

Data Fuels Digital Business

The Nokia Data Marketplace Solution at Equinix



One Platform. One Solution.



Simplify Data Exchange and Monetization

Easy to share. Easy to consume. Easy to sell. Easy to buy.
All in one platform.



Ensure Data Traceability and Integrity

Fully secured. AI models and Data Lineage Tracking.



Bring Algorithms to the Data

Run analytics wherever data resides. No need to move data.



Enjoy Proven Trust and Neutrality

Turnkey data center and technology capabilities in ONE solution.
No need to integrate point solutions.



Global Distributed Solution

Support for data sharing in different regions/markets
for data residency/compliance.

Contents

Section 1

Introduction	4
--------------	---

Section 2

Data Marketplace as a Solution	5
--------------------------------	---

2.1 The Different Types of Data Marketplaces	6
--	---

2.2 Data Marketplace Overview	8
-------------------------------	---

Section 3

Use Cases and Requirements	9
----------------------------	---

Section 4

Data Marketplace Functionality	12
--------------------------------	----

Section 5

What Makes the Nokia Data Marketplace Solution at Equinix Different	14
---	----

5.1 Support for Multiple Data Sharing and Trust Archetypes	14
--	----

5.2 Support for Federated, Geo-distributed and Heterogeneous Analytics	16
--	----

5.3 Multi-Zone Security Architecture	18
--------------------------------------	----

5.4 Blockchain-based Lineage Tracking	20
---------------------------------------	----

5.5 Data Exchange at an Interconnection Hub	20
---	----

Section 6

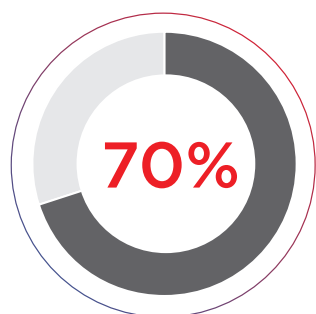
Data Marketplace Operational Model	23
------------------------------------	----

Nokia Data Marketplace Solution at Equinix	25
---	-----------

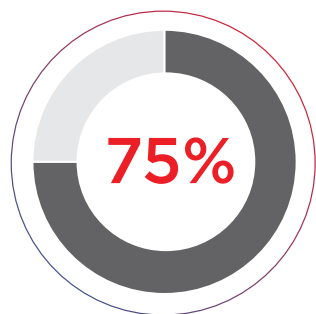
References	25
-------------------	-----------

1.0 Introduction

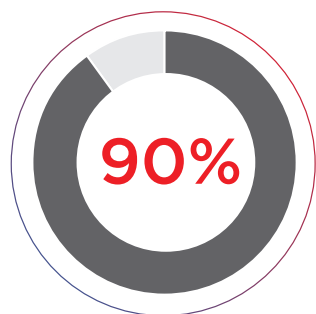
Artificial intelligence (AI) has become mainstream and is being leveraged by enterprises to improve customer service, automate decision-making, predict trends and optimize business processes. The amount and diversity of data accessible to AI-based algorithms determine how functional and accurate they will be. Data fuels digital business – and that’s increasing the demand for external sources of data exponentially.



of enterprise applications will leverage AI in some form by 2021.¹



of enterprise applications use 10 external data sources, on average.²



of large enterprises want to monetize their company-generated data.³

Most large enterprises want to use their data for financial gain, but they may hesitate to do so because they don’t want to lose control of it. Even with legal agreements in place, they aren’t confident their data and/or algorithm will be used for authorized purposes only. For good reason: up to 40 percent of data is shared between enterprises using easy-to-compromise files, spreadsheets and emails⁴. As well, in many large organizations data remains siloed within groups who are reluctant to share it with each other. Data marketplaces allow providers and consumers to share – or buy and sell – data and algorithms privately and securely without violating any government regulations (e.g., GDPR).



In data marketplaces – also known as digital data marketplaces, or DDMs – governance models regulate how the members of the marketplace interact. Their technology stacks facilitate legal contracts and asset registration that enable third-party services (data anonymization, conflict arbitration, analytics tools, etc.) and payment management.

There are many data marketplace solutions already in place, but they aren’t fully addressing enterprises’ data sharing challenges and concerns. In this report we discuss:

- The basic tenets of a data marketplace
- The concerns and requirements of data providers and consumers
- How Nokia Data Marketplace and Equinix™ have partnered to provide a unique data marketplace solution

2.0 Data Marketplace as a Solution

Data marketplaces allow buyers and sellers to exchange goods or services via an established structure that enables them to interact and transact. An exchange between buyers and sellers is called a transaction. The transaction consists of interactions that ultimately result in an agreement involving payment or other compensation. The actual exchange is completed via some type of medium or infrastructure that resides within the structure - a set of components/elements that make up the marketplace.



The value of a data marketplace is to facilitate the trading process as efficiently and securely as possible while removing obstacles that prevent buyers and sellers from performing their activities.

Before buyers and sellers are willing to interact and transact within a data marketplace, they must trust it. That means reducing risk by establishing rules for:

Market admission

Enforcement

Dispute Resolution

Ensuring Fairness and Competitiveness

The Consequences of Non-Compliance

Buyers and sellers are expected to comply with public laws. Stealing goods - although a transfer of goods takes place - is not a legal transaction, and a data marketplace must be able to organize theft prevention and have the ability to manage its consequences. When operating on behalf of specific communities, the data marketplace's governance must be able to provide additional community rules to ensure, for example, that a seller is not allowed to operate outside established hours to avoid unfair competition.



2.1 The Different Types of Data Marketplaces

Organizations increasingly depend on data to support and automate decisions and actions. While traditional methods based on statistical analysis and programmed reasoning support such decisions, AI-based approaches – where data is needed to learn how to reason – are rapidly gaining popularity. However, if we are unable to address the challenges around data sharing, we face the danger of an AI winter – starving algorithms with too little data for accurate and meaningful results. The data marketplace is a structure that organizes trust. It will accelerate competitive AI development by digitizing interactions and transactions using a programmable, community-owned, safe and secure infrastructure.

How does a data marketplace provide and implement governance, and what means are available to enforce rules and agreements? This is where an underlying neutral exchange infrastructure can play a key role.

There are several types of data marketplaces



Hub Model

In this model, a central party organizes the data marketplace and determines how and under what rules and conditions interactions and transactions are performed. The hub organizes and owns the underlying platform infrastructure, thus organizing trust. This is an increasing concern, because the hub can gain knowledge from interactions, enabling monopolistic behavior and questionable data use.



Network Model

This model allows sovereign organizations to interact and transact to access and use data assets based on community rules, where the data marketplace is governed by a community. It is gaining popularity because it does not require a centralized data marketplace, which prevents monopolistic behavior and democratizes the use of resources and data due to interactions between community members.



Open Model

In this model, parties find each other on open data marketplaces. The hosting site usually asks consumers to sign licenses before giving access to data. Consumers typically don't pay fees. The model suits open source projects where datasets are published and shared without concern or constraint. There is no guarantee regarding dataset quality, however; it is usually buyer beware.

Broker Model

In this model, participants directly purchase datasets from the broker who manages the marketplace. Brokers may aggregate information from multiple sources and provide value-added services. Some broker-controlled marketplaces also facilitate (for a fee) bilateral data trading between participants that is governed by the broker's marketplace rules. Broker models are popular for consumer-driven data.







