
Information technology — Artificial intelligence — Process management framework for big data analytics

Technologies de l'information — Intelligence artificielle — Cadre de gestion des processus pour les analyses des megadonnées

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024



Reference number
ISO/IEC 24668:2022(E)

AVA Documents
Document Sharing Hub
© ISO/IEC 2022

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abbreviated terms	3
5 Overview of process reference model	4
6 Process reference model	6
6.1 General	6
6.2 Organization stakeholder processes	6
6.3 Competency development processes	8
6.4 Data management processes	9
6.5 Analytics development processes	11
6.6 Technology integration processes	13
7 Overview of process assessment model	14
7.1 General	14
7.2 Process dimension	15
7.3 Process capability dimension	15
7.4 Assessment indicators	16
7.5 Process attribute rating scale	16
8 Processes and their performance indicators	17
8.1 General	17
8.2 Base practices (BPs) and information products (IPs)	17
8.2.1 Organization stakeholder processes	17
8.2.2 Competency development processes	22
8.2.3 Data management processes	26
8.2.4 Analytics development processes	30
8.2.5 Technology integration processes	35
9 Process capability indicators (Levels 0 to 5)	37
9.1 General	37
9.2 Process capability levels and process attributes	37
Annex A (informative) Mapping of indicators to process attribute outcomes:	38
Annex B (informative) Information product characteristics	41
Bibliography	50

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see <https://patents.iec.ch>).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

This document provides a process management framework for using big data analytics (BDA) across most functions of an organization. The quantum of data, the collection, storage, utilization, technology, the speed of data generation, structure and variety of data cannot be handled by the conventional data handling methods and frameworks.

This document provides a BDA process reference model (BDA PRM) and then provides process assessment model (BDA PAM). The BDA PAM are composed of two dimensions: process dimension that includes processes based on a set of PRMs including the BDA PRM and capability dimension based on process measurement framework (PMF).

This document defines a PRM and PAM as part of the framework for big data analytics, in accordance with the requirements of ISO/IEC 33004:2015 and ISO/IEC 33020:2019, for use in performing a conformity assessment in accordance with the requirements of ISO/IEC 33002:2015.

Primary audiences of this document are implementers of BDA in organizations as well as BDA capability assessors. This document provides five process categories such as organization stakeholder, competency development, data management, analytics development, and technology integration.

This framework can be used for:

- managing the processes that are considered to be best practices;
- enabling risk determination and process improvements of the incumbent organization.

Value delivered through automation, either prediction, or decision-making support or both using BDA is valuable to organizations. Implementing, improving, and assessing BDA processes based on this document expect benefits such as:

- competitive advantages;
- better decision-making;
- improve customer experiences;
- sales improvement;
- responsiveness to opportunities and threats;
- mistakes and errors reduction;
- cost reduction.

[Clause 5](#) provides an overview of PRM and [Clause 6](#) details out the specific processes under each process categories for the PRM. [Clause 7](#) provides an overview of the PAM and [Clause 8](#) provides details of process attributes and process performance indicators and [Clause 9](#) provides process capability indicators.

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

Information technology — Artificial intelligence — Process management framework for big data analytics

1 Scope

This document provides a framework for developing processes to effectively leverage big data analytics across the organization irrespective of the industries or sectors.

This document specifies process management for big data analytics with its various process categories taken into account along with their interconnectivities. These process categories are organization stakeholder processes, competency development processes, data management processes, analytics development processes and technology integration processes. This document describes processes to acquire, describe, store and process data at an organization level which provides big data analytics services.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 33001:2015, *Information technology — Process assessment — Concepts and terminology*

ISO/IEC 33003:2015, *Information technology — Process assessment — Requirements for process measurement frameworks*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 33001:2015 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1

big data

extensive datasets, primarily in the data characteristics of volume, variety, and either velocity or variability or both, that require a scalable technology for efficient storage, manipulation, management and analysis

Note 1 to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

[SOURCE: ISO/IEC 20546:2019, 3.1.2, modified]

3.2

data analytics

composite concept consisting of data acquisition, data collection, data validation, data processing, including data quantification, data visualization, and data interpretation

Note 1 to entry: Data analytics is used to understand objects represented by data, to make predictions for a given situation, and to recommend on steps to achieve objectives. The insights obtained from analytics are used for various purposes such as decision-making, research, sustainable development, design, planning, etc.

