Paper SAS1684-2018

Adopt a Pet (Elephant?) – Are You Enjoying You're Apache Hadoop Investment?

Rex Pruitt, SAS® Institute, Indian Trail, NC

ABSTRACT

Hundreds of companies have embraced Apache Hadoop as a component of their data storage and analytics strategy. But they have yet to realize the return on investment (ROI). Many have been focused on the savings from a move to cheaper hardware and software infrastructure. Others thought having large volumes of unstructured distributed data would lend itself to new and innovative ways to analyze that data. Their hope is to drive out valuable business insights, resulting in better decisions for the company.

Seriously, have companies really just gone to an adopt-a-pet strategy to save a few bucks?

What are you doing to achieve "full adoption?" Come and learn the path to success!

INTRODUCTION

In this paper, you will learn about materials specific to Hadoop (an open-source platform) and how to be successful in delivering value from that environment. The basic flow is as follows:

1. Hadoop overview – What is it? Where do I put it? Why would I want it?
    - Quick explanation of the Hadoop Distributed File System (HDFS)
    - High-level architecture picture
    - ROI considerations
2. I've bought it and installed it. Now what do I do with it?
    - Where does SAS® software fit?
3. Do I really save and/or make money with Hadoop?
    - Path to success: Leveraging SAS to achieve quick success
    - Moving data and advanced analytics
    - Adoption stories

The conclusion refers to additional material where you can continue your educational process and/or increase your knowledge of the Hadoop environment and the data it contains.

HADOOP OVERVIEW – ADOPTION (WHAT? WHERE? WHY?)

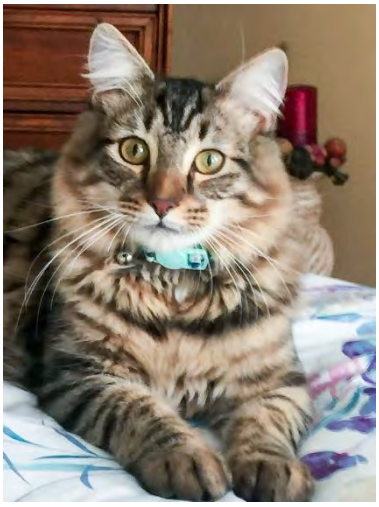
Personally, I've had my share of experience with adoption (pet or technology…LOL!). From a human standpoint, we are all built with an affinity for adoption driven by compassion. Recently, we (my wife and daughter particularly) acquired a kitten that had lost its mother to vicious dogs. The kitten was feral and extremely malnourished. Of course, stricken with compassion, we decided to adopt A pet kitten with minimal knowledge of the future expense or challenges. The deal was sealed when my daughter named the kitten Mordecai. Naming something generally establishes ownership, commitment, and accountability. (See Displays 1 and 2.)

Display 1. Mordecai 1 Day After Being Found

Display 2. Mordecai 3 Weeks After Being Found

Subsequently, over a very stressful period that included long hours (mostly by my wife, Kathy) of tender care and a few costly trips to the pet clinic, we **achieved successful adoption**. Of course, compassion was the underlying driver, but we only achieved success by **executing on our plan** to help this helpless kitten survive and live a healthy and satisfying life. (See Displays 3 and 4.)

Display 3. Mordecai 3 Months After Being Found



Display 4. Mordecai 6 Months After Being Found

So, why the pet adoption metaphor? Well, let's see if we can draw a few correlations that are relevant to Hadoop. (See Table 1.)

Condition/Commitment	Feral Kitten Adoption	Hadoop Adoption
Feral (Wild, Untamed, Uncontrolled)	No mother, living in wild	Open Source, No real oversight, Managed by Community
Acquisition	Free – Just pick it up…	Free – Just download if from GitHub
Implementation	Integration with other pets	Integration with other systems
Maintenance	Pet Clinic Expenses	Version changes, security, & stability
Education	Dr. Google	Dr. Google & GitHub
Usage	Patience, Play & Enjoy	Patience, Play, Develop, & Enjoy
Adoption	Healthy Family Contributions (Hopefully)	Production Contributions (Hopefully)

Table 1. Correlations between Adopting a Feral Kitten and Installing a Hadoop Cluster

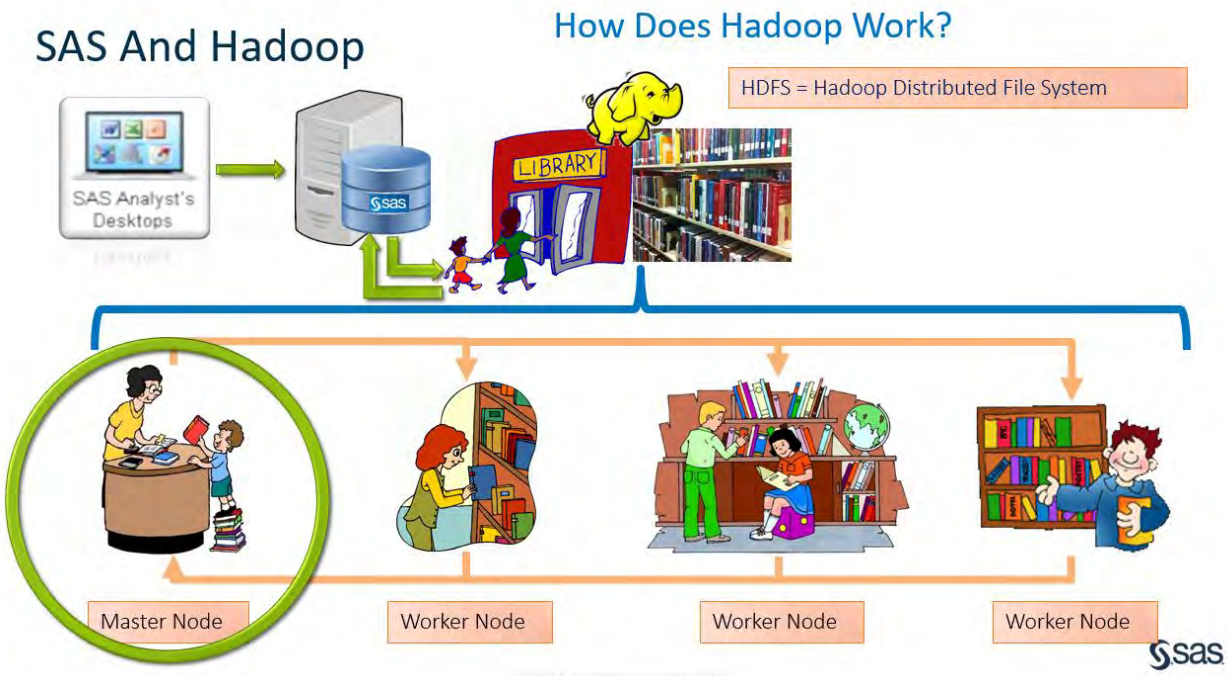
As promised in the outline of this paper, I hope to provide a relatively simple overview of Hadoop and how it works. There will be some architecture discussion because that is very important in understanding the nuances that affect performance. I'm even going to make an attempt at helping Hadoop data users with the overall justification of leveraging Hadoop interactions in the context of return on investment (ROI).

