Gartner Data & Analytics Summit
Summit 2018
22 – 23 May 2018 / São Paulo, Brazil

Gartner Summits

Data Management Solutions for Analytics: Going Beyond the Data Warehouse

Adam M. Ronthal

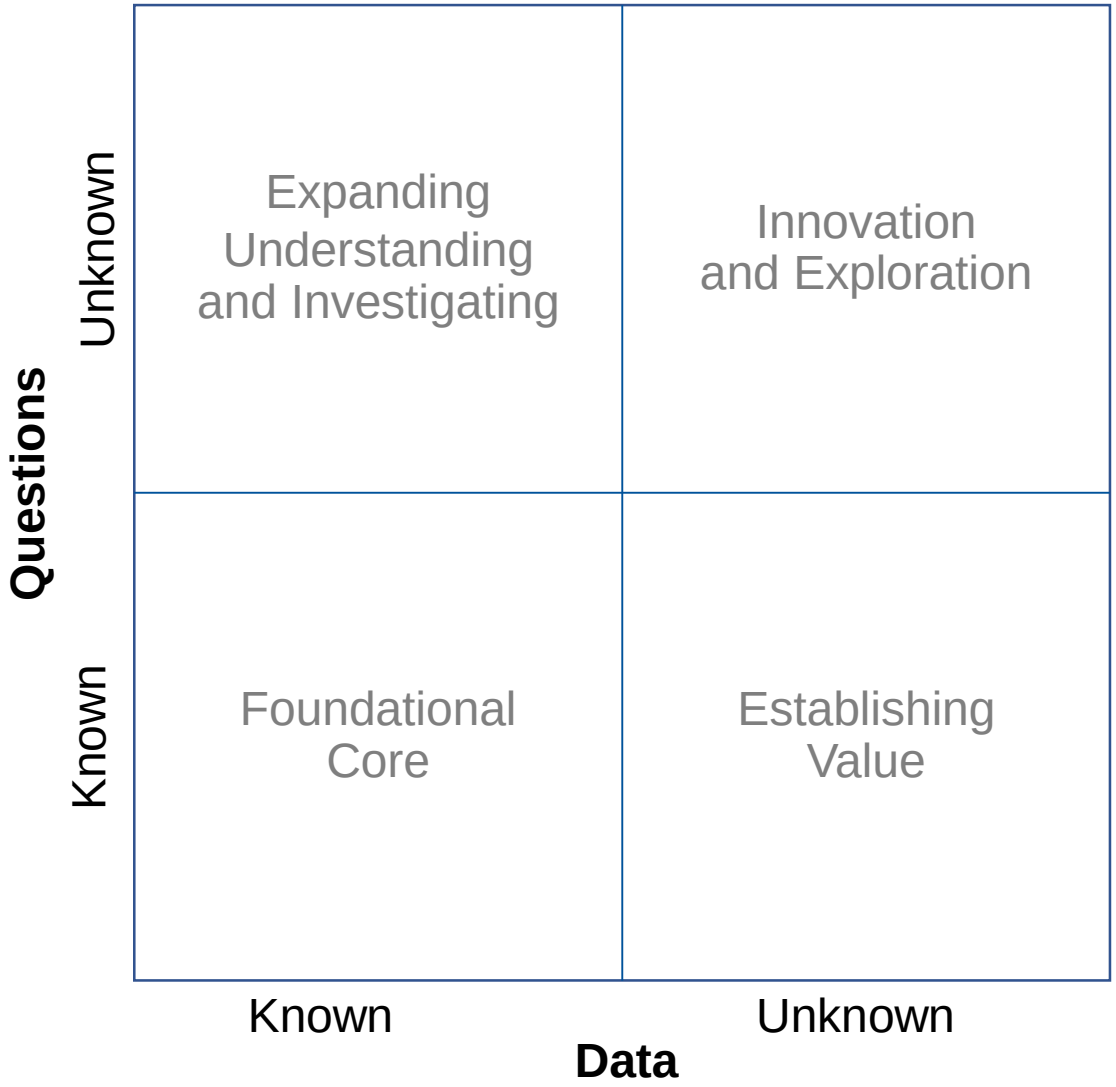
Key Issues

1. How can you address the increased scope and breadth demanded by DMSA platforms today?
2. Do traditional data warehouse platforms still have a place?
3. How do you build an architecture that leverages the traditional while using innovations?

