What is eXtollo?

Welcome to eXtollo!

This page gives an introduction to what eXtollo is, the history of it, and how it is usually used.

What is eXtollo? (in a nutshell)

eXtollo helps Daimler's business units to perform advanced analytics and AI use cases on very large amounts of data (big data) with high flexibility. eXtollo's main building blocks are the Data Lake, which serves data from the whole company, and the eXtollo Instances, which provide a Azure-based toolset for Big Data and Machine Learning use cases. Customers using eXtollo have a multitude of benefits, including

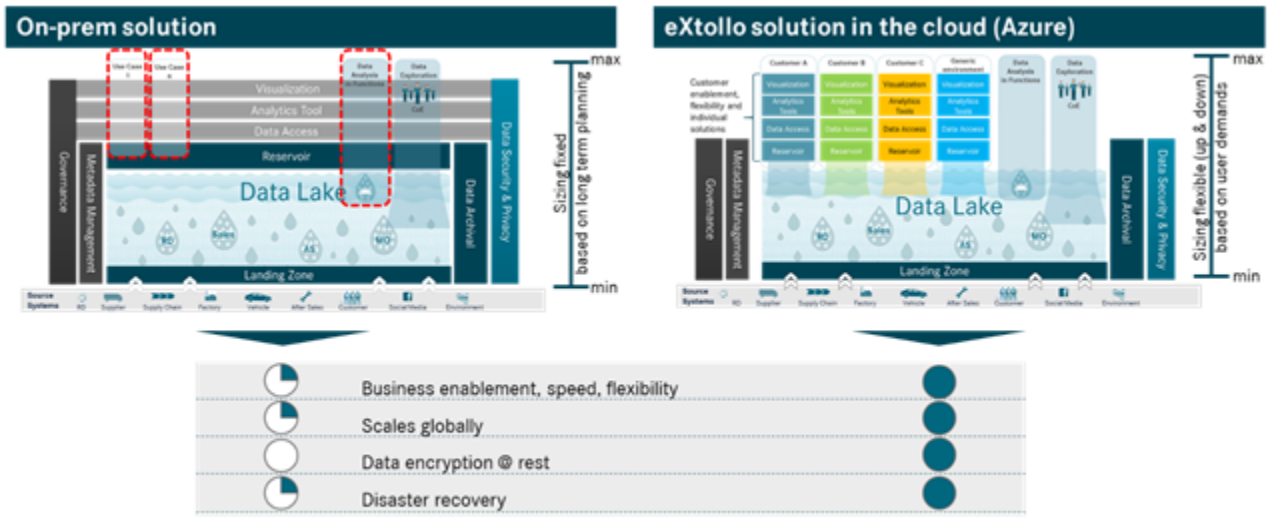
- a ready-to-use, secure and flexible platform with no need for the customer to go through a Data@Cloud before usage,
- access to the extensive eXtollo Data Lake,
- a well-experienced operations team supporting customers in managing and supporting their environment(s).

Why A New Big Data Platform?

Data Is The New Oil, so Daimler IT makes an effort to find, refine and retain it in big quantities as well as draw the most beneficial business decisions from it. State-of-the art solutions to process big data, like Hadoop, have been employed at Daimler IT for years, but the current on-premise implementations displays certain limitations.

About eXtollo

With our eXtollo cloud solution we avoid the shortcomings of on-premise Big Data & Advanced Analytics set-ups and unleash full value generation and digital enablement within Daimler. Moreover, eXtollo comes with a signed Cloud Risk Process (CRP) – customers do not have to run through an individual CRP (explained in the CRP section below) and can work with data for analytics use cases classified as up to confidential.



Focus on value creation by offering a flexible Big Data platform which provides future-proof advanced analytics and artificial intelligence solutions with global scaling capabilities in the Microsoft Azure cloud

eXtollo and Cloud Risk Process

With eXtollo customers around the world do not need to run a cloud risk process (CRP) for moving their projects in the cloud.

What our customers say

By using eXtollo my team was able to start directly in the cloud without running through a time consuming and effort intensive CRP. This saved us a lot of time and money.

Operating on a global level with different projects and use cases was essential for my department. eXtollo was very convenient because we did not had to run a single CRP and started in the cloud immediately.