Once one has an understanding of the architecture and how Hadoop works, the focus will turn toward how SAS interacts with Hadoop, particularly why SAS is a GREAT choice in leveraging Hadoop data. However, in order to be truly successful and save or make money in the adoption of Hadoop, certain best practices must be considered. After the challenges are addressed, the adopted pet elephant (Hadoop) will be relatively satisfying. The data scientists and business analysts will be able to leverage SAS to prepare, visualize, and model (depending on the SAS license) their data in-database and realize the value of the SAS platform's diversity, scale, and trust [see Display 5].

Display 5. SAS Platform's Diversity, Scale, and Trust

QUICK EXPLANATION OF THE HADOOP DISTRIBUTED FILE SYSTEM

Display 6 is a simple metaphor of how Hadoop and SAS work together. Everyone has been in a library at least once in their lifetime (however, the internet has stunted this in recent years). Think of the distributed nature of the information in a library. All books are categorized and numbered generally using a system known to many as the Dewey Decimal System. This system makes it easier for library visitors to retrieve books from and return books to their original place.



Display 6. Metaphorical Depiction Using a Library to Illustration How SAS Interacts with Hadoop

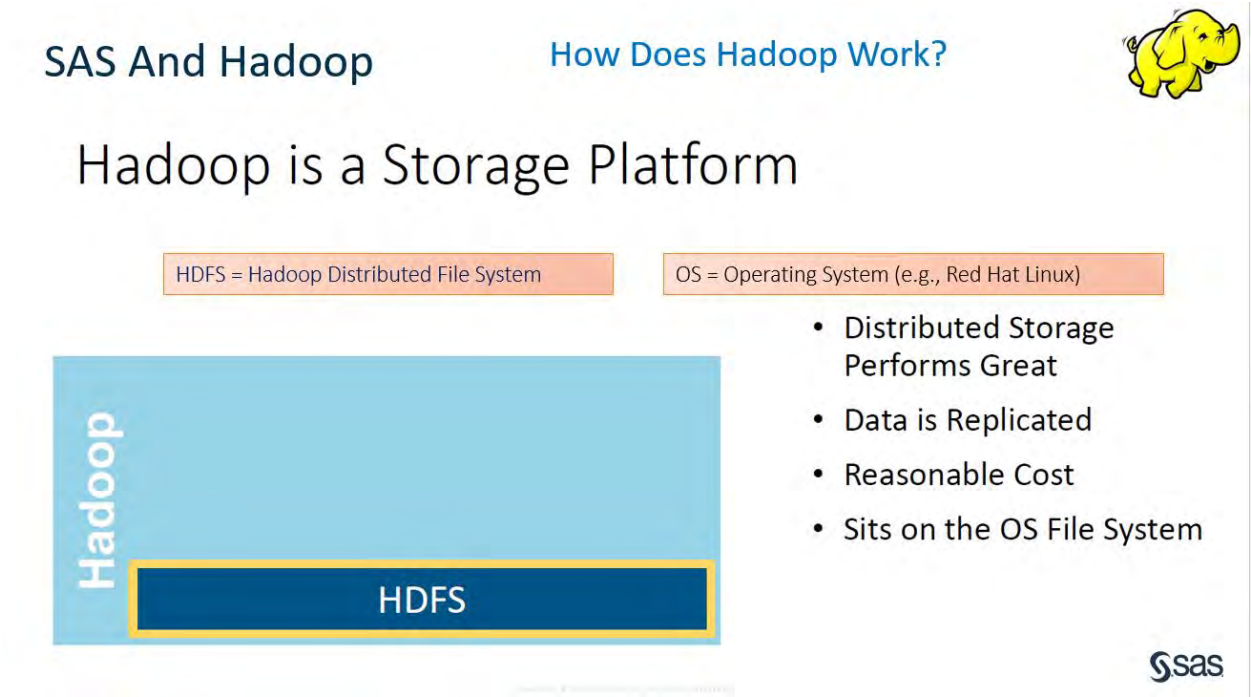
Now think of the library without a system of placement and organization that supports self-service. In order to remove the cost of maintenance, the books are now randomly organized using cheaper shelf space. The only individual in the library who has knowledge of where everything is located happens to be the librarian. On top of that, the librarian has placed three copies of each book on separate shelves in case one of the shelves is broken. In the event of a broken shelf, the librarian can have the workers get you a book from a good shelf. Oh, by the way, the librarian has workers who are allocated to each bookcase to receive the requests for books.

Let's break down the metaphor with respect to Hadoop. The librarian is the master node and the bookshelves are the worker nodes. When companies migrate to the open-source Hadoop platform, they are basically moving to an environment that is distributed across several computer servers or worker nodes that are controlled by one or two master node servers (two in the case of redundancy). The master node controls the mapping of all distributed data in the form of native operating system data files according to the particular system specifications (for example, CSV, ORC, Parquet, and so on). In its functional form, this environment is often referred to as a *Hadoop cluster* or *data lake*. (The architecture components are illustrated and explained in the next section.) I've actually worked with customers who have named their Hadoop cluster very specifically with their purpose in mind. This seals the deal and establishes their ownership, commitment, and accountability toward adoption.

In order for SAS, or any other software, to interact with the data in the Hadoop cluster, certain technical components must be installed and configured properly. This is not a simple easy button methodology. Regardless, if all components are working properly, the SAS user can submit a standard set of Base SAS® code from the client interface, have the code execute in-database, and return the results back to the SAS client interface. In respect to the library metaphor, this is like walking in to the library door and asking the librarian where your piece of information is located while not going past the librarian's desk.

