Get Data Insights Out into the Open

Expanding to open source data visualization software
Open source data visualization tools are changing how companies consume data, apply advanced analytics and solve business problems. Innovators that build these capabilities into their business intelligence and analytics processes will have more flexibility, interactivity and customization options—leading to deeper insights from myriad data sources.

To take advantage of the explosion of available data, enterprises need to become more open minded about open source software (OSS), which offers publically available source code so that anyone can copy, modify and redistribute it. By 2016, Gartner predicts at least 95 percent of all IT organizations will be leveraging open source tools and methodologies for their software development projects.1

Open source tools are especially disruptive in the space of data processing, analysis and visualization. In the past, companies typically developed custom internal data software to meet specific business purposes. The open source movement turned that model upside down, spreading the development effort across large communities of non-employees. Now startups like Zoomdata are building tools based on open source data visualization and becoming viable competitors.

Data visualization, in particular, has made it easier to extract meaning from big data to drive business intelligence. Instead of getting bogged down in advanced analytics and complex algorithms, executives can focus on the data displayed in graphical formats, which allow them to quickly recognize patterns, understand the story the data is depicting and derive deeper insights. (For more information, see the Accenture Technology Labs point of view, "Accelerating Understanding through Data Visualization."²)

Given the importance of data visualization, it is critical to expand in-house data software development skillsets to include open source data visualization tools such as D3.js, which has an array of JavaScript tools and frameworks built on top. These open source, web-based tools can provide companies with more flexibility, data interactivity and customization options than proprietary vendor data visualization solutions.

At the same time, enterprises must recognize that open source projects are not turnkey solutions. Unlike vendor relationships, there is little to no direct accountability for open source data visualization software. Certain open source platforms also require more advanced technical expertise to use effectively. To produce the best data visualization results, enterprises may find a blended approach works well, using a vendor solution for data exploration while turning to open source platforms for data explanation.
The case for open source software

Advances in proprietary software development have been largely disjointed. Initially, companies assigned large groups of developers to code software for operating internal business functions or for selling as a product. These were long-horizon projects with hefty budgets. As new development approaches took shape, companies increasingly turned to waterfall and agile methodologies to speed the development cycle and reduce the cost. And when the outsourcing trend kicked in, they began allocating software projects to third-party companies to expand the development pool and take advantage of labor arbitrage.

However, the most innovative advancement in software development in the past 20 years has been the open source movement—think of it as a more accessible and affordable development ecosystem. The open source software community includes coders spread across the globe, who use distributed project management tools to coordinate and streamline their development on projects. The most well-known example of these is the Linux operating system.

In recent years, the open source community has been particularly active in the data space with tools created for:

- Data processing—Hadoop and Apache Kafka, Storm, Spark and Samza
- Data analysis—Plyr, Dplyr, Scipy, Pandas and Pig
- Data visualization—D3 and Leaflet

With the ever-increasing importance of data processing, analysis and visualization in day-to-day businesses intelligence and analytics, companies must find ways to add these open source methodologies into their software development cycles. Developing open source skillsets is particularly important in data visualization because it melds specialization with control. Companies can use JavaScript and other web-based frameworks to interact more effectively with various data sources and to customize visual elements. It is also easier to maintain compatibility with emerging standards for the web and to participate in organic platform improvement using open source software.
Benefits of open source data visualization tools

Open source visualization tools are generally more cost effective than proprietary vendor tools while providing greater innovation, adaptability and speed to market. Companies are not dependent on the features the vendor creates, nor limited to the vendor’s software roadmap and timelines. Open source also encourages experimentation. Whether it is an enterprise looking for deeper insights from the data lake or a start-up relying on cloud-based infrastructure, companies of any shape and size can experiment with free open source data visualization solutions before committing to a product.

Open source software licenses are also more flexible than vendor licensing models. (See sidebar for a chart of frequently used open source licenses.) The MIT license, for example, has minimal requirements for how the software can be redistributed. This means that if a company uses the MIT license in its internal data visualization software, it will still be able to retain the software’s proprietary nature.

The D3 open source platform, which has been adopted by companies such as The New York Times and Square, owes some of its success to the MIT license because it has contributed to the viral growth of D3 examples. Companies can use these examples to speed data visualization development and then contribute their specialized use case back to the D3 community, creating a multiplier effect.

