Data and Analytics Essentials: Data Lake — Presentation Materials

Published 17 August 2021 - ID G00755941 - 2 min read

Philip Russom

Initiatives: Data Management Solutions

Data and analytics leaders must help their peers and stakeholders understand key value propositions and best practices associated with data lakes and their use cases in analytics. This presentation on the fundamentals of data lakes will drive aligned and effective investment choices.

Overview

The data lake (DL) takes a different approach to data organization and database design, as compared to traditional practices for data warehouses and operational data. The essential characteristics of the DL are:

- Data in raw, native form
- Some means of processing that data, usually in support of exploratory and data science initiatives

For example, when a data lake integrates with a data warehouse, the data lake provides large volumes of detailed source data for advanced analytics and data science, which differs from the optimized data that is typical of a data warehouse. In this way, the data lake and the data warehouse complement each other to address a broad range of use cases in reporting and analytics. Most data lakes collect data from multiple operational sources and organize it on semantically flexible data platforms. This enables discovery-oriented analytics (e.g., mining, graph and machine learning), which are not possible with the same data in their original systems.

The data lake's overlap and integration with other data and analytics approaches has led to some confusion over data lake design and use, and the confusion is magnified by the continuing evolution of the data lake. For example, the earliest data lakes were only for data science; they continue to support that use case, but now also support others, including various forms of analytics, self-service data access, and operational uses in marketing, supply chain, and logistics. Early data lake designs were organized as a single large analytics sandbox on Hadoop, typically on-premises. Current implementations are usually organized as a series of data refinement zones on a cloud object store.

Data and analytics leaders should avoid these problems by educating their organizations about the capabilities and value points of the data lake. They can use this presentation to level-set the understanding and expectations of team members and key stakeholders, as well as gain their buy-in for data lake initiatives.

These presentation slides will be most useful to data and analytics leaders who are working to educate their stakeholders. Many Gartner clients working to get buy-in for their business case to modernize the data management infrastructure will use a simple presentation format to help executives understand the role and importance of each of these concepts.

Recommended by the Authors

Best Practices for Designing your Data Lake

Building Data Lakes Successfully — Part 1 — Architecture, Ingestion, Storage and Processing

Building Data Lakes Successfully — Part 2 — Consumption, Governance and Operationalization

Metadata Is the Fish Finder in Data Lakes

Data and Analytics Essentials: Data Warehouses, Data Lakes and Data Hubs

Market Guide for Analytics Query Accelerators