From: www.cio.com

When Your CRM System Passes 1 Million Records

– David Taber, CIO

March 02, 2010

There isn't a sales force in the world that says it has enough Leads. And you won't find many marketing VPs who want to do fewer campaigns. So there's a never-ending stream of new leads, prospect interactions, and conversations to be stored in the CRM system. At companies in consumer markets, open source software, and other categories it's not unusual to find a million leads or more. But that's just the beginning: if you're using the latest marketing automation system, every e-mail, web download, and prospect response is recorded in the CRM system. And if you have a large call center, every call and e-mail exchange should be recorded well.

[For timely data center news and expert advice on data center strategy, see CIO.com's Data Center Drilldown section.]

This can mean millions of records, with thousands of new ones every day. No other system in most enterprises needs to deal with this kind of data flow, particularly with the number of simultaneous users that a CRM may have. If you're growing the scope of your current system, or consolidating several division-level CRM systems, here are things to consider as the data scales upward.

CRM Definition and Solutions

Platform Performance

Nearly any real CRM system will have reasonable performance for most use cases even with data sets of this size. But there are some features that will inevitably start to slow down: reports, dashboards, or any functionality that involves scanning all records. While some of the problem can be solved with the usual workarounds (de-normalizing data, creating analytic digests during off-hours, pre-joining views, etc.), anything that involves an ad-hoc select or upsert can't fundamentally be helped without throwing hardware or coders at the problem.

User Interface Performance

Under most conditions, the responsiveness of the CRM user interface will be fine with large data sets because users are typically working with only one record at a time. But if your CRM system runs inside a browser, watch out for UI features that loading of large lists for the user. Some systems may try to load thousands of names into a scrolling list, with appalling performance consequences (particularly on Firefox).

Third Party Apps

While the core CRM platform will probably handle really big data sets, we've seen cases where supposedly enterprise-scalable third party apps couldn't even load the data, let alone process it properly. The combinatorial explosion that occurs within some applications' algorithms can lead to corner cases that are tough to anticipate. As always, run pilots with the complete data set before you put third-party apps into production.

Duplicates and Corruption

Both of these data quality issues can happen in any size company, but they are magnified with every new system that feeds into your CRM. Since the really big data sets go hand in hand with external integrations, de-duping and data cleansing are an absolute must with large scale CRM. Of course you'll be using tools on a regular basis to keep these issues at bay, but the tools will need to have several parametric controls and thresholds for their matching and cleaning algorithms. Your

administrative staff will need to develop specific sequences of cleaning and deduping passes, and maintain logs of the settings for each pass. Over time, they'll discover patterns in data quality problems that will provide clues about the underlying causes and possible fixes in the CRM system.

Backup and Archive Policy

Most CRM systems can be configured to run an online backup, and most SaaS CRM systems do this for you. But recovering data from a purely incremental backup is in no way straightforward, and can be quite the project. So having a periodic baseline backup of your own is really a necessity for a CRM system of any size.

The problem, of course, is how much data you can pull out of the system during the backup window. For almost any company, the CRM system can be quiesced for several hours each weekend. So you should be able to pull dozens of gigabytes out of the system, more than sufficient for all the parametric records. But you won't have enough time to pull over all the attached documents and e-mail threads. Consequently, you'll need to develop a backup data partitioning policy that fits with the way your business works.

In addition, you'll need to develop a detailed policy about when and how you move data into archival storage. I have yet to find a CRM system that knows what to do with nearline or offline storage, and your archival policy will need to take analytics tools into account as well. As a matter of policy, I can't think of a reason to keep CRM data (even summaries) on line for more than 7 years. But I can think of lots of good reasons to keep at least the last 2 years worth of data transparently available, no matter what your industry. Although these may seem conceptually simple issues, your business requirements can involve a surprising amount of complexity to enforce application-level data integrity, so put some smart business analyst on this problem for a while before you ask for recommendations.

David Taber is the author of the new Prentice Hall book, "Salesforce.com Secrets of Success" and is the CEO of SalesLogistix, a certified Salesforce.com consultancy focused on business process improvement through use of CRM systems. SalesLogistix clients are in North America, Europe, Israel, and India, and David has over 25 years experience in high tech, including 10 years at the VP level or above.

Follow everything from CIO.com on Twitter @CIOonline.

© 2009 CXO Media Inc.