Datameer
stay curious

Best Practices for a Successful Big Data Journey

Big Data = Big Impact

Big data analytics has the potential to be transformative. Companies such as Amazon, Facebook and Uber have built their kingdoms on big data analytics. Other companies such as General Electric, American Express and AT&T have completely altered the way they compete in the marketplace because of big data analytics.

With big data analytics, an organization can:

1. **Increase revenue** by dramatically improving their effectiveness in the marketplace and growing customer relationships.

2. **Reduce operating costs** by improving the efficiency of business processes, production, supply chains, networks and more.

3. **Avoid potentially debilitating losses** by eliminating fines from lack of compliance and reducing the risk in various assets.

The Journey Will Have Challenges

To quote a common saying, “If it were easy, everyone would be doing it.” But when it comes to big data, a more applicable adaptation to this saying would be:

“If it were easy, everyone would be successful at it!”

Many companies have stalled in their big data journeys. The majority of the time, technology is not the issue. It’s very possible to become successful at big data. But it requires a commitment to cultural changes, business model adjustments, new process and additional skills.

Gartner

Through 2017, 60 percent of big data projects will fail to go beyond piloting and experimentation and will be abandoned.

Through 2018, 90 percent of deployed data lakes will be useless as they are overwhelmed with information assets captured about uncertain use cases.

Gartner: Seven Best Practices for Your Big Data Analytics Projects (2015)

Only 27 percent of executives described their big data initiatives as successful.

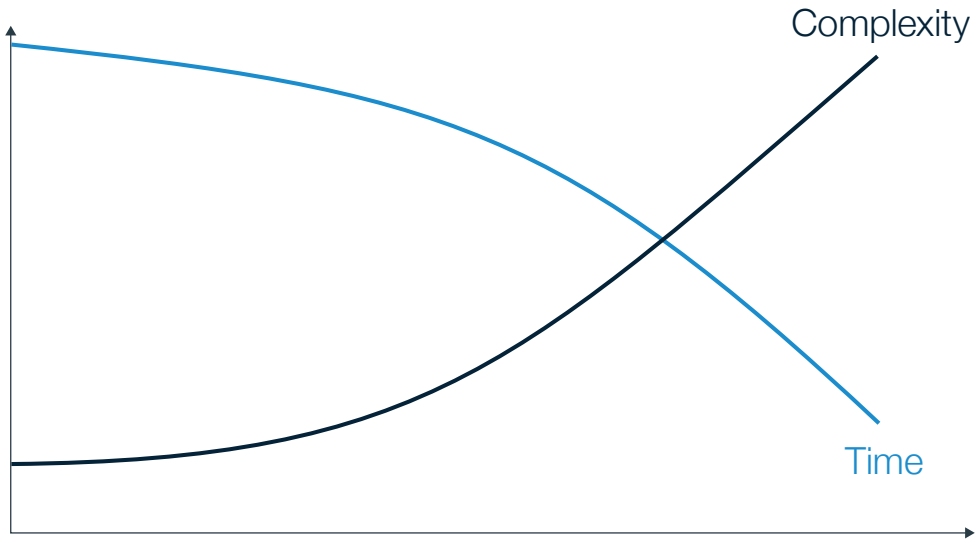
Cap Gemini: Big & Fast Data: The Rise of Insight-Driven Business (2015)

Time versus Complexity

There is increasing pressure on analytics teams to reduce time to insight and answer questions faster. Yet, big data has introduced many new forms of complexity:

1. Complex, multi-structured data
2. Sophisticated analysis and correlation
3. More data and result sets to govern

Big data analytic platforms can help you handle these new forms of complexity and reduce your time to insight. Yet, factors unrelated to the software can put a drag on your analytic processes and limit the time to insight gains your organization sees.



