Giant Flying Saucer

…coming to you from the great state of Texas

« Flash Player 10.1 – On-demand HTTP Dynamic Streaming (Tutorial)

Book Review: Amazon SimpleDB Developer Guide

Today I'm going to review "Amazon SimpleDB Developer Guide" by Prabhakar Chaganti and Rich Helms.

SQL Server Query Tool
Improve SQL Server Performance 65%. New Ignite 8, Get a Free Trial Now!

Get Labor from the Cloud
Connect to qualified workers without picking up the phone.

Chapter One: Getting to know SimpleDB
This chapter is a very simple and concise introduction to just what SimpleDB is and how it compares in terminology to more familiar things like the traditional RDMS database or even a spreadsheet. There is some discussion on how you will be interacting with SimpleDB via Java, Python, PHP and even using the Firefox "SDBtool" plugin as well as the S3Fox plugin for interacting with Amazon S3 storage. To understand SimpleDB though you have to first understand the terminology of what a "Domain" is, what and "Item" is, etc. This chapter is probably the best introduction I've seen yet for explaining this.

Chapter Two: Getting Started with SimpleDB
The authors now turn their attention to creating an AWS (Amazon Web Service) account which is critical to interact with SimpleDB, S3, EC2, or any of Amazon's Web Services. The walk through on setting up your account is thorough so you should have no problems following along. Moving into the chapter you get a good introduction on how to install and use the Firefox SDBtool which can make working with SimpleDB much easier.

Chapter two moves into using code to interact with SimpleDB and they introduce the basic coding concepts in Java first. They use the Java SimpleDB open source library "Typica" showing you how to ensure your code will work as you follow along with the book. I guess my first complaint is something I see a lot of books and articles do which is to leave out the "imports". I know this is probably for space considerations but I've pulled my hair out with other books or sample code on the Internet more than once trying to get all my imports right before my code would run. The source code is provided by the publisher (and the PHP code samples can be found here) and of course you can find the Java "imports" in the source code if you have any trouble. The Java examples are very good and simple to understand and you should have a good foundation after trying out the examples for moving forward.

Next on the list came the PHP examples. I was curious as to why the authors chose to use the "SDB-PHP" library by Dan Myers rather than the one supplied by Amazon. I emailed Rich Helms to find out and got a reply back right away. Rich essentially said that the Amazon provided library was spread across several files and not as straight forward as SDB-PHP. Fair enough. Looking over the PHP sample code I'd have to agree

that the SDB-PHP based code is as easy as it gets. Rich worked with Dan to expand the API for SDB-PHP and ended up with a library that had complete SimpleDB functionality. For S3 support they picked the S3-PHP library by Donovan Schonknecht which interestingly enough is what SDB-PHP was based on. So with the PHP library chosen and setup they walk you through the examples used with Java a few pages prior. Now, some people will complain here wondering why they are repeating these examples with Java, PHP, and Python. Typically developers just want to see their own pet language listed and nothing else. Myself though I like to see a mixture of languages and I really like the fact they took their time to show three popular languages and how to use each against SimpleDB.

The Python examples use the "Boto" library originally built by Mitch Garnaat. The Python examples that follow are very quick and easy to work with. A small personal preference here (for me) is I think I would have put the Python examples into individual files and run them that way instead of through the Python Interpreter on the command window, but, I do know many developers would rather tinker with Python code this way than with files, so I guess it comes down to personal preference.

Overall you basically get a lot of solid information in chapter two especially if you've never work with SimpleDB. I think they did a great job in showing how simple the code is to work with SimpleDB.

Chapter Three: SimpleDB versus RDBMS
This is great chapter if your use to working with traditional databases (Oracle, SQL Server, etc.). You learn here why some traditional databases might not be a good fit for some projects anymore and how SimpleDB can alleviate that issue. You can also learn why SimpleDB is not always the solution you might want each and every time. The pros and cons are laid out in this chapter and summed up well. I found the information on "eventual consistency" to be of interest as well as the updated information from the authors on eventual consistency supplied by Amazon in February 2010.

Chapter Four: The SimpleDB Data Model
What I found of great use in this chapter was how Amazon goes about calculating the costs of your SimpleDB instance based upon numerous criteria. There is some good discussion and examples on how to go about pulling out metadata and what exactly you can find out from the metadata as well as discussion on the constraints for domains, items and attributes.

Chapter Five: Data Types
This chapter starts to get deeper on how SimpleDB works and stores your data and how you need to go about saving it as well as retrieving it. Lexicographical comparison is discussed which is essential to know how your comparisons are going to work. This knowledge becomes very important when working with things like numbers which of course is carefully and well explained and fortunately they also show you some examples and you can breathe a sigh of relief that the libraries do the hard work for you already regarding the encoding/decoding. So the tricks on storing and retrieving numbers, dates, booleans, etc is all covered and reading this chapter is critical if you going to be using SimpleDB as this is the foundation you'll need going forward. Its all there in the chapter but you might find yourself like me and reading it over a few times since there is a lot of information tossed at you to keep straight and understand. The chapter ends off with some Base64 discussion.

