The Role and Functionality of an Independent Data Layer - IDL

Providing Unified, Federated, Distributed Data Access to Enterprise Apps of All Types

SkyFoundry

Table of Contents

The Role and Functionality of an Independent Data Layer

The concept of an **Independent Data Layer** is to elevate data to be its own independent entity without dependence on the applications that may use it. An IDL makes data available to any and all applications via a standard, open interface. Put another way, an IDL "frees the data" from being restricted to any single application.

The concept of an IDL is especially important as new Proptech applications rapidly emerge, all of which need access to diverse, existing data sources. The IDL vision: clients should be able to use their data with any new application that comes along without having to duplicate the work to access it. **We at SkyFoundry agree!**

In fact, a **data-first design,** providing **open access** to any external application via a standard, open, fully supported API has been fundamental to SkySpark from day one.

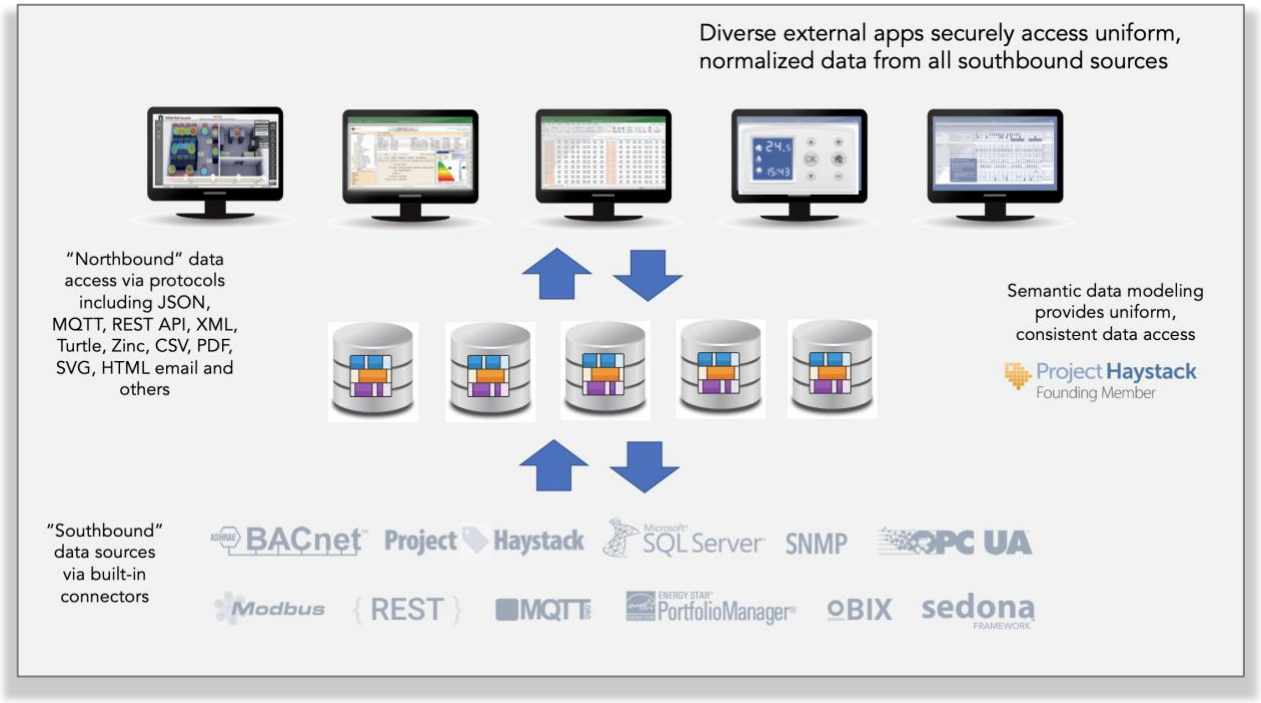
Our View on the IDL (and Clarification of Some Terms)

The term **independent** represents an important characteristic of an IDL, but to make data useful and truly interoperable, data needs to be "normalized". By this we mean applying metadata to contextualize data from diverse sources so that applications accessing the data see it as one uniform data model, no matter where the data comes from. SkySpark utilizes the Haystack standard (https://project-haystack.org/) to apply semantic tagging to create a uniform, standards-based data source.