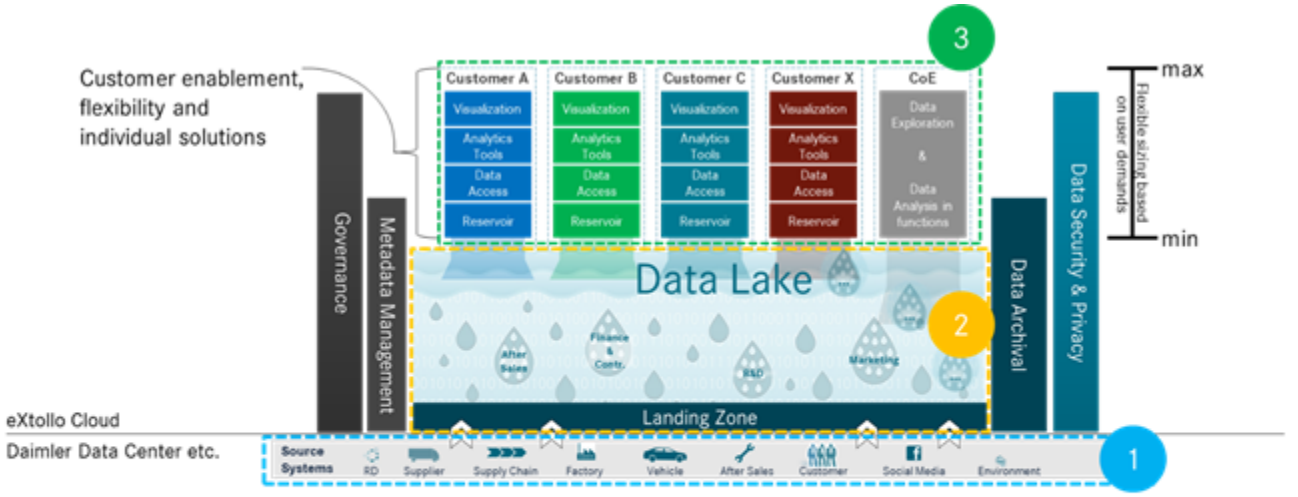
How does CRP in eXtollo work

- eXtollo is an established and robust working big data solution based on Microsoft Azure
- eXtollo has a signed CRP in place for all the components & functions offered within eXtollo
- Customers are able to go directly into the cloud with eXtollo without running though an individual CRP
- As new components/services are requested by many customers the CoE Big Data and Advanced Analytics Architecture Council will take care of all CRP related actions and support the customers

More speed, insights, cost savings and flexibility for all Daimler cloud projects with the Daimler Big Data Lake and the eXtollo solution where no CRP is required by the customers

eXtollo building blocks and component overview

To establish a basic understanding of eXtollo, it is important to understand the three major building blocks:



1 Data Source

Data Sources are owned by the users (BU/ITx) and have to be ingested into the eXtollo Data Lake.

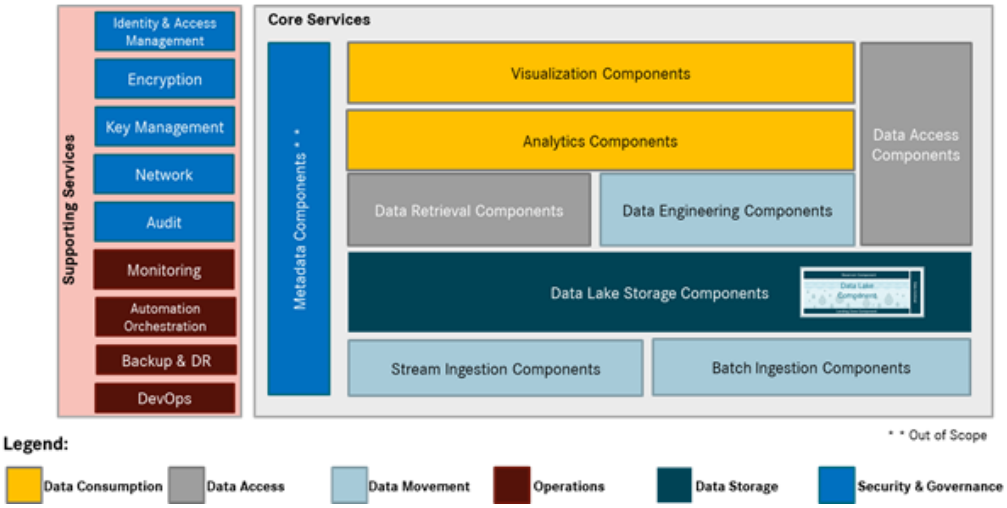
2 eXtollo Data Lake

Data Lake and its Landing Zone are managed by the CoE and based on Microsoft Azure. The Data Lake holds ingested Data by the BU and ITx units.