Key Issues

1. How can you address the increased scope and breadth demanded by DMSA platforms today?
2. Do traditional data warehouse platforms still have a place?
3. How do you build an architecture that leverages the traditional while using innovations?

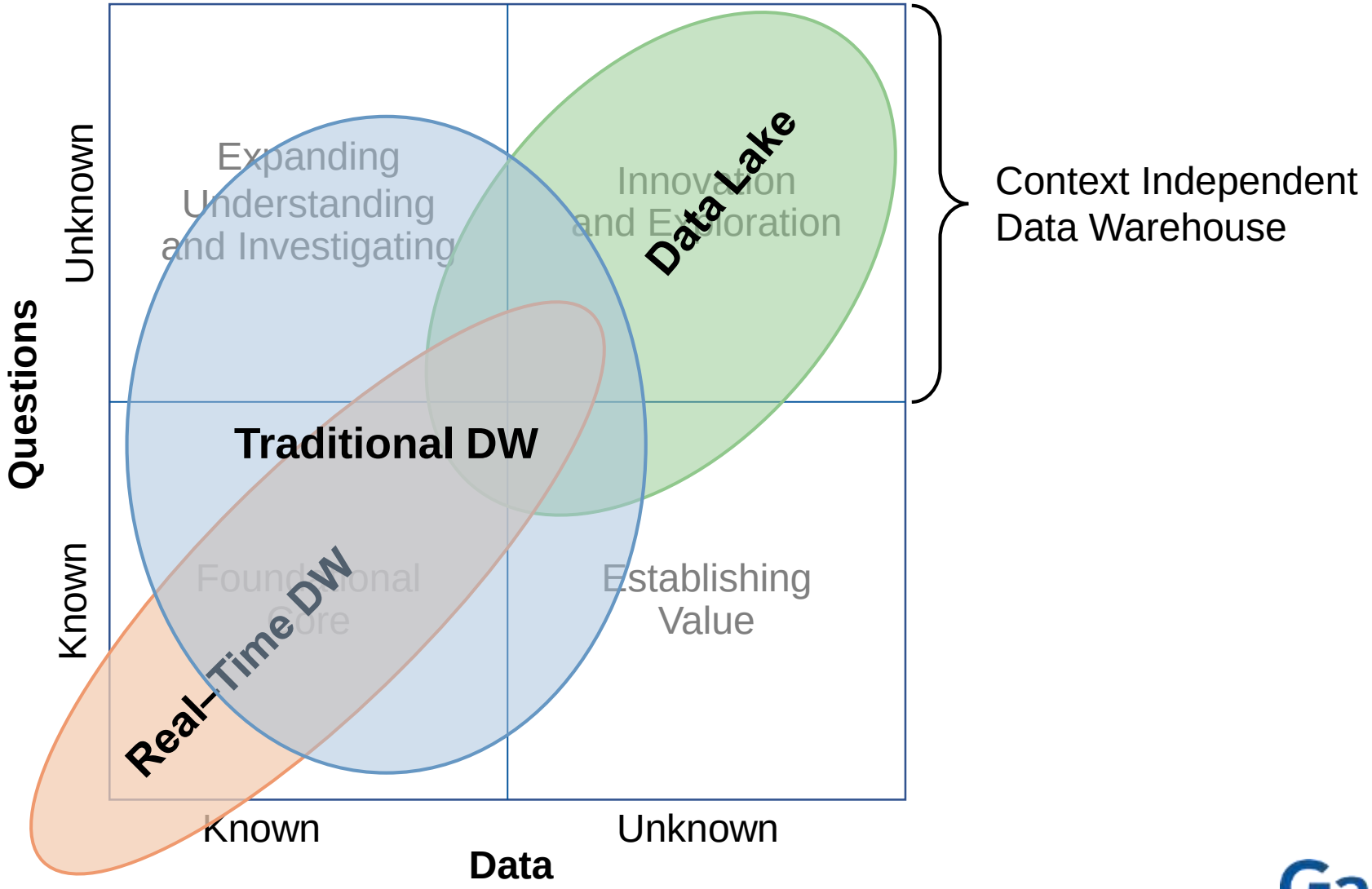
We Need a New Way to Think About Data Challenges



Source: "Solve Your Data Challenges With the Data Management Infrastructure Model," (G00336474)

