Gartner Data & Analytics Summit
18 – 21 March 2019 / Orlando, Florida

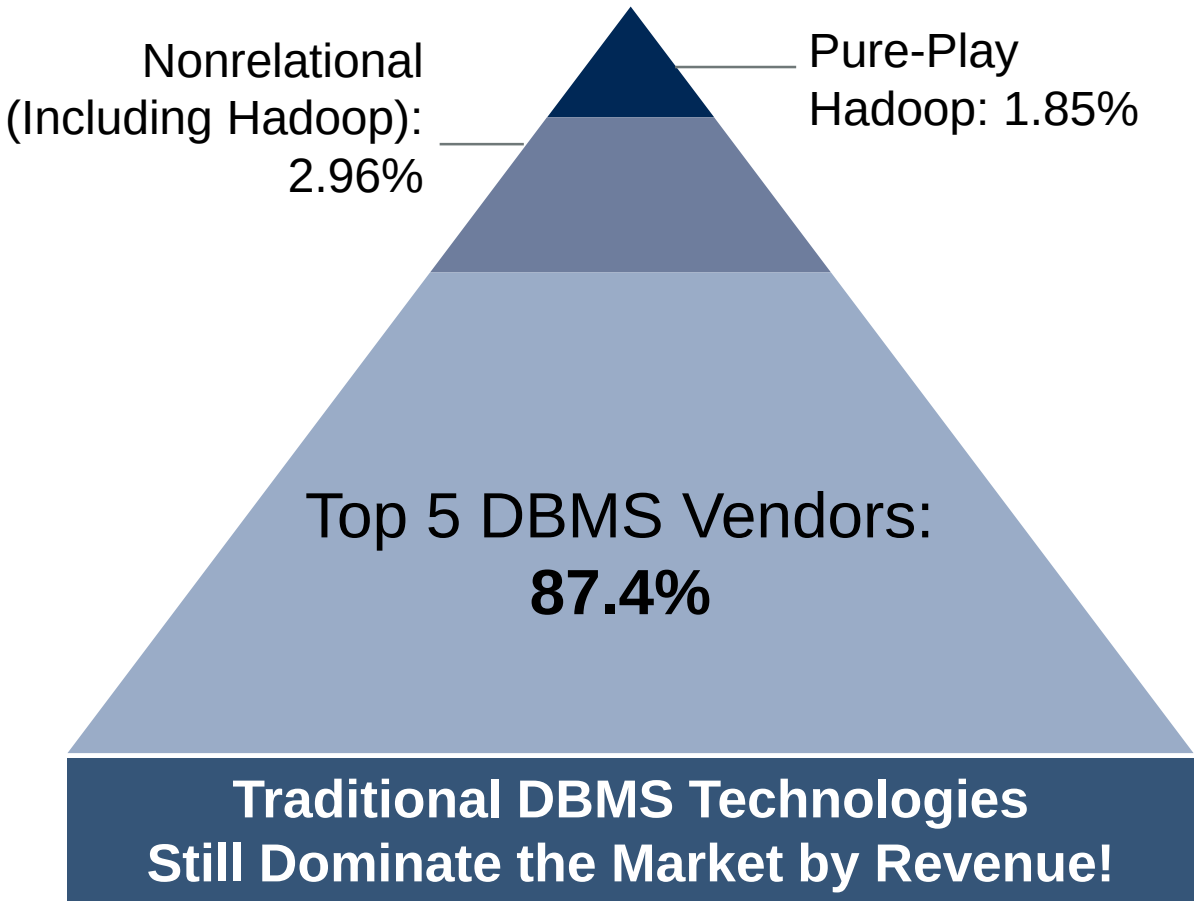
Data Management Solutions for Analytics: Why Hadoop Won't Replace Your Data Warehouse Anytime Soon