Chapter Six: Querying
Here you start off populating SimpleDB with some music related data with which you will soon query against. You start off with the well known "Select" and move on from there into more complex queries. All the tips and tricks are laid out clearly in tables or done via examples. This whole chapter is laid out well for moving from simple queries onto more advanced queries and I love the hints and tips that roam about the book such as the one found in this chapter regarding "getAttributes uses about 41% of the box usage of the SELECT query". Since your paying for SimpleDB and how much CPU, bandwidth, etc. your using this kind of information is critical to getting the most bang for your buck. This is the kind of thing you probably won't get just from the SimpleDB docs alone or might easily skip over. The chapter ends off with some examples using "getAttributes" which is great if you really only want one item back.

Chapter Seven: Storing Data on S3
As the authors build on the music data structure from the previous chapter and we are challenged in this chapter as to where we want to store our MP3 files and of course the best choice is Amazon's S3. So just like

in chapter one we are given the rundown of what S3 is as well as the terminology used by it as well as the constraints. There is some brief discussion on the S3 pricing scheme which is helpful since we always need to be aware that almost everything with Amazon Web Services costs us money.

The chapter defines S3's terminology like buckets, objects, and keys and then moves on how to create an S3 bucket as well as metadata. From here we are given some sample code now to upload the MP3 files up to S3 as well as retrieving it back. This code is surprising simple and as usual everything is explained.

Chapter Eight: Tuning and Usage Costs
Here is where we begin to get a little deeper into how to tune our code as well as what this is all going to cost us in the end. "BoxUsage" (the amount of system resources that were utilized but not bandwidth or storage fees) is explained. The authors do show the current costs for these services but as we all know these fees are subject to change at any time so make sure to always refer to the AWS website for the latest pricing. They give a nice introduction into how to go about getting your usage report from Amazon as well.

The sample code shows you how BoxUsage information can be retrieved and what you can do with this information as well as some tips on how best to optimize your code. The chapter ends with a brief discussion on "Partitioning" which only if your planning to hold a lot of data (10 GB or more) you probably won't ever need to worry about this.

Chapter Nine: Caching
Nowadays caching for most high traffic websites comes down to using something like memcached. Not only do the authors show you how to use memcached they quickly explain that doing so will save you money if you do use SimpleDB in future projects. This book scores high in the sense "the bottom line $$$" is always discussed and the author's concerns to keep your costs to a minimum via such tips and are very welcome especially if your new to SimpleDB and haven't yet learned which API calls are more expensive and which are not.

So we are introduced to memcached and how to install it on a Linux system. It would have also been nice had they mentioned there are Windows ports of memcached (or better yet Microsoft's Project Velocity) however I think many readers of this book are probably be using Linux based servers. The chapter closes off with some good samples of how to use memcache (and Cache_Lite) with SimpleDB.

Chapter Ten: Parallel Processing
The final chapter is full of examples that show you how to make use of batching up your requests including a good Java threaded example that I found very interesting. The chapter ends off with a few "gotchas" you need to be aware of as well as some discussion on each one. Again, another example of the authors passing on their experience onto the reader.

So overall I'm impressed with the book. The few minor issues I found with the book are just that: minor. The vast amount of information squeezed into 235 pages (my eBook edition) and code examples outweigh any of my nitpicks. I think even if your not using PHP, Python or Java to interact with SimpleDB you'll still get plenty of good solid information out of the book and you'll see that most other languages and libraries you will use to interact with SimpleDB follow a similar path as the code examples in the book.

If your going to work with SimpleDB or are simply curious then I'd suggest this book. Its recently published (June 2010) at the time of this blog post and the most up to date information is all inside.

Good luck with your SimpleDB projects!

www.microsoft.com Ads by Google

- Bookmark on Delicious
- Digg this post
- Recommend on Facebook
- Buzz it up
- share via Reddit
- Share with Stumblers
- Tweet about it
- Subscribe to the comments on this post

This entry was posted on Sunday, June 27th, 2010 at 9:48 pm and is filed under Amazon Web Services, Java, PHP, Python, Reviews. You can follow any responses to this entry through the RSS 2.0 feed. You can leave a response, or trackback from your own site.

One Response to "Book Review: Amazon SimpleDB Developer Guide"

1. *Webmaster in Residence » First Review of my SimpleDB Book* says:
   June 28, 2010 at 6:04 am

   [...] http://giantflyingsaucer.com/blog/?p=1225 No Comments [...]

Leave a Reply

Name

Mail (will not be published)

Website

Submit Comment

Giant Flying Saucer is proudly powered by WordPress
Entries (RSS) and Comments (RSS).

- follow:

- RSS