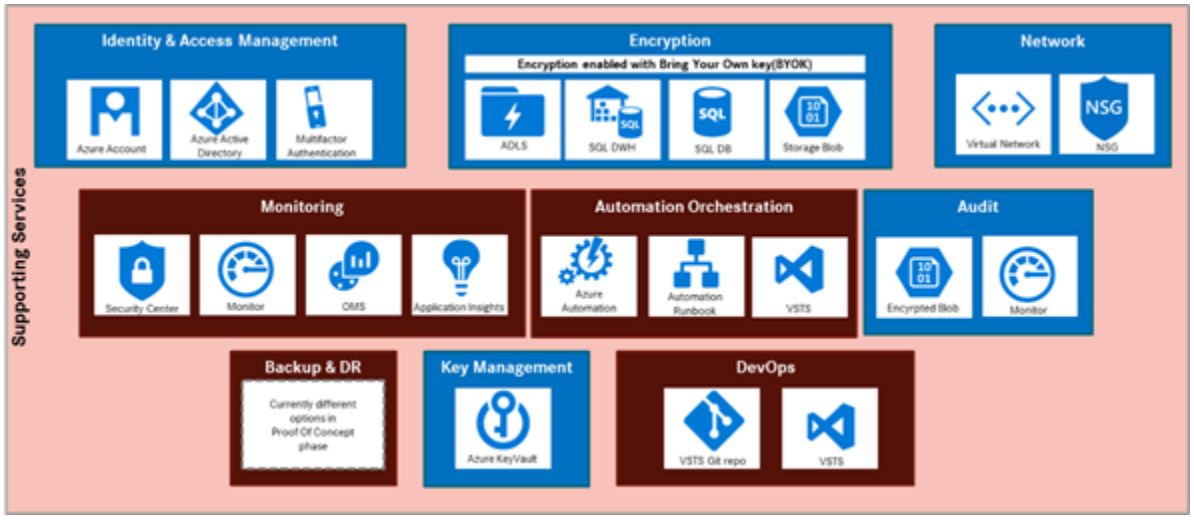
3 eXtollo instance

eXtollo instances provide ITx and BU users a high level of flexibility (e.g. sizing, compute, etc.) and individual solutions (e.g. freedom to install tools of choice). Managed by the users.

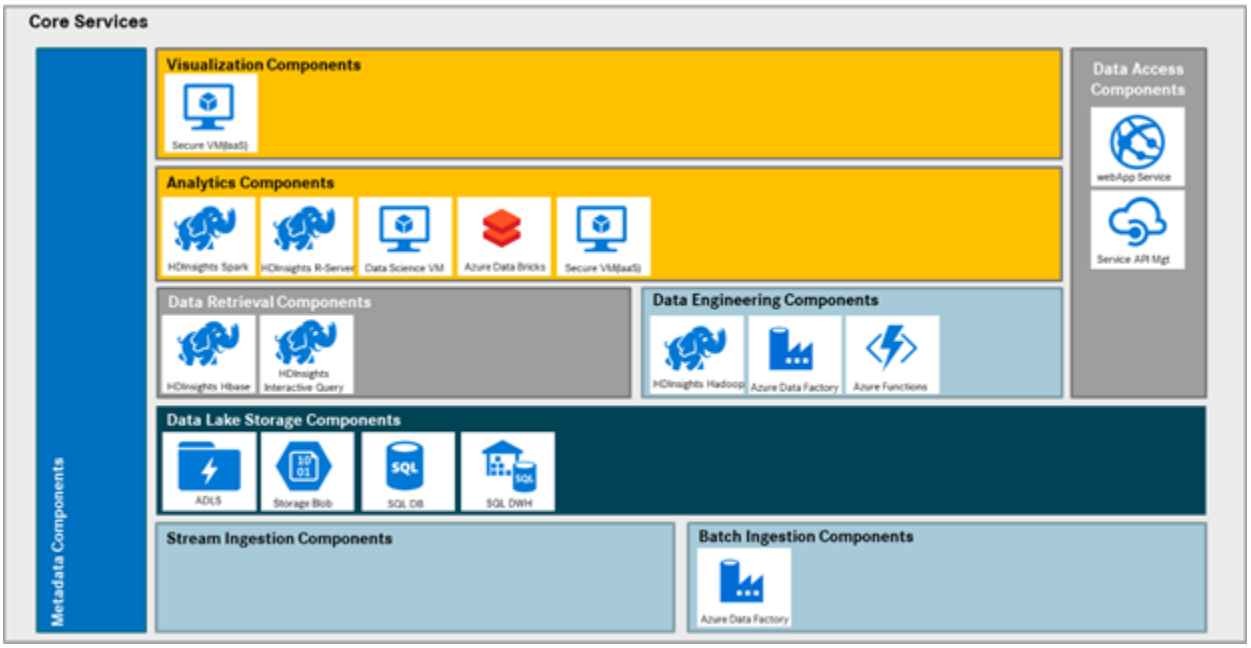
For providing a comprehensive and easy to understand high-level overview of the eXtollo components, we introduced two types of categories: The first category is “supporting services” which consists of security and governance as well as of the daily operations components. The second category is named “core services”. This category contains Data Movement, Data Storage, Data Access and Data Consumption components which are highlighted in the below figure.