Structure	Interaction Governance	Transaction Governance	Remarks
 Hub	Platform (single owned)	Platform (single owned)	Parties have a contract with a platform that brings them together and facilitates transactions.
 Network	Platform (community owned)	Neutral Infrastructure (community governed, provided by data center industry)	Parties are a member of a consortium platform bringing them together for interaction. Transactions are facilitated by a neutral exchange infrastructure with consortium governance. Such exchange infrastructures reside inside neutral data centers.
 Open	Bilateral	Bilateral	Parties find each other via public channels.
 Broker	Platform (single owner)	Bilateral	Parties are brought together by a broker platform for a fee. Processing transactions are agreed bilaterally.

Table 1: Different types of data marketplaces

The basic role of the data marketplace is to organize and facilitate interactions between data suppliers and algorithm developers to explore, select, and agree to create, execute and complete data science transactions. A data marketplace is a global structure enabling sovereign organizations, which require absolute control of their data assets, to offer and under strict conditions make assets available to achieve mutual benefits that no single organization could achieve on its own.

2.2 Data Marketplace Overview

The high-level architecture of a data marketplace is depicted in Figure 1. It shows the data marketplace as an entity owned and operated by a membership organization through which data providers and consumers can interact and transact based on community rules and individual agreements.

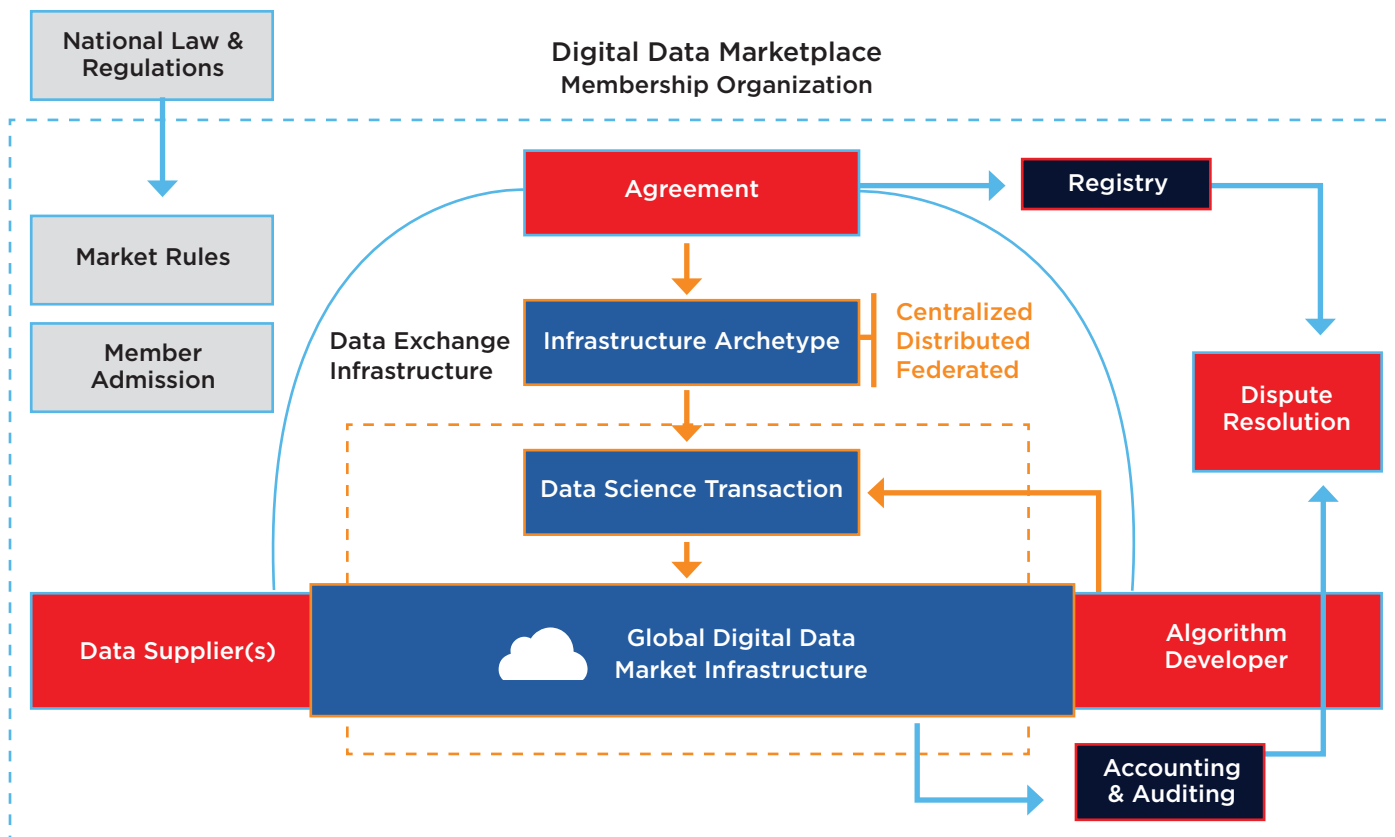


Fig. 1. Data marketplace architecture

A data marketplace must have a process for creating membership rules, and its process for admission must require a prospective member to agree to comply with all membership rules. A member might be a data supplier, algorithm developer or data services provider. After being admitted, members can decide which other members they want to interact with based on an established understanding or agreement. Member groups can compete with other member groups.






Members create agreements with other members to enable more interaction and subsequent transactions that trade data assets and/or services. Within a data marketplace, data and algorithms are considered assets that can be traded and used to achieve benefits by providing a trusted infrastructure for interaction and transaction. Members may decide to collaborate with other members for one purpose and compete for another.

Once an agreement to collaborate is established, a member can specify additional agreements about how they want to create, execute and complete transactions that perform data science workflows, enabling algorithms to train on one or multiple data sources. Agreements between parties will be digitized as smart contracts that are used to orchestrate and authorize all necessary steps needed to access and use the data. This will be further explained in Section 5.

3.0 Use Cases and Requirements

Data sharing between organizations is nothing new. It has been happening via bilateral agreements for decades. There are many data brokers and data aggregators who buy and sell data and specialize in various vertical markets. Industry analysts IDC and Gartner have defined industry verticals where data sharing is prevalent^{4,5}.

Here are a few examples of the value a data marketplace affords data consumers and providers:

Sector	Use Case	Consumer and Provider Benefit
 Smart Industry	Predictive Maintenance	Increase equipment availability by optimizing maintenance planning, spare parts stock levels and employee resources.
 Logistics	Predictive Logistics	Optimize logistics network performance, mitigate delays proactively and predict demand to plan capacity.
 Health	Personalized Interventions	Enable clinical decision support systems to tailor treatment to individuals or patient groups.
 Smart City	Predictive Traffic Management	Improve urban traffic management strategies to avoid congestion and pollution.
 Agriculture	Precision Agriculture	Increase yield, minimize herbicides and optimize the use of fertilizers depending on actual (non-uniform) soil conditions by algorithm-driven control of farm equipment.

In many cases, the data marketplace solutions available in the industry today do not adequately satisfy the demands of data providers. Thus, there is a need for solutions that give data providers the confidence to share data and algorithms as part of data marketplaces.