A Single Central Data Lake? IDL functionality does not have to be dependent on a single, centralized "data lake". Much of the recent industry hype appears to imply that an IDL is singular in nature AND that all data from all systems must be pre-aggregated into a single repository. This approach is not cost effective, efficient, **or necessary** to achieve IDL functionality. Furthermore, it cannot achieve real-time operational requirements. Attempting to pre-aggregate all facility data from diverse systems and devices runs into numerous limitations including - latency, network reliability, a potential single point of failure and cost – data transfer is not free.

SkySpark – The Data Unification Platform for Operational Data

If we accept that we will have multiple different data sources and repositories AND want the functionality of an IDL to make data freely available to external applications, there is an additional term we need to introduce – "federated". By this we mean connecting distributed data sources/repositories into a trusted, shared network, accessible by external applications as if it was a single, unified data set. Federating distributed data sources eliminates the need to attempt to aggregate all data in a single repository, while still providing the functionality of an IDL to serve data to all applications that need it.



By combining SkySpark's industry leading capabilities for data acquisition from real world equipment systems, the highly efficient storage of its Folio database, its high-speed industrial historian, analytics processing engine, and efficient, open API for data access, SkySpark provides the essential capabilities to achieve the vision of an IDL – delivering normalized data from sensors, equipment systems, databases, and web-services to external applications.

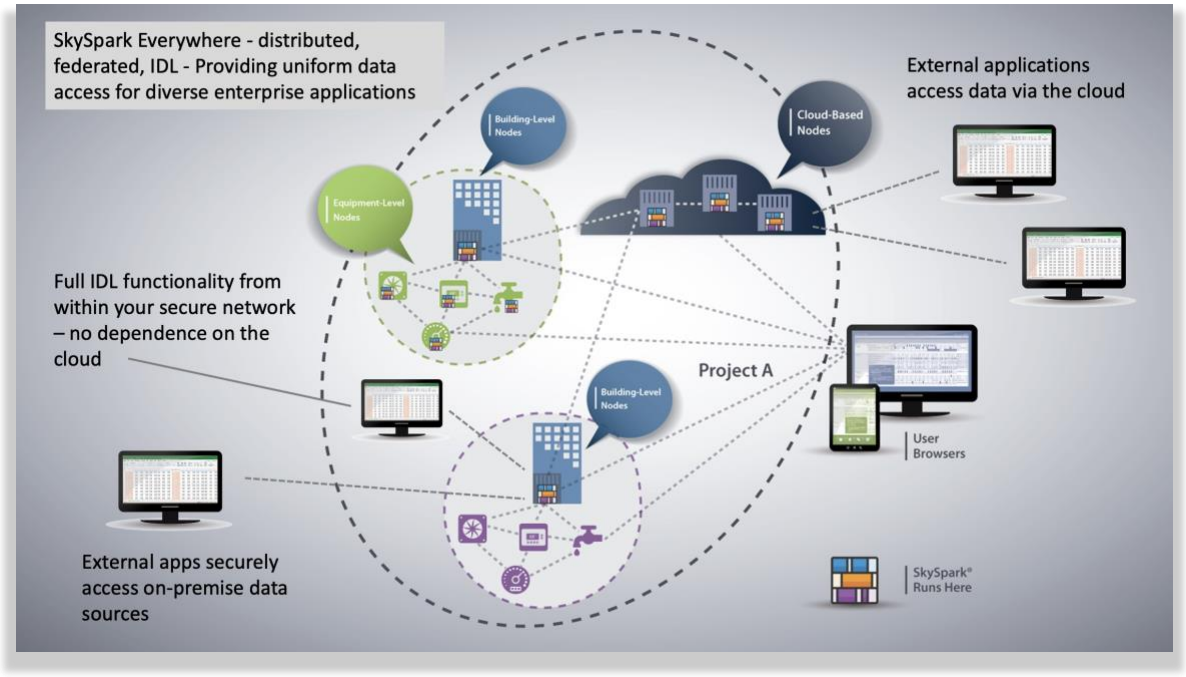
Unique Capabilities Set SkySpark Apart as an IDL Solution

