Workstream Big Data

Big Data Guide

v2.0

Authors:
Barbara Engerer
Jörg Hetterich
Frank Cersovsky
Jürgen Nguyen
Dr. Alexander Hildenbrand

cross business architecture lab

Management Summary

- Mission Statement:
  The digital world allows the availability of vast amounts of data (internal and external) concerning new insights und business models. This raises the question of what architecture and frameworks are needed to control and utilize the amount of data?

  In addition, this workstream clarifies the various concepts regarding the topic Big Data / Industrial Analytics, the architectures, frameworks and skills that are needed. Existing reference architectures as well as the needs of the company are analyzed.
- Goal:
  Definition of a big data guide, which includes the following content

Overview

- The purpose of this "how-to guide" is to provide a comprehensive approach for setting basics of flexible and adaptable big data solution.
- This guide based on a modular structure. The modules are grouped into four distinct planning dimensions:

- In general, the planning procedure traverses the four dimensions from left to right. However, an parallel execution is possible to a considerable extent.

Overview – how-to guide

Steps

Module 1: Align with business drivers, customer expectations and behavior

Module 2: Classify use cases & identify data

Module 3: Assess your org. capabilities, IT landscape and maturity / Assess the viability

Module 4: Build an IT-strategic vision and roadmap for big data computing

Module 5: Design big data reference architecture

Module 6: Establish and enable big data platforms, infrastructures and knowledge (run time)

Module 7: Establish and enable big data development environments and expertise

Module 8: Extend IT-/ Data Governance Model

Module 9: Enable big data development processes & policies

Module 10: Enable big data operation processes

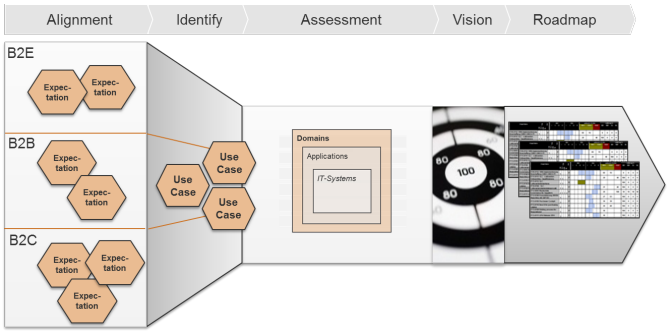
Module 11: Enable big data support processes

Module 12: Assess additional risks and impacts

Module 13: Review legal obligations and contractual status

Introduction: Strategy dimension

- **Guidance on how to create a vision and strategy for big data scenarios out of different expectations from various stakeholders.**
- **Alignment of the expectations to the current organization and IT-Landscape is necessary to get a reliable roadmap.**



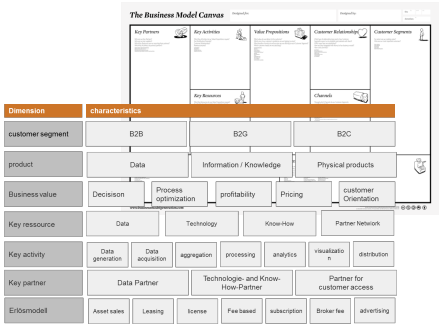
Strategy

The following modules describe important aspects to consider:

- Align with business drivers, customer expectations and behavior
- Classify use cases & identify data
- Assess your org. capabilities, IT landscape and maturity
- Build an IT-strategic vision and roadmap for big data computing

Module 1: Align with business drivers, customer expectations and behavior

Customer expectations and behavior: enables people to self-serve on the data and acquire insights they need to make informed decisions



Involved Roles

Business Leaders, ...

What to do?

1. Subject matter experts (SME) assemble a working group of data subject-matter experts (SMEs) across all of the functions and divisions within the business.
2. Identity the right data stakeholders e.g. Chief data officer, Data governance program leader, Business data steward, Enterprise/information architect, Business data analysts
3. Create Awareness of Big Data,
4. Gather expectations and ideas based on business canvas model

Stakeholder analysis (1/3)
Who addressed? What is his position?

cross business architecture lab

Stakeholder	Influence (H/M/L)	Support Defecit		Type of personality and conviction strategy	How would he benefit from the change?	Who can influence him / her?	Required stept to involve stakeholder?
		Current Mood*	Support, that we need from him+				