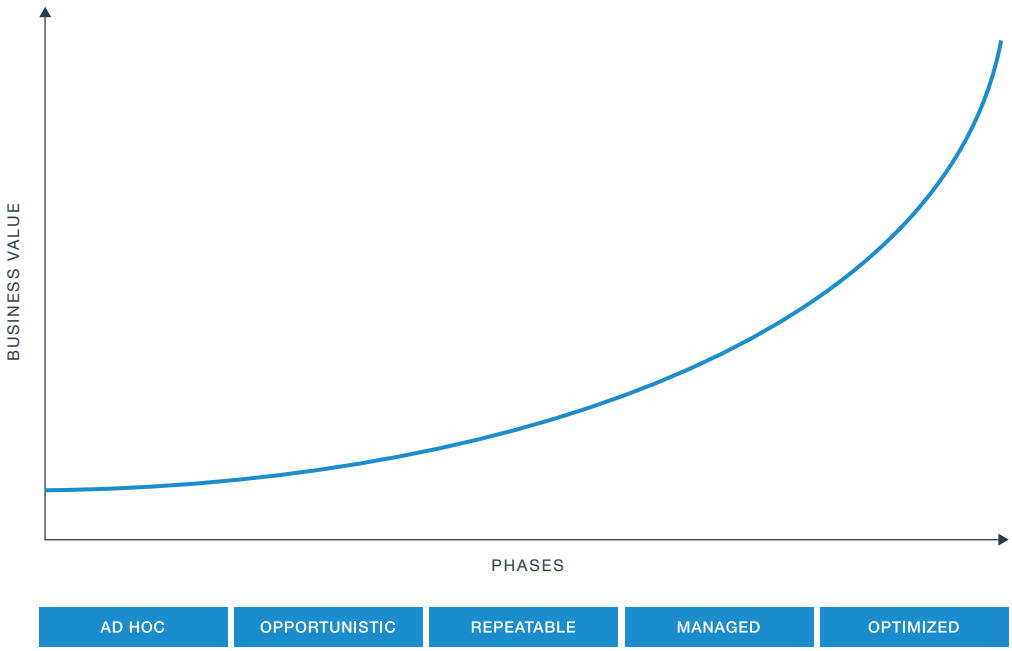
Data Complexity	Analytic Complexity	Governance Complexity
Varying forms of data Large volumes of data Many sources of data Complex blending	Graph and Path Analysis Finding hidden patterns Identifying behavior Getting more granular	More raw data Managing prepared datasets Securing private data Data democratization

Five Phases of a Big Data Journey

Big data is not simply a project. It will be a journey that involves many different projects to continuously gain more value from your big data. Over time, your big data journey will evolve through five different phases:

1. **Ad-hoc** — the earliest phase where organizations experiment and learn about their big data needs.
2. **Opportunistic** — the second phase where an organization starts to deliver value to the business, building their skills and knowledge.
3. **Repeatable** — the phase where a company creates a replicable model for big data projects and starts to operationalize.
4. **Managed** — a phase where big data analytics becomes a managed service that starts to spread across the organization.
5. **Optimized** — where big data becomes a well-oiled machine, continuously delivering new projects and exponential value to the business.

As your company evolves through these phases, you will see an exponential increase in value and momentum with your big data journey.



The following sections will delve into the details of each phase, and this table provides a summary level detail description.

Summary Table of Big Data Journey Phases

	Ad Hoc	Opportunistic	Repeatable	Managed	Optimized
Scope	• Unbudgeted projects • Hero culture	• Departmental • Repeatable success	• Business unit level • Documented process	• Cross business units • Measure success	• Enterprise • Continuous improvement
Data	• Missing data • Manual effort	• Multiple sources • Structured and unstructured	• Consistent data • Little governance	• Data managed • Timeliness	• On time and Trusted • Enterprise access
People	• Limited skills • Lacks mgt. interest	• Teams have skills • Lack intra-organizational coordination	• Skills govern by strategy • Skills augmented externally	• Executive support • Centralized COE • Decentralized skills	• All skills exist • Executive priority
Process	• Siloed information • Lacks collaboration	• Data analysis • Data prep	• Monitoring and documenting decision process	• Metrics for evaluating process, quality and success	• Processes are properly staffed and funded
Tools	• Custom scripts • Legacy systems/ tools	• New tech required for specific purpose	• Multiple “fit for purpose” tools deployed	• Broadly adopted and wide range of “purpose fit” tools	• Optimized with a high-level of automation

The Ad-Hoc Phase

Many companies start off with ad hoc, experimental big data efforts. And there is nothing wrong with starting here. At this point, organizations are typically trying to figure out what big data is, and more importantly, how it can impact the company.