Two unique capabilities set SkySpark apart as an IDL solution:

1. **SkySpark is NOT cloud dependent.** Some proponents suggest that the only way to implement an IDL is to store all data in a single "data lake" in the cloud. This approach restricts an IDL to be cloud dependent with the associated reliability and latency risks and other limitations mentioned earlier.

   With SkySpark, clients that need true on-premise deployment - with no external connections - can still have all the benefits of an IDL to serve data to any applications desired via SkySpark's open API.

2. **Distributed Architecture** – SkySpark provides a distributed data and compute platform enabling data acquisition and application processing to be performed across multiple distributed computing devices - from IoT edge devices to PC's, servers and cloud platforms. This unique capability enables SkySpark to provide a distributed, yet federated, IDL. (See diagram below).



The Importance of a True, Distributed Data Architecture

One of the most fundamental characteristics of the IoT is that it is a *distributed computing challenge*. Similar to how computer-based automation systems started in the 1980's with centralized computers and then moved to distributed-control architectures, the IoT is distributed by its very nature and requires software technology that supports that reality. *SkySpark® Everywhere™ provides that technology.*

We hear lots of talk about the "cloud" as it relates to the IoT. In many cases it seems like the "cloud" is being presented as the solution to all things IoT. The reality is that it is not possible, cost effective or desirable to transmit every piece of data from our devices to the cloud in order to gain value from the data. A true IoT architecture needs to recognize and embrace the highly distributed nature of the IoT and enable data processing at the edge as well as the cloud. An IoT platform must provide an architecture that matches that reality.

The IoT is a Distributed Computing Challenge

"The reality is that it is not possible, cost effective or desirable to transmit every piece of data from every IoT device to the cloud in order to gain value from that data.

An IoT technology platform needs to recognize and embrace the highly distributed, non-hierarchical nature of the IoT and support that with a corresponding software architecture."

Software functions need to be able to reside at every level of the architecture from the "edge" – for example, on a piece of equipment in an electrical closet or equipment room, to the building level where data from multiple smaller nodes can be collected, aggregated, and analyzed, to the cloud where data analysis can be performed at a portfolio level. SkySpark Everywhere provides this edge-to-cloud capability.

SkySpark Everywhere allows systems consisting of multiple SkySpark nodes to be connected in unified, near-seamless systems. Data collection, storage, management, computation, analytics and visualization functions are performed by the individual nodes at whatever level they are deployed, AND all data is available to external applications via the SkySpark REST API as if the system were one single, seamless system.

Data Unification: Combining, Unifying, Normalizing Data from Diverse Systems and Sources

Even in modern buildings with "smart" devices, equipment systems are often siloed due to a variety of factors. Some examples:

- Different systems use different protocols and data formats creating barriers to interoperability
- Building automation system data is often not combined with utility rate data, which is essential to calculate actual energy costs
- In many cases, data comes from external websites via "web-services" requiring integration via API's (Application Programming Interfaces) – this involves some level of software development work, which can range from simple to complex
- Software applications often utilize proprietary databases that do not provide documented "schemas" to describe the meaning of the source data in a standardized format, this creates a barrier to easily query and utilize data in other applications
- Even the newest IoT devices and sensors are often "islands" with their own separate networks, user interfaces, communication protocols, and "closed" data repositories which are often available only via API's that require "pay to play" to access the data

The message – no one system or data source has it all and there is great diversity in the data formats used by different systems. A fundamental requirement to achieve an effective IDL is the ability to provide uniform access to data from diverse systems. SkySpark accomplishes this through its ability to directly communicate with external systems, devices and applications of all types, its standards-based semantic data modeling, and its open API which provide uniform data access.