Data Provider Requirements	What This Means
<p>Full Control and Auditability</p>	<p>Data and algorithm providers want full transparency when it comes to how many copies of their data are being maintained in the underlying infrastructure (for example, how many copies AWS S3 makes of a particular object). They also want to control which data can be taken out of the marketplace and which algorithms can be run on their data. In some cases, they only want the data to be in the marketplace while it is being shared; subsequently, they want it deleted or made inaccessible to others.</p>
<p>No Cloud Lock-in</p>	<p>In many cases, data providers do not want to have their data stored with a cloud provider for confidentiality and cost reasons (egress fees for moving data out).</p>
<p>Support for Different Trust Models</p>	<p>Providers want different data sharing models for different types of data. Sometimes they don't want extremely sensitive raw data to leave their premises, so the consumer algorithm must be brought to the data. Sometimes they are willing to send data and let the consumer algorithms operate on it in public clouds. In other instances, the consumers and providers want a neutral third-party location in which to share their data and algorithms.</p>
<p>Distributed Solution at the Edge</p>	<p>132 countries have regulations or are in the process of formulating regulations with respect to data residency. Data providers need a distributed data marketplace solution that allows certain data to be traded without leaving a region. Additionally, data consumers may want to access and process data at the edge to decrease latency and costs. Moving massive data sets to far off data centers gets expensive, and doing model inferencing at a remote data center slows things down. This can be mitigated by inferencing on data stored at edge locations.</p>
<p>Data Licensing and Governance Options</p>	<p>Data providers want the marketplace to support different data licensing models for different types of data. Data providers and data consumers also want the marketplace to support different governance models with respect to the operation of the marketplace (i.e., hub model or network model).</p>



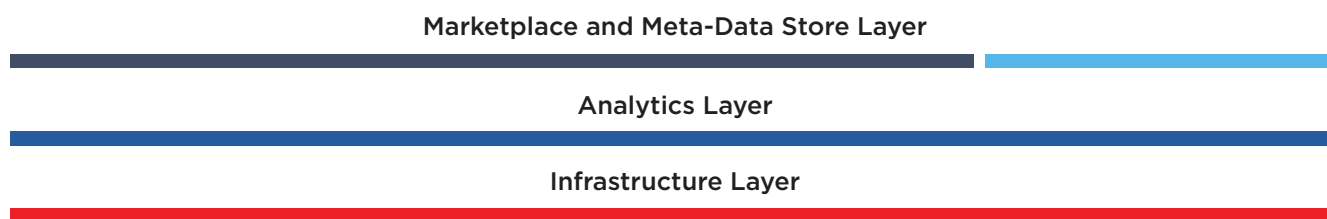
Similarly, data consumers also want specific functionality from data marketplaces.

Data Consumer Requirements	What This Means
<p>Usage Privacy</p>	<p>In many cases, data consumers want privacy with respect to how they are using the data to create their AI model assets. They do not want data providers to have detailed knowledge of how they are using the data they have purchased.</p>
<p>Data Lineage</p>	<p>Data consumers want to know the source of their data, because people often use incorrect data that leads to biased or inaccurate models.</p>
<p>Provider Reputation and Data Quality</p>	<p>Data consumers are concerned about data provider reputation, data quality and data certification because they affect the quality of their AI models. For example, data consumers may want data providers to ensure that the data being sold is GDPR compliant (i.e., it does not contain any private citizen data).</p>
<p>Choice of Analytics</p>	<p>Data consumers do not want to be forced to use an analytics vendor that is being supported by the marketplace. Different analytics vendors have expertise in different vertical domains. Data consumers would like a choice of AI/machine learning (ML) frameworks and tools. In some cases they want to be allowed to bring their own analytics frameworks in docker containers.</p>
<p>Support for Data Scientists and Production Workflows</p>	<p>Data scientists typically want to experiment with sample data on their own laptops or private workstations. Subsequently, they want to build models with real data in a secure location where the data providers want to have control over their data. Finally, they want to take the built model out of the data marketplace and use it in a production environment. The data marketplace needs to be flexible enough to support this mode of operation.</p>



4.0 Data Marketplace Functionality

The Nokia Data Marketplace architecture consists of three layers (see Figure 3):



Unlike many data marketplace solutions that employ a monolithic architecture where customers are locked into an analytics provider or infrastructure provider, the Nokia Data Marketplace solution at Equinix decouples these layers with an interaction and transaction layer. This architecture gives marketplace operators their choice of analytics and infrastructure providers. Here is a quick look at the functionality of each layer and how it compares with other solutions in the market.

Data Marketplace and Meta-Data Store Layer

The data marketplace layer provides digital interaction and transaction functions to members of a consortium by:

- Providing membership management
- Allowing providers of data and algorithms to register their assets
- Letting participants find and explore each other's assets
- Maintaining a catalog of the resources being traded in the marketplace
- Helping participants create smart contracts to trade their assets
- Handling token-based payments
- Helping orchestrate analytics pipelines on behalf of data scientists and production teams
- Providing trusted audit trails to allow a consortium to resolve disputes by maintaining a Hyperledger-based blockchain ledger to keep track of all the user contracts and actions in the marketplace

The meta-data storage layer stores all the meta-data associated with the above actions.

Analytics Layer

This layer is integrated with the open source ML platform Kubeflow. It allows data scientists to create AI models using the analytics tools of their choice. A user can have many Kubeflow pipelines, with each running on a different Kubernetes cluster in a different infrastructure cloud. A data scientist can choose to experiment with different AI frameworks and tools. The layer also allows for the importing of pre-built Kubeflow pipelines. Finally, this layer supports learning frameworks running on distributed data sharing infrastructure locations being managed by the Nokia Data Marketplace control plane. By virtue of separating these three architectural layers, Nokia Data Marketplace is capable of being integrated with multiple analytics layer solutions.

Infrastructure Layer

This layer provides the basic, software-definable compute, storage and networking infrastructure for storing data and running analytics frameworks. It contains east-west and north-south networking equipment, as well as security firewalls, denial of service and load-balancing services. The compute equipment can comprise CPUs and GPUs (for AI model training). The storage hardware can be file, block or object storage. The Nokia Data Marketplace solution at Equinix gives marketplace operators a choice of whether to host the infrastructure layer in public clouds, private data centers or bare metal-as-a-service located in a colocation data center such as Equinix.