[SOURCE: ISO/IEC 20546:2019, 3.1.6]

3.3

data governance

development and enforcement of policies related to the management of data

Note 1 to entry: ISO/IEC 38500 specifies six principles of information technology governance: responsibility; strategy; acquisition; performance; conformance; human behaviour. These principles also apply to data.

[SOURCE: ISO 8000-2:2022, 3.16.1]

3.4

benefit

advantage to the organization of the actionable knowledge derived from an analytic system

Note 1 to entry: It is often ascribed to big data due to the understanding that data has potential benefit that was typically not considered previously.

[SOURCE: ISO/IEC 20546:2019, 3.1.1]

3.5

strategic plan

document specifying how data management is to be aligned to the organizational strategy

Note 1 to entry: This term has the same meaning as strategic asset management plan (SAMP) defined in ISO 55000:2014 with data point of view.

3.6

base practices

BP

activities that contribute to achieving a specific process purpose and fulfil the process outcomes when consistently performed

3.7

process attributes

PA

process features that can be evaluated on a scale of achievement to provide a process capability measure

3.8

capability level

set of process assessment indicators that together describe an ability to operate and perform a process at a given capability level

3.9

process attribute rating

judgement of the degree of achievement of the process attribute for the assessed process

3.10

process

process attribute rating

means of assessing the capabilities addressed by the defined process attributes

3.11**outcome**

observable result of the successful achievement of the process purpose

4 Abbreviated terms

BDA	big data analytics
PRM	process reference model
PAM	process assessment model
PMF	process measurement framework
BDAP	big data application provider
BDFP	big data framework provider
BDSP	big data service partner
PaaS	platform as a service
SaaS	software as a service
DevOps	development and operations
IP	information product
PoC	proof of concept
MDM	master data management
EDW	enterprise data warehouse
API	application programming interface
FMEA	failure modes and effects analysis
ER	entity relationship
SIPOC	supplier input process output customer
CTQ	critical to quality
KRA	key result area
KPI	key process indicator
BSC	balance score card
RACI	responsible accountable consulted informed
CRM	customer relationship management
ERP	enterprise resource planning
PoS	point of sale
HRMS	human resource management software

PIM	product information management
MSE	mean squared error
MAPE	mean absolute percentage error
MoM	minutes of meeting
BFSI	banking financial services and insurance
AMC	annual maintenance contract
CSM	customer service management
OSP	organization stakeholder processes
CDP	competency development processes
DMP	data management processes
ADP	analytics development processes
TIP	technology integration processes
GP	generic practices

5 Overview of process reference model

ISO/IEC 33001:2015 defines process reference model (PRM) as a model comprising definitions of processes described in terms of process purpose and outcomes, together with an architecture describing the relationships between the processes. To define a process reference model, the requirements specified in ISO/IEC 33004:2015 should be met. ISO/IEC 33004:2015, Annex A provides detailed requirements. ISO/IEC 33004:2015 requires that processes included in a PRM satisfy specific requirements.

A process description shall meet the following requirements:

- a process shall be described in terms of its purpose and outcomes;
- the set of process outcomes shall be necessary and sufficient to achieve the purpose of the process;
- process descriptions shall not contain or imply aspects of the process quality characteristic beyond the basic level of any relevant PMF in accordance with ISO/IEC 33003:2015.

[Figure 1](#) details the key process categories for advancement of data analytics in an organization. These key process categories interplay with each other in terms of the readiness of an organization to implement and deploy big data analytics.

There are 5 process categories such as organization stakeholder processes, competency development processes, data management processes, analytics development processes and technology integration processes which stand on the foundation of technology infrastructure and are guided by the organizational leadership strategy and culture. The big data analytics processes and their categories are not based on any particular organization and it is not mandated to implement them.

Organization stakeholder processes - organization top management is a key facilitative force in various ways – starting from creating shared understanding for the requirements of data analytics to mapping the benefit of embarking on such projects to strategic goals of the organization. Assignment of decision rights and accountabilities to stakeholders is key to overall success of long-term data analytics projects. The top management should also enable identification of the key data providers, data consumers, application requirements and data quality and management rules to kick start the data analytics

journey. Defining a data centric culture and mitigation of resistance to change in such a situation should also be addressed.