Adam Ronthal
@ARonthal

Gartner®


CHIHULY
R•M

DBMS Market Dynamics



2017 cloud revenue exceeded $5 billion

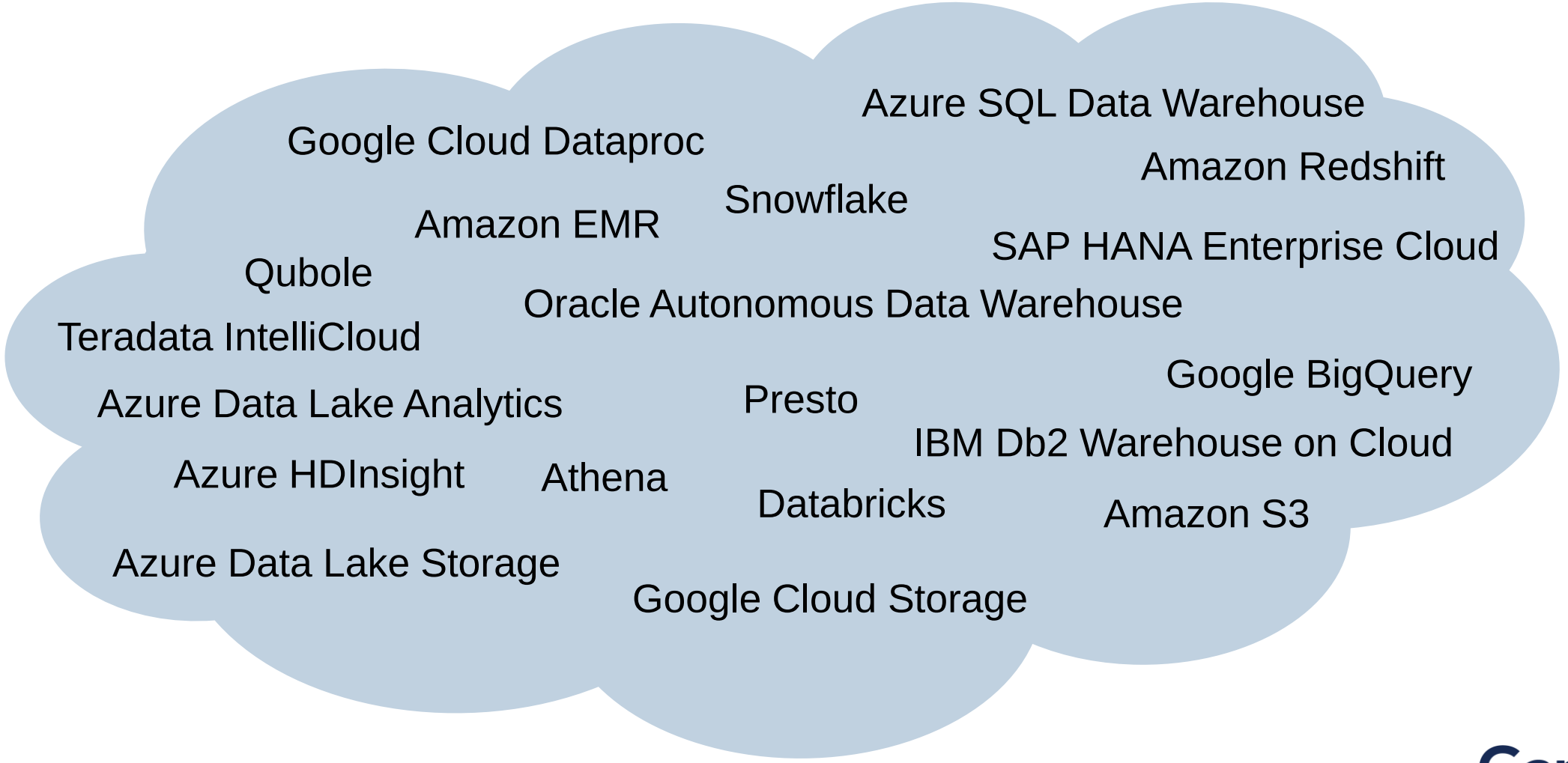
more than 13% of the market

USD **$38.8** billion market