Then the librarian interprets your request and assigns the request task to the workers, and the fulfilled request for information is delivered to requestor. Easy, right? Well, there are a ton of rules that must be considered when submitting the request in order for the librarian to interpret it correctly. These rules, or best practices, will be addressed in a later section.

HIGH-LEVEL ARCHITECTURE PICTURE

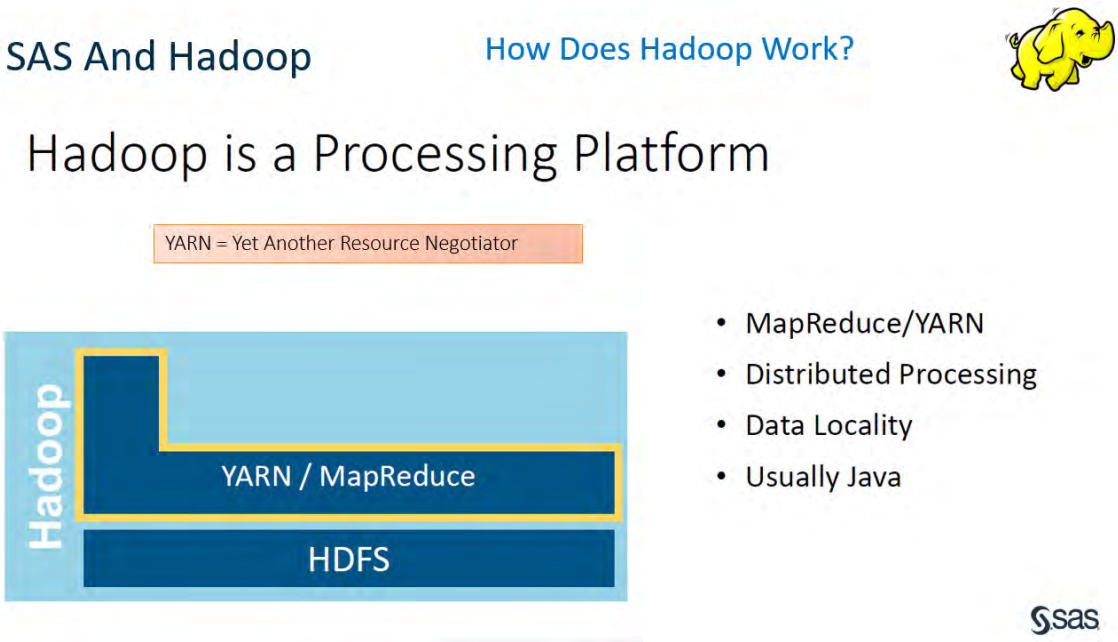


Display 7. Base-Level Hadoop Architecture

Display 7 is the library with very simple redundant storage. At its lowest level, the installation is the Hadoop platform with an open-source operating system (typically Linux) that supports a Distributed File System (DFS) and, since this one is on Hadoop, it is called HDFS. This environment runs very lean and leverages several computers networked together delivering great performance and redundancy. Because all of the software is open source and can be downloaded from GitHub, the hard dollar cost (or capital investment) is very low. The greatest measurable Hadoop-related expenses are the computer hardware, data center location cost, and human technical resources required to perform the installation and configuration. This is assuming that only free open-source software is used to configure and leverage the Hadoop environment.

BEGIN SIDEBAR COMMENTS: There are many soft dollar costs associated with moving an enterprise environment to a true open-source environment. This paper is not focused on that specific consideration, but I encourage a serious Total Cost of Ownership (TCO) assessment through an unbiased lens before going down the open-source path. I would not be the one who has to eat crow after sponsoring such an endeavor.

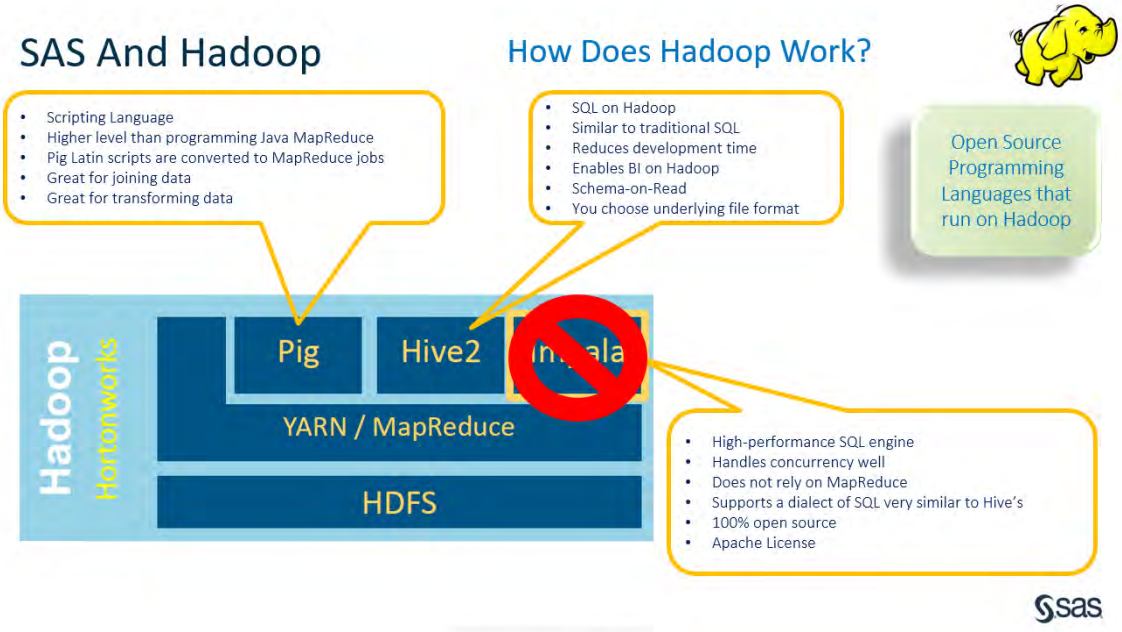
What qualified me to make these observations? Am I just a fake news junkie regarding open source spreading a bunch of FUD (fear, uncertainty, and doubt)? Let's take a brief commercial break and establish my credibility regarding Hadoop. I've set up a Hadoop cluster from scratch once on my home computer and twice in a SAS training lab. I also took a Hadoop Administration course delivered by Cloudera. I've interacted with the Hadoop clusters at SAS Institute more often that I can count. This requires Hadoop and SAS administrative-level capabilities. Moreover, I have experience working with Operating System (OS) level programming using EDLIN in 86-DOS (1983) and Linux. So, I'm speaking from experience when I say that this requires a pretty experienced technical resource to provide consistent and reliable support. **:END SIDEBAR COMMENTS**



Display 8. Utility Layer of Platform Management Software on Hadoop Architecture

After the operating system and HDFS are installed, the layers of supporting software are added to help manage the Hadoop environment. Basically, controlling software layers are, metaphorically speaking, added to help the librarian and workers to be more efficient by doing more data processing faster (MapReduce) and managing the workloads better (YARN), as depicted in Display 8.

