The Big Data Journey: Where do I start?

Robert Morris
Rackspace Data Services
Where to start?

“I know so much that I don’t know where to begin.”
Many want to start the journey

In late 2014 a Gartner Survey revealed that 73% of organizations have invested or plan to invest in Big Data in the next two years.

In 2015 IDG found that Enterprise organizations they surveyed will invest 13.8M on data-driven initiatives in 2015.

Global spending on Big Data exceeded $31 billion in 2013 and is expected to hit $114 billion in 2018.

Cap Gemini noted that nearly 60% of senior executives believe that Big Data will disrupt their industry within the next three years.
But getting to the destination…

Only **27%** told IDG they **had already deployed/implemented** data driven projects. **42%** were still planning implementations.

Cap Gemini found only **27%** of their respondents described their Big Data projects as “**successful**” and only **8%** described them as “**very successful**”.

**Proof of concept success rate? 38%!**

Only **3 in 10** business that plan to implement Hadoop have done so.
The problem?
Where do the journeys begin?

Engage with Vendors

Start of Project

Project completed to point where value obtained
Where do the SUCCESSFUL journeys begin?
First steps to take towards a successful Big Data project

- Business Objectives
- Inventory
- Use Cases
Learn the objectives of:
• the business overall
• any business unit that might be a part of the project

As you navigate though the journey this information will be your compass point to follow and keep you headed towards success – aka the Why
Inventory: Data & Stakeholders

Who and What are vital to finding the best use cases and/or honing the use case to help achieve business objectives.

Projects are easily derailed by surprise BUs, people or data sources.

This takes a lot of work and, often, cross department/BU communication. This is where champions are vital!
Stakeholders – Identify:
• champions of the project, ideally overall and in each BU
• budget owners
• those who might benefit and consume the outcome in each BU – both IT and NON-IT
• who own/controls the data in each BU
• those responsible for security/compliance
Inventory: Stakeholders

Stakeholders – Interview/survey:

• Which of the business objectives is most important to them?
• What is their personal business objective this year/quarter?
• What data do they use now?
• What data do they want to use and can’t & why?
• How do they like to consume data, e.g. graphs, spreadsheets, software, SQL, code, APIs, etc.
Inventory: Data

Taking a data inventory is vital for any organization for which data is a mission critical part of the organization (and for most it is).

Check and see if anyone else has already started. Often security has done an inventory for access controls.

Think of it to mapping the streams that will feed your data lake.
Inventory: Data

Use your stakeholder inventory and interview/survey as a starting point.

IMPORTANT! Don’t just ask DBAs! Talk with analysts, developers, and sysadmins/devops.

Once you have identified the data start learning more about it:

- Who owns/who uses
- Source
- Access methods
- Sample data, documentation
- Way to obtain if desired but not there
Use Cases

After you have a good grasp of the business objectives, stakeholders and data sources you can now, really look at use cases.

Use cases are not just a technical problem you are solving – they are a business problem you are solving.

You likely already have or had a few in mind but you are significantly better positioned to have a successful POC and more importantly KNOW WHAT THAT MEANS!

If you project is stalled, slow, or sluggish then review what you’ve done. Have you taken the first 3 steps?
Once you’ve take the first 3 steps you should be able to easily communicate what a successful outcome will look like and what is needed from the business side.

A POC should always have a defined desired outcome that you and the vendor/partner agree upon. If you don’t know what success looks like – don’t start!

Now Reach Out to Vendors & Start a POC!
Real World Examples of Success

Large Electrical Equipment Manufacture

• **Business Objective**: Solve their customers’ biggest challenges. Increase revenue.

• **Key Stakeholders**: Service providers, IT, developers, product, data scientists

• **Data**: Information from devices, existing information about devices and failures, service provider data.

• **Use Case**: Use existing knowledge of how devices fail to monitor current devices in order to help service providers stock and staff for maximized profit. Also allow for forecast of demand for manufacturing and new possible products.

• **Outcome**: Came to us with all three steps completed. Were ready to test but didn’t want to run the complex environment/software required. POC was successful and they are moving into production – POC allowed right sizing environment and honing down to specific tools.
Global Professional Services Firm

- **Objective**: Increase efficiency – fraud detection. Making working world better.

- **Key Stakeholders**: CTO, auditors, IT, product, security, developers

- **Data**: Customer accounting databases, E-mails, key word knowledge

- **Use Case**: Take manual task and make it automated, faster and intelligent via full text analysis with machine learning. Give auditors better data.

- **Outcome**: Did not start out with all 3 steps. We helped guide back around to stakeholder inventory. Now have a very solid POC scope and expect great success.
Questions?
Thank you!

Robert Morris
Rackspace Data Services
Robert.morris@rackspace.com