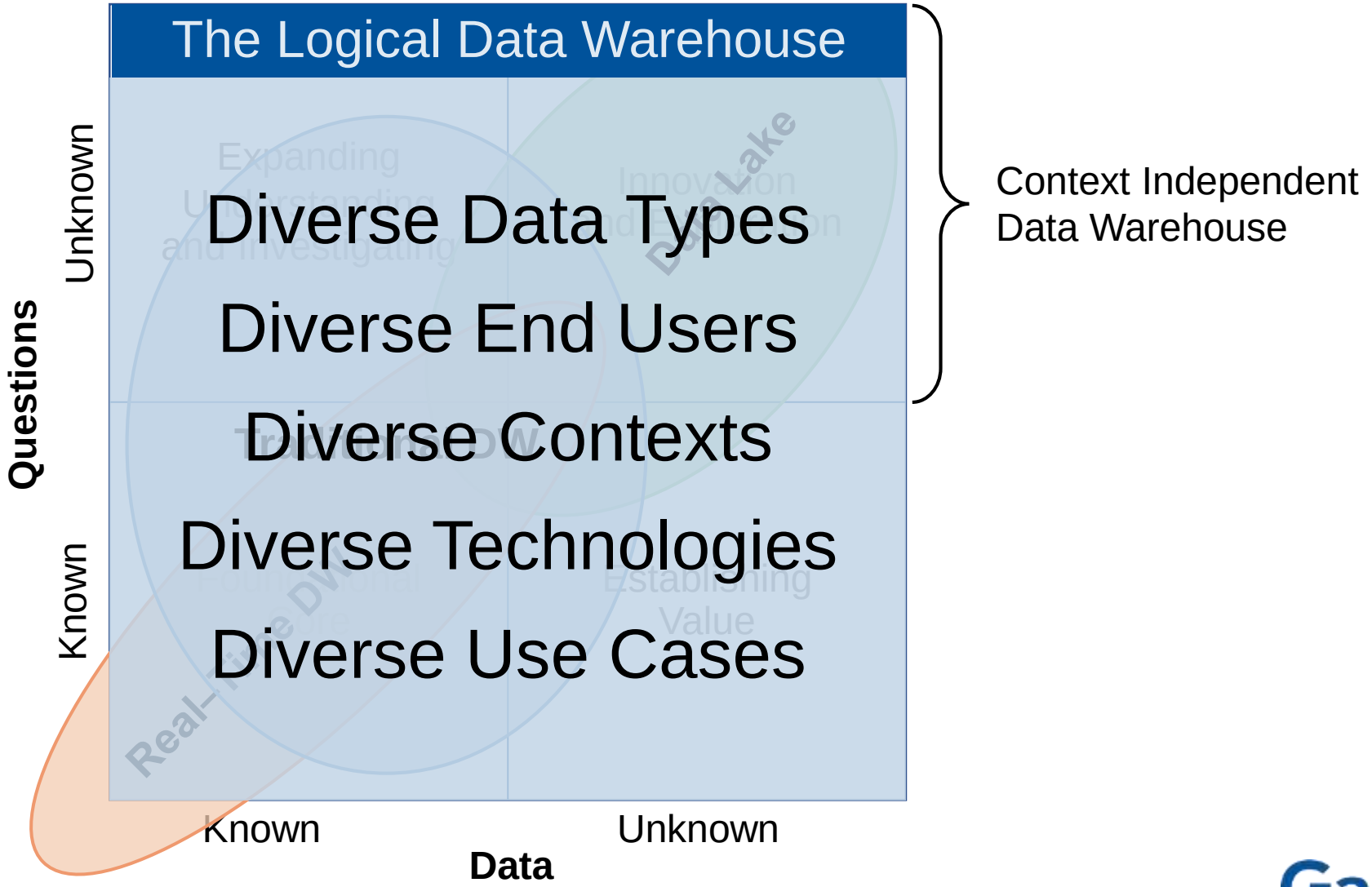
Gartner

The Data Management Infrastructure Model and the LDW



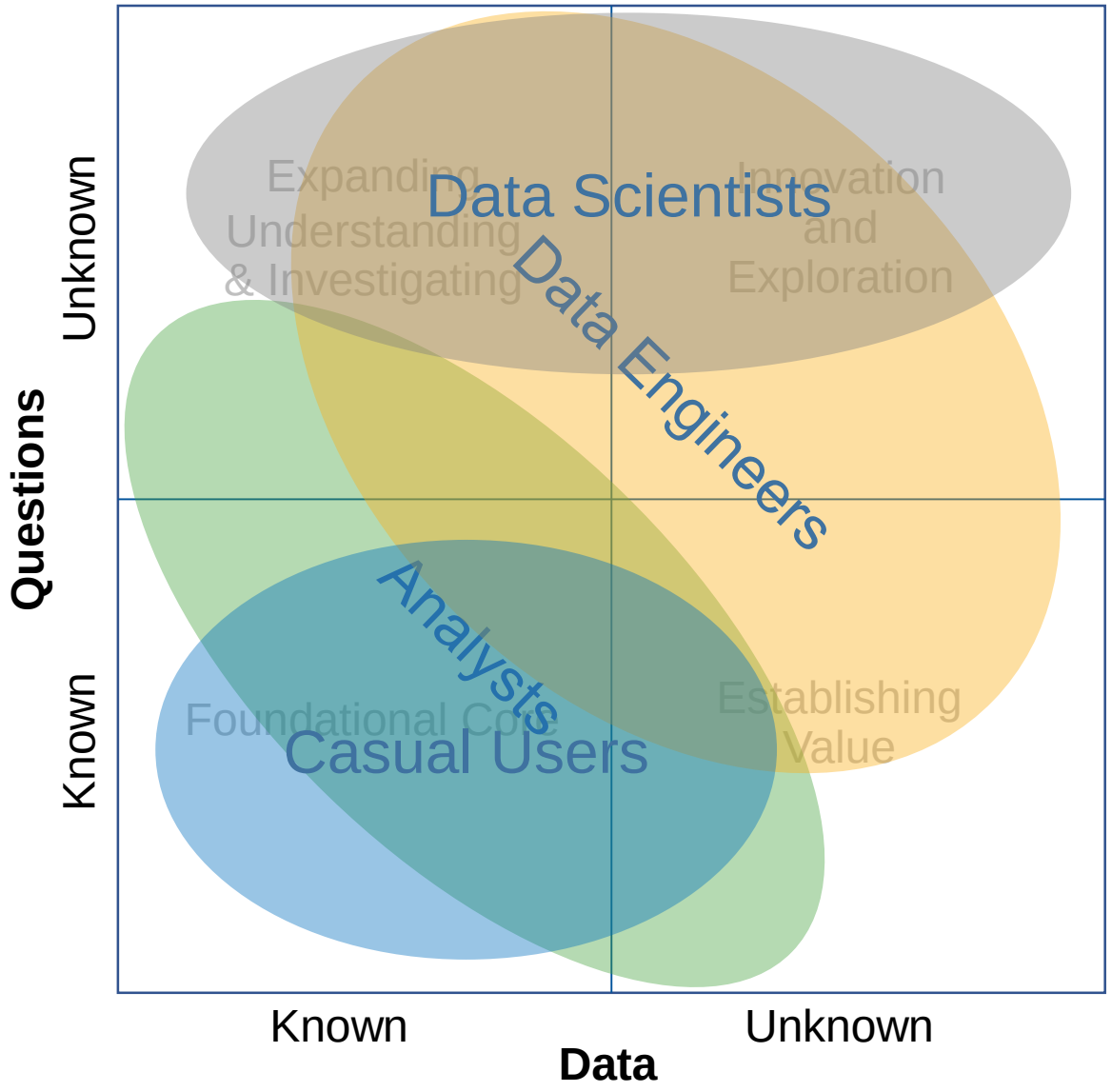
Gartner

The Data Management Infrastructure Model and the LDW



Gartner

We Can Use It to Align Roles and Skills …