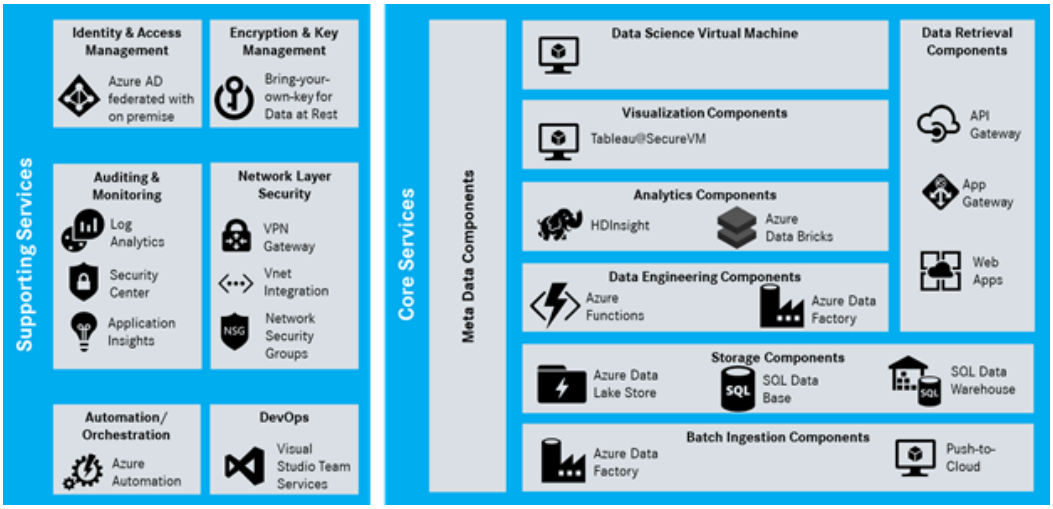
Supporting services in greater detail:



Core services in greater detail:



Furthermore, with ‘eXtollo’, Daimler is lifting the full potential of data & analytics capabilities in Azure by shopping the Azure supermarket. Here is an overview of Azure components which eXtollo is utilizing:



What is an eXtollo Instance?