Scope

During this phase, the scope of the initiative is small and often unbudgeted. The team is simply trying to get an understanding of the assets available, including data and people. Much of the time is spent identifying, evaluating and gathering the assets.

Data

At this initial stage, there will be many problems with the data – missing data, cleanliness, shaping, fit, suitability, etc. Learning the data is an extremely important aspect of this phase, specifically learning the suitability of the data for the analytics at hand.

People

Your team will be converts from the traditional analytics world at this stage, and thus may lack some of the skills necessary for big data analytics. These skills may not just be technical ones, as big data requires new knowledge in the art of data preparation and discovery. In addition, management knows this stage is experimental, and thus won't show serious interest until value is proven.

Process

There's also a tendency to rely on manual labor to process data, and analysts are building their analyses by hand, learning about the data and what they can discover from it along the way. Often times, analysts will uncover what data is missing or what more data they need to complete their analysis. There is little formal collaboration, which is fine because the team is small and typically exploring individual items.

But to be successful, companies need to move beyond this "hero" type approach to big data, gain alignment with the business and embrace methods to increase productivity.

The Opportunistic Phase

Being opportunistic means taking advantage of immediate circumstances without necessarily having a grand plan in place. Of course, the long-term goal of any big data program should be to develop that grand plan. But a great first step is to take advantage of the opportunities in front of you.

Scope

At the opportunistic phase, your big data initiative will focus on one or two very specific, departmental-level solutions. The team will try to answer questions that were deep in the backlog due to a project-long analytic cycle.

Performing a Use Case Discovery workshop where your team will work together to align the big data initiative with the needs of the business, will help you identify places to start. You also want to identify a key means to measure the value of the analysis delivered, so you can start to capture Return on Investment (ROI).

Data

Your initial efforts to get more organized around your data will start in this phase. Previous analytics cycles were long because data was spread across different silos and in different forms (structured, semi-structured, unstructured). A key outcome from this phase is beginning to break down these silos and barriers so you can use more of your data.

People

In this phase, your team will start to gain many of the skills needed for big data analytics. They will learn what the shape of the data is and how to prepare it. They will also learn how to perform data discovery to look for answers. In addition, cross-functional coordination will grow between the IT, analyst and business teams.

Process

The process at this phase is very data-centric, focused on how to transform the complex pile of raw big data into something meaningful for this specific project. At this stage, the objective isn't to create repeatable processes. The team is still learning. But you do want to identify the key characteristics of the process that can be made repeatable.

The Repeatable Phase

In the Repeatable phase, your objective is to gain momentum and increase scale with your big data initiative. Creating a replicable model for your analytic process is crucial to gaining scale, and is an important output from this phase.

Scope

The scope of your initiative expands to cover multiple departments within a single business unit, each with distinct problems but having some commonality (e.g. multiple analytics using customer behavior or customer-360 data). In this phase you will also begin to operationalize your insights to gain everyday use within the business.

Data

Gaining data consistency is important in this phase. This means gaining a better idea of where data comes from, how to curate it, and how various analysts use it. This helps your team gain repeatability at the data level. Within a single business unit there should also be fewer barriers to sharing data, so you can begin to experiment with your governance models.

People

Gaining repeatability is also about developing consistency and reliability within your teams. It is important to establish a coherent, dedicated and coordinated team at this phase. The team can be small still, but requires core skills around data ingestion, curation, stewardship and analysis. From this core, the team should continue to grow and evolve.

Process

As indicated by the title, establishing a repeatable process is an output of this phase. This means the basic approach, structure and communications processes should be put in place and be used repeatedly. The process needs to be fully documented and articulated to all the teams involved: IT, analysts and business.

The Managed Phase

Finally, after a lot of team effort your big data program is growing up right before your eyes. It has grown into a valuable, organized and repeatable part of your company. Now it's time to give your big data the proper management structure to ensure success.

Scope

With the big data program servicing multiple business units, measurement is now more vital than ever before. Quantifiable goals should be in place, and progress towards these goals should be measured. Process, quality, and success should each be measured and tracked individually. ROI must also be closely considered and maximized.

Data

Data should be flowing in at a regular rate and being processed consistently. The data is well managed with consistent retention policies. There is a wide range of raw data being used, and an extensive catalog of curated data for use by the business analysts. The management protocols and governance for this data should be in place, or at least in the process of being put in place.