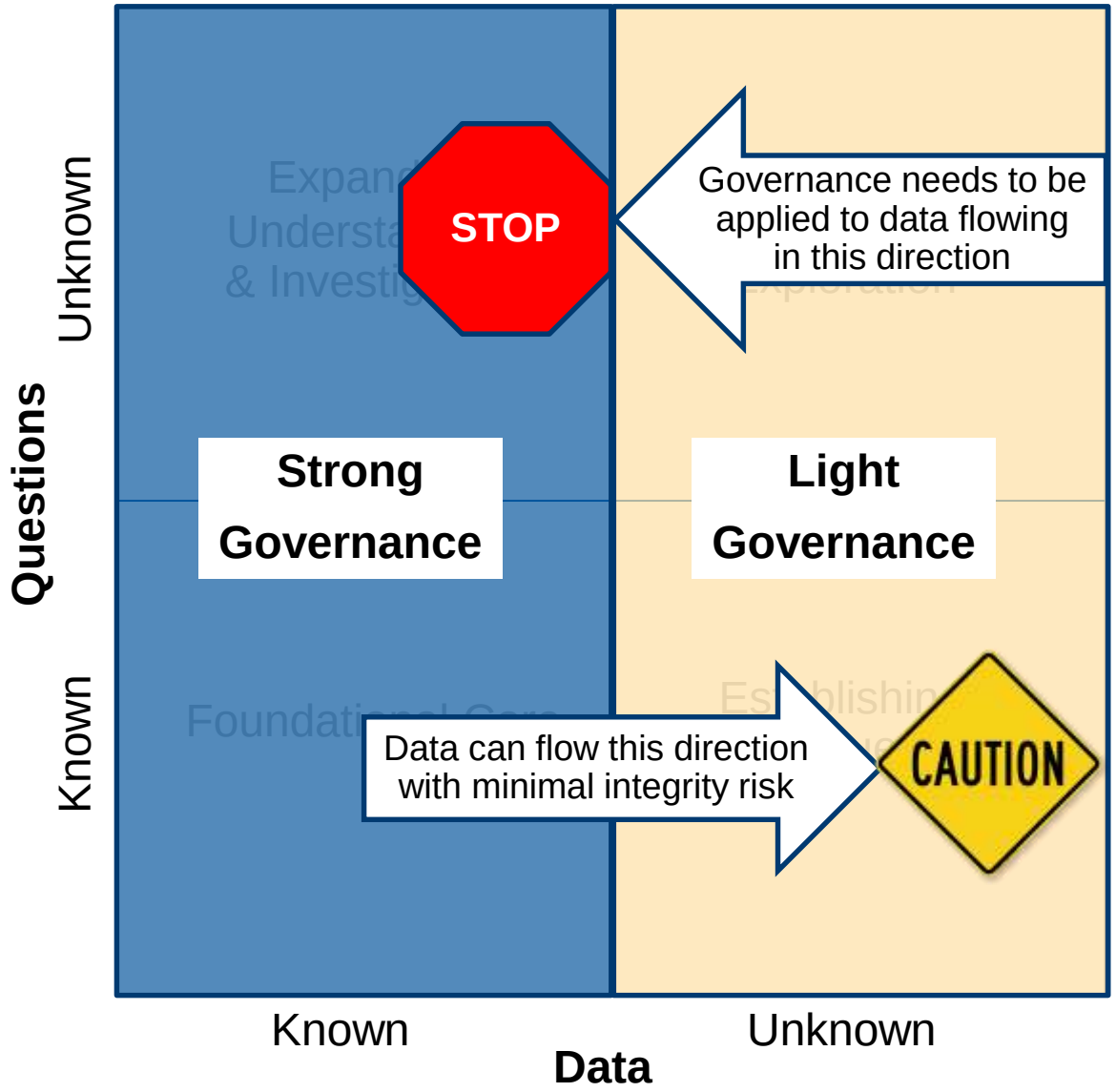


	Analytics Center of Excellence	User Distribution
Casual (Apprentice)	Reports, dashboards Level 1 support, possibly Level 2	1,000
Business Analyst (Journeyman)	Create new reports Needs technical assistance Level 2 and Level 3	90
Engineer (Master)	Ops process, systems analyst, data architect, reliable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

Source: "Organizing Your Teams for Modern Data and Analytics Deployment," (G00280461)