As shown in Figure 1, a large number of JavaScript frameworks and chart libraries have grown up around and on top of D3, and span functionality from Crossfilter for data manipulation to Leaflet for geographical mapping.
Open source code often evolves through large developer communities that organically develop under the leadership of the project originator, who chooses a licensing scheme for the code.

There are hundreds of different open source licenses in use today, but the majority of open source is licensed under just a few different license types.

Here is a list ranked by percentage of projects where a license is in use.4

<table>
<thead>
<tr>
<th>Rank</th>
<th>License</th>
<th>Percent of projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIT</td>
<td>44.69%</td>
</tr>
<tr>
<td>2</td>
<td>Other</td>
<td>15.68%</td>
</tr>
<tr>
<td>3</td>
<td>GPLv2</td>
<td>12.96%</td>
</tr>
<tr>
<td>4</td>
<td>Apache</td>
<td>11.19%</td>
</tr>
<tr>
<td>5</td>
<td>GPLv3</td>
<td>8.88%</td>
</tr>
<tr>
<td>6</td>
<td>BSD 3-clause</td>
<td>4.53%</td>
</tr>
<tr>
<td>7</td>
<td>Unlicense</td>
<td>1.87%</td>
</tr>
<tr>
<td>8</td>
<td>BSD 2-clause</td>
<td>1.70%</td>
</tr>
<tr>
<td>9</td>
<td>LGPLv3</td>
<td>1.30%</td>
</tr>
<tr>
<td>10</td>
<td>AGPLv3</td>
<td>1.05%</td>
</tr>
</tbody>
</table>
Making open source work in an enterprise setting

Expanding to an open source data visualization platform and building the accompanying skillsets offers many upsides, including flexibility, interactivity and customization. At the same time, companies should be aware of some potential drawbacks of open source and proactively address to get the most benefit from the expansion.

**Increased innovation opportunities**—The creative nature of data visualization and virtually unlimited ways in which the same data can be displayed in different pictorial formats lends itself to open source. (See Figure 2.) Enterprises can easily search for examples and templates created by a global community of data software developers—and share their own designs for others to use. This will not only speed data visualization development time, but may also spark new ideas for how to apply data visualizations in the business.

**Broader compatibility**—Using open source software can enable companies to leapfrog vendor-imposed usability limitations that might pose future issues. For instance, the JavaScript library Leaflet affords compatibility with open web standards such as HTML5, which makes it possible to render on any browser. This kind of compatibility is important for companies looking to engage a global online audience on a wide variety of mobile and desktop platforms. It also enables delivery of the right content, in the right context and at the right point in time across any device. The broad compatibility makes it easier to take advantage of technological improvements as well. If enterprises choose to “fork” or adapt existing Leaflet code for their visualization practices, compatibility with future versions of the framework often requires little to no modifications to production code. However, this widespread compatibility can make the design process more complex because developers must think about how to extend the visualization to multiple devices and resolutions. Technologies such as Bootstrap can help manage this complexity.

**Accessible documentation and support**—Companies might be concerned about the lack of vendor support for open source data visualization technologies. Fortunately, the various development communities post documentation, resources and best practices, as well as keep developers up to date via mailing lists. In some cases, open source software creators actively participate in question and answer forums, too.

**Accountability for code issues**—In general, defect density is lower in open source software than in private software. Unlike a vendor relationship, however, most open source licenses do not give enterprises a clear organization or individual to hold responsible for usability issues or code flaws.

**Required technical expertise**—Open source software development requires highly skilled employees who are familiar with programming languages such as JavaScript. With a fiercely competitive job market for technical skills, it can be difficult for companies to find and retain the right talent. Since many universities adopt cost-effective open sources solutions, it is more likely that companies will be able to find these skillsets in college graduates.

Given these considerations, some enterprises may want to take a blended approach, using vendor-created data visualization tools for certain tasks and open source for others.

---

Figure 2: Innovative examples of open source data visualization.

In this data visualization created for an audience of Security professionals, Accenture leveraged the “onion security model” metaphor to visually narrate the underlying story using industry-specific language.

Accenture used this forced diagram data visualization to help a finance organization see and think differently about how clusters of its portfolio changed in response to external market conditions.

Copyright © 2015 Accenture. All rights reserved.
Matching the data visualization tool to the need

Traditional business intelligence vendor tools are generally good for producing regulatory reports, such as financial metrics. Companies that need to run hundreds or thousands of finance reports every month—perhaps with bar charts configured with analytics that are widely understood and accepted—should continue to use those vendor tools for the purpose.