People

At this stage, a centralized Center of Excellence (CoE) has emerged, but the team itself will become decentralized and embedded further in the business. A full range of skills will be available on the team. Executive support will remain vital helping to ensure that business units are big data informed and driven.

Process

As you decentralize, however, your Center of Excellence will take on an increasingly vital role, setting the tone and providing guidance for your big data program. The CoE will be essential for codifying and documenting processes, and ensuring that your company is always following best practices.

The Optimized Phase

Finally, your big data program is up and running, and it's in good hands too. You also have the proper software and tools in place, along with a Center of Excellence that can provide constant insight and guidance. Time to kick back and relax, right? Wrong.

Scope

Even as you further expand your big data program across the enterprise, you must continue to strive for continuous improvement. You will continue to uncover more use cases and encounter new challenges along the way. It's essential to constantly measure of how the program is performing and understand how to make it better.

Data

Data should be flowing in at a regular rate and being processed consistently. The data is well managed with consistent retention policies. There is a wide range of raw data being used, and an extensive catalog of curated data for use by the business analysts. The management protocols and governance for this data should be in place, or at least in the process of being put in place.

People

Your team will be fully flushed out and distributed as needed across the organization. All the skills necessary will be available across the team, and many team members may become highly specialized in certain areas – curation, stewardship, discovery, advanced analytics, etc. The CoE should be well organized, and working regularly to stay abreast of any developments and to continue to refine best practices.

Process

You are a well-oiled machine at this point, with highly repeatable, well-documented and streamlined processes. The effort for each new use case can be accurately estimated and projects are consistently delivered on time. Many analytic processes are optimized with a high degree of automation.

Three Steps to Big Data Success

As your big data journey unfolds, there are three major hurdles that will present challenges to the organization moving between different phases:

1. Proving big data can deliver real value
2. Putting insights to work so value is obtained
3. Influencing the organizational change needed for success

Answering these challenges will require strategy shifts in key areas, including:

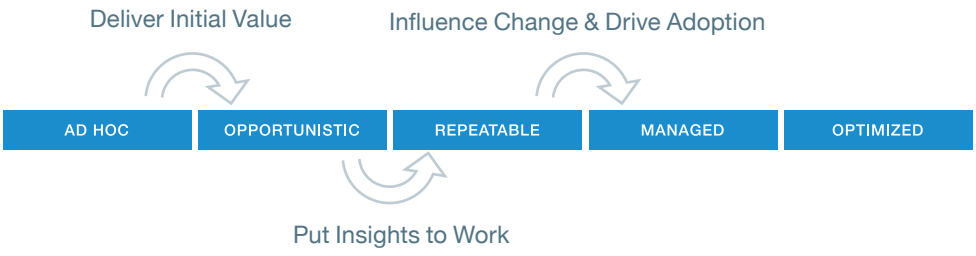
- Tools
- Approaches
- People and skills
- Partnering with IT and the business

The solution to each of these challenges is:

1. Delivering Value — Organizations evolve from the Ad-hoc to Opportunistic phases by starting small and focusing on specific solutions. By identifying these early wins, the big data initiative can deliver initial value to the organization.

2. Putting Insights to Work — You can move from Opportunistic to Repeatable phases by partnering with the IT team to operationalizing your insights. Deployment aspects for scalability, security and governance will enable key information to flow to the business teams regularly.

3. Influencing Change — Lastly, teams can reach the Managed phase by driving broader adoption of the platform, tools and information. Establishing a Center of Excellence (CoE) that services the wider organization can help drive this adoption.



1. Deliver Initial Value

One of the keys to ensuring successful implementation in the early stages is to set your sights on reasonable goals. Start small and identify achievable ways to deliver value. Ensure you can prevent scope creep.

You should focus on departmental level wins and identifying key users where value can be delivered effectively. Identify use cases which are on the edges of data between systems or involve new data never used before.

The big data heroes and business analyst team need to work closely to clearly identify the objectives, and initiate an agile prototyping process. Work with key power users closely, and make sure you're delivering timely, relevant and valuable results. This process should deliver first insights quickly and efficiently. Don't get discouraged, as discounting an idea is as important as validating one! Think of it as questions answered and showcase it to management.