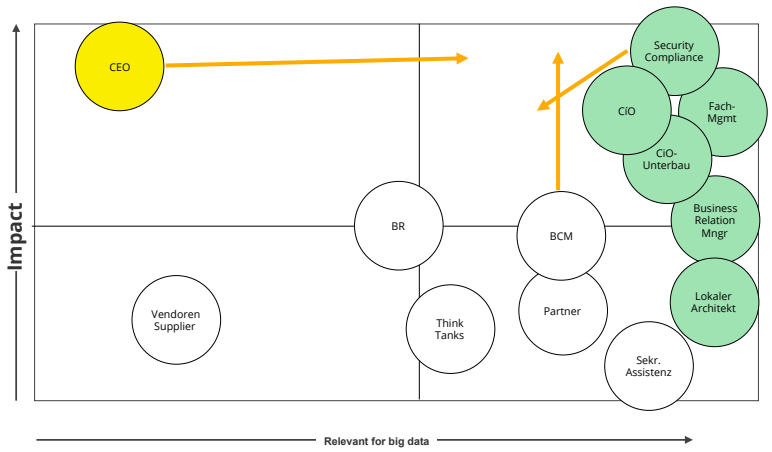
*** resistant (R), neutral (N), acepting (A), support (S), committed (C)**
+ resistant (R), neutral (N), acepting (A), support (S), committed (C)

Stakeholder analysis (2/3)
Who addressed? What is his position?

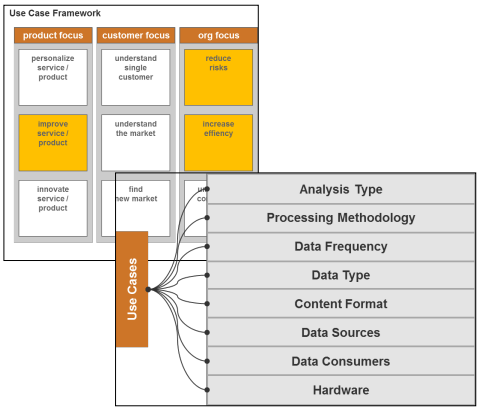
cross business architecture lab

- **Resistant**
  He clearly shows that he is against the change; gathers reasons why this will not work;
- **Neutral**
  Shows neither pro or contra against the change; is not seen as a barrier; is often not interested, is hard to «read»
- **Accepting**
  Is willing to accept the change; does what needs to be done; he will be cautious
- **Support**
  Positive attitude towards the change; he will acknowledge the need for change.
- **Committed**
  He will promote the initiative actively with his time and resources. He is visible and active.

Stakeholder analysis (3/3)
Who addressed? What is his position?



Module 2: Identify use cases and requirements for big data



Involved Roles

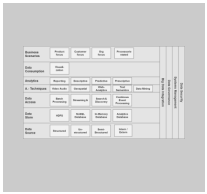
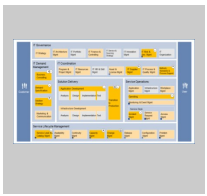
Business Leaders, ...

What to do?

1. Identify use cases
2. Classify and prioritize the use cases regarding
3. Identify characteristics of big data scenarios
   - Analyze Requirements:
   - Device/App. capabilities
4. Architecture capabilities / Infrastructure capabilities
   - Operational capabilities

Module 3: Assess your org. capabilities, IT landscape and maturity (1/3)

cross business architecture lab

AS IS IT land-scape		Evaluating the actual landscape creates transparency regarding the current big data capabilities of the company
Org. Capabi-lities		Evaluating the organization creates transparency regarding the current capabilities of the IT organization

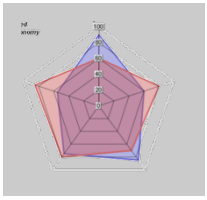
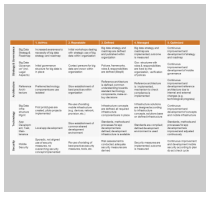
What to do?

Prior to the development of an IT strategy, the different elements of the IT landscape have to be evaluated in relation to the envisaged use cases. The following steps are necessary:

1. Develop an overview: assessment of the current landscape (see elements)
    1. Development plan (domains, applications)
    2. Systems, interfaces, processes, devices
    3. Development infrastructure
    4. Organization chart
2. Organization / responsibility
    1. Clarify responsibilities
    2. Evaluate skills of employees

Module 3: Assess your org. capabilities, IT landscape and maturity (2/3)

cross business architecture lab

Gap-Analysis		The gap analysis regarding the required use cases defines the changes of the IT landscape.
Maturity Level		The current capability to build the big data scenarios can be assessed by determining the degree of maturity.

