GIGASPACES

Data Hub vs. Data Lake vs. DIH

Data Hub

A Data Hub is an innovative architecture that streamlines data integration, and serves as the central repository in a hub-and-spoke design, powered by a multi-model database. The data hub offers a more cost-effective and efficient alternative to the traditional "relational database plus ETL" approach.

A data hub helps to eliminate the need for multiple copies of the same data, reducing the risk of errors and inconsistencies. It makes it easier to ensure data security and compliance, as all data can be centrally managed and secured.

Data Architectures Compared

	Data Hub	Data Lake	DIH
Purpose	Centralized repository of data for simplified data integration & management	Large scale repository of raw, unstructured data for storing & processing data at scale	Integrated digital systems & data sources to create a unified view of data
Data Format	Multi-model database	Raw, unstructured data	Structured & unstructured data
Data Quality	High-quality data due to ingestion of raw data & focus on data management	Raw data with varying quality	Improved data quality through data integration & management processes
Performance	Faster performance compared to traditional relational database & ETL processes	Performance can be slow due to large volumes of unstructured data	Improved performance through integration of multiple digital systems & data sources
Integration Type	Ingestion of raw data for immediate processing	Batch processing of raw data	Integrating data from multiple sources & systems
Data Governance	Focus on data governance & security for mission-critical data	Lack of focus on data governance	Focus on data governance & management for unified data view

Data Lake

A Data Lake accepts and retains all data from all data sources, holding data in an unstructured way with no hierarchy or organization among the individual pieces of data. A data lake tends to ingest data very quickly and prepare it only when people access it. Schemas are applied only when the data is ready to be used. Data lakes offer centralized storage, scalability, flexibility and are well-suited for MI and AI applications, providing a rich dataset for model training and development.

DIH

Digital Integration Hub (DIH) architecture is designed specifically to rapidly deliver new digital services while ensuring high throughput, low latency, and always-on service availability for operational and transactional workloads. A DIH typically utilizes functionalities such as data integration, data management, and data governance for the specific purpose of accelerating the delivery of new services to business applications. While a data hub describes various data management architectures that can address specific use cases, a DIH could include a data hub as one of its components.

GigaSpaces SmartDIH offers an out of the box DIH that enables you to quickly launch highly demanding real-time apps easily, at scale and cost-effectively.

About GigaSpaces

GigaSpaces is a global pioneer in in-memory computing. Smart DIH, the company's operational data hub, powers real-time modern applications, unleashing the power of customers' data by transforming data silos into assets, turning organizations into digital-first, data-driven enterprises. For more information visit gigaspaces.com