Welcome to Obsrv

Obsrv was conceived and incubated within the open-source initiative Sunbird in 2016. Obsrv is built to operate with highest levels of reliability with bare minimum operations effort at scale.

One of the early and significant applications of Obsrv is within **DIKSHA**, the national platform for schools and teachers in India. In the DIKSHA environment, Obsrv handled data volumes ranging from a few million events per day to **2 billion events per day** at peak. The current average volume is 500 million events per day and the system has not experienced any downtime in the past 2 years.

This microsite provides a comprehensive overview and guides to understand and use Obsrv for various data use cases.

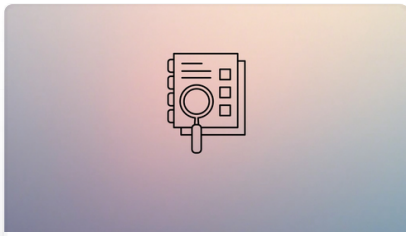
Getting Started

Introduction

Intro to Obrsv

Core Concepts

Understand Obsrv Fundamentals



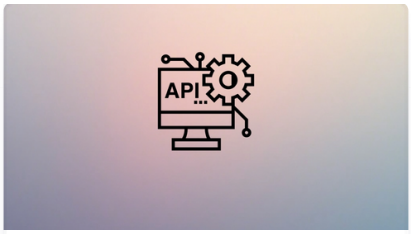
Case Studies

Real World Use cases

Guides

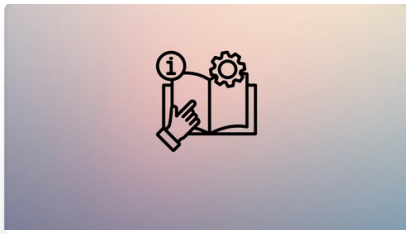
Installation

Guide to setup Obsrv



APIs

API Specification



Developer Guide

Inputs for the Developers to setup and contribute

The Value of Data

Importance of Data and its use-cases

In today's digital age, data has become the cornerstone of innovation, driving decision-making processes, and revolutionizing industries across the globe. Below are a few data-driven approaches that drive significant organizational growth:

- **Informed Decision-Making**: Data empowers organizations to make informed decisions backed by evidence rather than intuition alone. By analyzing patterns, trends, and correlations within data sets, businesses can optimize operations, mitigate risks, and capitalize on emerging opportunities.
- **Enhanced Customer Experiences**: Understanding customer behavior through data analytics enables businesses to tailor products, services, and marketing campaigns to meet evolving consumer preferences. Personalization based on data insights fosters stronger customer relationships and increases brand loyalty.
- **Predictive Capabilities**: Advanced analytics and machine learning algorithms leverage historical data to predict future trends, behaviors, and outcomes. By anticipating market shifts, demand fluctuations, and customer needs, organizations can stay ahead of the curve and adapt proactively to changing circumstances.

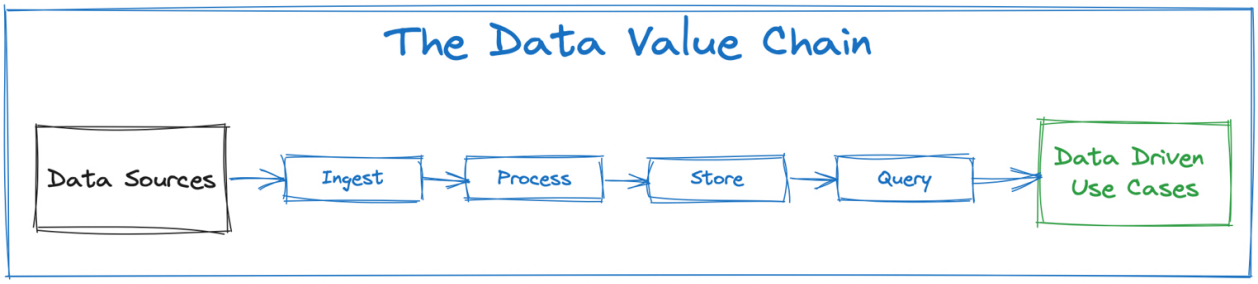
The value of data lies not in its abundance but in its transformative potential to drive innovation, inform decision-making, and create tangible value across diverse industries. By embracing data-driven methodologies and harnessing the power of advanced analytics, organizations can unlock new insights, capitalize on emerging opportunities, and chart a course towards sustainable growth and success in the digital era.

> ***What's fundamentally needed is a robust "Data Value Chain" capable of unlocking the full potential of an organization's data assets and driving sustainable value creation.***