Actually, it's best to go to the Wikipedia page to read up on this as a better authority than me. See https://en.wikipedia.org/wiki/Apache_Hadoop.



Display 9. Example of the Open-Source Software Additions to Help with Data Interactions

Now we add extra search and query capabilities to help get the stored information out of the Hadoop library, as depicted in Display 9.

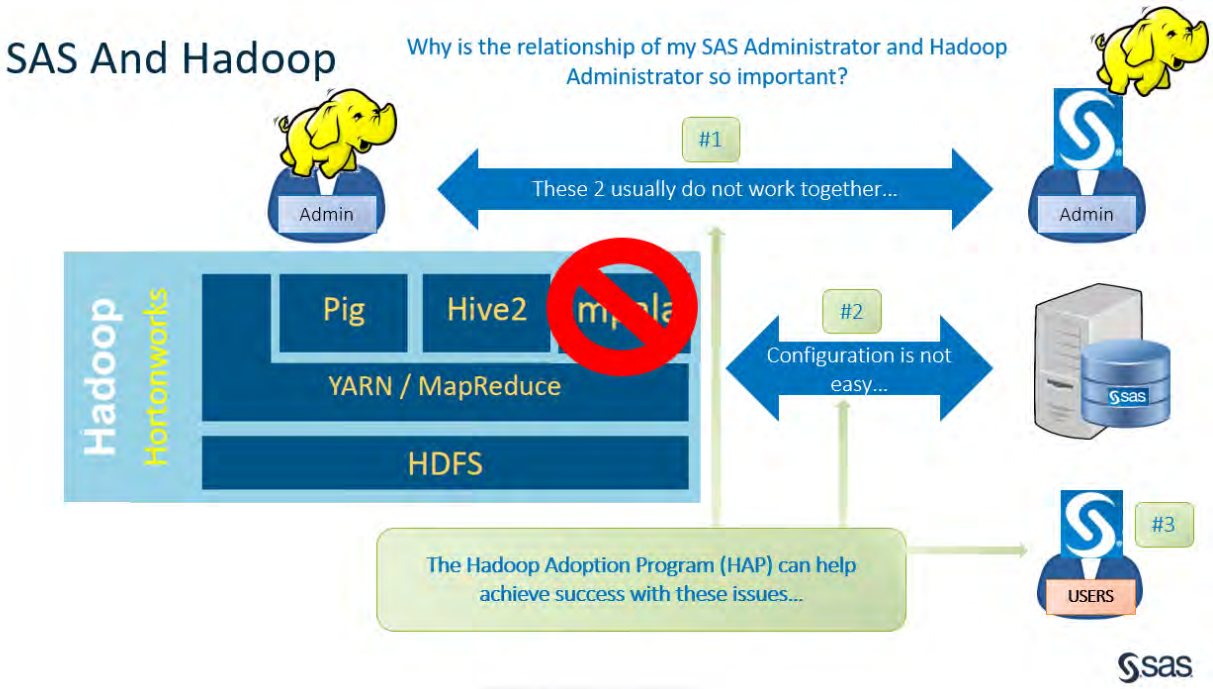
Pig is a scripting language that helps with basic data movement, joining, and transformation.

HiveQL leverages the Apache Hive data warehouse software architecture and performs SQL programming code similar to traditional SQL in other database environments (that is, Teradata, MySQL, and so on). It provides data summarization, query, and analysis functionality.

If the Hadoop distribution platform is Cloudera, Impala would be a possible addition. But Display 8 is a depiction of a Hortonworks distribution, so Impala is marked out. Regardless, Impala adds additional in-memory concurrent processing of data that eliminates the MapReduce processing, delivering a more performant interaction. Not requiring the MapReduce step is what helps deliver the faster processing speeds.

The SQL processing with Hive and/or Impala is what the SAS/ACCESS® Interface to Hadoop and/or the SAS/ACCESS® Interface to Impala does for SAS users. Their existing PROC SQL code will run in Hadoop as long as the functions and formats leveraged in the SAS code are compatible with the Hive and/or Impala functions and formats. (See section: "BEST PRACTICES FOR MOVING DATA AND ADVANCED ANALYTICS PREPARATION.") This will be covered again later in the spirit of repetition for emphasis!

RETURN ON INVESTMENT CONSIDERATIONS



Display 10. Depiction of How Collaboration Is a Key to Success

Display 10 is really about internal collaboration at your company to ensure successful adoption of using SAS Software to interact with Hadoop. This is the foundation to achieving ROI. My recommendation during the delivery of the Hadoop Adoption Program Workshop is for SAS and Hadoop Administrators to meet over lunch at least once a month to discuss their respective environments. This will stimulate comraderie that will push significantly toward Hadoop adoption.

IMPORTANT:

1. Administrators have to collaborate and work together.
2. Configuration is not easy, but it is critical before any value can be realized.
3. SAS users must be patient with the administrative process and procedures.

But once the configuration is complete, there are many ways that SAS can interact with Hadoop!

