

big data

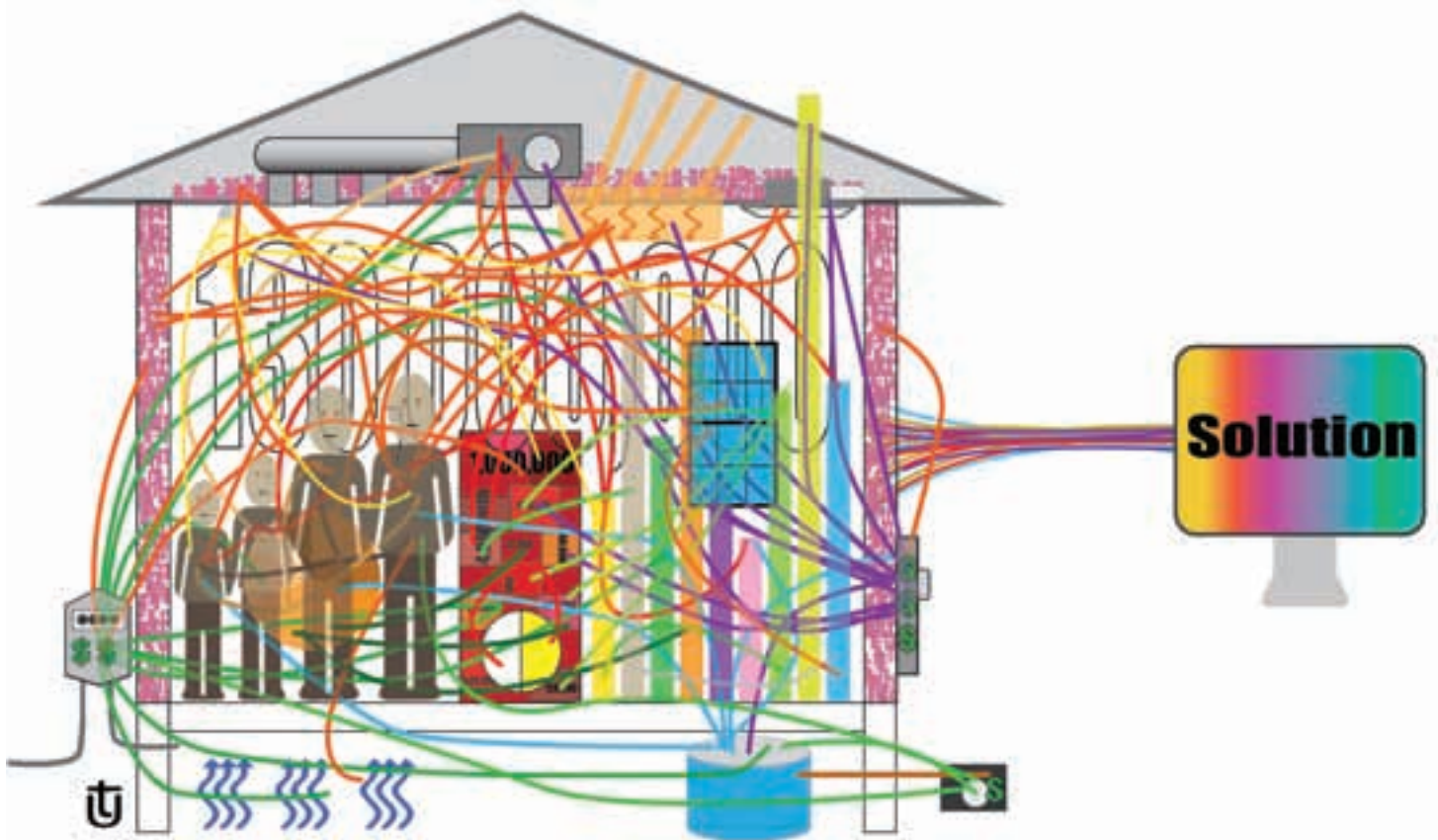
Big Data Offers Big Solutions

We see great opportunities with the use of big data in the home performance industry. This topic is complex, so we've sorted through some of the details, identified key points, and described ways in which we believe this technology relates to the home performance industry.

The idea of big data sometimes gets a bad rap, mostly due to concerns about privacy. We believe that discussion about privacy is very important, but in this article we set our sights on the opportunities available to the industry. We're interested in ways we can use these large data sets to improve the performance of home performance.

Synergizing Existing Sets of Data

We are already immersed in the collection and use of large amounts of data. We now have the opportunity to apply new systems of analysis to these existing data sets to create solutions that were not previously available, and in some cases, we can now synergize data analysis across industries. Large data sets, for example, are the heart of any Customer Relationship Management system (CRM). They're embedded in existing Quality Assurance (QA) systems. They're widely utilized by the real estate industry. Some organizations have large amounts of latent data—the content—but not the technology needed to manage it



and synchronize it with modern systems. A few forward-looking organizations have identified and seized the true power of big data to solve problems. But this much is clear -- for the home performance industry, big data provides important new tools by both capturing information, and by connecting parties who have until now worked separately.