Data Value Chain

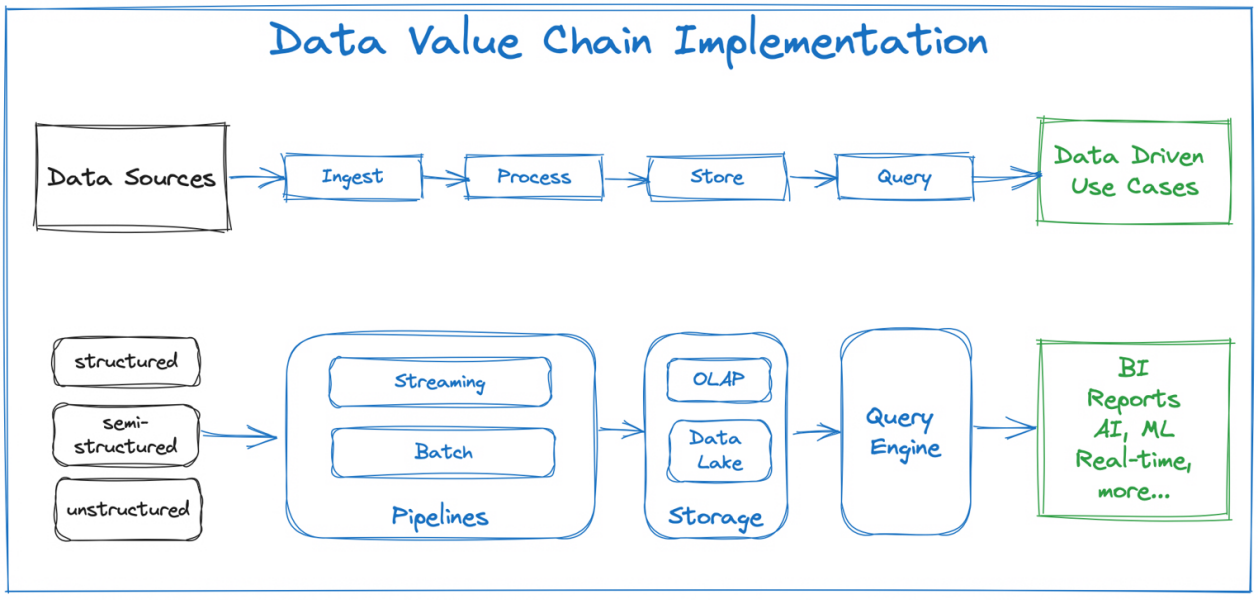
Stages of a Data Value Chain & typical implementation

The Data Value Chain represents the journey of data from its raw form to actionable insights and strategic decision-making. The data value chain typically encompasses the stages shown in the below diagram:



Stages of Data Value Chain

Efficient management of the Data Value Chain is crucial for all companies seeking to derive maximum value from their data. The figure below shows the typical technology components used in implementing a data value chain.



Typical Technology Components in a Data Value Chain

> ***There is growing demand for a robust data value chain in organizations seeking to extract maximum value from their data…***

Challenges

Challenges in implementing and managing a robust Data Value Chain

Despite the availability of numerous scalable and dependable technologies in the data space, the combination of these technologies often results in a fragile end solution. Only select big-tech companies have successfully mastered the processes of generating, consuming, and utilizing data reliably at any scale—examples include Google, Facebook, Netflix, Amazon, and LinkedIn. The existence of reliable data platforms facilitating a robust Data Value Chain has served as a significant distinguishing factor for these companies. And none of these companies have shared their end-to-end solutions with others, only a handful have released some of the tools they internally use, such as Facebook's contribution with Cassandra.

> *A pressing issue for many organizations is the substantial effort required to develop, operate or maintain an end-to-end data solution reliably.*

The challenge of reliably managing the data value chain is growing for numerous companies, particularly as contemporary products generate substantial data volumes, even with a relatively modest user base.

Key Challenges

The challenges faced by most of the organizations in operating a data value chain mainly falls into one of the following four categories:

1. **Time**: Significant amount of time mis-spent in managing the solution rather than in leveraging data's full potential
2. **Cost**: High upfront CapEx and running costs
3. **Capability**: Challenges to build, manage and operate complex data technologies & systems
4. **Risk**: Business & technical risks due to propietary & fragmented solutions, rigid & less reliable systems

Listed below are some challenges that are often faced by organizations with data platforms & solutions:

- Comprehensive, ready-to-use solutions for implementing the entire data value-chain are scarce. In many instances, organizations resort to employing extensive teams of Data and DevOps engineers to build these solutions.
- Data Analysts consistently grapple with challenges related to data integrity, quality, and accessibility, primarily due to the dynamic and evolving nature of data.
- Data Engineers spend substantial time resolving reliability issues due to the agile nature of data and interoperability challenges between components of data platforms.
- Inherent complexities of data platforms, when exposed, increase timelines for new pipeline creation, increasing the lead times for generating data insights.
- Organizations find themselves compelled to transition to a new data solution as they grow and have diverse data use-cases.
- Majority of existing managed solutions, if not all, bind users to proprietary data tools and formats, creating a vendor lock-in.

The Solution: Obsrv

Obsrv: A Resilient and Reliable Data Value Chain Orchestrator

The ideal solution to address the challenges in creating & operating a data value chain should have the following characterstics:

- Retrieve data from diverse sources, comprehend various formats, and adjust to any changes.
- Handle and store data without requirement of scripting or coding.
- Facilitate data utilization for all scenarios.
- Function reliably at any scale without necessitating modifications.
- Orchestrate the optimal data value chain...

Obsrv brings together the best data tools and technologies, and seamlessly orchestrating their integration through extreme automation techniques. The outcome is an end-to-end low-code data platform that is not only reliable but also resilient across diverse data requirements. Over the course of seven years, Obsrv has evolved to effectively tackle a broad spectrum of challenges in data analysis and data engineering.

> ***Functioning as an orchestrator of the Data Value Chain, Obsrv connects data to its inherent value.***

Key Benefits of Obsrv include:

- **Unified Data Infrastructure:** Obsrv is an end-to-end low-code data platform, facilitating integration of data from diverse sources, adaptable to changes and generating valuable insights.
- **Built-in Observability:** Obsrv possesses innate observability, knows when the data breaks and avoids data down times, ensuring a continuous and reliable data flow.
- **Reliability by Design:** Obsrv is engineered with extreme automation, ensuring seamless & reliable operation irrespective of the scale at which it is deployed.
- **Instant Data Utilization:** Obsrv facilitates configuration for input data sources, transformations, and the entire pipeline effortlessly, without the need for coding.
- **Diverse Applications:** Obsrv empowers use of data across a spectrum of scenarios, including real-time applications, ensuring organizations stay ahead in the era of rapid data-driven decision-making.
- **Freedom:** Obsrv uses open technologies & formats and its core engine is fully open source, guaranteeing zero lock-in and complete freedom to operate & exit.