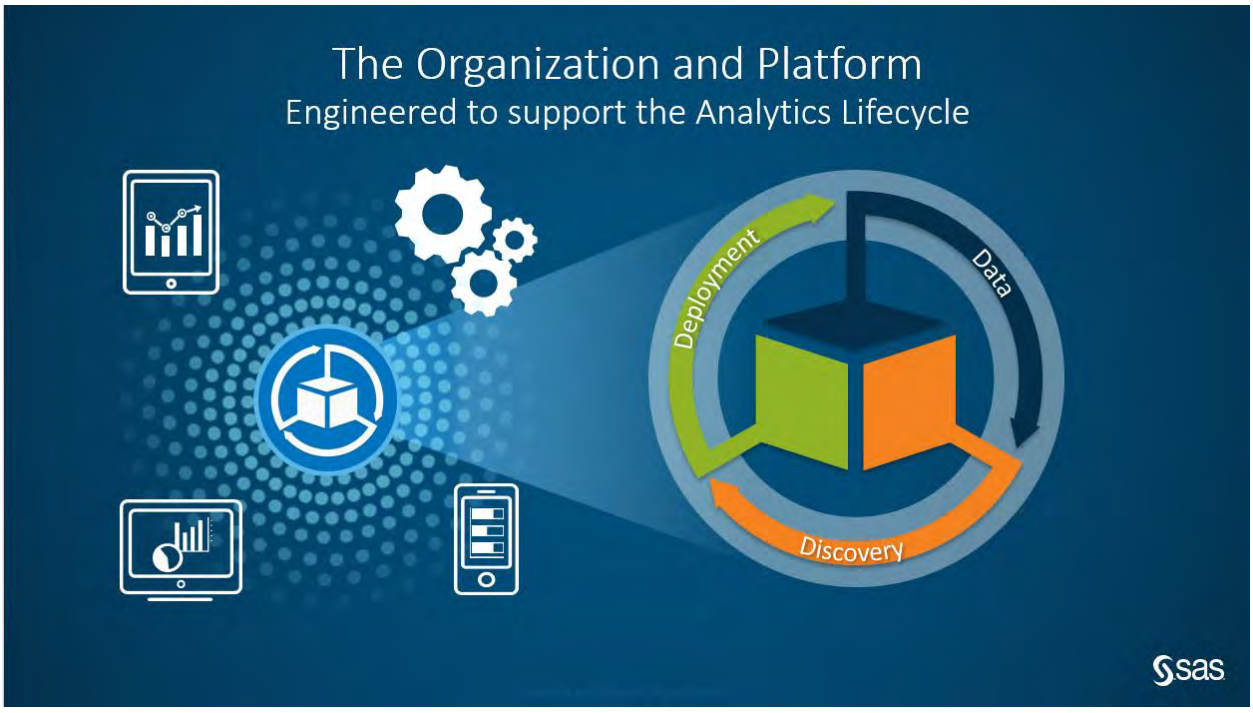
Now, with SAS fully enabled, as a SAS user, you can leverage the Hadoop information (remember the library metaphor) without having to know anything about how the library works, only that the information you want is there and accessible through your familiar SAS user interface.

While the collaborative efforts are a key factor in achieving reasonable ROI, the two administrative roles are partucularly focused on enablement of the functional capabilities. Once the Hadoop and SAS environments are is in place and fully functional, attention needs to turn toward a practical approach to returning value from the adoption of Hadoop processing with SAS software. So what SAS products can be used to generate value from the Hadoop data?

I'VE BOUGHT IT AND INSTALLED IT. NOW WHAT DO I DO WITH IT?

WHERE DOES SAS SOFTWARE FIT?

In reality, the ROI component is more targeted at the structure and quality of the associated business model. Successfully executing on data, discovery, and deployment is the key differentiator when it comes to applying the best available technical capabilities. SAS makes this end-to-end process become a reality. (See Display 11.)



Display 11. The SAS Platform

There is much more to be discovered regarding SAS capabilities and support for Hadoop, which is found at this link: https://support.sas.com/en/documentation/third-party-software-reference/9-4/support-for-hadoop.html

Here's a list of supported SAS products, offerings, and technologies (see Table 2) that can be enabled on the SAS Platform depicted in Display 11. These SAS capabilities will allow a SAS user to interact with Hadoop data and deliver value to the organization.

SAS Products, Offerings, and Technologies
Base SAS®: FILENAME Statement for Hadoop Access Method
Base SAS®: HADOOP Procedure
Base SAS®: SQOOP Procedure
SAS® Scalable Performance Data Engine
SAS® Scalable Performance Data Engine SerDe
SAS® Scalable Performance Data Server
SAS/ACCESS® Interface to Hadoop
SAS® In-Database Code Accelerator for Hadoop
SAS® Scoring Accelerator for Hadoop
DATA Step Processing in Hadoop
SAS® Grid Manager for Hadoop
SAS® High-Performance Analytics Environment
SAS® LASR™ Analytic Server
SAS® Data Loader for Hadoop

Table 2. SAS Products, Offerings, and Technologies Enabled for Hadoop

Well, it's all great that SAS has enabled users to fully adopt utilization of SAS software to interact with Hadoop data. But how can success be achieved as quickly as possible? I'm glad you asked! Everyone can benefit from the experience and knowledge of those who have dealt with adoption before. In the case of my introductory example, we leveraged the experience of anyone we could. One huge advantage was that my wife had adopted pets before and we were able to circumvent potential hazards.

The same is true with Hadoop adoption. Leveraging the experiences of many others, SAS has developed a program that I've been referring to, the Hadoop Adoption Program. Through this program, SAS has a resource that will pass along best practices for using SAS to interact with Hadoop in a workshop setting. This is designed to help SAS users optimize their experience and deliver immediate value using their use case in their environment. This is a practical application of what has likely been learned during typical education and knowledge transfers. This approach helps SAS users on their path to success and enables them to deliver value.

DO I REALLY SAVE AND/OR MAKE MONEY WITH HADOOP?

PATH TO SUCCESS: LEVERAGING SAS TO ACHIEVE QUICK SUCCESS

Based on the experience I've had in delivering Hadoop Adoption Program services to more than 30 different SAS customers in the last year or so, here is a prescription for successful Hadoop adoption leveraging SAS software:

1. Establish a solid use case that can be leveraged to pay for the disruption impact.
2. Properly size the Hadoop environment with consideration for the long-term data strategy. This requires a collaborative approach with your SAS Account Team, IT Support Team (SAS Administrator and Hadoop Administrator), and Business Technology Resources (for example, data scientists, business analysts, and so on).