Competency development processes - data analytics projects need capabilities related to big data application provider (BDAP), big data framework provider (BDFP), big data service partner (BDSP) defined in ISO/IEC 20547-3. These capabilities can either be outsourced or developed from within. If outsourcing is preferred, then one needs additional competency to manage the outsourcing. Hence, relevant capability build-up and continuous maintenance and enhancement is critical for data analytics projects success.

Data management processes - data requires strong management and governance, preferably integrated with IT, information and information security management and governance which includes monitoring of emerging data sources, measurement of data quality metrics and data ownership roles. Privacy, security, policy compliance should be ensured. Specific legal requirement can apply.

Analytics development processes - data analytics development processes include data exploration, data diligence (outlier and missing value), algorithm adjustment and customization, algorithm development, algorithm validation, algorithm fine-tuning, evaluation of population stability index, etc. Analytics development processes, throughout their life cycle, rely on close cooperation with IT function of organization.

Technology integration processes - relevant technology infrastructure is required to implement data analytics. Ensure the results are formatted and optimally presented to target consumers or stakeholders. Capability should be integrated with the functional architecture. Processes to select these functional components and integrating them into the overall data analytics architecture are crucial. These processes include technology maturity evaluation, implementation approach (e.g. leveraging PaaS or SaaS) definition, and configuration or version management (e.g. DevOps).

[Figure 1](#) shows process categories and processes of big data analytics.

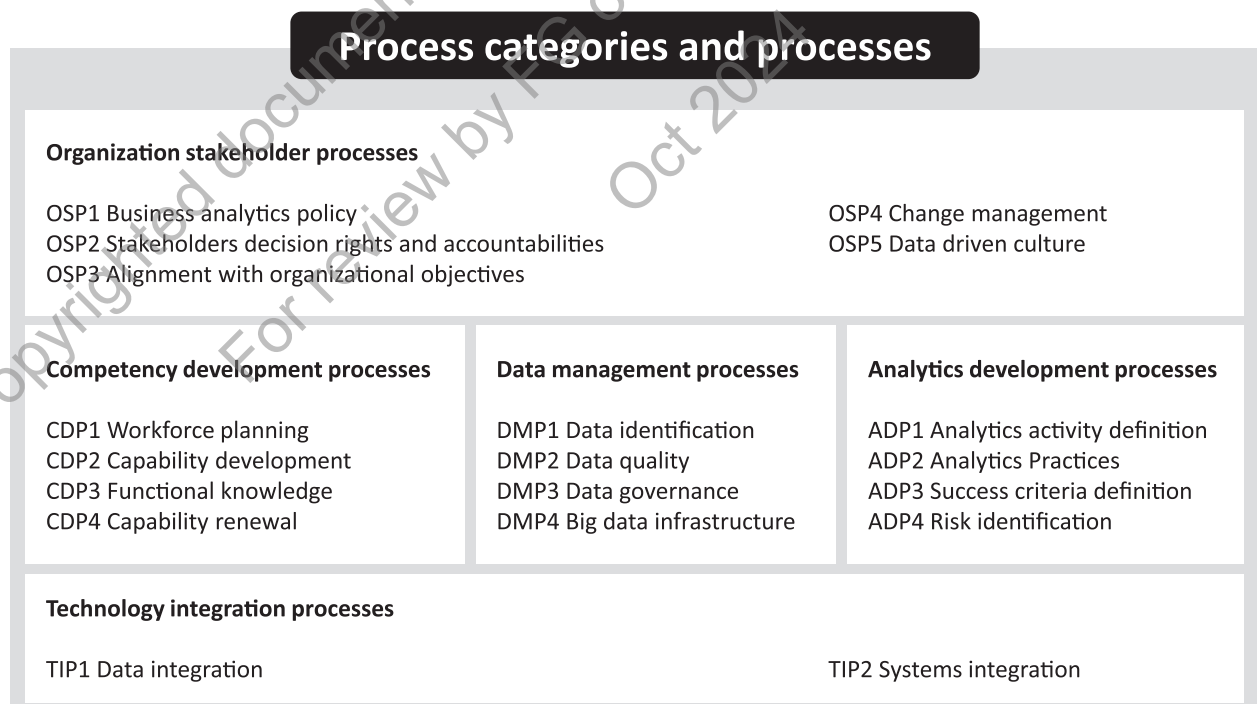


Figure 1 — Big data process categories and processes

6 Process reference model

6.1 General

[Tables 1](#) to [19](#) contain the following descriptive elements according to ISO/IEC/IEEE 24774 for each process in the PRM. The individual processes are stated in terms of process name, process purpose and process outcomes:

- a) name: the name of a process is a short noun phrase that summarizes the scope of the process, identifying the principal concern of the process, and distinguishes it from other processes within the scope of the PRM;
- b) context: for each process, a brief overview describes the intended context of the application of the process;
- c) purpose: the purpose of the process is a high level and overall goal for performing the process;
- d) outcomes: Outcomes are measurable, tangible technical or business results that are achieved by a process. They are observable and assessable.

6.2 Organization stakeholder processes

[Tables 1](#) to [5](#) contain the relevant processes related to organization stakeholders:

- [Table 1](#): OSP1 Business analytics policy;
- [Table 2](#): OSP2 Stakeholders decision rights and accountabilities;
- [Table 3](#): OSP3 Alignment with organizational objectives;
- [Table 4](#): OSP4 Change management;
- [Table 5](#): OSP5 Data driven culture.

Table 1 — OSP1 Business analytics policy

ID	OSP1
Name	Business analytics policy
Context	This process covers establishing business objectives and strategies for the organization specific to big data analytics. This involves analysing the external environment and finalizing the strategic goals and business objectives of the organization.
Purpose	The purpose of the OSP1 process is to define a policy for big data analytics initiatives, roadmap and a guideline to implement these initiatives.
Outcomes	The outcomes of this process include: a) business objectives, direction and strategies are defined and shared to the organization and relevant stakeholders; b) strategic roadmaps are developed within the constraints of the service provider resources.

Table 2 — OSP2 Stakeholders decision rights and accountabilities

ID	OSP2
Name	Stakeholders decision rights and accountabilities
Context	This process covers the assignment of stakeholders who are responsible, accountable, consulted and informed for successful implementation of big data analytics projects and initiatives.
Purpose	The purpose of the OSP2 process is to identify and assign specific responsibilities to key stakeholders.
Outcomes	The outcomes of this process include: a) key stakeholders are identified with expertise in big data technology and process or domain knowledge; b) assignment of roles and responsibilities are done; c) accountability of the stakeholders is defined; d) succession plan of the roles based on responsibilities are defined.

Table 3 — OSP3 Alignment with organizational objectives

ID	OSP3
Name	Alignment with organizational objectives
Context	This process covers the alignment of the big data analytics with the overall organization objectives. This is to ensure proper mobilization of resources, planning and arrive at actionable from the recommendations out of analytics outcomes or inferences.
Purpose	The purpose of the OSP3 process is to align an organization's big data analytics initiatives with its business strategy.
Outcomes	The outcomes of this process include: a) big data analytics initiatives, specific to relevant departments or processes are arrived; b) each of the initiatives is aligned with the stated objectives of departments or processes; c) the high-level initiatives are published across the organization with relevant stakeholders.

Table 4 — OSP4 Change management

ID	OSP4
Name	Change management
Context	This process covers the change management amongst the internal stakeholders of the organization.
Purpose	The purpose of the OSP4 process is to identify and manage people who are impacted by business analytics initiatives and manage the changes, including resistance and workarounds.
Outcomes	The outcomes of this process include: a) big data analytics initiatives or projects are subjected to progress monitoring and reviewed against expected outcomes; b) progress is communicated to stakeholders; c) the impact of the changes, issues and improvement is analysed and reported; d) awareness sessions and trainings are organized across the organization for different roles of stakeholders regarding big data analytics.