Two cloud vendors are driving more than **70%** of the **$4.3b growth**

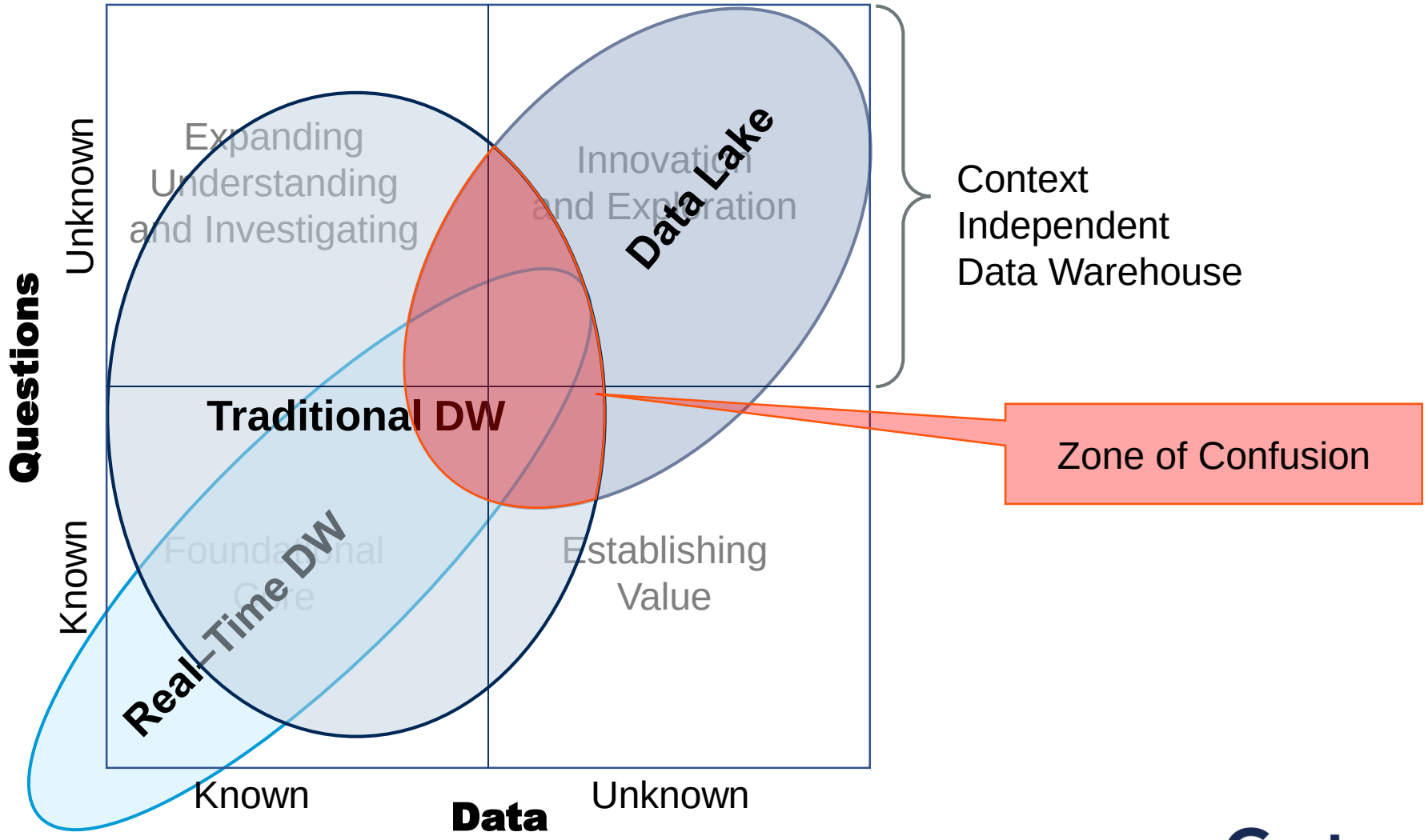
Source: "Market Share: Enterprise Infrastructure Software, Worldwide, 2017" (G00365968)

Cloud Is Driving the Market … What Role Will Pure Play Hadoop Vendors Have?