What to do?

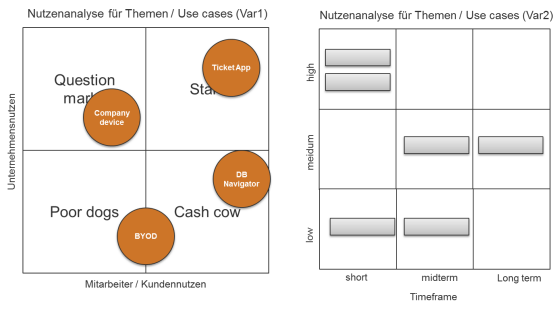
3. Gap analysis regarding the identified use cases of Module 2
    a. Capabilities of the company
    b. Required organizational structures
    c. Identify and assess needed action
4. Maturity level: Determination of the capability to build the big data scenarios
    a. Identify and assess needed action

Module 3: Assess your org. capabilities, IT landscape and maturity (3/3)

		1. Ad-Hoc	2. Repeatable	3. Defined	4. Managed	5. Optimized
Strategy & Governance	Big Data Strategy& Roadmap	Increased awareness to necessity of big data strategy and roadmap	Initial workshops dealing with strategic use of big data within organization	Big data strategy and roadmap are defined and published within organization	Big data strategy and roadmap are implemented, outcome is measured	Continuous improvement and development of strategy and roadmap
	Big Data Governance* (incl. Legal Aspects)	Initial governance structure for big data is in place	Contact persons for big data are known within organization	Polices, frameworks, roles & responsibilities are defined (Step9)	Gov. structures with roles & responsibilities are lived by the organization, verification of polices	Continuous improvement and development of mobile governance
Architecture	Reference Archi-tecture	Preferred technology components are use isolated	Establishment of best practices within organization; initial re-use of architecture	Reference architecture is defined, common understanding towards standard technology components, make-or-buy decisions	Reference architecture is implemented, mechanism to check compliance is implemented	Continuous improvement and development reference architecture due to internal and external changes (e.g. technological progress)
Technology	Big Data Infra-structure Mgmt	First prototypes are created, pilots projects implemented	Re-use of existing big data infrastructure (e.g., network, provision, etc.)	Infrastructure concepts are defined, all required infrastructure components are in place	Infrastructure solutions are designed according to infrastructure concepts, solutions base on defined infrastructure	Continuous improvement and development of concepts and mobile infrastructure
	App Develop ment, Test, Main-tenance	Local app development	Establishment of common shared development environment	Standards, methods and processes for app development are defined; development infrastructure is available	Standards are complied; defined development environment is used	Standards, methods and processes for app development are improved and adjusted continuously

Module 4: Build an IT-strategic vision and roadmap for big data computing (1/2)

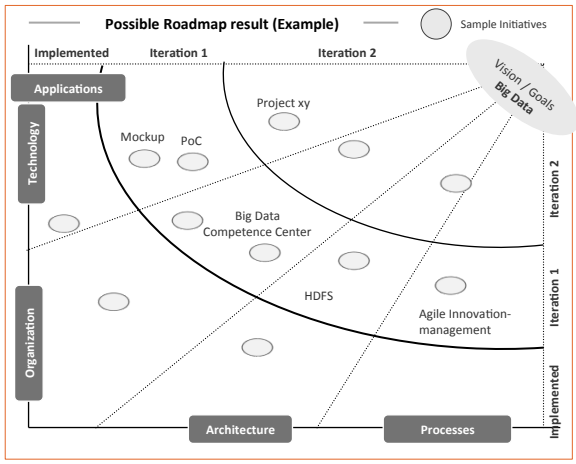
Different ways of performing value benefit analysis



What to do?

1. Carrying out the value benefit analysis / portfolio matrix of each use cases

Module 4: Build an IT-strategic vision and roadmap for big data computing (2/2)

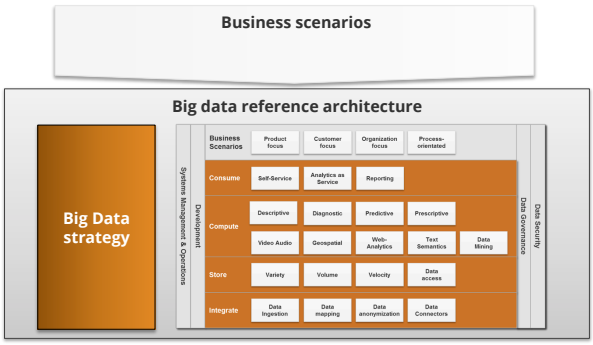


What to do?