Table 5 — OSP5 Data driven culture

ID	OSP5
Name	Data driven culture
Context	This process covers the organizations shared values and mission statements depicting the data driven decision making.
Purpose	The purpose of the OSP5 process is to create decision making processes based on data, analytics and related set of fact-based systems to attain better strategic intelligence capability.
Outcomes	The outcomes of this process include: a) process or business performance always measured in metrics; b) the metrics comprise of leading and lagging indicators; c) quantitative analysis is encouraged with possible statistical correlations; d) possible big data analytics initiatives or projects are discussed and explored on problem or opportunity areas during reviews; e) quick wins or successful initiatives on big data analytics should be rewarded and communicated; f) a framework for piloting ideas exploring big data analytics for current processes or new areas of business should be rolled out (similar to a Kaizen framework in many organizations).

6.3 Competency development processes

Tables 6 to 9 contain the relevant processes related to competency development:

- [Table 6](#): CDP1 Workforce planning;
- [Table 7](#): CDP2 Capability development;
- [Table 8](#): CDP3 Functional knowledge;
- [Table 9](#): CDP4 Capability renewal.

Table 6 — CDP1 Workforce planning

ID	CDP1
Name	Workforce planning
Context	This process covers the talent forecasting and resource estimation of the organization for executing big data analytics projects and initiatives.
Purpose	The purpose of the CDP1 process is to arrive at plans to ensure availability of workforce and other resources for executing big data analytics projects and initiatives.
Outcomes	The outcomes of this process include: a) identify future leaders in big data analytics; b) align relevant responsibilities and craft succession plans in critical roles; c) recruit the right talent.

Table 7 — CDP2 Capability development

ID	CDP2
Name	Capability development
Context	This process covers the empowerment and training of employees of the organization for executing big data analytics projects and initiatives.
Purpose	The purpose of the CDP2 process is to support people throughout the organization to achieve their plans, objectives and targets through capability enhancement.
Outcomes	The outcomes of this process include: a) understand and develop the underlying capabilities of the organization; b) evaluate the set of results achieved to improve future performance and provide sustainable benefits to all their stakeholders; c) recognize their efforts and achievements in a timely and appropriate manner

Table 8 — CDP3 Functional knowledge

ID	CDP3
Name	Functional knowledge
Context	This process covers the identification and deployment of industry domain workforce as part of the big data analytics teams at all levels.
Purpose	The purpose of the CDP3 process is to encourage the BDA leaders to learn quickly and rapidly respond with accountability in their individual enhanced role.
Outcomes	The outcomes of this process include: a) encourage stakeholders to participate in activities that contribute to the wider acceptance to big data analytics initiative across the organization; b) use approaches to understand, anticipate and respond to the different needs and expectations from big data analytics implementation team; c) promote a culture which supports the generation of new ideas and new ways of thinking to encourage innovation through big data analytics.

Table 9 — CDP4 Capability renewal

ID	CDP4
Name	Capability renewal
Context	This process covers identification, analysis and understanding of external indicators, such as global and local economic, market or societal and technology trends, which can affect the organization and translate these into potential future scenarios for big data analytics initiative.
Purpose	The purpose of the CDP4 process is to use a structured approach for generating and prioritizing creative ideas and allocating resources to execute them innovatively within appropriate timescales with updated tools and techniques of big data analytics.
Outcomes	The outcomes of this process include: a) identify, evaluate and develop new and emerging technology portfolio to improve the agility of organization; b) establish and manage learning and collaboration networks to identify opportunities for creativity, innovation and improvement in technology and human resources.

6.4 Data management processes

Tables 10 to 13 contain the relevant processes related to data management:

— Table 10: DMP1 Data identification;

- [Table 11](#): DMP2 Data quality;
- [Table 12](#): DMP3 Data governance;
- [Table 13](#): DMP4 Big data infrastructure.

Table 10 — DMP1 Data identification

ID	DMP1
Name	Data identification
Context	This process covers the key step of identifying data elements. This involves identification of data elements so that the team will not lose sight of any important dimension or factor that plays a key role in the outcome analysis.
Purpose	The purpose of DMP1 process is to identify, define, classify, and collect data for all data elements available for the information flow in the context of the project or department or function.
Outcomes	The outcomes of this process include: a) data elements relevant to the process or function or department are identified; b) data elements can be classified into categories, such as unstructured, transactional, hierarchical, reference; c) meta data (operational definition) such as units, frequency, source or sources, functional definition, range, possible functional correlation, producer or consumer or ownership or steward is derived; d) new data sources are identified; e) data collection is performed.