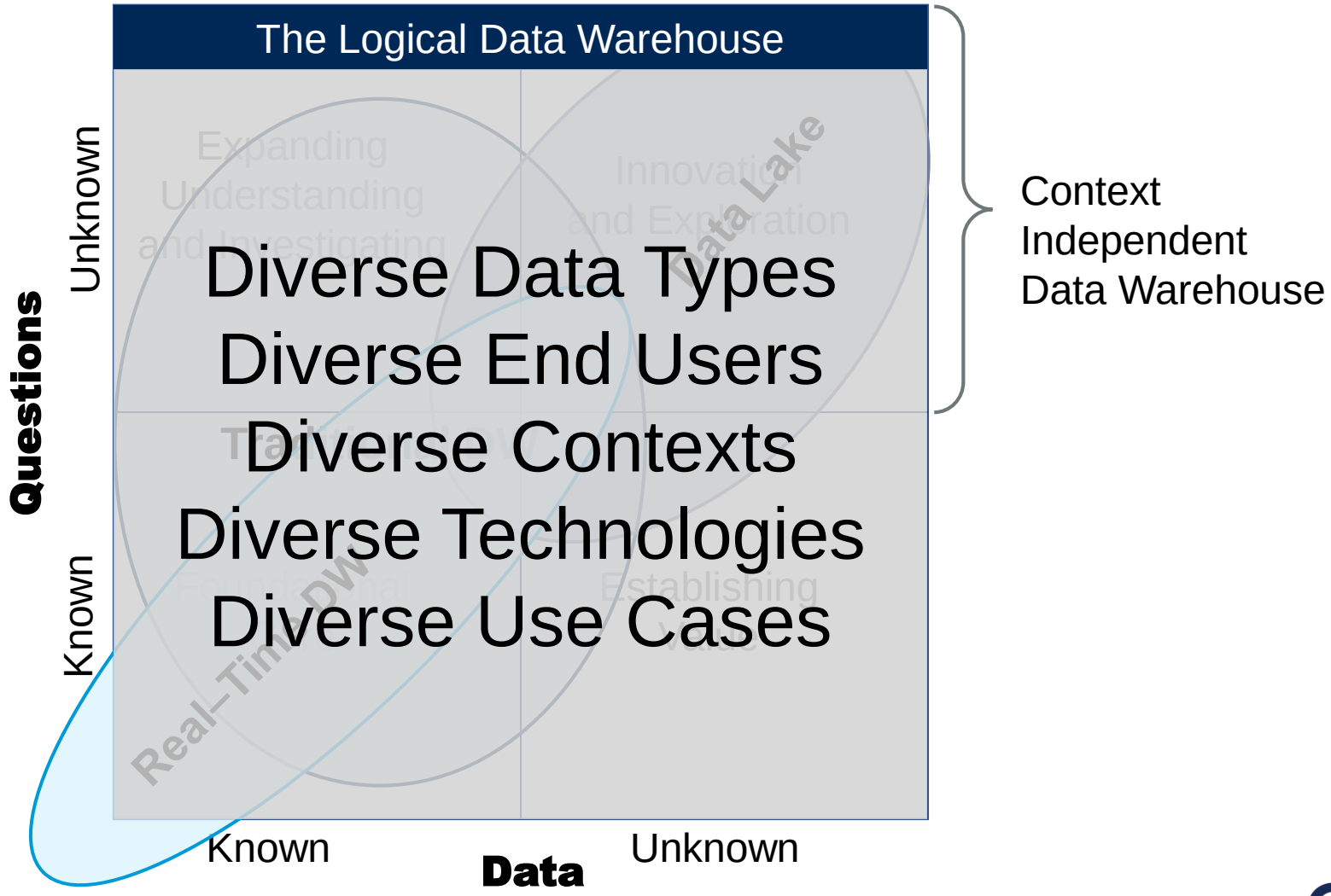
Gartner®

The Data Management Infrastructure Model and the LDW



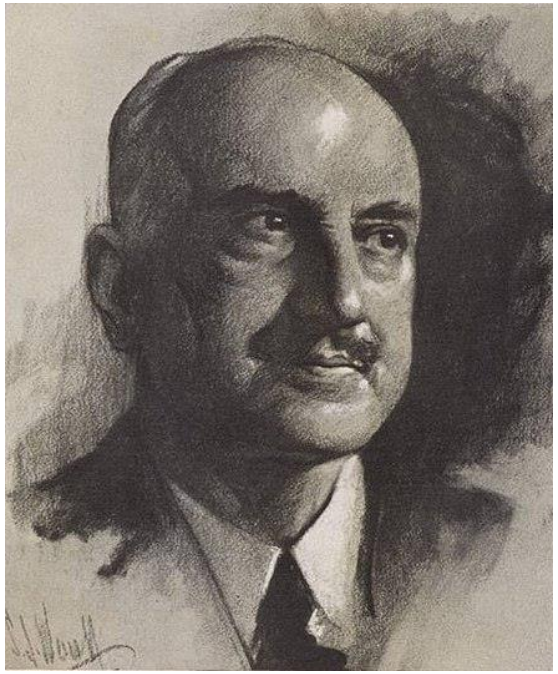
Gartner

The Data Management Infrastructure Model and the LDW



Why Make the Same Mistakes Again?

- Data warehouses do not have to be in a relational database; data lakes do not have to be in a nonrelational database.
- Data warehouses are optimized for consistency across aspects of performance, repeatability and integration.
- The data warehouse was pushed, pulled and prodded into roles that were beyond the mission…
- … because there were other missions! And some were misuse cases.
- The data lake, and its supporting technology stack, does not need to make the same mistakes.