Delivering Data to External Applications

Because all of the data contained in SkySpark, including all analytic results, KPIs and other calculations, are directly available to other applications, SkySpark provides the data foundation to serve the wide range of applications used by owners and operators of facilities.

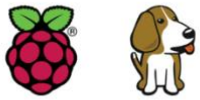
- Scaling for Very Large Enterprise Systems
  Multiple nodes of SkySpark can be deployed to support extremely large enterprise systems that collect store and process billions of data samples from equipment systems and IoT devices. SkySpark has numerous deployments operating at enterprise scale: https://skyfoundry.com/file/437/Case-Study-Proven-at-Scale---Enterprise-scale-Deployments.pdf

- Greater fault tolerance
  - By collecting data, processing computations and analytics, and creating visualizations as close to the data source as possible, SkySpark provides greater fault tolerance
  - Data replication features enable in-building personnel to have full access to their data – and analytic results – *even when they cannot communicate to an external cloud-based application or data source*

Flexibility and Choice - Deploy Where You Want – From the Edge to the Cloud

- Low latency
  - Computing at the edge provides near real-time data acquisition, analytics and control actions that cannot be accomplished with the latency of sending data to a cloud

- Support applications with "constrained networks"
  - IoT devices and equipment are often connected to slower, bandwidth limited or intermittent networks, or use cellular connections with high data transfer costs. By processing analytics at the edge, network traffic can be reduced by as much as 1000:1 !!!

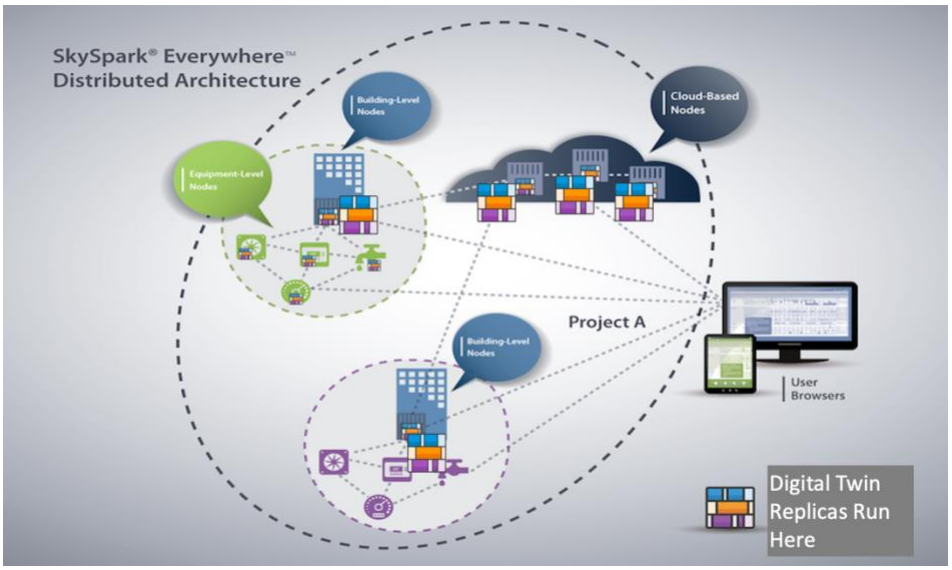


- Security
  - Many projects require data be kept on premise. SkySpark meets requirements for projects that cannot send data to an external cloud or are subject to regulations for data storage location
  - Isolating in-building systems from the Internet – SkySpark acts as a security barrier for connected equipment with its Arcbeam, WebSocket-based protocol

- Save Engineering Costs
  - Using SkySpark from the edge to the cloud means you engineer your application once with one uniform set of tools from the edge to the cloud.

- Reduce hardware costs
  - SkySpark at the edge can eliminate the need for additional gateway and security appliances in many applications.

