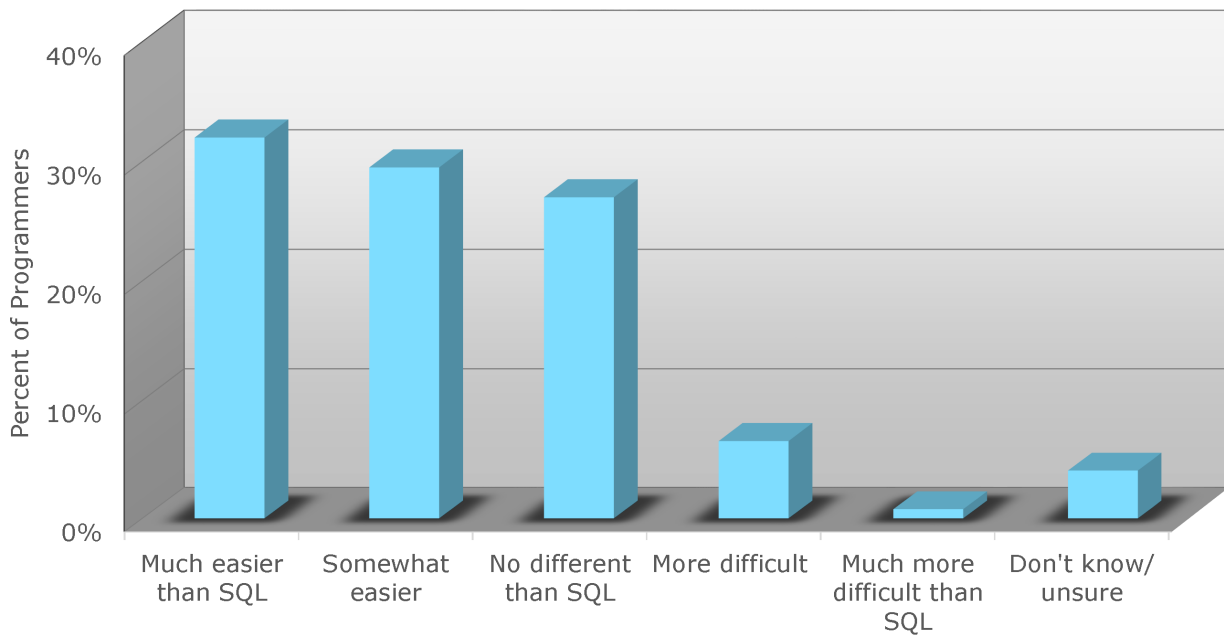
Neo4j State of the Graph Market Survey © 2016 Evans Data Corp.

Particularly in academic institutions and small businesses, where organizations aren't always able to tie up resources in training, an easy to learn, easy to use query language can come as a tremendous boon. Furthermore, knowledge acquisition is not always a rigid and predictable process; data scientists need queries that are as adaptable as they are, and developers expect greater flexibility and ease of use than SQL can provide.

The Benefits of Graph Databases

Many technologies promise performance gains, and graph databases are no different. However, developers and end users alike also value reliability. Having a fast database is essential for delivering actionable insights as they are needed, but those high performing databases also need to be both highly reliable and available. Regardless of industry or company size, downtime inhibits end users' ability to gain quick insights, and ultimately will end up costing their organizations money. As Dr. Jim Webber pointed out in his GraphConnect 2015 keynote, "there is a significant tension between having a system that is available and having a system that is reliable," and you cannot have both. Whereas reliability, highly valued by developers, means that even in the presence of faults, a system returns the right answer, availability guarantees that the system returns some answer. As Dr. Webber states, "for graphs, reliability is far, far, far more important than availability."

Graph databases are about relationships and the means of mapping and illustrating those relationships. When many database users think relationships, they think relational databases, but their queries may be better suited for graph databases, instead. Relational databases require explicit instructions about how to explore relationships

"(Graph databases) give me a competitive edge over traditional databases."

App Developer
Helsinki, Finland

through inferred connections, but graph databases allow queries and traversals to occur through the data already gathered in the model, and data is not only expressed by figures and factoids, but also by the many relationships that those data points have to other data points.

Nevertheless, graph databases need to interact with the databases around them, and relational databases are among the most common. Sixty-nine percent of Neo4j's sample use relational databases alongside graph databases. However, half of respondents report that a graph database is their primary data store, and only 32% report that graph databases are secondary to a relational database in this respect.

Which have you used alongside graph databases? (top five)
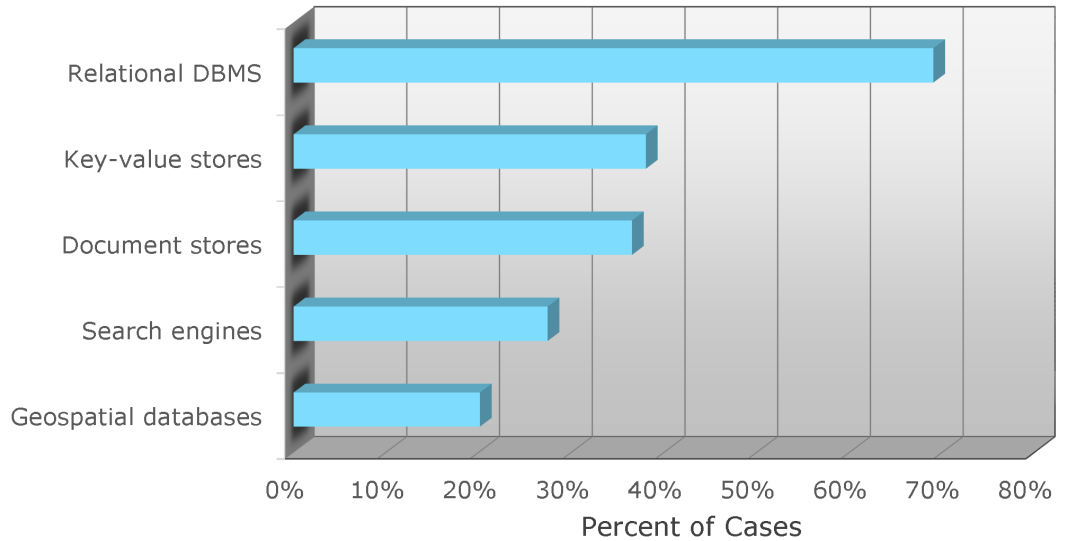
Neo4j State of the Graph Market Survey © 2016 Evans Data Corp.