New data visualization vendor tools are more useful for open-ended data exploration and self-services. Enterprises can use these tools to find fresh value in available data, identify patterns, generate initial insights or build reporting dashboards. It should be noted that some data visualization vendors are responding to the open source movement by adding extensions to support open source tools or by adding storytelling functionality into their software.

Open source data visualization software is best suited for data explanation that requires a more complex view or nuance of the data that can only be achieved with precise visualizations, such as every pixel representing a data point.

Open source is also the right tool for creating highly interactive experiences that allow large internal or external audiences to manipulate the data and make their own discoveries. For example, the New York Times publishes interactive data visualizations to illustrate some of its articles. Using simple narration features, the Times walks readers through the key data points or storyline. Readers can then further explore the information using interactive functionality, including tabs to display different views of the data and tooltips to show details of selected data.

Furthermore, the customizable user interaction capabilities make it possible to showcase the advanced analytics behind certain visualizations, which makes the business insights more accessible and pervasive across an organization at all levels. For example, Figure 3 depicts a state-the-art graph sampling algorithm utilizing empirical probabilities and transition information to the create a smaller graph of uniform sampling (green) across multiple probability densities present in the original graph (blue) and aggregated in the depicted graph.

In some cases, companies will opt to combine new vendor data visualization tools with open source software to create a more comprehensive solution. This can take two different forms:

- Tools are used side by side where open source tells a story answering a few questions with a defined data set, and the new vendor platform allows for data discovery and exploration of an entire data set. This approach can help an audience use the open source software to explore the data and ask questions they might not originally have asked, and then leverage the vendor tool to dig deeper for answers to the questions.

- Tools are used for their best purpose with open source data visualization software sitting on top of a new vendor tool. The vendor software manipulates the data and completes the analytics while the open source software is used for the visualization.

Figure 3: Open source data visualizations can be configured to show the underlying analytics in support of a guided exploration of the data.

Copyright © 2015 Accenture. All rights reserved.
Use this chart to help determine where and how to implement open source data visualization skillsets and tools across the company.

**Business demand and need**

Catalog how data is currently being collected in the organization to drive business decisions. Understand how having more data insights about daily operations could improve business outcomes.

Identify data silos within the organization (e.g., HR, finance, engineering), along with corresponding data needs that are currently unmet across the business.

Define and prioritize the ways in which the organization engages with data. Brainstorm ways to deliver compelling user experiences that offer new insights into their decision-making.

**Skills**

Identify any existing open source efforts or areas where company might be recruiting new talent with open source skillsets.

Consider assets from the open source software community that are usable for core IT functions and begin planning how to integrate them.

Develop a strategy for identifying which workers to skill-up.

**Tools and workflow**

Design and implement a pilot to leverage the selected open source data visualization tool.

Pilot a few use cases and measure usability across each.

Based on pilot results, create a roadmap for partnering with the business to build and deliver high priority visualizations.
Conclusion

With the increasing importance of data visualization in business, enterprises should evaluate and expand their data software skillsets to include open source data visualization tools. Using these flexible and cost-effective tools, in combination with new vendor data visualization tools, provides a more robust platform for both data exploration and explanation—and ultimately deeper business insights.
Contact us

Nathan Shetterley
nathan.shetterley@accenture.com

Acknowledgements

We’d like to thank Diana Munoz, Jeffrey Catania and Nikhil Mulani for contributing to this paper.

References


About Accenture

Accenture is a global management consulting, technology services and outsourcing company, with more than 323,000 people serving clients in more than 120 countries. Combining unparalleled experience, comprehensive capabilities across all industries and business functions, and extensive research on the world’s most successful companies, Accenture collaborates with clients to help them become high-performance businesses and governments. The company generated net revenues of US$30.0 billion for the fiscal year ended Aug. 31, 2014. Its home page is www.accenture.com.

About Accenture Technology Labs

Accenture Technology Labs, the dedicated technology research and development (R&D) organization within Accenture, has been turning technology innovation into business results for more than 20 years. Our R&D team explores new and emerging technologies to create a vision of how technology will shape the future and invent the next wave of cutting-edge business solutions. Working closely with Accenture’s global network of specialists, Accenture Technology Labs help clients innovate to achieve high performance. The Labs are located in Silicon Valley, California; Sophia Antipolis, France; Arlington, Virginia; Beijing, China and Bangalore, India. For more information, please visit www.accenture.com/technologylabs.