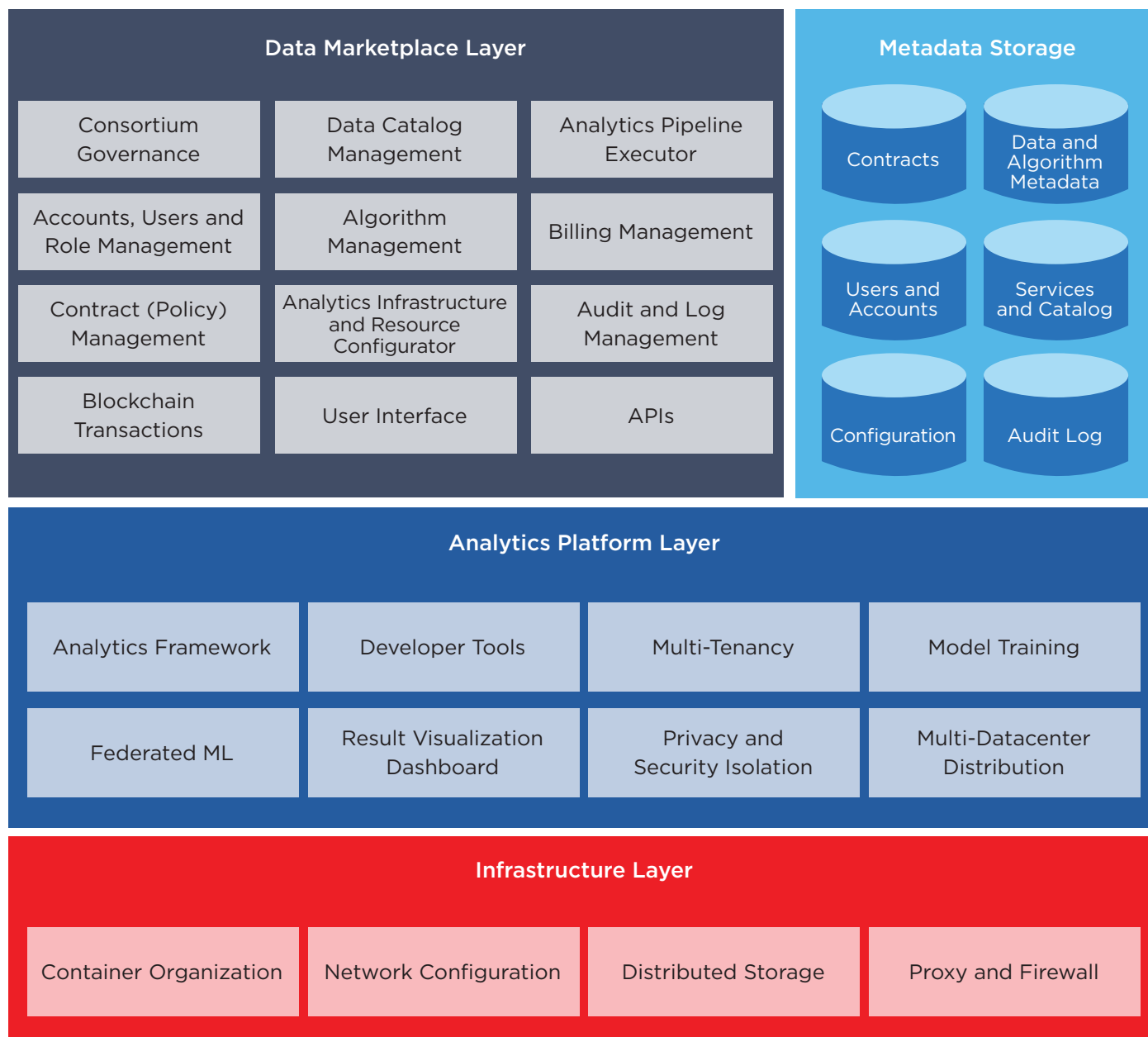


Fig. 2. Nokia Data Marketplace architecture

5.0 What Makes the Nokia Data Marketplace Marketplace Solution at Equinix Different

As discussed earlier, not all data marketplaces are created equal. Here's why we think the Nokia Data Marketplace solution at Equinix is right for most use cases.

5.1 Support for Multiple Data Sharing and Trust Archetypes

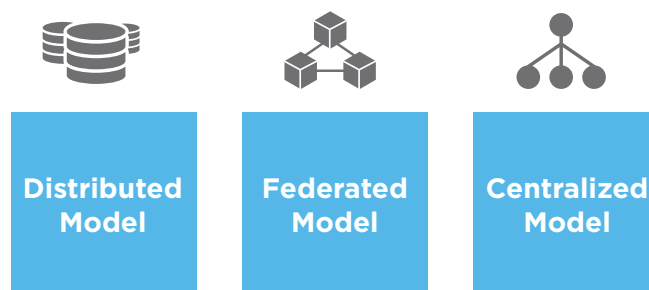
The underlying data marketplace infrastructure can offer a range of data science processing archetypes (some examples are shown in Figure 3). It is up to the data marketplace membership organization to select processing archetypes that are suitable to serve their exchange data assets using a data science workflow, they will agree to select an infrastructure archetype.

Once a model is selected and the contract determines which assets are available for processing, a data science workflow is constructed. The selected algorithm is then deployed and orchestrated by the distributed analytics platform. The contract also controls whether the results (trained model) of the data science workflow can be moved out of the shared infrastructure.

Most data marketplace solutions support the centralized data sharing model shown in Figure 3, where transactions bring data together in a central place where developers will train their algorithms. However, this mode of data sharing is not acceptable to many enterprises because, even with legal agreements in place, they are afraid that the data science organization will use the raw data for purposes other than those agreed upon.

KEY INSIGHT

Different datasets warrant different data sharing and trust models. The Nokia Data Marketplace solution at Equinix supports all three of these data sharing mechanisms:



Distributed

Data providers are not willing to let many types of data leave their premises due to confidentiality or intellectual property concerns. In these situations, instead of moving the data to a data marketplace for trading, it is necessary to move the model training algorithm to the location where the data resides. This sharing model assumes that the algorithm provider is willing to let their algorithm run in the security domain of the data provider. There are numerous privacy-preserving training frameworks that facilitate moving the compute to the location of data. Nokia Data Marketplace supports distributed, privacy-preserving AI/ML frameworks. It is important to note that in many of these AI/ML cases, training operations require GPU-based hardware stacks that consume a lot of power (20kW-30kW per rack). These AI model training racks cannot be hosted in a private data center because they cannot handle the higher power and cooling requirements. Nokia Data Marketplace customers can utilize a private cage at Equinix to host their AI training stack and run federated, privacy-preserving algorithms that require higher power density requirements.



Federated

When the data provider and algorithm provider prefer that their assets remain inaccessible in each other's locations, they need a neutral, consortium-governed exchange location like Equinix. In this model, consortium rules and auditing ensure compliant resource access and usage via a blockchain ledger. In many use cases, enterprises do not want to use public cloud infrastructures to trade their assets because they do not get full visibility into the number of copies being made of their data assets. Equinix provides interconnected, secure, neutral exchange infrastructure cages inside their data centers that enable data trading and algorithm use. The governance of this neutral exchange infrastructure is negotiable. Options include single-owner (hub model), bilateral (peer-to-peer) and multi-party consortiums (network model). In this sharing model, the raw data and algorithms are never allowed to be taken outside the neutral shared cage.