Managing a Distributed Data Environment – Clustering, Replication, and Provisioning



When you have systems that consist of multiple computing nodes distributed across a facility or the world, you have another challenge – the need to manage the data they collect, the computational results they produce, as well as the software they contain. SkySpark Everywhere's unique technology directly addresses these needs with its clustering, provisioning and replication features.

The term **clustering** refers to the ability of multiple, distributed SkySpark nodes to be connected into a unified system using SkyFoundry's highly efficient and secure Arcbeam protocol. Once connected into a cluster, users AND external applications interact with the data, analytic results, reports, and views as if they were interacting with a single database. The result is a seamless user experience even as data acquisition and processing are distributed across many computing nodes.

Replication - Going Beyond Simple Backup of Data

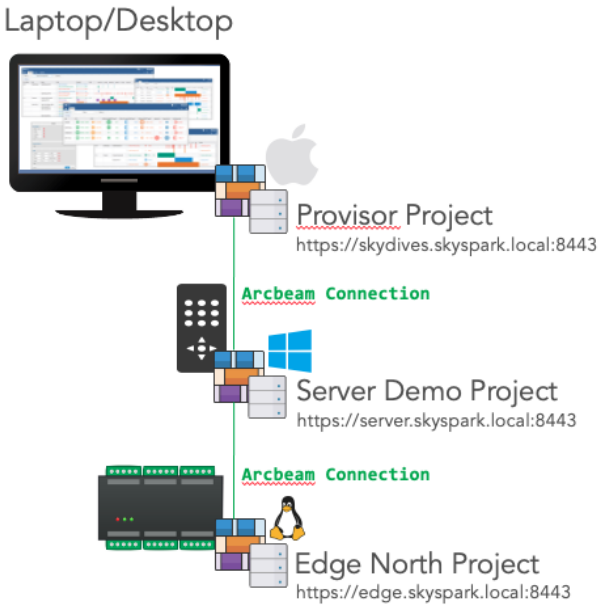
SkySpark's **Replication** feature enables copies of SkySpark databases from distributed nodes to be automatically copied (replicated) to one or more servers. Replicas are not simple data backups. Rather, SkySpark replicas are fully operational copies of individual nodes – providing full user interaction even when the original data source(s) are not available (offline). Replicas allow users **and external Apps** to work with the last available data and analytic results. And replicas continue to make all this data available to other applications, providing continued operations, even when live data feeds are interrupted. This capability is highly unique.

Replication features allow databases from edge nodes to be synched (copied) to one or more servers on-premise, or in the cloud, on a configurable basis. The result are replicas that allow users and external applications to fully interact with data from systems that may be offline or only intermittently connected.

Provisioning – Managing Software Versions Across Highly Distributed Systems

Another requirement for systems that consist of multiple distributed computing nodes is the need to manage the software they contain.

SkySpark's **Provisioning** features provide the ability to upgrade and deploy files to SkySpark nodes connected in a cluster. Provisioning makes it easy to manage SkySpark software on distributed nodes as well as other software, including third-party software and even operating system files.



Clustering, Replication, and Provisioning

- SkySpark provides automated replication of nodes in distributed systems
- Provides a fully operational replica of each node that users can interact with
- User queries work with last available data in a replica – can save significant data transfer costs
- Replication also provides a full automated backup of individual nodes
- Saves time, engineering, and IT support effort

Clustering, Replication, and Provisioning provide the depth and range of features essential to provide clients with true, distributed processing Independent Data Layer capabilities whether deployed on-premise, at the cloud or both.

ABOUT SKYFOUNDRY

SkyFoundry's mission is to provide software solutions for the age of "the Internet of things". Areas of focus include:

- Building automation and facility management
- Energy management, utility data analytics
- Remote device and equipment monitoring
- Asset management

SkyFoundry products help customers derive value from their investments in smart systems. Contact us to learn more.

https://skyfoundry.com/

info@skyfoundry.com

SkyFoundry