

April 2014

Smart Data
STRATEGIES

HOW TO DO A SMART DATA PROJECT

Guideline



www.altiliagroup.com

 **ALTILIA**
we make data smart

Summary

ALTILIA'S APPROACH TO SMART DATA PROJECTS	3
1. BUSINESS USE CASE DEFINITION	4
2. PROJECT PLANNING	6
3. TECHNICAL REQUIREMENTS ANALYSIS	7
4. IMPLEMENTATION AND ASSESSMENT	9

Organizations are racing to implement Big Data architectures by using different Big Data facilities currently available.

But this is not enough. What's needed is a **smart high-scalable architecture** that combines the ability to simply collect, join, consolidate, store and manage web and enterprise, structured and unstructured Big Data with the ability to process explicit data semantics (*e.g. semantic extraction and annotations, natural language processing*) and implicit data semantics (*e.g. machine learning, inference*). So developers can make better sense of data for applications and users can get faster business insights for more informed decisions and optimized business processes.

ALTILIA'S APPROACH TO SMART DATA PROJECTS

Your **Smart Data project** essentially depends from your specific business use case and company profile. Taking into account that a Smart Data Project is first of all a mean to catch a business opportunity and not a technological matter, there are **four general key steps** to successfully implement it and to give the right shape to your Smart Data business application:

1. Business use case analysis and definition – with clear identification of objectives that drive business value and criteria that enable to measure value creation.
2. Project planning – to define a project roadmap that can incrementally lead to successfully reach project objectives
3. Technical requirements analysis – to ensure you harness right data and build what you need to reach your objectives that drive business value.
4. Project implementation and assessment – to implement the plan that by fusing business case with technical requirements, and continuously evaluate the closeness between results and objectives by measuring the effects of the Smart Data projects on business variables.

A Smart Data Project is first of all a mean to catch a business opportunity and not a technological matter.

The general approach we suggest is to start small to show results with a 30 days Proof of Concept, then expand the business use case by iterating steps 2, 3, and 4. For these reasons ALTILIA shares, by this document, its experiences in planning and implementing Smart Data projects.