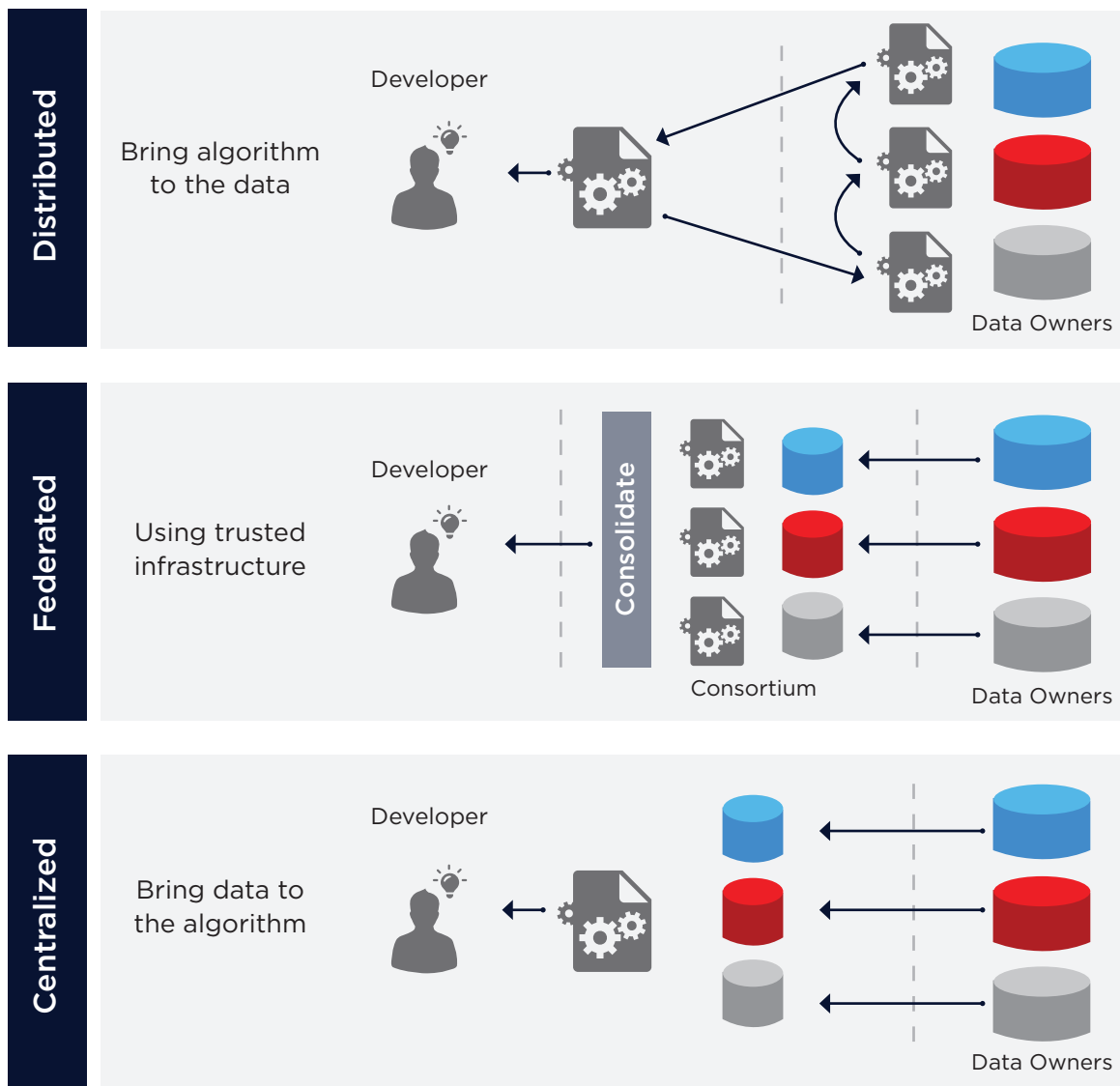


Fig. 3. Data science processing archetypes. **Distributed:** The algorithm is sent to the data owners. **Federated:** The data from each owner is learned separately on trusted consortium infrastructure near the data. **Centralized:** the data is sent to a developer for processing; results are consolidated, yielding accuracy comparable to a centrally trained model.



Centralized

This is the most common data sharing model supported by data marketplaces. In this model, the providers bring their data/algorithm assets to the marketplace as part of completing a trade with the data science organization that subsequently stores the data inside its infrastructure. Enterprises are mostly comfortable sharing low risk and non-confidential datasets. The marketplaces used to share these datasets typically reside in public clouds. The Nokia Data Marketplace solution at Equinix also supports this type of data sharing model. Additionally, as shown in Figure 4, it also supports a hybrid model where the data to be shared is stored in a persistent manner in a private cage at Equinix, then moved into the public cloud infrastructures used by data science organizations for sharing or model training purposes.

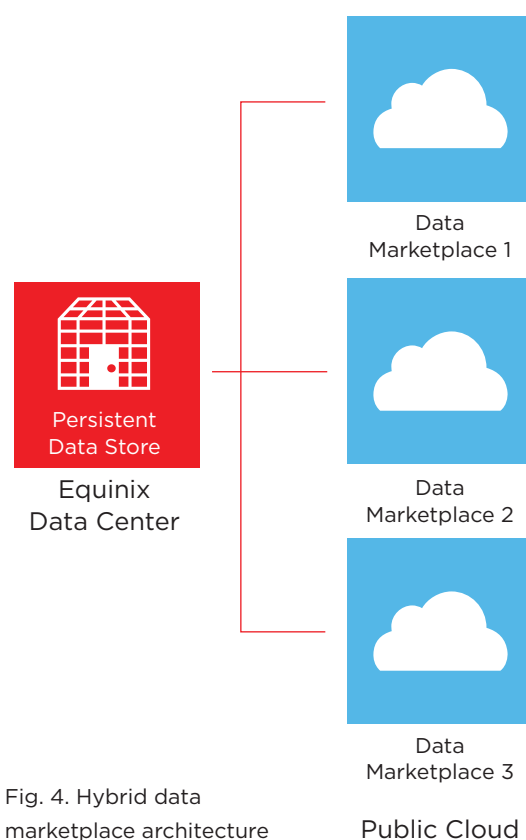


Fig. 4. Hybrid data marketplace architecture

5.2 Support for Federated, Geo-distributed and Heterogeneous Analytics

The Nokia Data Marketplace solution at Equinix makes it easy to bring data and algorithms into a secure, software-definable and geo-distributed data exchange sandbox. AI algorithms can be trained on data from different owners at different locations via the data marketplace. The analytics framework in the solution is unique in comparison to other data marketplaces in the following ways:

Support for Experimental and Production Analytics

As shown in Figure 5, we believe that data scientists first explore different analytics frameworks and do their initial model building and AI/ML pipeline building on their laptops or in public clouds. Subsequently, they export these analytics pipelines and do model training with sample data sets from many external sources in the data marketplace. Then they do production-level model training with real external datasets. Finally, once they have built their AI/ML model, they export this model and use it in their production workflows for inferencing. Typically, data scientists are not involved with the production workflows.

The Nokia Data Marketplace solution at Equinix supports the importing of AI/ML Kubeflow pipelines into the marketplace from a data scientist's laptop, then leverages these pipelines to train AI models on the data that has been brought in from geo-distributed and sovereign data providers. The model built by the data scientist can be used as part of a separate, production-level Kubeflow pipeline that is mostly API-driven by the operations team.