The database landscape may be changing. In the next two years, one graph database user, working for a utilities company in Brazil, anticipates that his organization will be "[replacing] RDBMS everywhere possible, as graph databases do not suffer (much) from data scaling." While it is true that graph database users also use a variety of databases in addition to their graph databases, it is clear that many think highly enough of graph databases to consider them sufficient for all of their database tasks. Forty-three percent of survey respondents do so, and those working in their organizations' IT and R&D departments are particularly likely to report that graph databases are sufficient for their database tasks.

Another key benefit to graph databases is the flexibility of modeling. Graph databases can handle complex data sets and allow end users to glean key understandings from diverse data. These data sets are not necessarily simple, in fact, the plurality of respondents turns to graph databases due to their ability to handle data complexity. Another common consideration is the intuitive data modeling that comes with graph database use. In this way, developers can accommodate the growing analytics and IT needs of their own organizations.

How do you foresee graph databases helping you in your work over the next two years?

"Primarily in modeling the data in a flexible way, so small changes allow new or unexpected categories of entries to be recorded."

App Developer, independent
Sidney, MT

Complementary Technologies To Graph Database Development

When it comes to analyzing and visualizing data generated from graph database projects, the plurality of respondents report they use / have used off the shelf tools without customizations, whereas only one in ten respondents have used off-the-shelf tools with some form of in-house or third-party customization.

Graph data users work on graph database projects that are hosted in a variety of places, from their personal desktops to their company's own data centers and the Cloud. Graph databases often provide options for both on-premises and Cloud deployments. When respondents in this survey were asked where they host or intend to host their primary graph database, the majority of respondents intended a host for graph databases that includes their own infrastructure, whether on-premise or at their organization's own remote data center.

Conclusions

Graph databases can be used to address a variety of use cases across a multitude of industries, presenting intuitive means for analysts and data scientists alike to gather actionable knowledge. Graph search is the most widely used implementation, but social media applications and real-time recommendations are also strong. A key benefit to graph databases is the flexibility of modeling and the ability to handle complex data sets to glean key understandings from diverse data.

Neo4j's respondents tend to evaluate their graph databases based on performance and reliability, and most frequently turn to graph databases due to the new technical capabilities that they present. SQL languages are gradually giving way to easier to use languages tailored specifically for graph databases, such as Neo4j's Cypher language, and use of Cloud and visualization tools is becoming more common.

About this Study

The Neo4j 2016: State of the Graph report conducted by Evans Data and commissioned by Neo Technology, reflects a survey of individuals known to Neo4j from GraphConnect, from downloading O'Reilly's Graph Databases book courtesy of Neo4j, or who follow Neo4j on social media such as Twitter, Facebook etc. As such these results reflect the opinions of Neo4j users and familiars, but can not necessarily be projected out to the universe of developers as a whole.

Respondents were required to use graph databases or have used or evaluated graph databases in the past two years; 58% are current graph database users. Respondents were also required to interact with graph databases; 64% do so as developers, architects, or IT professionals. The online survey was conducted in English, with no geographical restrictions; 42% are within North America, another 42% are within EMEA. This survey began fielding on October 21st, in conjunction with Neo4j's GraphConnect, and closed on December 8th. A total of 445 respondents were successfully recruited by Neo4j, yielding a margin of error of 4.65%.

About Neo Technology

Neo Technology is the creator of Neo4j, the world's leading graph database, that brings data relationships to the fore. From companies offering personalized product and service recommendations; to websites adding

social capabilities; to telcos diagnosing network issues; to enterprises reimagining master data, identity, and access models; organizations adopt graph databases as the best way to model, store and query both data and its relationships. Neo Technology researchers pioneered the modern graph database and have been instrumental in bringing the power of the graph to numerous organizations worldwide. Neo4j is the world's leading graph data base. It is an embedded, disk-based, fully transactional Java persistence engine that stores data structured in graphs rather than in tables. A graph (mathematical lingo for a network) is a flexible data structure that allows a more agile and rapid style of development.

About Evans Data Corp

Founded in 1998, Evans Data Corporation was created to fill the demand for market research, market intelligence, and strategic planning in the software development industry. Since then we have become the industry leader in market intelligence focused on all areas of development from software to hardware to mobility.

At EDC we have in-depth and focused experience working with high-tech professionals, and we specialize in conducting market research in the IT and development community. We are experts in analyzing technology trends and attitudes. We know how to speak the language and ask the right questions and more importantly, we can understand the answers.

Evans Data Corp

340 Soquel Avenue
Santa Cruz, CA 95062
Phone: 800-831-3080
www.evansdata.com