3. Select skilled administrator(s) who can work collaboratively across the Hadoop and SAS architecture. The technical skills should be commensurate with the Hadoop and SAS platform infrastructure complexity.
4. Verify that SAS users can connect to the Hadoop cluster from their preferred SAS user interface (for example, SAS® Enterprise Guide, SAS® Studio, Display Manager System, SAS® Data Loader for Hadoop, SAS® Visual Analytics, SAS® SAS® Visual Statistics, SAS® Visual Data Mininig and Machine Learning, and so on).
5. Enroll SAS users in SAS education courses. I Recommend SAS user attendance of **Introduction to SAS and Hadoop** and/or **Working with SAS® Data Loader for Hadoop** and/or the free An Introduction to SAS® Viya® Programming for SAS 9 Programmers course from SAS Education and Training.
6. Help the SAS users translate their use case to representative SAS code that will leverage processing in-database wherever possible.
7. Ensure that all SAS code addresses Hadoop specific functions and formats where appropriate.
8. Beware of the 32,767 Hadoop STRING format issue and how to address the problem with appropriate SAS options and/or data format adjustments.

MOVING DATA: HOW ABOUT A LITTLE SAS CODE TO GET STARTED?

This is where it gets really serious for the SAS User when I'm visiting a customer during a Hadoop Adoption Program workshop.

1. Set SAS logging options and connect to the Hadoop cluster:

```
/* Options to help with debugging macros.  */
Options symbolgen mlogic mprint mrecall;

/* Options to help with determining when execution occurs in-database.  */
Options sastrace=',,,ds' sastraceloc=saslog nostsuffix;

/* Set up connection to HiveServer2. */
Libname hdp HADOOP port=10000 schema=default host="sasdata1.demo.sas.com"
user=sasdemo;
```

2. Move data using the DATA step:

```
/*** BEGIN - Use DATA Step - Move SAS Data into Hadoop and check out the
new Hive table and metadata. ***/
/* Bulk Load Cars data set to Hadoop from the sashelp library.  */

/* Check for existing table and delete if exists to mitigate error. Hadoop
is write once, read many.*/
%macro checkds(dsn);
   %if %sysfunc(exist(&dsn)) %then
      %do;
         proc delete data=&dsn;
         run;
      %end;
%mend checkds;
%checkds(hdp.cars);
```

```
/* Execute the Bulk Load DATA Step. */
data hdp.cars;
    set sashelp.cars;
run;

/*Examine the available metadata. */
proc contents data=hdp.cars; run;

/* List data from Hive, and use a where clause to create a MR job.*/
proc print data=hdp.cars (where=(Upcase(Make)="ACURA")); run;

/*** END - Use DATA Step - Move SAS Data into Hadoop and check out the new
Hive table and metadata. ***/
```

BEST PRACTICES FOR MOVING DATA AND ADVANCED ANALYTICS PREPARATION

Before executing advanced analytics, make sure that steps 6-8 in the path to success are addressed in respect to the data and in-database techniques. I'm not going to get into all the details, but here are a few of the key points that I emphasize during Hadoop Adoption Program workshops:

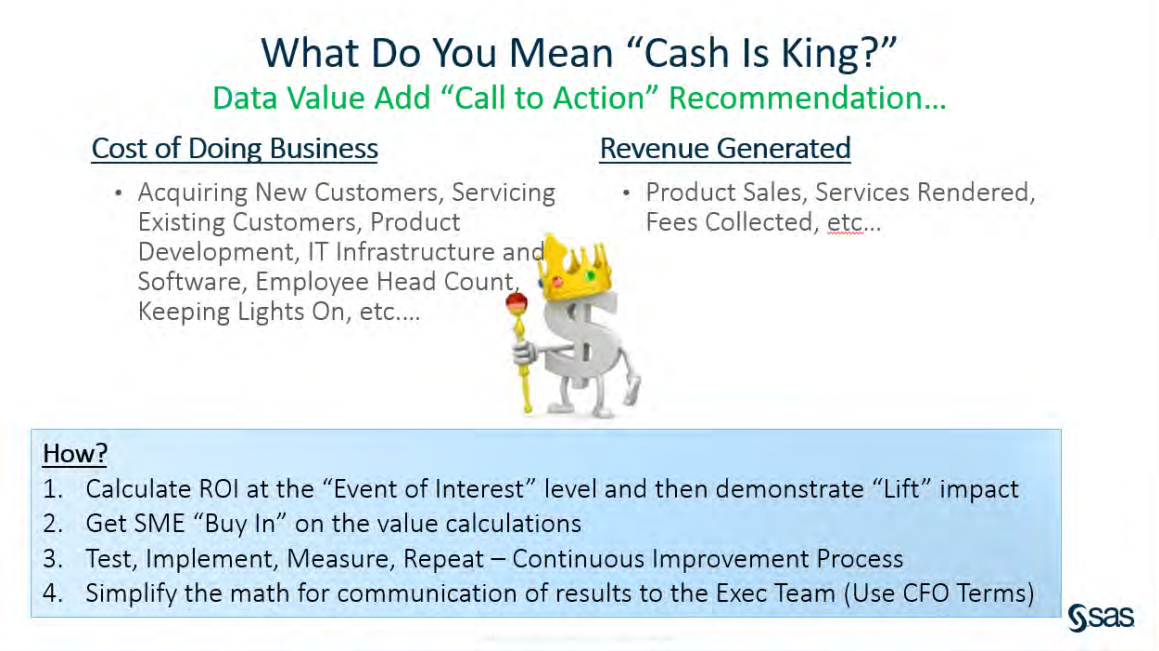
1. Have the data formatted for SAS compatibility at the time it is placed in Hadoop (for example, VARCHAR or CHAR data type instead of STRING in Hive).
2. Perform as much data manipulation as possible in Hadoop and do not move it to SAS (for example, change any WORK librefs to Hadoop librefs).
3. When developing SAS projects (or migrating code that was previously used for other data sources), make sure the proper functions and formats are used:
    - See **Passing SAS Functions to Hadoop.**
    - See also **Data Types for Hadoop.**
4. Use DBMAX_TEXT= and/or DBASASTYPE= LIBNAME/data set options to explore data when it is unknown if the columns in Hadoop (Hive) are STRING (32,767 bytes).
5. Make sure your **DBCREATE_TABLE_OPTS= LIBNAME option [or data set option]** is compatible with how your data is stored in Hadoop (for example, ORC, Avero, and so on).

There are several others, but these are probably the ones that will make the quickest impact.

ADVANCED ANALYTICS VALUE: NOW THAT BEST PRACTICES HAVE BEEN CONSIDERED, LET'S TURN DATA INTO DOLLARS

At this point, the data scientists and/or analyst should be thinking about the next step in creating a data set with their event of interest (also known as a target, null hypothesis, and so on) and the problem they are planning to solve with Hadoop data.