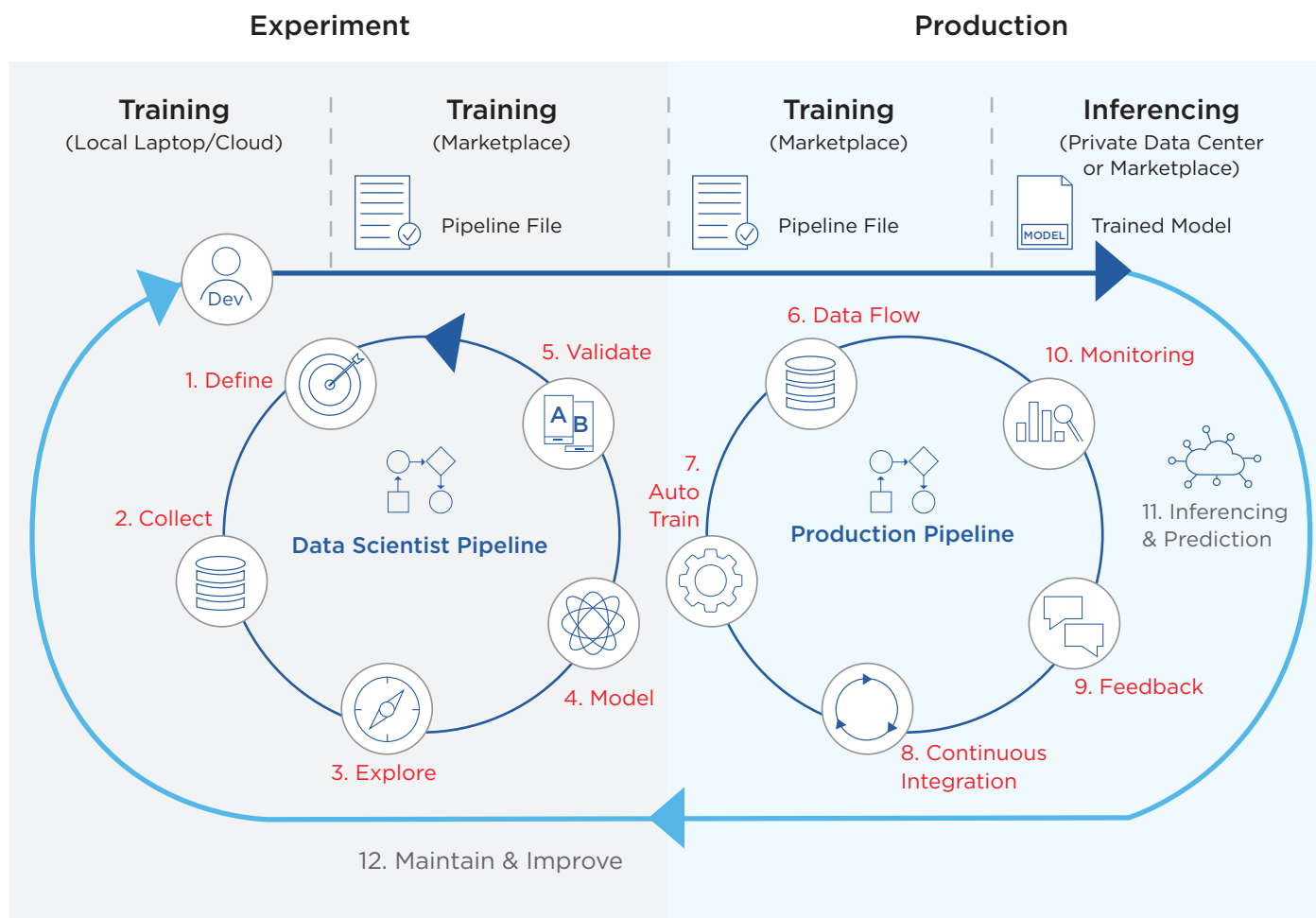


Fig. 5. Analytics workflow

Support for Heterogeneous AI/ML Frameworks

The industry-leading Kubeflow AI/ML open source orchestration framework has been natively integrated into the Nokia Data Marketplace solution at Equinix. This allows data scientists to choose their analytics tools and frameworks and instantiate them via the Kubeflow framework in the data exchange location. The secure data exchange location maintains a secure container repository that contains the data marketplace-approved docker images of the various analytics tools and frameworks. Thus, data scientists have a choice of their favorite analytics tools and frameworks.

Federated Analytics

The Nokia Data Marketplace solution at Equinix is distributed, meaning that a marketplace can simultaneously manage multiple geo-distributed data exchange locations at any given point in time. This allows it to support federated learning frameworks where local AI models on private infrastructure stacks (as shown in Figure 6) can be built, then aggregated into a global AI model at a mutually trusted neutral location like Equinix. The Nokia Data Marketplace solution at Equinix allows data scientists to invoke third-party federated learning frameworks via Kubeflow. A federated learning approach is primarily useful for two reasons:

- **Privacy-Preserving AI** - When data providers want algorithm providers to ship their algorithm to the data location because they do not want to let raw data out of their security domain, federated learning can be leveraged to build a model locally, then share the anonymized model with the consumer.
- **Efficient Handling of Large Datasets at the Edge** - Federated learning is also useful when the size of the dataset being generated at the edge is large. In these situations, rather than sending it to a far-off central core location, one can build a local AI model and ship the model (typically kilobytes) to the central location rather than the raw data, which can reach into the terabytes. Since traffic doesn't have to be backhauled from the edge to a core location, it costs less.

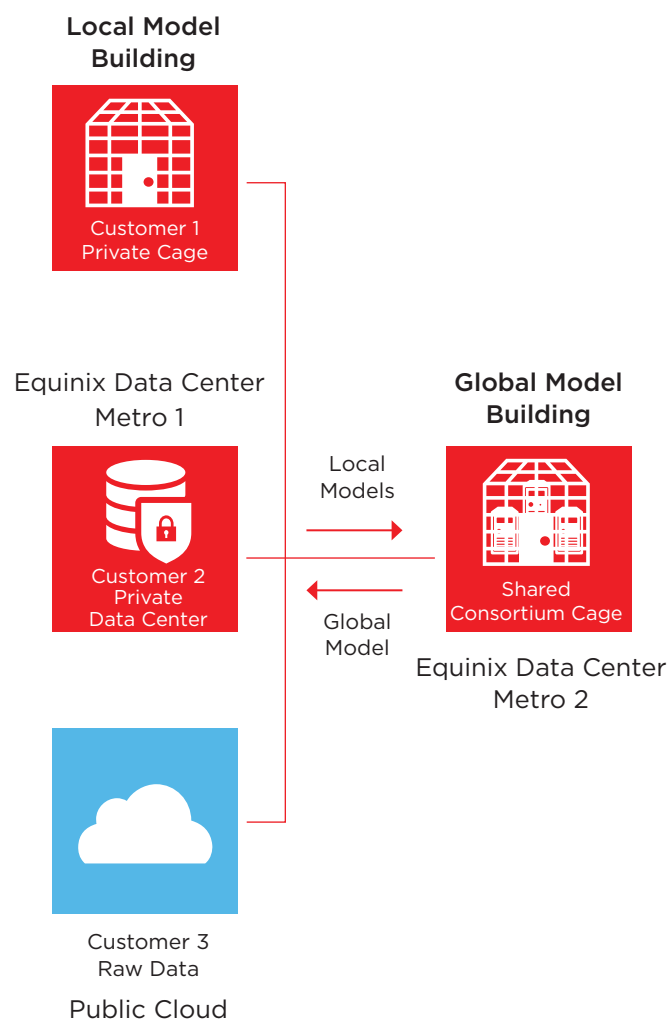


Fig. 6. Federated learning across multiple data centers

5.3 Multi-Zone Security Architecture

Most enterprises are reluctant to trade their data and algorithms in marketplaces because they fear losing control over their data. To alleviate this concern, the Nokia Data Marketplace solution at Equinix employs a multi-zone security architecture. As shown in Figure 7, in this architecture the data marketplace control plane software, the software-definable exchange data plane location (where the actual data/algorithm exchange takes place), and the permanent location where data assets are located (a secured repository belonging to a data provider) are all in different security domains.