As you move through this step, you should see emerging:

- New ways to deliver insights quickly
- The ability to enable power users in the business
- Initial buy-in from key stakeholders

Finally, as you perform this evolution, start to document what worked and what didn't. This will help feed internal Communities of Practice (CoP) and Centers of Excellence that can provide support, guidance and shared learning, as well as outlining measurements and governance.

2. Put Insights to Work

Once you have showcased value to the executive team and the business, you need to operationalize your results, so a constant stream of value is derived from your big data efforts. This is where you need to partner with IT create a secure, reliable deployment.

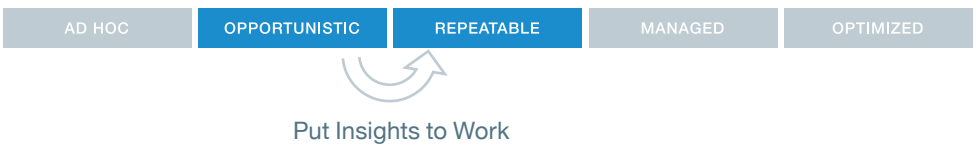
Having a strong relationship with your IT team is critical for this phase. If you involve them too late, you may find your plans backlogged.

Integration with existing systems both in the business and in IT is essential. Use your analytic platform's robust APIs to ensure data comes from critical sources and is delivered to the applications the business teams use. Use an Event Bus to allow critical information from the runtime execution of the analytic platform be integrated with other key IT systems where this information is gathered.

The outcomes of this transitional step include:

- Your initial power users will become champions for your big data program, helping to drive further adoption
- A nascent CoP will start to emerge with power users and emerging experts exchanging tips and tricks
- The systems delivering the new-found insights will become increasingly performant and scalable

The growing CoP will truly start to enable adoption among more power users and drive best practices within your big data program.



3. Influence Change and Drive Adoption

Big data is more than just technology. It helps an organization see things in a whole new light — customers, operations, risk and more. But with this new vision comes organizational change. New ways of using data, interacting with customers and executing the business.

The final leap in your big data journey is moving from the Repeatability to Managed phase. This involves helping the organization recognize the change needed and helping to drive not only the adoption of big data, but also the adoption of change.

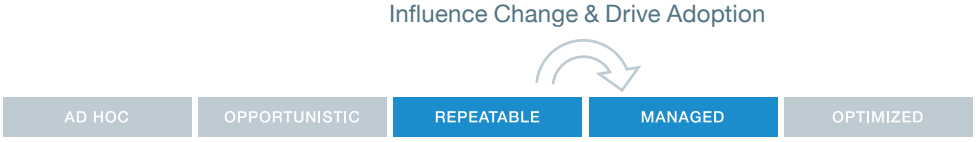
First you need to evolve your CoP into a CoE. This requires documenting and codifying your processes so they can be applied to multiple areas of the business, and making the CoE the true heart of the big data initiative. This turns repeatable processes into easily adopted ones.

Then you need to become true champions of big data, and drive adoption with numerous business stakeholders across the organization. This ability to scale and continuously drive new use cases will steepen the value curve big data is driving into the organization.

Lastly, to get more parts of the organization involved, facilitate scalable, formal training processes so anyone who wants to adopt your big data initiative in their department can do so. This training involves both pieces of the technology stack that are exposed to these new users and processes the CoE utilizes to facilitate big data adoption.

As a results of this leap, you will see the following outcomes:

- The big data program becomes self-sustaining, enabling more power users and analysts to engage with the platform.
- Adoption moves to an enterprise scale, driving more use cases and more value across the organization.
- ROI is accurately measured for each project and ROI for the overall initiative grows exponentially.



Conclusion

A big data journey is wrought with challenges. Technology. Data. Skills. People. Processes. But as with any journey full of challenges, they can be overcome with the combination of organization, preparedness and best practices.

Best practices are simply a set of guidelines to help you along your path. Your big data journey will remain tailored to the needs of your organization, with these best practices to help you create a unique journey. Follow these guidelines to evolve and mature your big data initiative and see the exponential value big data can bring to your organization.

FREE TRIAL
datameer.com/free-trial

TWITTER
@Datameer

LINKEDIN
linkedin.com/company/datameer