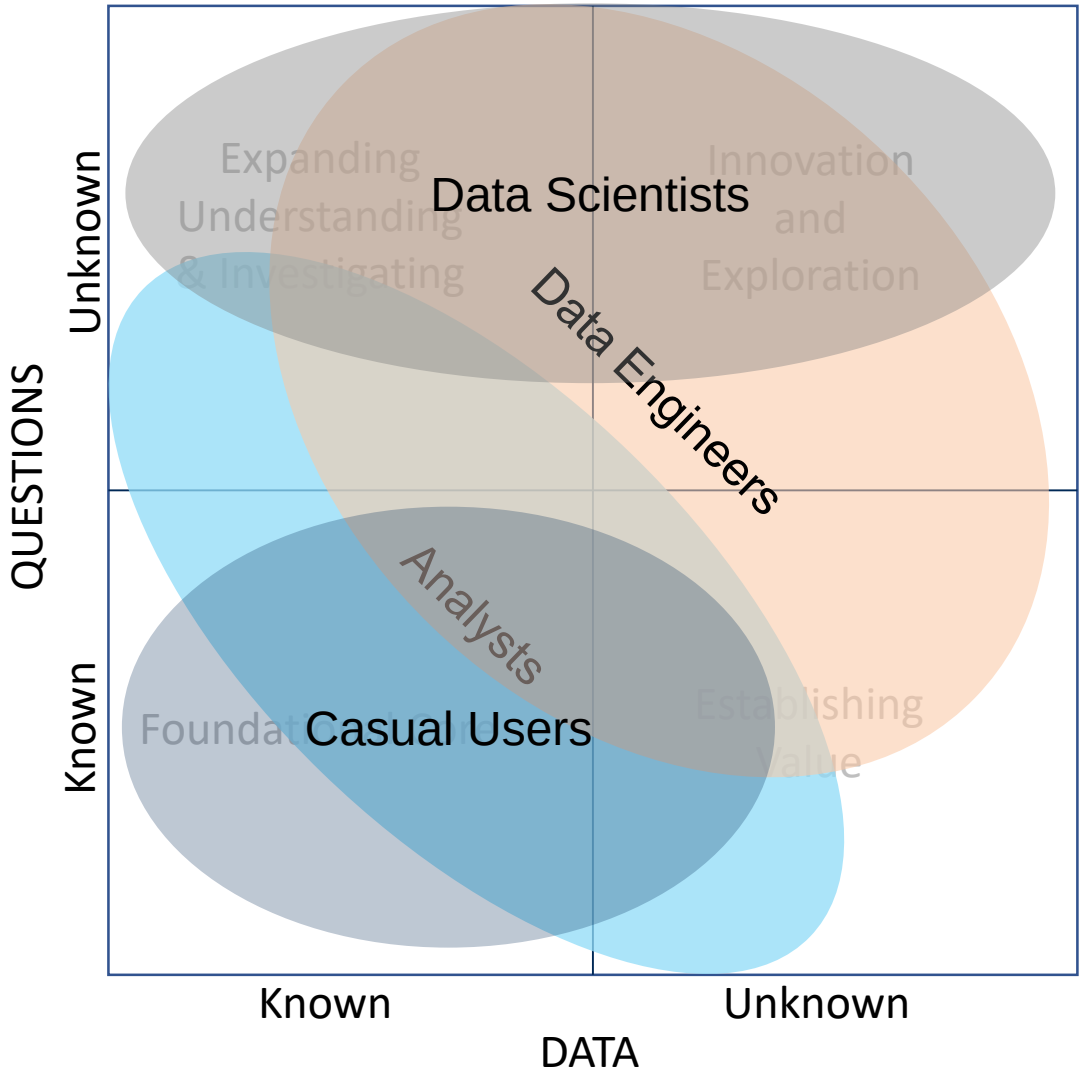


"Those who cannot remember the past are condemned to repeat it."

— George Santayana, The Life of Reason: The Phases of Human Progress (1905-1906)

Gartner

These Environments Serve Different Roles and Skills



	Analytics Center of Excellence	User Distribution
Casual (Apprentice)	▪ Reports, dashboards ▪ Level 1 support, possibly Level 2	1,000
Business Analyst (Journeyman)	▪ Creates new reports ▪ Needs technical assistance ▪ Level 2 and Level 3	90
Engineer (Master)	▪ Ops process, systems analyst, data architect, reliable tech	5
Data Science	▪ Modeling theory, graph theory, mathematics, program languages	1