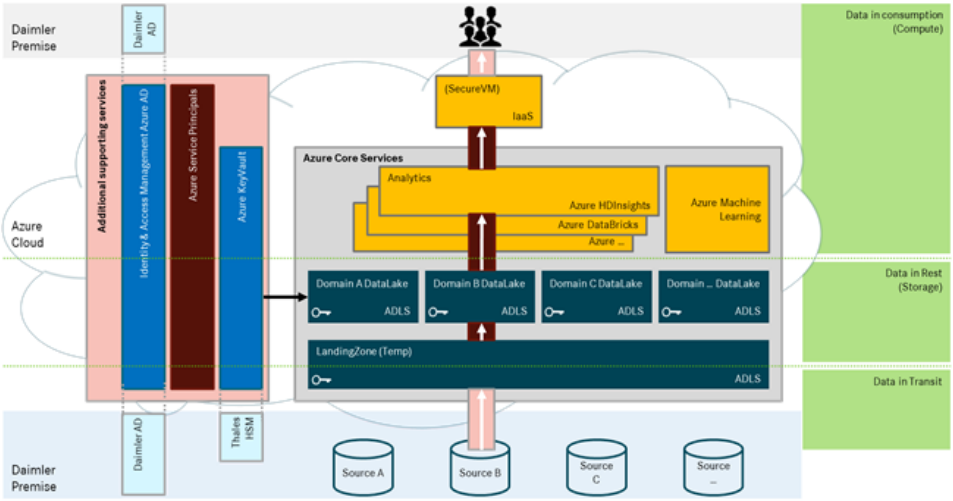
An instance is the workspace in eXtollo which is provided with default and on-demand Azure components for performing Big Data & Advanced Analytics use cases.

An eXtollo instance has clear boundaries, which cannot be changed on request. Please make sure that the solution architecture fits with eXtollo.

- Except for the Azure Key Vault which is used to hold the BYOK encryption, all the default components in eXtollo do no cause any cost unless used. See table below for a component list.
- All components of an instance are **bound to a specific region** and can only use data that was ingested from that region. Instance users are responsible to make sure that data is not transferred outside the eXtollo instance without the approval from the Data Lake team..
- There is **only one bill** for the instance. The bill cannot be split for individual use cases running on one instance.
- All instance admins have the **same permissions** inside their instance. All instance users have the same permissions inside their instance. Instance Service Principals will have the same permissions as the instance admins. Note that we can give individual access rights on the reservoir storage for external applications interacting with your instances.
- Instance admins can order consumer groups for interactions between instances and interactions with external applications. Both **a consumer user group and a consumer service principal** will be configured with the same permissions. Consumer groups can also be used to manage access for non-developers or external stakeholders to show dashboards and other visualizations.
- Some Azure resource types can exist only **once per instance**, while other resource types might be allowed multiple times in parallel.

- Users requiring multiple Environments such as Development, Integration, Staging and Prod should make sure to order a separate instance for each environment
- Data Science Virtual Machines (DSVM) acts as a Data Science Workstation and as a jumphost to other eXtollo components. The user should have at least one DSVM (independent of the number of eXtollo instances) in-order to work with eXtollo instances. DSVMs can be used across eXtollo instances.

High level architecture blueprint & data flow in eXtollo cloud components



eXtollo key design principles and Terms of Use for blueprint

By using eXtollo it is important to understand that:

- eXtollo is a template-based blueprint which incorporates Azure Big Data and Analytics components
- All storage components within eXtollo are encrypted by BYOK concept and managed by CoE
- During analytics data is de-crypted and the users do not need to care about
- eXtollo blueprint and all generated instances are CRP approved and governed by the CoE
- Terms of Use must be accepted by data owner (L3) before engaging eXtollo
    - Analytical usage only
    - Data classification up to confidential possible
    - Non-business critical processes