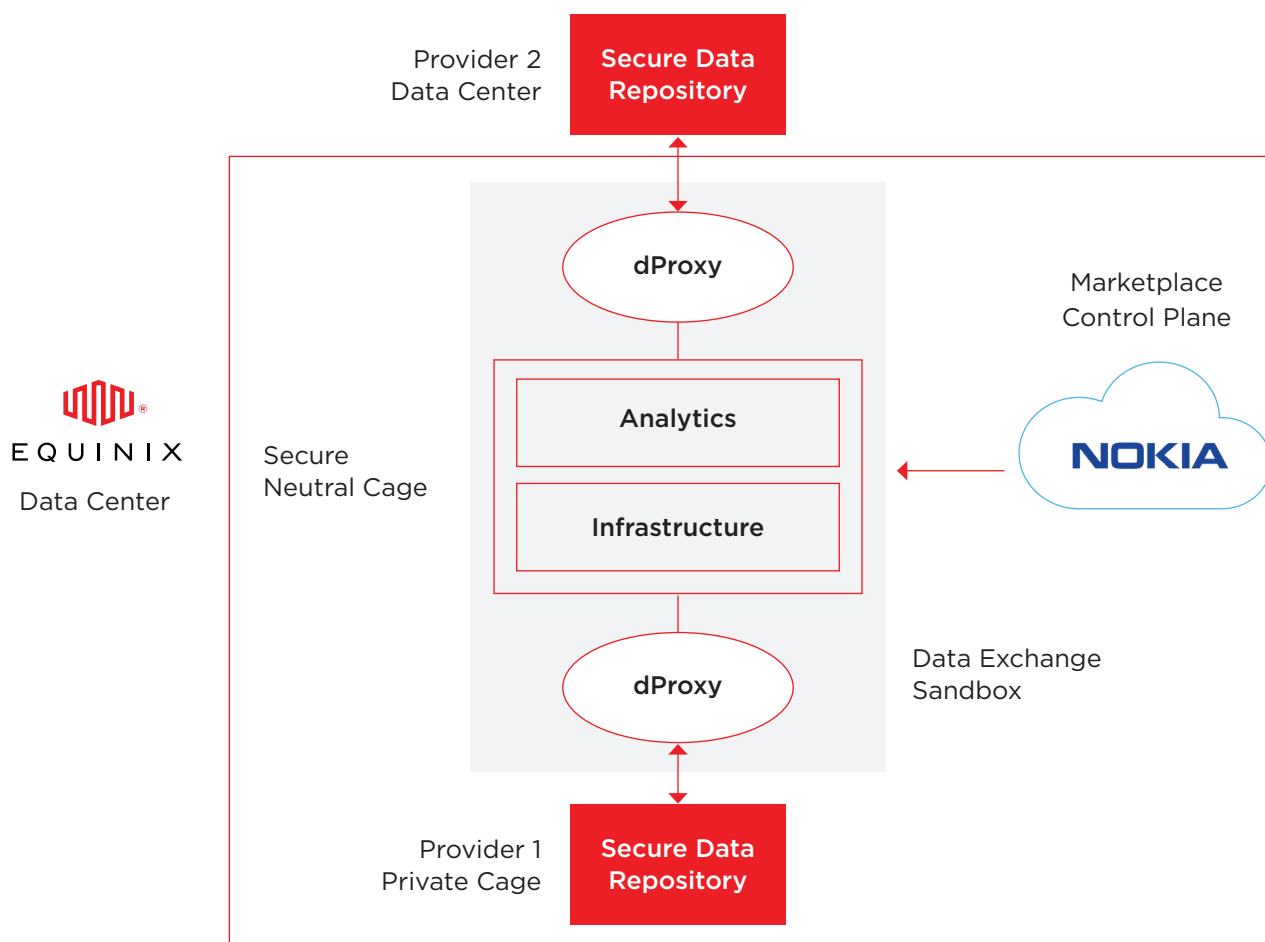


Fig. 7. Multi-zone architecture

The key benefits of this architecture are:

Flexibility in Data Exchange Location and Infrastructure

Depending upon the privacy/security level agreed upon by the data provider and data consumer for a particular type of analytics pipeline, the Nokia Data Marketplace solution at Equinix allows one to select the location where the analytics pipeline will be executed, provided that the marketplace operator has procured and configured those data exchange locations. The pipeline can be executed in the public cloud; in a neutral, secure cage at Equinix; or in a private data center. Furthermore, the marketplace operator has the flexibility to choose the desired type of infrastructure at each data exchange location. For example, the operator could choose secure enclave CPU technology if the data provider never wants their data to be in the clear (when data needs to be encrypted in the shared location at all times). The cloud type, hardware type and analytics stack can also be chosen. If a customer never wants their data in the clear, they can install hardware with secure enclave technology.

Disaggregated Control and Data Plane

In the Nokia Data Marketplace solution at Equinix, the data marketplace software never stores or accesses the data. Even if the data marketplace software gets compromised, the hacker will not have access to the data exchange or the persistent data locations. A conscious decision has been made to reduce the threat attack surface by decoupling the control and data exchange/execution locations.

Proxy-Based Access to Secure Repository

Data is stored persistently in the providers' respective secure repositories. As shown in Figure 7, these secure repositories can reside in a private, secure cage belonging to the data provider; in the customer's private data centers; or even in public clouds. Data is brought into the shared cage for exchange via a proxy. After the completion of the analytics pipeline execution in the data exchange location, the proxy removes access to the data in the provider's secure repository. Thus, the programs executing in the shared data exchange location do not have permanent access to persistent data in the secure repositories. The proxy provides a level of indirection and helps separate the secure repository and data exchange zones. The proxy can also run special marketplace-provided programs for data anonymization or personally identifiable information (PII) data removal to ensure that the data brought in for trading complies with government and marketplace regulations.

Secure Sandbox

In this architecture, a secure sandbox across single or multiple data exchange locations prevents programs that are executing in the sandbox from making unauthorized calls outside of it. This sandbox exists at the network firewall level and does not allow the executing programs to directly move containers or files elsewhere. The executing programs have to get the permission from the data providers via the marketplace in order to move their AI/ML models out of the sandbox. They are never allowed to move raw data out of the sandbox.



5.4 Blockchain-based Lineage Tracking

Most enterprises do not build their AI/ML models from scratch. For example, companies send their image or audio clips data to a public cloud, then get a customized model in return for model inferencing. Similarly, enterprises buy data from various data aggregators/brokers. In many cases, they don't have visibility into the lineage of the data. Thus, they don't know whether the data is relevant for their context, whether it satisfies government regulations, or whether the dataset has any biases.

The Nokia Data Marketplace solution at Equinix supports the data sharing model, where data providers and data consumers bring their data and algorithm into a secure and neutral sandbox. Thus, the consumers of data know exactly who is selling the data, and their model training actions are logged in a Hyperledger-based permissioned ledger by the data marketplace. By signing a membership agreement, everybody is expected to understand community rules that provide standards for data quality, timeliness, availability and more. If the model does not give accurate predictions, the lineage of the model can be checked and a causal analysis can be performed.



5.5 Data Exchange at an Interconnection Hub

Equinix was founded on the principals of data sharing and exchange more than two decades ago. The company began as a neutral location for organizations to exchange their routes to internet destinations, enabling content access and sharing. Equinix has evolved into the world's leading platform for protecting and connecting the digital economy. Equinix is home to a wide range of ecosystems that simply would not function efficiently (if at all) if the global Platform Equinix™ didn't exist. A shining example is the electronic financial trading business, where real-time, low latency access to market data is paramount.

Data is the fuel for a world where AI is in demand. Businesses need a trusted platform on which data can be securely exchanged. Here are the key reasons Equinix is a desirable platform for data exchanges.

Key Value

Why This is Important



Neutral Interconnection Hub

Equinix data centers provide neutral hub locations across the globe, where 10,000+ companies – ecosystems of clouds, networks, finance companies, media companies, enterprises and more – interconnect and exchange network traffic between each other. In many cases, the data to be traded already resides in the enterprises' private cages at Equinix, or at a data marketplace hosted at Equinix that has secure, high speed access to data located in clouds, data brokers, private data centers and edge devices. Thus, it makes a lot of sense for data aggregators, data providers and data marketplace operators to be hosted at the interconnected hubs.



Control and Audit

Data providers, data science organizations and marketplace operators want full auditability over the underlying infrastructure. In case of a dispute or as part of agreed-upon audit procedures, they want to know how many copies of a particular object have been created and who has access to their logical and physical infrastructure. The bare metal infrastructure in a private cage gives data marketplace operators full transparency into the number of instances of their software in use, any copies of their data being made, and who has physical access to their infrastructure. This is not the case in public clouds.



Global Interconnected Presence

Today's digital landscape demands adherence to strict data privacy and sovereignty regulations. In many use cases, data cannot leave a particular region, so it has to be exchanged in that region. Equinix provides 200+ edge data centers in 53 markets across the globe with unmatched, 99.9999% availability. Data providers and data marketplace operators can deploy their services in multiple geographies, and Equinix provides a consistent security and operational model across all locations globally. ECX Fabric™ connects these distributed data centers, allowing customers to create distributed applications spanning geographic locations.



Edge Presence

Data is generated in the cloud or at the edge. When it's generated at the edge and the datasets are large, it is more efficient to trade and process that data at the edge for compliance, cost and performance reasons. Equinix data centers are within 10ms of endpoint IoT devices in most metros.



OpEx Model

Equinix provides robust, interconnected bare metal service across the globe through its subsidiary, Packet, to support customers who want to quickly spin up a data marketplace in an OpEx consumption model.