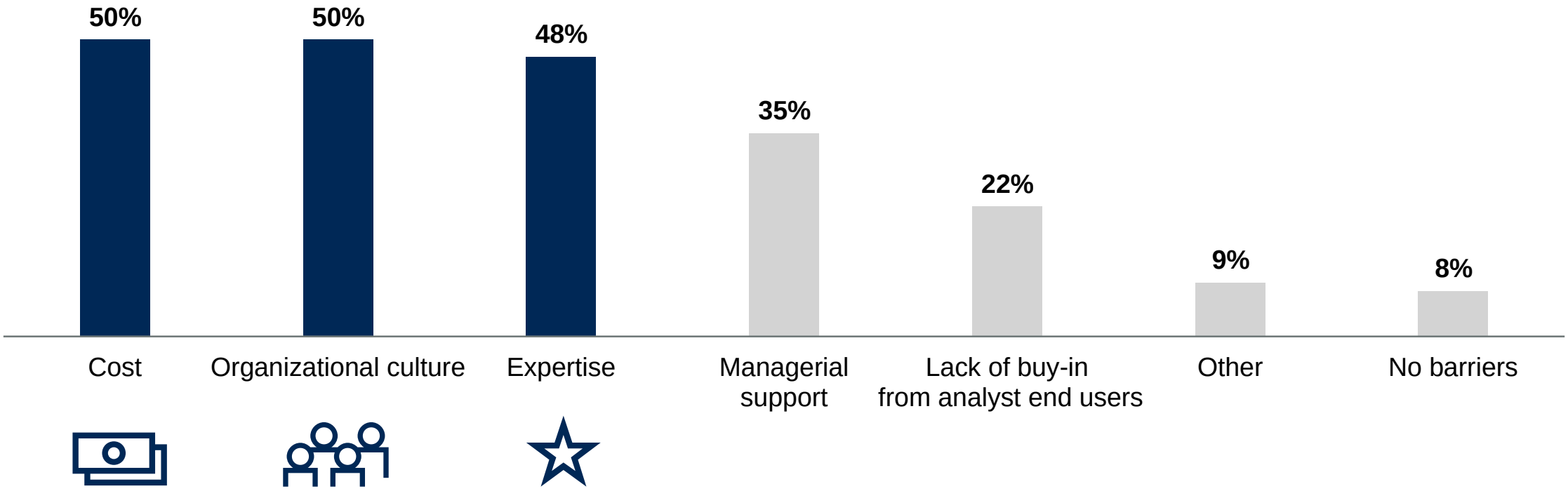
Source: "Organizing Your Teams for Modern Data and Analytics Deployment" (G00374036)

Gartner

Cost, Culture, and Expertise Represent the Barriers to Enhancing Data Availability for Analytics

Barriers to Enhancing Data Availability for Analytics

Percentage of Respondents



Base: n = 208 Gartner Research Circle Members External Sample; Excluding 'Don't know'
Q. What are the barriers to enhancing data availability for analytics at your organization? Please select all that apply.

Gartner®

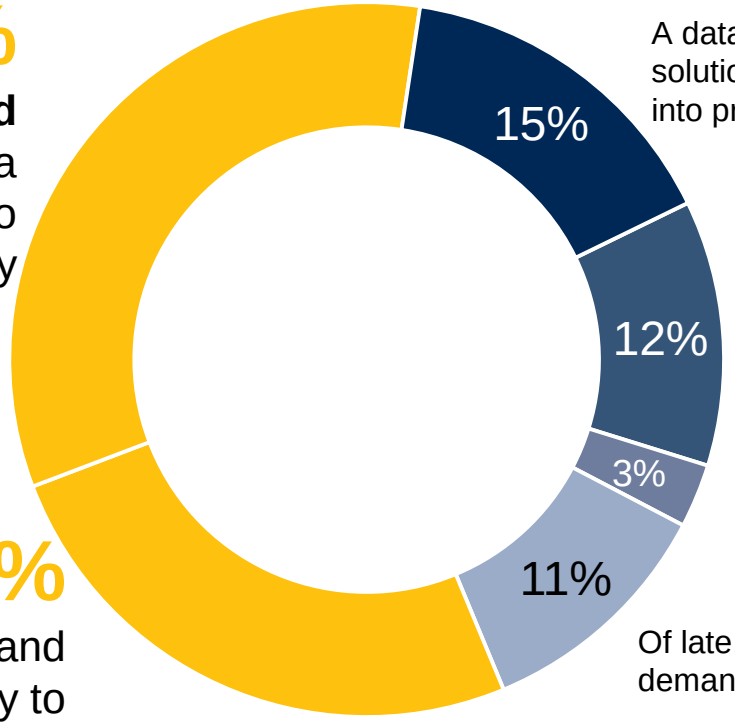
Increased Demand for New Data Types and Data Modeling Is Coming From Highly Skilled Users but Also Broadly From Throughout the Organization

What scenario best describes how your organization's data capabilities meet current demand?

Percentage of Respondents

33%

There is **increased demand** for new data types and the ability to model it independently



A data science team designs new solutions and helps IT put them into production for everyone else

Our business intelligence and reporting capabilities meet current demand

Of late, some users want to model their own data

Of late, some users are demanding more data types

25%

Demand for existing and new data types and the ability to model independently is coming **from highly skilled users only**

Base: n = 208 Gartner Research Circle Members/ External Sample; Excluding 'Don't know'
Q. Which scenario best describes how your organization's data capabilities meet current demand?

Gartner

Invest in Skills

Gartner's 2017 and 2018 surveys on the adoption and deployment of logical data warehouse architectures show a **troubling gap** between demand for new data and analytics and the skills available to address that demand.