2. Definition of goal and vision.
3. Define initiatives / analyze internal (cloud, social etc.) dependencies.
4. Prioritization (based on iterations)
5. Use Case / Make or Buy Decision
6. Allocate Budget & Responsibilities
7. Create strategy document
8. Visualize and present gaps and needs for actions to the management

Technology dimension Introduction

- **Guidance on how to establish and enable big data strategy, reference architecture, management of big data platform, infrastructure and development into the organization**
- **The big data reference architecture is used as a basis**

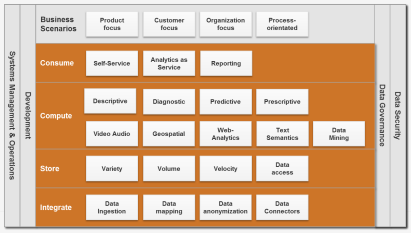


Technology

The following steps describe important aspects to consider when

- Design big data reference architecture
- Establish and enable big data platforms, infrastructures and knowledge (run time)
- Establish and enable big data development environments and expertise

Module 5: Design Big Data reference architecture



Big Data reference architecture comprises documents with directive character to select :

- Technology decision (e.g. data store, data virtualization)
- Products decision (e.g. vizualization)
- Security principles (e.g. encryption)
- Data integration / Policies
- Bodies / responsibilities
- Governance board and process

What to do?

1. Use existing BI architecture as starting point to derive Big Data reference architecture
2. Consider dependencies towards other reference architecture (e.g. cloud reference architecture)
3. Conduct gap analysis based on the capability model
4. Conduct a cost / benefit analysis to close gaps (business case, PoC)
5. Develop architecture principles / pattern for big data use cases

Goal:

- Reference architecture improve communication with stakeholders by providing an visual overview with common terminology

Module 6: Establish and enable big data platform, infrastructure and knowledge (1/2)

Big Data Reference Architecture

Checklist

- Decentral vs. central Big data instance?
- Type of data management (Data Lake, data silos, etc.)?
- Availability of big data service?
- Type of data interfaces to data sources?
- Number of users?
- Licenses?
- Make of buy?
- Lifecyle cost?

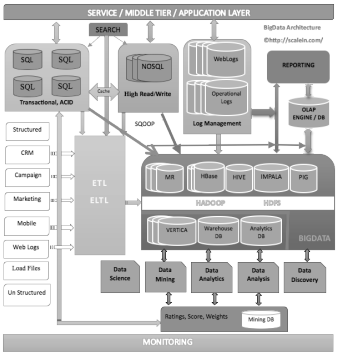
What to do?

Required Role: enterprise architect, procurement, system support, developer,

1. Create Checklist (central vs. Decentral big data instance, type of data
2. Conduct gap analysis
3. Conduct market analysis of big data solutions and providers
4. Evaluation / Proof of Technology incl. decisions
5. Define technology roadmap
6. Procure big data solutions
7. Build big data infrastructure / big data skills
8. Define big data operating model (Documentation, SLAs, etc.)

Module 6: Establish and enable big data platform, infrastructure and knowledge (2/2)

cross business architecture lab



With strategic requirements and mobile reference architecture a concrete solution for big data platform, infrastructure and knowledge can be defined (building blocks).

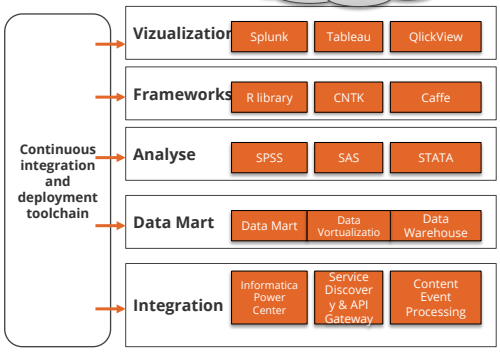
What to do?