Whether it's trying to find bad actors or the next best offer to give a customer, the idea is the same. There is an operational problem that is either costing the company a bunch of money or resulting in missed revenue opportunities. Regardless of the industry or the objective, the end result is to turn data into dollars.



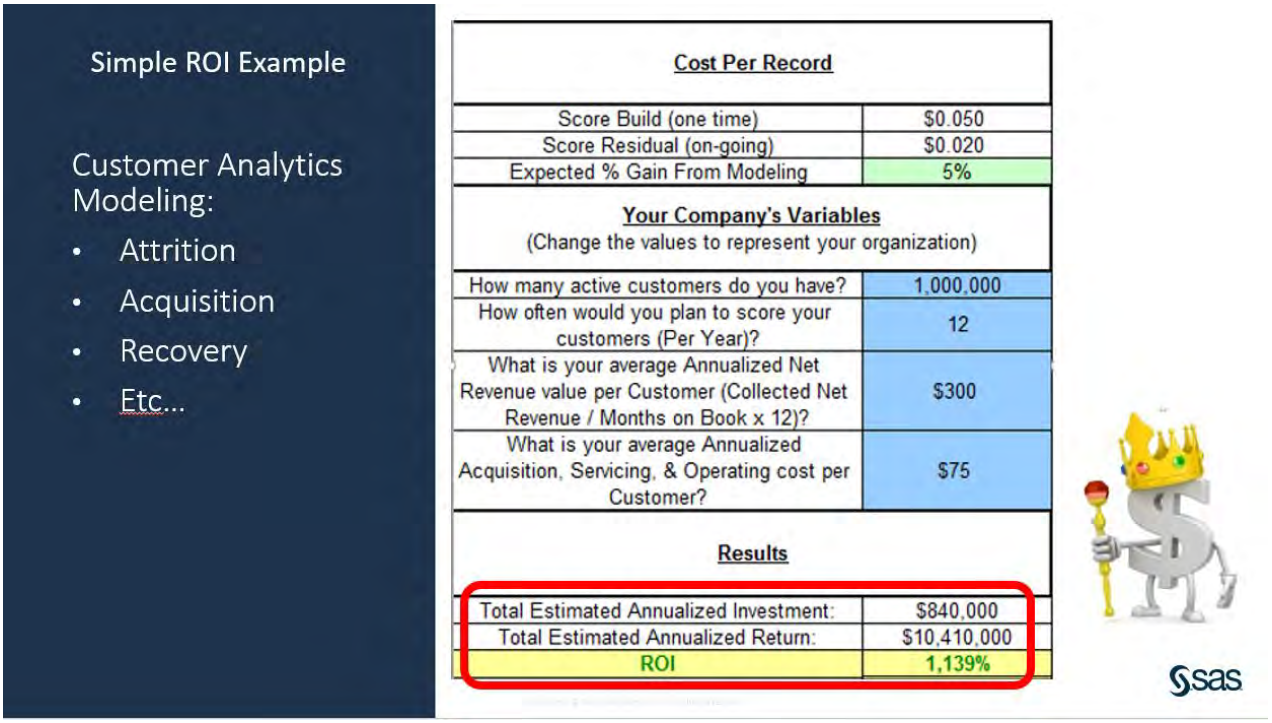
Display 12. How To Turn Data into Dollars

In respect to Display 12, there is a specific call to action, as follows:

1. Identify the relevant event of interest and the incremental value impact (Lift).
2. Have the Subject Matter Expert (SME) develop and/or agree with the value calculations.
3. Follow a process for continuous improvements.
4. Keep it simple to achieve executive commitments to your analytics platform investments.

Note: This is not the end all, be all, but it is the relevant approach that has worked for me over he past 32+ years of my career of turning data into dollars.

Always tie strategic initiatives to projects and explicitly express the value in relevant specific event of interest terms that can be understood by Executive Management (for example, cost savings/customer; revenue/customer; net income/customer). See Display 13 for a simple ROI example of how to present a specific use case calculation:



Display 13. Simple ROI Example

There are several ways to measure the value of success. ROI happens to be one that is fairly simple and commonplace.

To calculate ROI, the benefit (or return) of an investment is divided by the cost of the investment, and the result is expressed as a percentage or a ratio.

Read more: Return On Investment (ROI) http://www.investopedia.com/terms/r/returnoninvestment.asp

The business person (that is, the data scientist, business analyst, and so on) owns the purse strings and needs to build a case for strategic investments toward being an analytics-driven company.

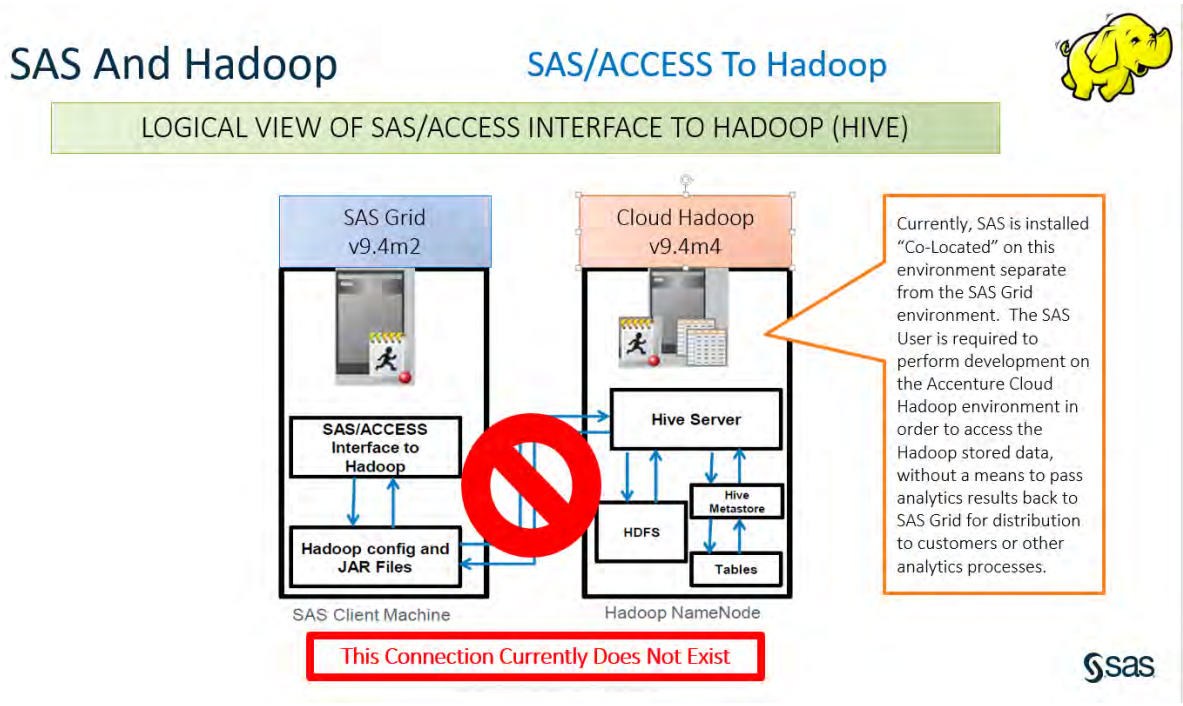
I recommend the following books by Tom Davenport to help as appropriate:

1. *Competing on Analytics*
2. *Analytics at Work*

Being able to articulate the value of an analytics project to executive management firmly establishes the continued viability of an ongoing Hadoop adoption investment foundational to the organization's success.

ADOPTION STORIES

Horror Story!!! There are several customers who are choosing to move their analytics platform to the Cloud. I recently had a discovery call with a major insurance company where their IT department had moved forward with the cloud strategy without proper consideration and input from the business stakeholders. Here's a logical architecture diagram view of the environment. (See Display 14.)



Display 14. Example Where the Architecture DOES NOT Support the SAS User

NOTE: Display 14 is my rendition of what was described to me by SAS users and administrators.

In Display 14, SAS is installed co-located on the environment separate from the SAS® Grid environment. The SAS user is required to perform development on the Accenture Cloud Hadoop environment in order to access the Hadoop stored data, without a means to pass analytic results back to SAS for distribution to customers or other analytics processes.

My specific recommendations for solving this architecture issue were:

1. Establish a SAS connection from the SAS® Grid to the Accenture Hadoop Cluster enabling the SAS/ACCESS® engine to interact via the LIBNAME connection with the Hadoop data.
2. Consider the need to upgrade the SAS® Grid from the second maintenance release of SAS® 9.4 to the same version as the Accenture Hadoop Cluster (v9.4m4).
3. Possibly engage HAP Configuration Validation or Test Services.

The jury is still out on this one, but it is critical that SAS users get in front of this issue. When IT is the sole driver of platform decisions, the potential is huge for a disaster. Business owners must respect IT for their role, but the accountability for delivering value and hardware/software adoption rests with those who have the capability to solve real business problems with the technology.

I have several customer adoption stories, but my paper is already too long. I hope you get some value from this documentation built from my experience as a longtime SAS user as well as the actual successful experiences of the Hadoop Adoption Program offered by the SAS Customer Success Organization (formerly Customer Loyalty and Retention – CLR).

Here are a few quotes from satisfied customers of the HAP:

"Overall I think the information provided in the workshop is invaluable for the users, and even more so for the Hadoop and SAS administrators. For example, your guidance on how the users and administrators should collaborate to resolve technical issues will help greatly."

- Jason, SAS Administrator
Large Health Insurance Company

"Using the suggestion you made in the workshop, our processing time has dropped from 51 minutes to 5 minutes! That is an outstanding productivity savings! I actually showed a colleague the same bit of code and it took his process from 30 minutes to less than a minute. I would highly recommend this program to your HADOOP community!"

-Business Intelligence Analyst
Leading Cereal Company

CONCLUSION

FINAL MESSAGE FOR ADMINISTRATORS

It is imperative that Hadoop and SAS administrators communicate and collaborate consistently and persistently. The connectivity between systems is heavily dependent upon the JAR and XML (config) files that allow the two environments to connect. It is also critical that the security permissions allow the SAS users based on their roles and privileges to interact with Hadoop from SAS without Read or Write permissions errors.

Here's a link to the administrators guide list: https://support.sas.com/en/documentation/third-party-software-reference/9-4/guides-papers-for-hadoop.html.

FINAL MESSAGE FOR SAS USERS

Be persistent at communicating your areas of impact. The SAS user is an individual contributor, whether as a data scientist, systems analyst, business analyst, or SAS admin (that's right, admins can be SAS users ☺) or in any role that touches SAS software. At the basic level, there is a need to express the value of what you do in terms that the chief financial officer, or any executive, can relate to. Typically, that is a return on something. Once the value is effectively communicated, then the software's value will ensure that your job will be relevant for years to come. In fact, it will lead to career advancement if that is your goal.

FINAL MESSAGE FOR STAKEHOLDERS (INCLUDING EXECUTIVES)

Listen to the suggestions of those who are individual contributors. Whether SAS administrators or SAS users, they are the closest to the data and they will deliver huge value back to the company through their individual and team contributions. Be diligent about removing obstacles and enabling the technology they need to do their daily job. When office politics get in the way, remediate the situation and provide the direction necessary to keep the focus on the tasks that align with your strategic plans for the organization.

GO HOME AND EMBRACE YOUR HADOOP ADOPTION PROCESS AND FEEL GOOD ABOUT IT!

- SAS Cares Portal
- Self-serve on the support.sas.com site for Hadoop materials
- Sign up and take the Solutions for Hadoop Education Modules
- "How To" Videos (Data Management Tab)
- Informational links
- Hadoop Distribution Support is found at SAS 9.4 Support for Hadoop

REFERENCES

Davenport, Thomas H. and Harris, Jeanne G. 2007. *Competing on Analytics*. Boston, MA United States of America : Harvard Business School Publishing Corporation.

Davenport, Thomas H., Harris, Jeanne G., and Morison, Robert. 2010. *Analytics at Work*. United States of America : Harvard Business School Publishing Corporation.

ACKNOWLEDGMENTS

Jeff Bailey, Data Access Product Management Technical Lead, Data Management Product Line.

Clark Bradley, Manager, Pre-Sales Support, Global Technology Practice, Business Management.

Both Jeff and Clark have been extremely active in supporting my knowledge acquisition and assisting when I have customer challenges that need to be solved. Special thanks to both of you for sharing your knowledge of Hadoop with me in addition to the time it takes to provide responses to my questions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Rex Pruitt
Indian Trail, NC 28079
SAS Institute Inc.
rex.pruitt@sas.com
http://www.sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.