Hybrid/Multi-Cloud Model

Any data marketplace solution will have to interact with data and services stored in multiple public clouds. Data can persist in the marketplace at an Equinix location then get moved to the public cloud for sharing, or vice-versa. Equinix data centers are within 1-2ms from large public cloud data centers in most markets. Equinix data centers are connected via 100GB pipes to public clouds. Approximately 4,000 clouds have their edge presence in Equinix data centers.

Equinix data centers are the ideal location to host data providers, aggregators and marketplace vendors. It is important to note that even if these providers have their control planes located in their private data centers or in public clouds, it is still possible and beneficial for them to have their data exchange nodes at Equinix. To become a true global solution, additional data center providers must be invited to join data marketplaces, as enterprises, public and private clouds, and edge infrastructures are distributed across multiple and competitive data centers.

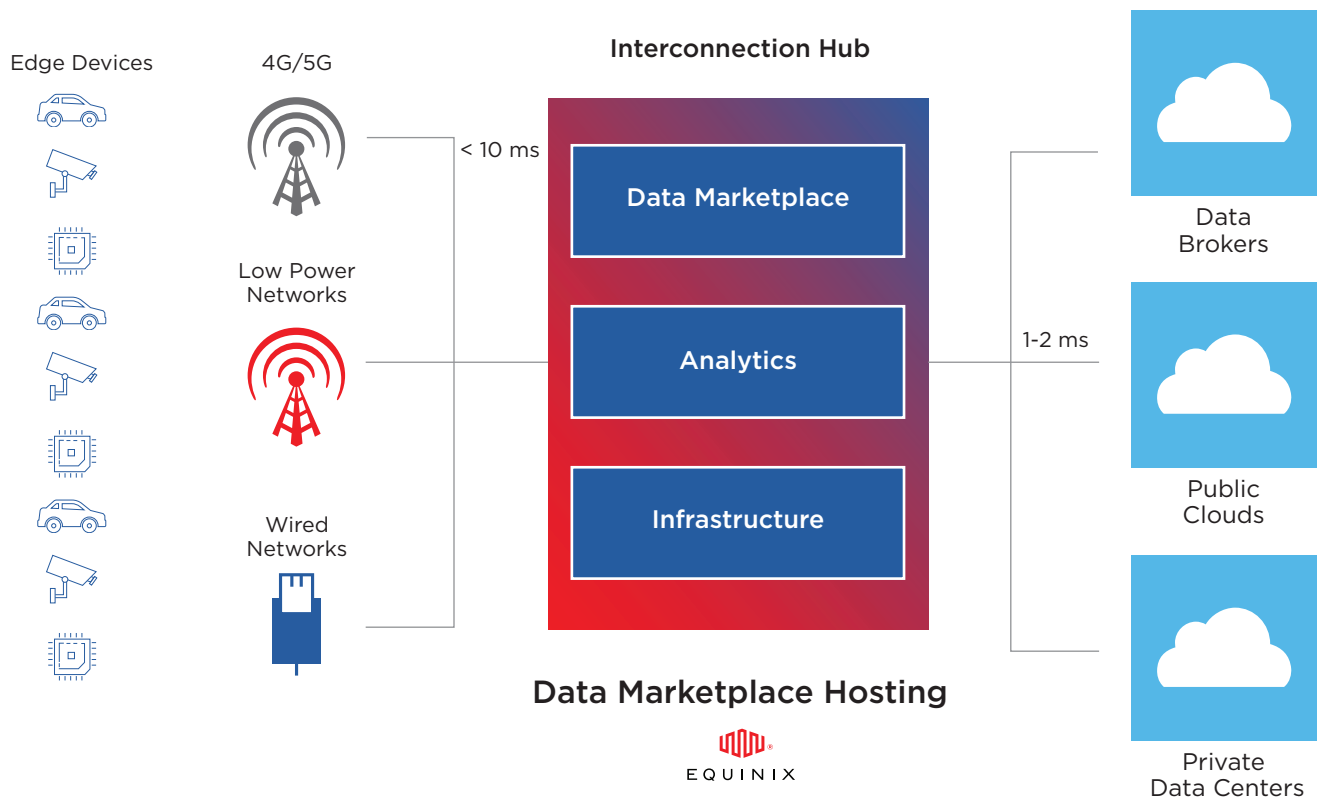


Fig. 8. Data marketplace hosted at an interconnection hub

6.0 Data Marketplace Operational Model



Fig. 9. Data marketplace usage workflow

Let's take a look at the steps involved in standing up a data marketplace.

Step 1

Consortium Setup and Membership Registration

The first step is for a consortium to set up a data marketplace. This consortium may elect to create and offer a shared data processing infrastructure governed and administered by the membership organization. The consortium specifies the location of consortium-shared infrastructure, most likely close to where member data is located. Members may choose to store their data permanently in a private section of the membership infrastructure, or in separate, private data cages offered by a neutral data center provider such as Equinix.

Alternatively, a high-speed connection from a private data center, located near the data center hosting the consortium infrastructure, can be used to transport data for processing. Algorithm developers often deploy initially in public cloud infrastructures. Therefore, the consortium infrastructure also needs to be in proximity to cloud infrastructure services. Many large data centers offer cloud exchange facilities, enabling the consortium infrastructure to be located close to public clouds (e.g., AWS, Google, Microsoft Azure). As shown in Figure 9, Equinix data centers are interconnection hubs that are close to public clouds and end user devices.

Step 2

Asset Registration and Trade Agreements

After setup, a consortium provides access credentials to its members so they can register themselves and be admitted to the data marketplace. Once admitted, members can complete their registration and start registering their assets. Members decide which information they want to publish publicly on the marketplace to attract interactions to do business with other members. Once members agree to explore sharing data with each other and establish a contract that arranges visibility of available assets, information describing the available data (meta-data) becomes visible to prospective members. This allows members to negotiate access and usage of specific datasets and/or algorithms.

Once members agree on which assets can be accessed and used, they create a data trade agreement. This agreement authorizes subsequent data science transaction execution that accesses and uses data. All agreements are stored in an immutable, distributed ledger, which provides auditability for compliance or dispute resolution. The Nokia Data Marketplace solution at Equinix offers a blockchain ledger (Hyperledger) that allows the data asset trade agreements between members to be specified via smart contracts. It also logs all transactions on the Hyperledger for auditing and lineage tracking purposes.

Step 3

Data Scientist and Production Workflows

Data scientists can experiment in the data marketplace. They can upload their data science workflows from their laptops or from public clouds, then establish contracts with providers and examine the quality of their datasets on a test basis. After they are convinced of the quality of the datasets, they can purchase them and train their models in the marketplace data exchange location. They can also bring their own private datasets into the marketplace. Once data scientists have successfully trained a model, they can do model inferencing in the data marketplace or take the trained model out, with permission, and use it for inferencing in their private clouds.

Nokia Data Marketplace Solution at Equinix

The Nokia Data Marketplace solution at Equinix can be deployed for:

Bilateral Data Exchanges
Between Companies

Consortium-Based
Data Marketplaces

Single Entity-Driven
Data Marketplaces

Data Sharing Between Departments
Within the Same Organization



To learn more, check out the data marketplace at www.nokia.com/networks/services/data-marketplace

References

1. Google AI Conference, San Francisco, 2019
2. 100 Data and Analytics Predictions through 2021, Gartner, 2017
3. IDC Predictions Provide a Blueprint ...for Becoming a Digital Native Business, IDC, 2017
4. Data as a Service, https://www.idc.com/getdoc.jsp?containerId=IDC_P31301, IDC, 2020
5. Gartner Data Broker Report, Gartner, 2017.