eXtollo key design principles focusing on compute

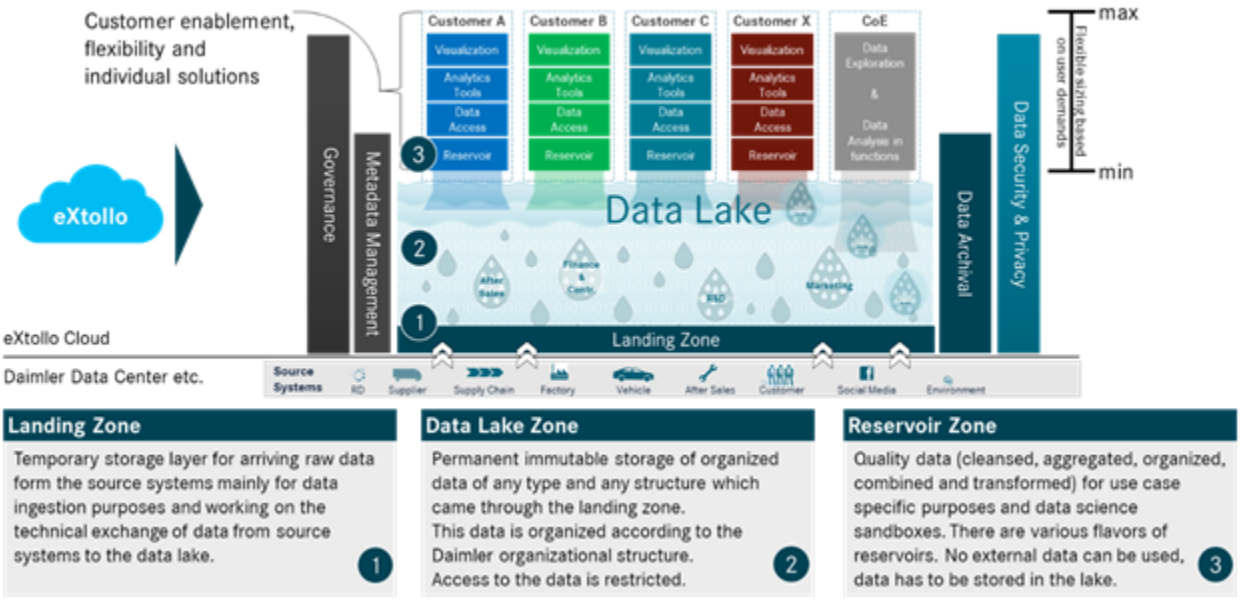
- Based on eXtollo blueprint, individualized instances for productive use cases can be generated (compute)
- Each instance provides
    - A flexible environment management (scale up, scale down) - individual compute
    - Continuous deployment (DevOps) via Azure DevOps
- Pay Per Use approach for individualized instances

eXtollo key design principles focusing on storage

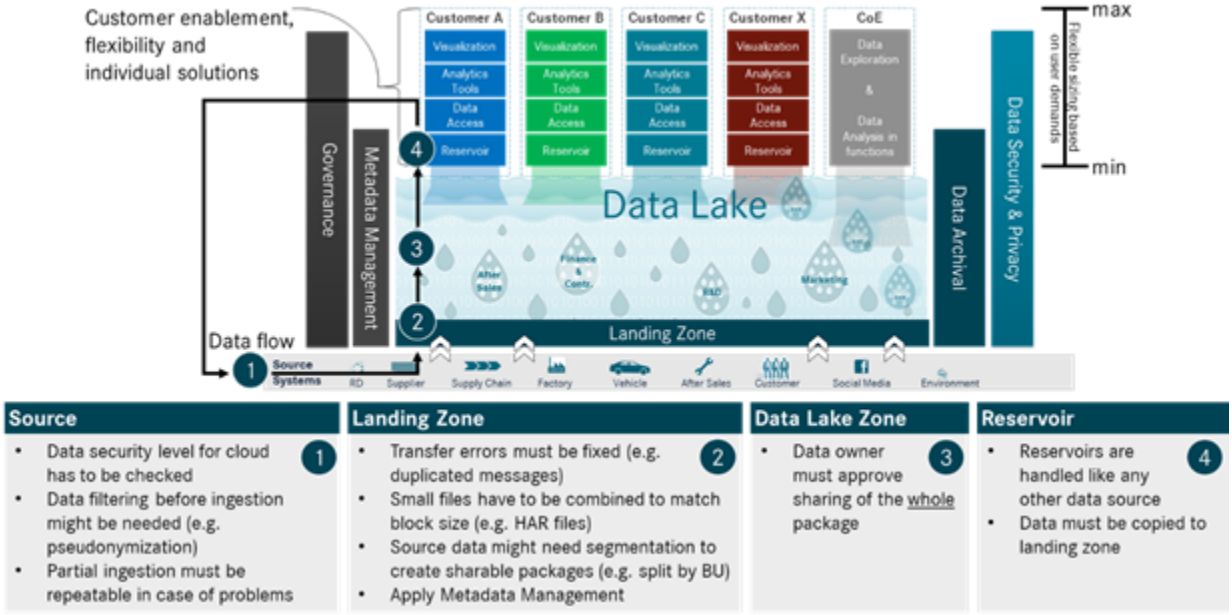
- The Daimler Data Lake concept remains to leverage cross-BU data sharing
- Data within domain data lakes can be shared across multiple instances (storage) and are governed centrally by the CoE
- The ingest factory approach is still valid, but now with focus on domain specific data lakes

High-level Data Lake overview, Data flow and Data sharing

The eXtollo Data Lake consist out of three major zones: Firstly, there is the landing zone. Secondly, the data reaches the data lake zone and lastly the reservoir zone.