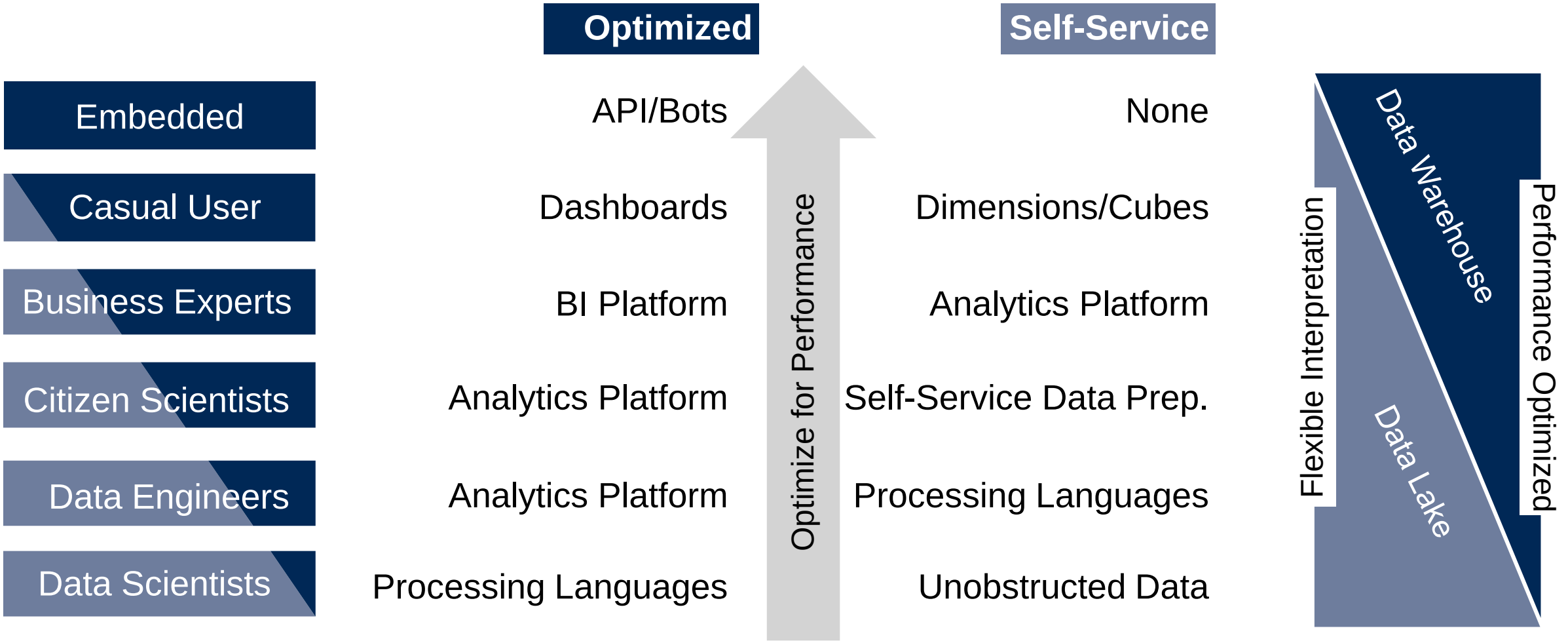
88% of organizations report that their current **data management needs** supporting analytics and reporting **remain unmet** and demand new data and data types.

Delivery capacity for prequalified data has almost doubled, **forcing** the **data management architecture** and infrastructure design to **support the rapid conversion** of data science discoveries **into production**.

Low-skilled analysts are demanding access to highly complex data use cases and **threaten to overwhelm the credibility of data use** in the digital business.