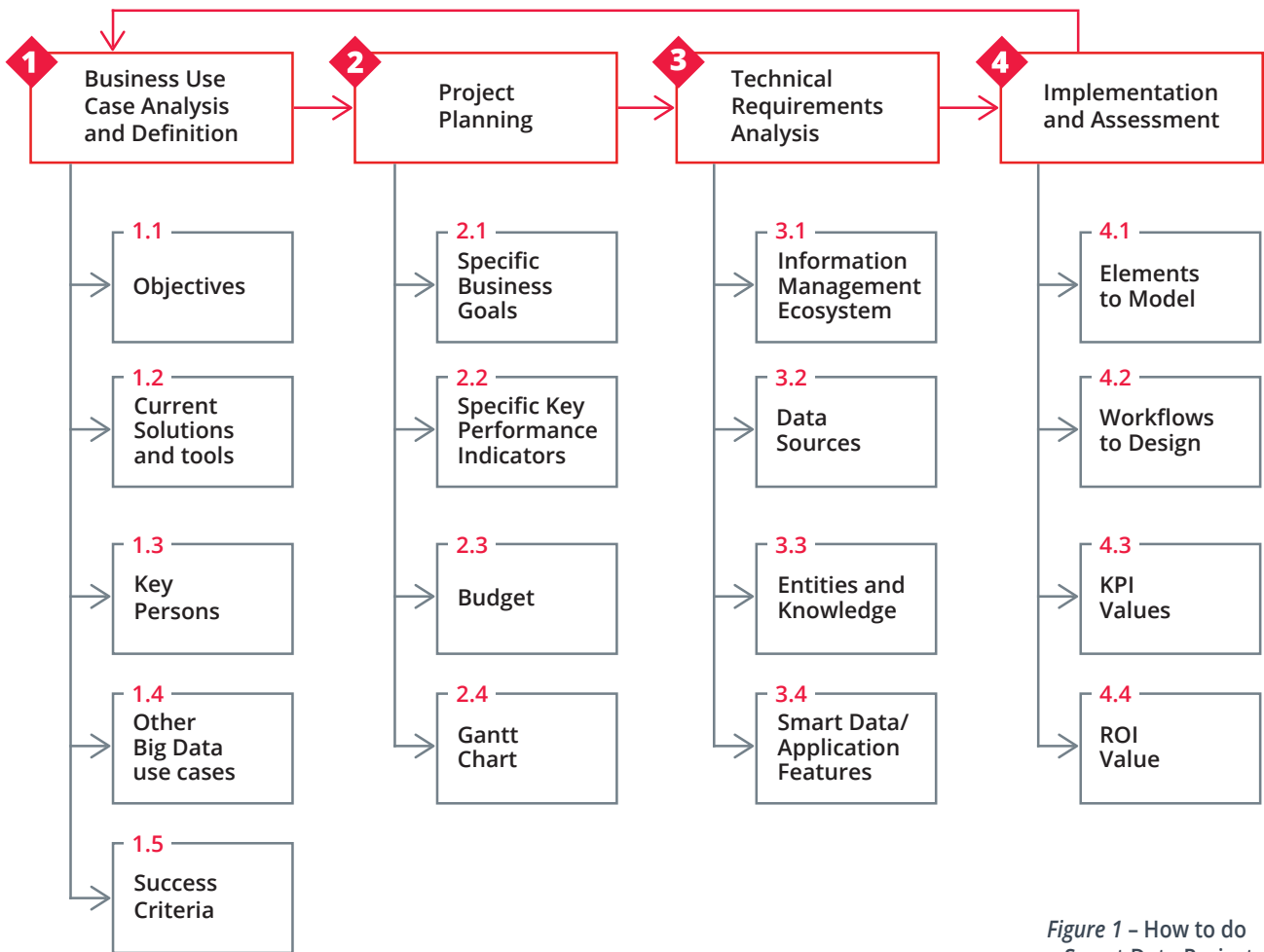


Figure 1 - How to do a Smart Data Project: Guideline with Steps.

1. BUSINESS USE CASE DEFINITION

The most important aspect to consider in defining the business use case is the objective of the Smart Data project. Business objectives may widely vary depending from company, industry and use cases. Frequent business objectives are related to:

- Revenue growth due to better customer satisfaction, new or better services and products.
- Cost saving due to more efficient and effective execution of specific business processes and operational tasks.

In order to define the business use case we suggest answering the following questions.

- **Which are the objectives of the Smart Data project?**
 - Identify the project application area such as marketing, operations, CRM, ERP, finance, sales automation, security, etc.

- Identify, understand and describe current business processes that will be influenced by the project.
- Identify application problems and define how the project may answer to application needs.
- Identify the project business goals in measurable business terms and assign priorities.
- Describe the business use case and motivations.
- **Why aren't current solutions and tools living up to expectations?**
 - Document existing obstacles, in business and technological terms, which currently hinder to reach business goals.
 - List what are deficiencies and disadvantages of the current solutions.
- **Who are the key persons (stakeholders, users, process actors, project team)? What are their roles and skills?**
 - Identify and describe business end-users and/or processes actors, their roles, skills and expectations.
 - Delineate executive sponsors, technical sponsors, and evangelists.
 - Describe the right project team composition (at ALTILIA we create team that include the following roles: project manager, application analyst/architect, software engineer, data analyst/scientist, knowledge/language engineer).
- **What are other Big Data use cases in which key stakeholders, users and project actors has been already involved?**
 - List and document problems, obstacles or success factors in other Dig Data use cases.
- **Which are the success criteria?**
 - Pick measurable criteria related to expected goals, then document.