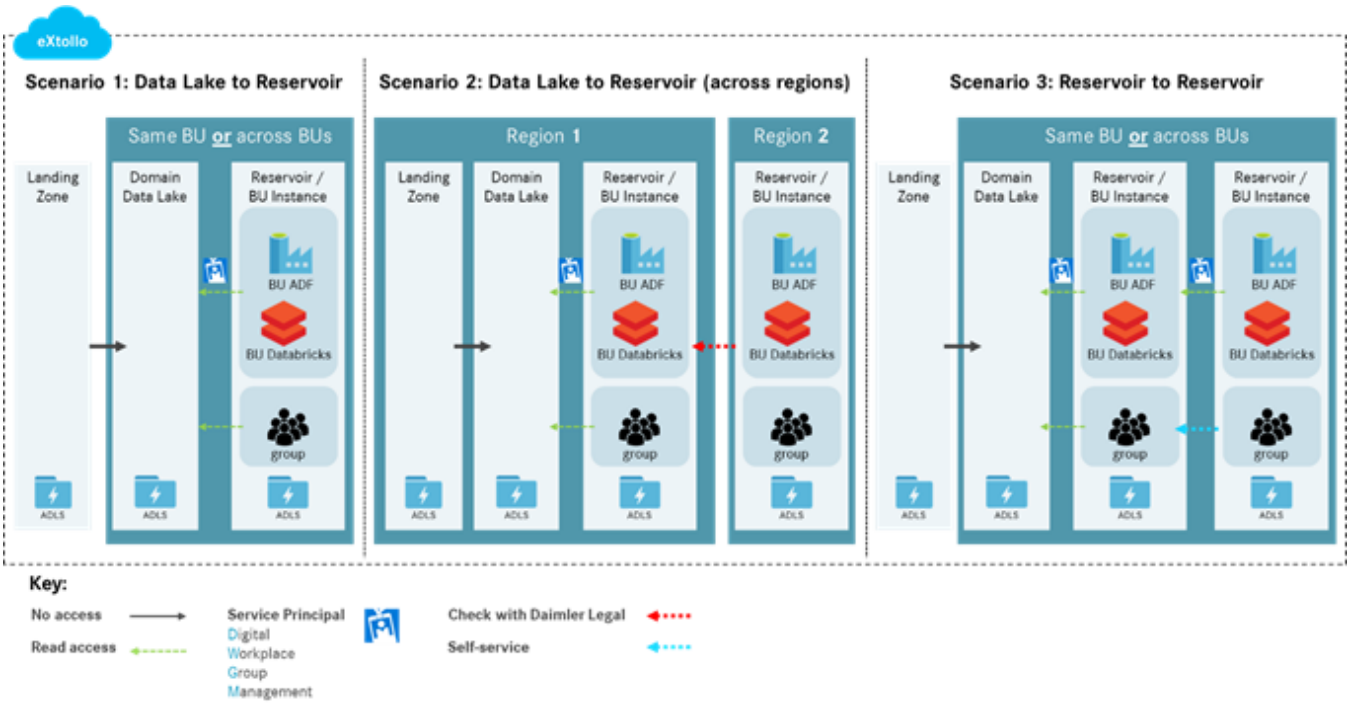
In eXtollo the data flow respectively the data consumption of a single instance can be described by four zones as highlighted in the figure below:



eXtollo customers have two options to share and consume data. This are the two possible scenarios:

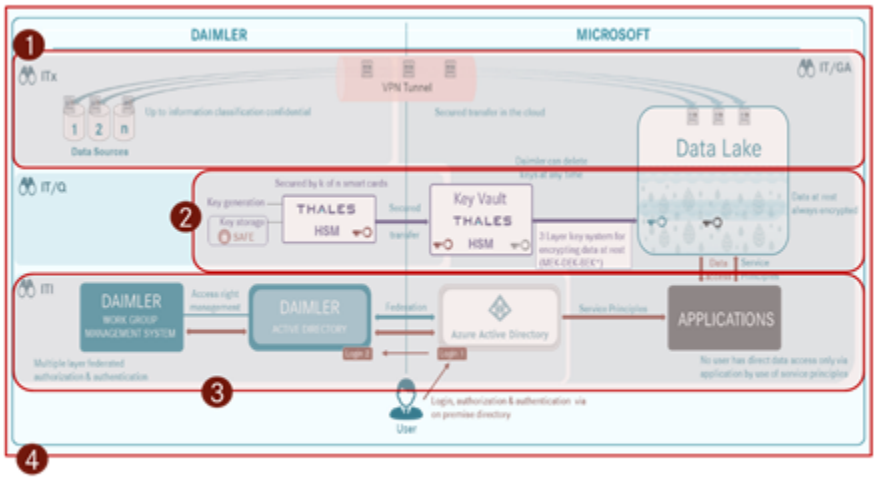
1. Data comes from the Data Lake into the Reservoir (same eXtollo region)
2. Data comes from the Data Lake into the Reservoir (across eXtollo regions)

No matter which data consumption/sharing scenario a customer prefers, the “Access Request Process” is binding and necessary to enable data flow within eXtollo (explained in the “how to use eXtollo” section). When data is shared across eXtollo regions (second scenario), customers should get in touch with Daimler’s Legal Department, as a legal approval is required to share data cross-country/regions.



Data Protection standards of eXtollo

In addition to the architecture and security measures a regional data center setup will be chosen to ensure data resides in the respective region of origin (E urope, US, China and Asia, etc.). No individual assessment of data stored in the Data Lake is feasible, but governance for analytic use cases is implemented to avoid legal risks and will be enhanced under the lead of Integrity and Legal.



1. All data which is moved into the cloud is at any point in transit and at rest secured by either transport layer security or **encryption** at rest with state of the art three-layer-key system with key rotation and block level encryption with max. 4 MB block size.
2. Key protection through **HSM of independent European supplier**. Keys generated for encrypting data are managed and backed-up from Daimler. Daimler has at all times the possibility to delete keys in the cloud.
3. High degree of control through **identity and access management** linked to Daimler on-premise systems (authentication and authorization): Access relies on the availability of the Daimler on-premise directories. User credentials, passwords and access rights are only stored on Daimler premise.
4. Permanent oversight by Daimler through **Logging & Monitoring** service provided by Microsoft log-files are transmitted real-time to Daimler.

Costs

Customers can get a variety of services on the eXtollo platform.

In the following the cross-charging model of the individual services is described in a transparent way.

eXtollo Instances

Customers can order an own instance on the eXtollo Cloud platform.

Find the cost information here Extollo costs - Data Ecosystem Wiki - Intranet Wiki (polygran.de)

The cost is cross-charged in the following way:

- Charging frequency: monthly
- Source of actuals: Billing Dashboard (Link to the Billing Dashboard)
- Billing scheme:

Billing month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actuals of month	No billing!	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct Forecast Nov Forecast Dec (FC based on actuals of October)	No billing!

eXtollo Power BI

Customers of eXtollo instances can additionally use the eXtollo Power BI services.

There are two flavours:

a. eXtollo Power BI - Shared premium capacity

- Charging frequency: monthly
- Charching base: Actual number of users of last month
- Billing scheme:

Billing month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actuals of month	No billing!	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct Forecast Nov Forecast Dec (FC based on actuals of October)	No billing!

b. eXtollo Power BI - Dedicated premium capacity

- Charching frequency: New ordering: one time with start of the service // Existing: one time per year at the beginning of the year
- Charging base: T-Shirt size of ordered premium capacity (P1, P2, P3)

Find the cost information here Extollo costs - Data Ecosystem Wiki - Intranet Wiki (polygran.de)

eXtollo Data Source Onboarding / Extraction - Batch

Customers wanting to connect their data source to eXtollo Data Lake have several possibilities. In addition, extracting the data from eXtollo to on-prem is also possible.

eXtollo Data Source Onboarding - Batch Ingest - New Source

- No cross-charging from 2021 onwards

eXtollo Data Source Onboarding - Batch Ingest - Change

- Charching frequency: Once per change after implementation
- Charching base: Complexity of change to data source (S, M, L)

eXtollo Data Source Onboarding - Batch Ingest - Extraction

- Charching frequency: Once per extraction after implementation
- Charching base: Complexity of extraction (S, M, L)

Find the cost information here Extollo costs - Data Ecosystem Wiki - Intranet Wiki (polygran.de)