1. Create a checklist (central vs. decentral big data instance, type of data
2. Conduct gap analysis
3. Conduct market analysis of big data solutions and providers
4. Evaluation / Proof of Technology incl. decisions
5. Define technology roadmap
6. Procure big data solutions
7. Build big data infrastructure / big data skills
8. Define big data operating model (Documentation, SLAs, etc.)

Module 7: Establish and enable mobile development environments and expertise

cross business architecture lab

Definition — CRISP-DM

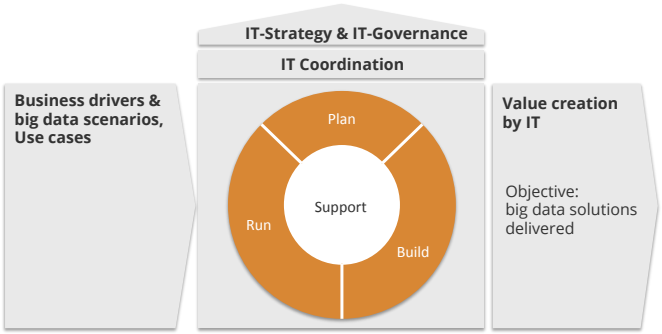


What to do?

1. Data security regulations
2. Consider guidelines & policies
3. Conduct data quality assurance
4. Build continuous integration and deployment toolchain
5. Establish and train standard tools / frameworks for analysis and visualization
6. Define Concepts based on existing BI concepts
   1. Testing concept
   2. Governance concept
   3. Lifecycle concept
   4. Training concept
   5. Support concept
7. Evaluate the ROI for use cases

IT-Processes & Policies Dimension Introduction

- **Guidance on how to <u>embed</u> big data strategy, management of technology & knowledge supporting processes, rules, roles and responsibilities into the <u>existing</u> organization**
- **A simplified and generic IT-Process Model based on the Plan-Build-Run-Support paradigm will be used:**



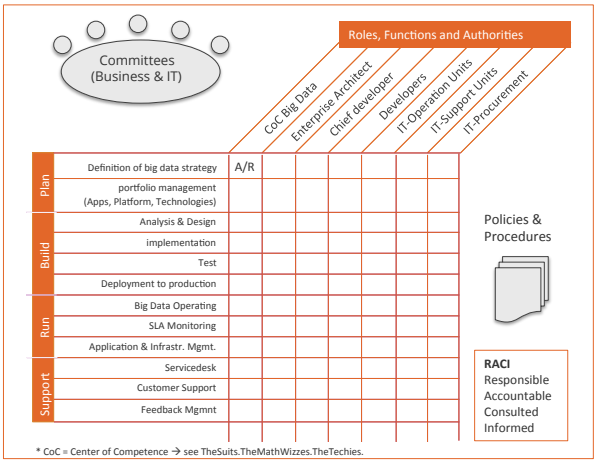
IT-Processes & Policies Dimension

The following modules describe important aspects to consider when

- Extending your existing IT-Governance Model
- Enabling big data IT-Development Processes (Plan / Build)
- Enabling big data IT-Operation processes (Run)
- Enabling big data IT-Support processes (Support)

In order to fulfill the objective of successfully delivering big data solution to your business.

Module 8: Extend IT-Governance Model (Rules, Roles, & Responsibilities)



Definition

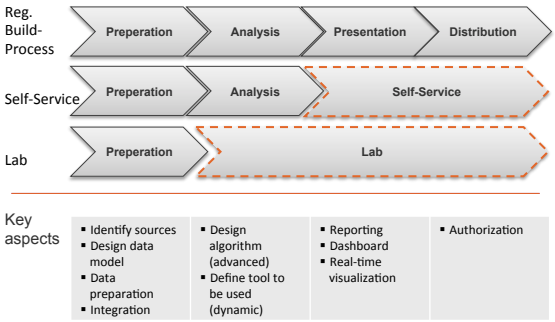
An IT-Governance Model defines roles, responsibilities and decision authorities

What to do?