General suggestions to take into account starting from this step are:

- Clearly identify the general impact of the Smart Data project on the organization. Does it aim at improving an existing business process or at creating a new one? Will the outcome be a new application for business end-users?
- Determine a proper scope, identify what is included in the expected outcome with dependencies, assumptions & constraints, and have a budget.

2. PROJECT PLANNING

After executing the step 1, you have a rough idea of your business use case and objectives, like “faster and smarter decision making”, “more effective business processes”, “avoid bottlenecks in data, applications and system integration”, “better customer segmentation”, and so on. Now you have to get detailed and specific. To do that, you need to answer the following questions:

- **What are specific business goals to reach?**
 - Identify and document specific success factors for each kind of target users (analyst, knowledge worker, business end-user), e.g. better user experience, less time in performing tasks, better understanding of a process, more clear and rich reports, less time to insight, less time to learn.
- **What are specific key performance indicators?**
 - Identify measurable business variables that can contribute to the evaluation of project short and long term results. For example, key metrics for different kind of target users, that measure tangible business value, are:
 - the “time to insight” which means for faster and smarter decision making.
 - Time and effort for learning how to use and maintain the application (minimal, no training?).
 - Level of ease-of-use, which means application enhancements, user experience improvement, etc.
 - Identify and document costs sources in maintaining the solution in order to be able to compute the TOC and relate this to real ROI.
 - What is the available budget?
 - Identify and define the available budget and elements that may influence it in short and long terms.
 - What is the Gantt chart of different project steps?
 - Identify and define short and long term goals so define time needed for the first iteration that lead to the first short term result.
 - Identify and document project iterations that expand the business case (Iterate on existing use cases and possibly find other use cases to improve short and long term business value).
 - Define time needed to complete each project step: technical requirements analysis, implementation steps (data and knowledge modeling, workflow design and test, workflow deploy and execution) and assessment (KPI evaluation).

At ALTILIA we plan by Gantt chart and implement by agile methodologies like SCRUM.

MANTRA Smart Data Management Platform allows user to be the subject matter experts, without need for complex implementation. Big Data scientists can concentrate their attention on data avoiding to spend time in managing the Smart Data Infrastructure.

3. TECHNICAL REQUIREMENTS ANALYSIS

To rightly achieve a successful Smart Data project you need to deep understand the digital ecosystem in which you are immersed that is composed by:

- Your internal tools, legacy systems, data/document sources, information management and business intelligence applications.
- External data/document sources (publicly available or purchasable) that you can exploit for enriching data you need to manage, analyze and explore for your business objectives and needs.
- Internal and External applications and systems you may need to integrate in your information management infrastructure.

In our experience Smart Data projects essentially can be split in two main categories aimed at obtaining Smart Data Application mostly focused on:

- *Volume* and *Velocity* of data, so applications mainly deal with structured information.
- Data *Variety*, especially referred to unstructured web and enterprise sources, and *Complexity*. In this case the project aims not only at integrate data but also at enrich data by means of information coming from unstructured sources and obtained by exploiting formalization of domain knowledge and extraction rules.

In both case Smart Data project aims at obtaining applications oriented to integrate disparate systems and data sources in which Smart Data are the lifeblood, enabling to execute better business processes and to build new, faster, and more informed decision making facilities.

Variety means that Big Data is any type of data - structured and unstructured - such as text, sensor data, audio, video, click streams, log files and more. New insights are found when analyzing these data types together.

To analyze technical requirements you have to answer the following questions:

- **How is composed the current information management ecosystem/technical infrastructure?**