Table 11 — DMP2 Data quality

ID	DMP2
Name	Data quality
Context	This process covers the key aspects of a practical, comprehensive, and well-managed data quality strategy that can eliminate scattershot efforts in different business units and help ensure that business users throughout an organization have access to consistent and accurate information.
Purpose	The purpose of DMP2 process is to identify that a program should address the root causes of data inconsistencies, fix errors through data cleansing, and unite separate data quality initiatives.
Outcomes	The outcomes of this process include: a) structured data has normalized relational mapping; b) unstructured data has valid references to structured data; c) data complies on the aspects of accuracy, completeness, timeliness, validity, consistency, integrity, etc; d) orphaned or inconsistent data should either be sanitized from the resulting dataset or filtered if a live dataset is used.

Table 12 — DMP3 Data governance

ID	DMP3
Name	Data governance
Context	This process covers the key steps of establishing data governance for the organization. An important aspect of data is correctly defining the ownership of the data. At times this can be a very difficult task. Many organizations think information technology (IT) should be in charge of the data because IT owns the system where the data are housed however, IT is rarely the actual owner of the data. When establishing the owner, it is important to understand who can answer questions about the data, provide definitions to the attributes, and determine the validity of the data. Those people are usually the true owners of the data. These people need to be involved in defining business rules for data cleansing, correcting the data, and matching and consolidation.
Purpose	The purpose of DMP3 process is to identify tools and define broad ranges of processes to implement an effective data governance across the organization.
Outcomes	The outcomes of this process include: a) a governing council is established; b) data stewards are identified; c) business rules associated with consolidating and updating data are defined; d) implementation and sustenance plans are defined.

Table 13 — DMP4 Big data infrastructure

ID	DMP4
Name	Big data infrastructure
Context	This process covers the key steps of implementing big data infrastructure for enabling data analytics. The big data paradigm is a rapidly changing field with rapidly changing technologies. Very few organizations operate solely on data organic to that organization these days. This means that systems that collect and analyse big data need to be able to securely and reliably interoperate and share data.
Purpose	The purpose of DMP4 process is to implement a big data infrastructure which is a system that leverages big data engineering and employs a big data paradigm to process big data.
Outcomes	The outcomes of this process include: a) Appropriate software and tools for distributed systems and storage and mining (NoSQL and relational databases, distributed file systems, and other distributed processing system) are identified; b) system implementation executed; c) big data architecture is identified and implemented; d) compliance with security and privacy aspect of big data is ensured; e) skill enablement is executed.

6.5 Analytics development processes

Tables 14 to 17 contain the relevant processes related to analytics development;

- Table 14: ADP1 Analytics activity definition;
- Table 15: ADP2 Analytics practices;
- Table 16: ADP3 Success criteria definition;
- Table 17: ADP4 Risk identification.

Table 14 — ADP1 Analytics activity definition

ID	ADP1
Name	Analytics activity definition
Context	This process covers planning the analytics activities in terms of enterprise-wide scope, targets and having a consistent perspective for analytics for the organization.
Purpose	The purpose of the ADP1 process is establish an analytics implementation plan to manage a unified big data and analytics platform.
Outcome	The outcomes of this process include: a) identification of small-scale analytics projects that suggest cross-functional or enterprise potential; b) identification of business areas that can have potential benefit from analytics; c) focus on high value and high impact targets; d) establish a priority matrix including criteria for executing projects; e) successful implementation of projects.

Table 15 — ADP2 Analytics practices

ID	ADP2
Name	Analytics practices
Context	This process covers the analytics implementation facets for the organization and develop a process to choose the combination of techniques for analytics development for or in the organization.
Purpose	The purpose of the ADP2 process is to establish consistent technology practices across the organization in implementing big data analytics projects. The process focuses on the analytics implementation and not on IT practices and data infrastructure needed for up-stream and downstream integration and deployment.
Outcomes	The outcomes of this process include: a) life cycle methodology selection criteria and guidelines; b) architecture considerations and selection guidelines; c) guidelines on analytics methods selection; d) tools selection criteria and guidelines.