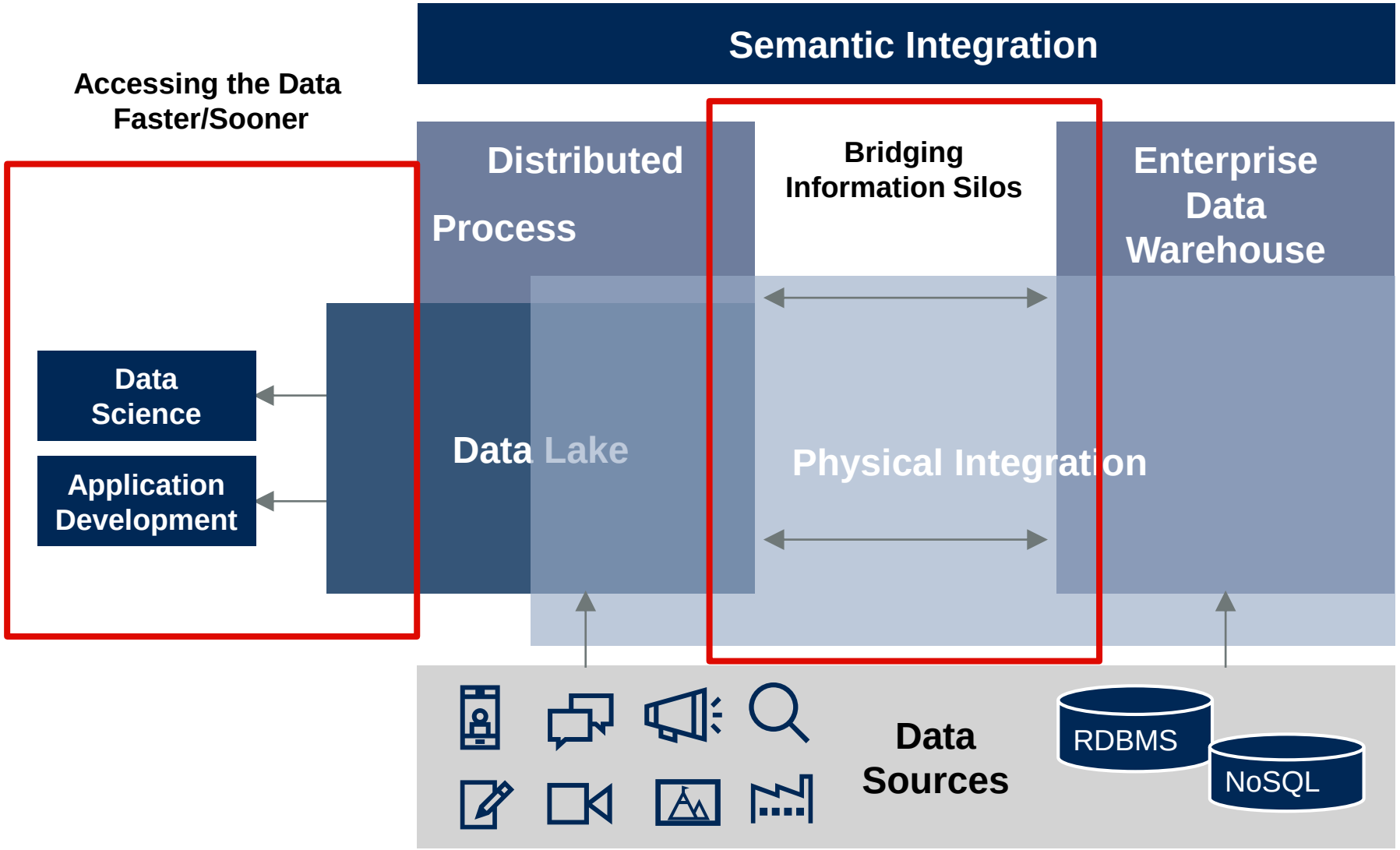
Gartner®

Infrastructures Must Balance Optimization Needs With Self-Service Demand



Gartner®

You Can Start "Light" From Either Direction — Lake or Warehouse, Simple Tools or Platform!



If you already have an enterprise data warehouse, you can extend it.

If you already have a data lake or Hadoop cluster that needs reuse optimization, you can extend it.

Gartner

Action Plan

Monday Morning:

- *Identify* classes of users and use cases present currently in the organization.
- *Catalog* different types of data management for analytics present.
- *Determine* which existing use cases/users are capable of self-support.

Next 90 Days:

- *Develop* timelines for when missing user classes and use cases are anticipated for support.
- *Identify* platform choices available from existing enterprise vendors and *perform* a gap analysis for capabilities that are missing.
- *Identify* a project in which initial work was done by data scientists and then convert to a performance optimized solution for lessons learned.
- *Map* your existing systems onto the data management infrastructure model. Many companies find that they already have 70% to 80% or more of the components.

Next 12 Months:

- *Target* a project to extend an existing warehouse, multiple marts or a data lake with new data **and** new use cases.
- *Evaluate* user experiences to create user qualifications for leveraging different infrastructure components.

Gartner®

Recommended Gartner Research

- **Magic Quadrant for Data Management Solutions for Analytics**
  Adam Ronthal, Rick Greenwald, Roxane Edjlali (G00326691)
- **The Practical Logical Data Warehouse: A Strategic Plan for a Modern Data Management Solution for Analytics**
  Adam Ronthal and Roxane Edjlali (G00342308)
- **Toolkit: Map Your Data Management Landscape with the Data and Analytics Infrastructure Model**
  Adam Ronthal, Rita Sallam, Carlie Idoine and Others (G00354009)
- **How to Avoid Data Lake Failures**
  Nick Heudecker and Adam Ronthal (G00367848)
- **FAQs on the Future of Hadoop**
  Merv Adrian, Arun Chandrasekaran and Adam Ronthal (G00363702)

For information, please contact your Gartner representative.

Gartner®