- Identify internal and external tools, systems and applications (DBMS, CMS, DMS, ERP, CRM, BI, SCM, etc.).
 - Make an inventory and sketch the structure of the current architecture.
 - Identify and describe which kind of connectors you need to interact with existing systems, tools and applications.
 - Sketch the proposed architecture with data sources i.e web, enterprise, streaming, private, transactional data, survey data, web logs, in order to identify which kind of gatherers you need, etc.
- **Which are the data sources relevant for the project?**
 - Identify and examine internal and external data sources (e.g. databases, documents, web sites, emails, social media, on-line news papers, blogs, forums, RSS feeds)
 - What are Data Inputs (e.g. HTTP Req./Stream, Syslog, Batch upload) and Data Types (e.g. JSON, XML, CSV/TSV/ Delimited, Fixed Width, SQL, Binary, Text)? Are there any firewalls? Verify access to external systems, Data masking/ Anonymization, resource utilization and scheduling.
 - What data elements can you use to link this data with existing data, who owns the data, can you have access to it, do security/regulations permit putting this data into the public cloud or you need private cloud, how can you get to the data (Push/Pull), loading volumes and schedule, personal identifiable information?
 - # of channels/URLs/documents, Avg. Events per sec and max sustained value per sec, Avg. data event size.
 - Aggregation? How much ETL vs Stream Processing with calls to external APIs/ data stores, complex algorithms/logic. Are you merging various data sources?
 - Consider whether something like a social media, web content and news stream would complement your current data to create additional project value.
 - How much data per time unit, how much time in units do I want to analyze, how much data do you plan to use each month, for how long.
 - **What are entities and knowledge relevant for the project?**
 - Examine the structure and content of data sources and define attributes, entities, objects and concepts that are relevant in relation to the business objective.
 - Identify and describe type and quantity of domain knowledge and extraction rules you needed to model for capturing data and information from sources.
 - Identify and define acquisition, extraction and normalization rules and tasks.

- **Which are Smart Data and application features that fit the business use case?**
 - Define the structure of final Smart Data needed to fit business use case needs.
 - Identify and document acquisition, normalization, transformation/integration and actuation APPs/Services needed for designing contextual processing workflow that constitute the final smart data application.
 - Define the features of the final Smart Data Application that fit business use case:
 - What are Data Stores (NoSQL/Hadoop, Cassandra, SQL, RDF triplestores).
 - Deployment options: On-premise (behind firewall) or Cloud then Public Cloud or Hybrid Cloud.
 - Consider using a cloud provider like Amazon/Windows Azure for fast provisioning.
 - Do I have an I/O-heavy or CPU-heavy workload i.e. Time to insight matters more than time per query!
 - Have you considered ongoing Software Maintenance costs?
 - What are the applications e.g. BI, data mining, knowledge collaboration, visualization, you're looking to better address and do you have Big Data connectors or proven interface methods?
 - Do you have the right people in place to deploy and use this solution? Will you need to train people?

4. IMPLEMENTATION AND ASSESSMENT

After the execution of step 3 you have clear and detailed idea of the Smart Data project you have to realize. Now you have to state working on the implementation by answering the following questions:

- **Which are knowledge elements to model?**
 - Identify and define the SCRUM sprints that enable you to model extraction and navigation rules, normalization rules, taxonomies, ontologies, connectors that enable you to access and manipulate data accordingly with business use case and objectives.
 - Execute SCRUM Sprints on the base of the plan.

- **Which are workflows to design?**

- Identify and define contextual processing workflow to implement by considering apps/service to use:
 - Acquisition APPs/Services - that connect applications, systems and data sources and acquire data in the MANTRA Smart Data Layer.
 - Normalization and transformation APPs/Services - that manipulate data by normalizing, transforming, and integrating data.
 - Actuation APPs/Services – that analyze data and/or create multidimensional data cubes that allow delivering data and/or analysis results by connectors and visual interfaces.
- Test obtained workflow.
- Define and execute SCRUM Sprints needed for implementing contextual processing workflows and testing them.

- **Which are KPI Values?**

- Measure business variable values and obtain the value of defined KPI.

- **Which the ROI value?**

- Measure TOC and value gain in order to compute the real ROI.



www.altiliagroup.com

info@altiliagroup.com

Phone: +39 0984 494277

Twitter: @altiliagroup

Piazza Vermicelli, snc
87036 Rende (CS), Italy

P. IVA 10780921002