1. Evaluate big data demands (inputs of all other steps)
2. Gap analysis with existing IT-Governance structures and determination of adaption needs
3. Establish required committees or decision bodies, obtain commitment and assign responsibilities
4. Assign responsibilities on a process level (RACI)
5. Communication of decisions
6. Periodic review and improvement

Module 9: Enable Big Data Development Processes & Policies (1/2)

Big Data Development Processes - Overview



Key aspects

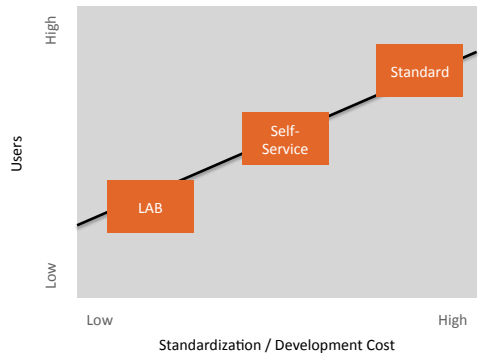
▪ Identify sources ▪ Design data model ▪ Data preparation ▪ Integration	▪ Design algorithm (advanced) ▪ Define tool to be used (dynamic)	▪ Reporting ▪ Dashboard ▪ Real-time visualization	▪ Authorization

What to do?

1. Identify specific challenges / distinguishing characteristics of big data development
2. Recommend and provide standardized solutions or frameworks
3. Enforce guidelines, policies & procedures, e.g. coding guidelines, quality standards for internal and external developers
4. Enforce quality assurance of mobile applications
5. Define and communicate big data computing development processes

Module 9: Enable Big Data Development Processes & Policies (2/2)

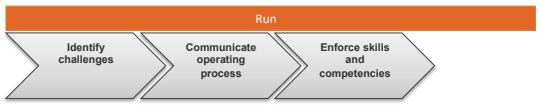
Big Data Development Processes – Cost



What to do?

1. Identify specific challenges / distinguishing characteristics of big data development
2. Recommend and provide standardized solutions or frameworks
3. Enforce guidelines, policies & procedures, e.g. coding guidelines, quality standards for internal and external developers
4. Enforce quality assurance of mobile applications
5. Define and communicate big data computing development processes

Module 10: Enable Big Data Operation Processes



Distinguishing characteristics from traditional operating

- Speed: Requirements need fast reaction time
- Resource management needs new concept for flexibility
- Close collaboration with projects
- Release management will get more complex
- SLA-Monitoring
- New skills might be required

What to do?

1. Timely involvement of operations within big data projects
2. Change resource management model to enable / satisfy ad hoc capabilities in terms of
   - Computing resources
   - Storage resources
   - Flexible licenses model
3. Build cross functional team with business department and development
   - Define clear responsibilities
   - Provide and integrate tools
   - Communicate errors & fixes
4. **Speed** is the challenge and reason for the requested close collaboration of basis, operations and project team.
   Remark:
   DevOps is one possible concept.

Module 11: Enable big data support processes

Big Data Support

- **Standard** — **1st Level (Standard User)**
- **Self-Service** — **2nd Level (Power User)**
- **Lab** — **3rd Level (Expert)**

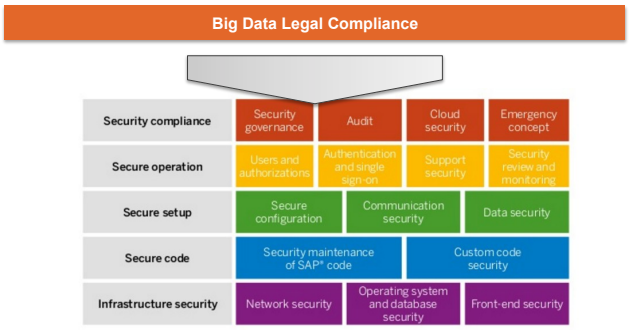
Standard	Self-Service	Lab	Support Areas
		X	Resource management
	X		Cost Management
	X		Integration3rd-Party (Cloud,..)
X	X		Data Model
X	X		Authorization
X	X	X	Training (tools, data privacy, security)
	X	X	Licenses

What to do?

1. Establish a big data support strategy
2. Establish technical support
3. Provide mobile devices for the first-level-support
4. Train your personnel in tools

Introduction: Security & Compliance

- **Guidance on how to assess risks and impacts and to review legal obligations and contractual status in the field of big data**



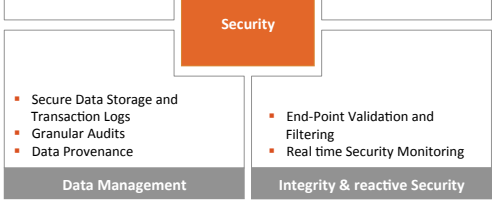
Source: SAP Securiry Solution Map

Security & Compliance

The following modules describe important aspects to consider when

- Assess additional risks and impacts
- Review legal obligations and contractual status

Module 12: Security



What to do?