And we have plenty of data to work with. We're not just talking about online search trends, or web browser histories. No, we have a lot of hard data from diagnostic tests (how many homes in a neighborhood have back-drafting water heaters?), records of materials installed (how many bags of insulation does it take to fill the walls of a 1000-square-foot home?), and observations made by people who live in these buildings (how often has the crawl space flooded over the last 50 years?).

(con't page 26)



How This Can Super-Charge Home Performance

Imagine how these inquires—these comparisons among data sets—could synergize the home performance industry. Most of these already exist in limited forms, but not with sufficient volume and connectivity among industries to be truly useful.

- Couple historical utility consumption with local energy code requirements to evaluate the effectiveness of these design guidelines.
- Track the value and marketability of efficient properties in the real estate market.
- Collect data generated by the Weatherization Assistance Program to track the non-energy benefits in effect in local economies.
- Cross reference data from related industries such as roofing, HVAC, and pest control.
- Cross-reference the academic performance of school kids with the attributes of their homes.
- Evaluate workplace records to compare the performance of workers with varying levels of natural light and indoor air quality.
- Evaluate the air-leakage rate of the homes of asthma patients, allowing caregivers to prescribe home ventilation or pollutant control as a medical intervention.

(con't from page 25)



These Four Vs Define Big Data

Volume, or scale

Volume of data is the V that causes the most controversy. Privacy advocates express concern about sharing personal information. Business interests imagine the new knowledge that will be created. We all compare the convenience of intuitive design against the intrusion of ubiquitous technology.

The home performance industry is, and should be, right in the middle of this discussion. The data we need is far-flung across the homes of the world, and gathering it will allow us to make more intelligent decisions. Yet that data, because it comes from within people's homes, and is inherently private. We'll need to achieve a balance in how we apply big data to home performance.

Velocity, or analysis

Advanced algorithms drive the analysis of big data sets. These sophisticated rules define the calculations and problem-solving operations. Algorithms running at a high velocity are the

difference between a big bunch of data like a huge Excel spreadsheet, and true Big Data applications.

Velocity is the part of big data that most of us take for granted. Velocity is the V that makes big data accessible and useful. It's also the aspect that will increase the most quickly over time.

Variety, or different forms

Variety stems from velocity and volume. When you have a big bunch of data that's easily accessible, you have options. You can perform advanced searches and you can establish large groups of metrics.

So you might say big data is really is data management on steroids. But it's not just about managing data -- it's also about gathering all kinds of data. Bigger data sets with less complex analysis provides better results. And the opposite is also true -- an in-depth and complex analysis of a small data set is often not statistically accurate.

Smart organizations expand their traditional types of data without extra effort. They extract it from social media data, browser logs, text analytics, and sensor data to get a more complete picture of their data set. The big objective, in many cases, is to create predictive models, and these are the key to making big data useful. But it takes creative, savvy people to identify and design predictive models.

Veracity, or uncertainty

Veracity is a V that poses a big problems for people creating predictive models. When you amass big bunches of data from everywhere, odds are good there will be a lot of bad data. As an example: try a Google search for your name and see how many pictures of other people show up. Bad data is one pitfall of big data. But time, and intelligent analysis, will cure this problem in most cases. In the world of home performance, there will always be a need for industry-wise people evaluate and confirm the veracity of information.

cloud data

In the world of home performance, access to field data is key. One way to achieve this is by storing all the data in a common repository. Retrotec has recently unveiled such a solution, known as rCloud.

Retrotec engineer Ben Walker first described the concept of storing home performance data in the cloud, adding geo-tags to the reports, and accessing the report through a user interface that tags test locations on a map. Their recently unveiled product does include this geographic front end. It captures results of blower door and duct tests, and other diagnostic information. Weather data, and age and square footage of the home are gathered from Zillow's database of existing homes. Users can add photographs, and can set access privileges for each report.

The system is accessed through an app that connects with the Retrotec DM32 wireless gauge. Users of smart devices will be able to add a geo-tag and date-stamp for the test. The use of wireless devices, with seamless transfer of data, is a key part of these cloud-based solutions.

Cloud-based systems such as rCloud will deliver important benefits to users, their organizations, and the industry. There will be less need for manual data entry. Data integrity will be improved. Calibration certificates for test equipment will be linked directly to the tests. Quality assurance professionals, HERS providers, and other program managers will benefit from having verified test data.

We look forward to seeing how this system, and others, will roll out. Retrotec has taken an important first step by capturing all these advantages in a single system. We anticipate that other manufacturers will join the movement toward creating platforms for the storage and management of data. We hope that one of them will develop a platform that can manage data from a variety of instruments made by a variety of manufacturers. Then we will have true universality of data.

Summary

The home performance industry, with its deep reliance upon accurate field data, will derive great benefits from the intelligent application of big data. But we'll be required to broker intelligent and realistic discussions among all stakeholders to identify the most relevant, advantageous, and fair uses of this new knowledge.

