



Data Ecosystem: An Introduction

Dr. Neeraj Tiwari

Professor & Head
Department of Statistics
Soban Singh Jeena University
ALMORA-263601, India
E.mail: neerajtiwari.amo@gmail.com

What is an ecosystem?

- An ecosystem is a geographic area where plants, animals, and other organisms, as well as weather and landscape, work together to form a bubble of life.
- Ecosystems contain biotic or living, parts, as well as abiotic factors, or nonliving parts.
- Biotic factors include plants, animals, and other organisms.

Data Ecosystem

- Organizations these days have a lot of data—**websites, customer information, general files, and overall enterprise infrastructure and applications.** The total sum of these components makes up the organization's data ecosystem.
- A data ecosystem refers to a combination of enterprise infrastructure and applications that is utilized to aggregate and analyze information.
- It enables organizations to better understand their customers and craft superior marketing, pricing and operations strategies.

Key components of data ecosystems



The people who use it



The technology that supports it



The processes that facilitate it

Base Components of Data Ecosystem

An infrastructural layer

- Which is the combination of hardware and software capture and manage data.
- This includes servers, databases, data warehouses, data lakes, and more.

Applications

- Which include the services and tools that act upon data to make it into usable sets.
- **Revelate** is an example of an application that acts upon data to transform it into usable data products that can then be shared or sold.

Analytics platforms

- Which provide the tools that decision-makers, data scientists, and analysts need to understand what the data is telling them, and how to act on that information.
- For example, Excel, SAS, R, MATLAB, STATA, Tableau, Jupyter, Tensor Flow, etc.

Data Eco Systems in context of Data Sharing

Closed data ecosystems

- Where data is shared in a closed environment consisting of several data providers with no additional outside access.

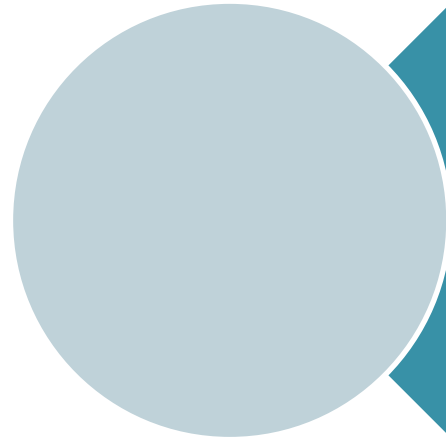
Strategic partnerships

- Where a small number of providers share data with each other for a specific purpose

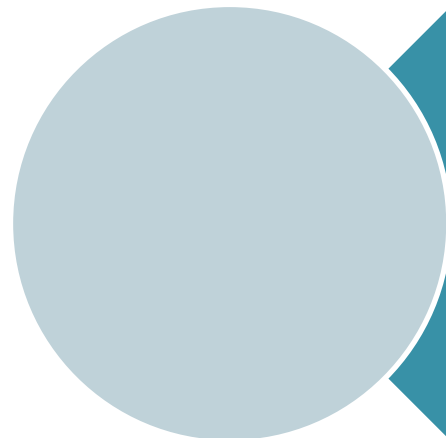
Open data ecosystems

- Where data is made publicly available.

A successful data ecosystem creates a mutually beneficial data sharing relationship between providers by:

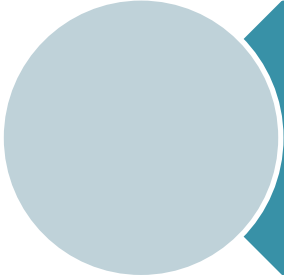


Allowing a low barrier to entry for relevant participants, with clear indications regarding how data is beneficial so that participants are less likely to want to leave the ecosystem and will continue to contribute to it

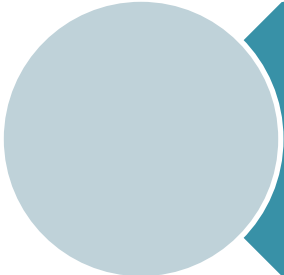


Motivating participants to join forces, typically because they have similar goals and interests and see the value of having different stakeholders in the system (such as users and developers)

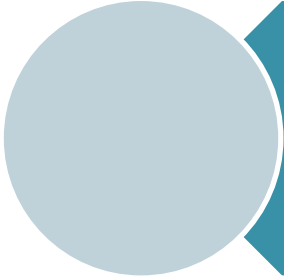
Real-life examples of how you can use data in your data ecosystem



You can use economic data and forecasts, as well as data from suppliers to improve demand forecasting and reduce instances of 'out of stock'.



Data from suppliers, social media data, and consumer data (such as purchase history and demographics) can be used by a telecommunication company to keep tabs on market changes and competition.



A transportation company can use geolocation data, traffic and routing data, and weather data to improve bus routing and ensure drivers arrive at stops on time.

Key Elements of Data Ecosystems

- **Source data**
- **Data storage ETL (Extract, Transform, Load)**
- **Data warehouses**
- **Data analysis infrastructure**
- **Data visualization**

Source Data

Internal Data Sources

- Internal sources are proprietary databases, spreadsheets, and other resources originating from your company.

External Data Sources

- External data sources are sources that originate from outside your organization.

Caution

- While identifying data sources for your project, you should evaluate its quality and accuracy.

Data Storage (ETL)

E

• **Extract**

T

• **Transform**

L

• **Load**

Data warehouses

Once the data is extracted and transformed during the ETL phase, it should be stored in a data lake or warehouse and eventually processed.

While a data lake holds data of all structure types, including raw and unprocessed data, a data warehouse stores data that has been treated and transformed with a specific purpose in mind, which can then be used to source analytic or operational reporting.

Storing data in lakes is different than storing it in warehouses.

Lakes preserve the original raw data.

Data stored in a warehouse is much more focused on the specific task of analysis.

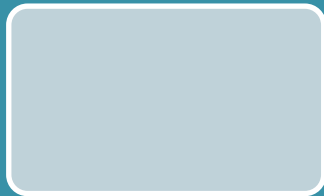
Difference between data lake and data warehouse

	Data Lake	Data Warehouse
Purpose	Big data storage and research	Big data analysis
Data Type	Unstructured, semi-structured & structured	Structured
Users	Data scientists and data engineers	Business analysts and data analysts
Cost	Low	High
Agility & Accessibility	High	Lower
Security	Lower	High

Data analysis infrastructure



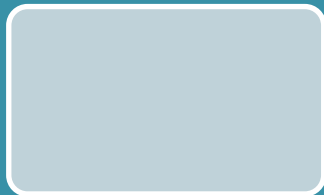
Diagnostic



Descriptive



Predictive



Prescriptive

Data visualization

- To make sure it is quick to understand, the data should be visualized as clean, clear charts
- Data visualization software helps users turn complex data into easy-to-follow charts and graphs.
- **Implementation of data analytics software** is a huge step toward data-driven, effective decision-making
- Data visualization tools include Looker, Tableau, Microsoft BI, etc.

Three fundamental elements of a modern big data ecosystem (technology stack)

a responsive data architecture

delivery at scale

AI-driven intelligent data management

Business benefits of a modern data ecosystem

**Higher
returns**

Cost savings

**Speed to
market and
to value**

**Customer
engagement**

**Process
improvement**

QUESTION



THANK YOU
ALL