1. Envolve Legal department per use case regarding combination of data sources, e.g. by developing a data privacy template
2. Create data security concept out of data privacy regulations based on country law and regional/local regulations
3. Ensure Legal ok for use of personal data (anonymization, aggregation, singling out, encryption…),
4. Ensure authority concepts on all data sources
5. Management of external data consumers and authority
6. Prevent from data leaks
7. Clarify open source challenges
8. Check data classification for cloud (private/public) usage

Module 13: Law Compliance

You should aware of these areas when dealing with customer related data…

What to do?

1. Identify relevant law & regulations
2. Conduct legal & risk assessment based on the organizational und business environment
3. Define tasks to ensure compliance
   a. Security Framework
   b. Reflect IT- and organizational processes
4. Establish governance and guidance
   a. Define policies & responsibilities
   b. Training & communication of policies and processes
   c. Continuously improvement

Module 13: Law Compliance

e.g. BYOD Checklist

Checklist	Checked?
Determine data ownership	
▪ Have you determined what portion of the data you own?	
▪ Have you classified the data you own as sensitive/personal or personally identifiable/non-personal?	
▪ Have you tailored your data protection strategy to specific data categories in an efficient manner?	
Gather customer consent	
▪ Is your communication with your customers on what data you collect transparent and unambiguous?	
▪ Are your mechanisms for gathering customers' consent active (e.g., a tick box)?	
▪ Is the request for consent highlighted in your terms and conditions (e.g., a different text color)?	
▪ Is the privacy policy available in local languages?	
▪ Are the subject rights included in the terms and conditions?	
▪ Are links to other privacy resources easy to find?	
▪ Can users — sharing a device — create different accounts with different privacy settings?	
Communicate the purpose	
▪ Is your communication on why you collect and process data transparent and unambiguous?	
▪ Have your given your customers meaningful opt-out options (e.g., opt out from data processing)?	
▪ Have you anonymized data processed for secondary purposes?	
Feed data into social media	
▪ Do you gather permission from customers each time you feed data into/derive data from social media?	

Module 13: Law Compliance

e.g. BYOD Checklist

Checklist	Checked?
Store and process data in the cloud	
▪ Have you communicated unambiguously about data storage and/or processing in the cloud?	
▪ Are you aware of the geographical location of your cloud provider's HQ and eventual subcontractors?	
▪ Have you obtained and reviewed the list of cloud subcontractors?	
▪ Have you reviewed your cloud provider's infrastructure security, security measures, and privacy policy?	
▪ Do you encrypt your data at rest and in motion? Do you hold the encryption keys?	
▪ If data is transferred outside the EU, does your company comply with international data transfer rules?	
▪ If data is transferred outside the EU, does your provider comply with international data transfer rules?	
▪ If sensitive data is transferred outside the EU, do you comply with specific regulations?	
Analyze the data	
▪ Are you processing only data for which you have gathered your customers' consent?	
▪ Are you processing data for the purpose you stated in your terms and conditions?	
▪ Are you collecting and processing only the data that you need to?	
Share data with third parties	
▪ Have you communicated unambiguously to your customers about sharing data with third parties?	
▪ Have you organized your data in dispersed data sets?	
▪ Have you given third parties access only to the portion of data they need?	
▪ Are third parties handling data for the purpose stated in your T&Cs or for compatible purposes?	
▪ Have you anonymized data shared with third parties for processing other than for the main purpose?	

Das **Cross-Business-Architecture Lab** ist ein Verband von Anwendern für Anwender.

Das CBA Lab erarbeitet mit und für seine Mitglieder innovative „Bausteine" für die Digitale Transformation, die die Architektur prägen und organisieren. In der Praxis erprobte Best Practices werden geteilt und damit weiter veredelt zu Leading-Edge-Ergebnissen, die belastbar und sofort nutzbar sind.

Das Cross-Business-Architecture Lab ist offen für die Aufnahme weiterer Anwenderunternehmen.

Weitere Informationen unter **www.cba-lab.de**.

Cross-Business-Architecture Lab e. V.
Artquadrat | Emil-Nolde-Straße 7 | 53113 Bonn

Tel. +49 228 26 73 179
Fax +49 228 976 720 7
E-Mail: info@cba-lab.de
Web: www.cba-lab.de
Community: https://cba-lab.gocoyo.com
Twitter: https://twitter.com/cba_lab