Designing and Building Your Hybrid Data Cloud

Juan Sequeda, PhD
Principal Scientist
@juansequeda

Bryon Jacob
CTO & Co-Founder
@bryonjacob

Dave Griffith
Principal Engineer
Takeaway

What does the future of Data Management look like?

● Short Answer: Hybrid Data Cloud!

● Long Answer: Enable People and Machines to find, access, and get value from data with its meaning and context, regardless of the data's location.
What data do we have?

Let’s catalog the data!

Start by cataloging data, metadata, and analysis across cloud systems, on-prem databases, and preferred tools.
How can I access the data?

Let’s access the data!

With a catalog unifying search and discovery, enable teams to access and seamlessly work with data no matter where it lives.
I don’t understand my data!

Let’s add knowledge to the data!

Develop a semantic layer based on the real-world concepts and relationships being represented in the data. Query the data via the semantic layer directly using the concepts and business terms that make sense to the end user.

1) Catalog
2) Access
3) Knowledge Graph
Amundsen — Lyft’s data discovery & metadata engine

Data Hub: A Generalized Metadata Search & Discovery Tool

Co-authors: Mars Lan, Seyi Adebajo, Shirshanka Das

As the operator of the world’s largest professional network and the Economic Graph, LinkedIn’s Data team is constantly working on scaling its infrastructure to meet the demands of our ever-growing big data ecosystem. As the data grows in volume and richness, it becomes increasingly challenging for data scientists and engineers to discover the data assets available, understand their provenances, and take appropriate actions based on the insights. To help us continue scaling productivity and innovation in data alongside this growth, we created a generalized metadata search and discovery tool, Data Hub.
The 2019 LinkedIn Top Startups Are Growing Fast

1) Snowflake
2) ...

Data Cloud

A data cloud is where your data is available to your people and your machines – it's where your data assets are leveraged.

+  

= Data Cloud
data.world's data cloud platform is built on semantic web standards - RDF and SPARQL at the core, and enriched with CSVW, DCAT, PROV, R2RML and more
Hybrid Data Cloud

The data cloud will evolve to be hybrid, combining elements of your on-prem systems and systems native to public clouds.
How do we start?

Knowledge Scientist (People)

Methodology (Process)

Hybrid Data Cloud (Tools)
Who is responsible of the data?

**Data Engineer**
Understand database schemas, including how the data are interconnected.

**Knowledge Scientist**
Serves as the communication and developer bridge between Data Engineers and Business Users

**Business User**
SME who understand the business

Data Access

“Geeky Person”

Business Modeling

“People Person”
A Pay-as-you-go Methodology to Design and Build Enterprise Knowledge Graphs from Relational Databases
Juan F. Sequeda, Willard Briggs, Daniel Miranker & Wayne Heideman
Best In-Use Paper Nominee
When: Monday Oct 28. 14:00-15:00
Session: 2A : Integration and Fusion
Room: 098

Business Question
THIS IS HOW WE MEASURE SUCCESS!

Self Service Analytics
8. Build Report
9. Answer Business Question
10. Move to Production

Enterprise Knowledge Graph
What does the future of Data Management look like?

Hybrid Data Cloud
Enable People and Machines to find, access, and get value from data with its meaning and context, regardless of the data's location

- Infrastructure & Tools
- Data Discovery
- Data Governance
- Semantic Understanding
- Collaboration
- Reproducibility

Enterprise
- Knowledge Graph
- Data Access
- Data Catalog

Enable People and Machines to find, access, and get value from data with its meaning and context, regardless of the data's location.