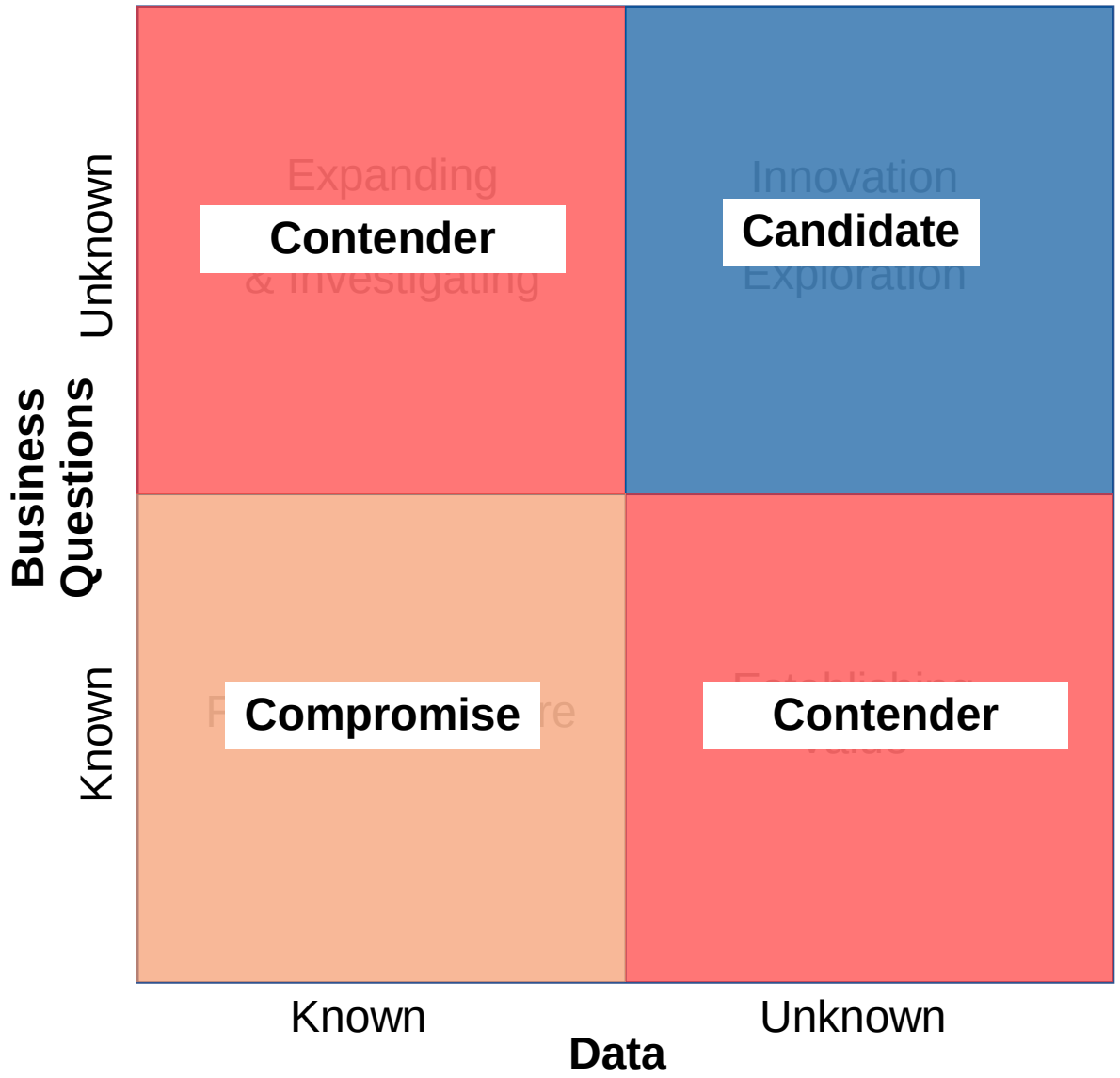
Gartner

… and to Address Governance Concerns

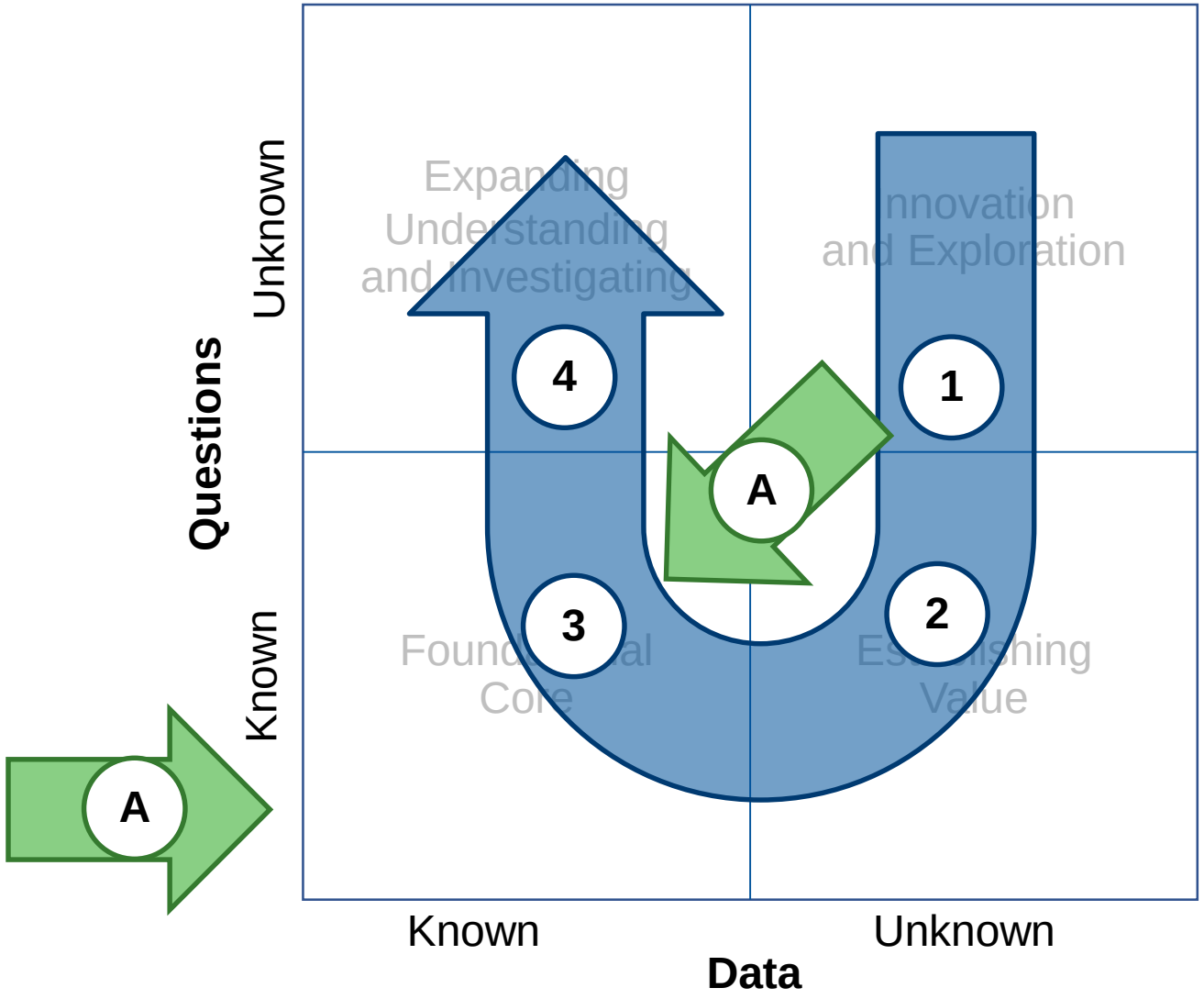


Gartner

It Aligns to the LDW SLAs …



… and Can Be Used to Map Changing Business Context



1 **Data Exploration:** The question development lab

2 **Testing Hypotheses:** What are the right data sources to answer my questions?

3 **Optimization**: Apply structure and optimize for the questions we are asking

4 **Exploration on Curated Data:** What else can I do with this?

A **ETL Load of Known Data for Known Purpose**

Gartner

Key Issues

1. How can you address the increased scope and breadth demanded by DMSA platforms today?
2. Do traditional data warehouse platforms still have a place?
3. How do you build an architecture that leverages the traditional while using innovations?

Does the Data Lake Replace the Data Warehouse?

- 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.
- Now, it's time to adapt, evolve and add **new** things.

No!*

The data warehouse is part of the past, present and future.

* At least not yet!

Gartner

Strategic Planning Assumption

By 2020, 30% of data lakes will be built on standard relational technology at equal or lower cost than Hadoop.

Why It Will Happen:

- RDBMSs are the enterprise standard
- Most RDBMSs support nonrelational data in multiple formats, and can support a schema-on-read approach
- Not all "native format" data is nonrelational
- Most data going into data lakes is relational from operational systems
- RDBMS ecosystem is very mature
- RDBMSs are not more expensive

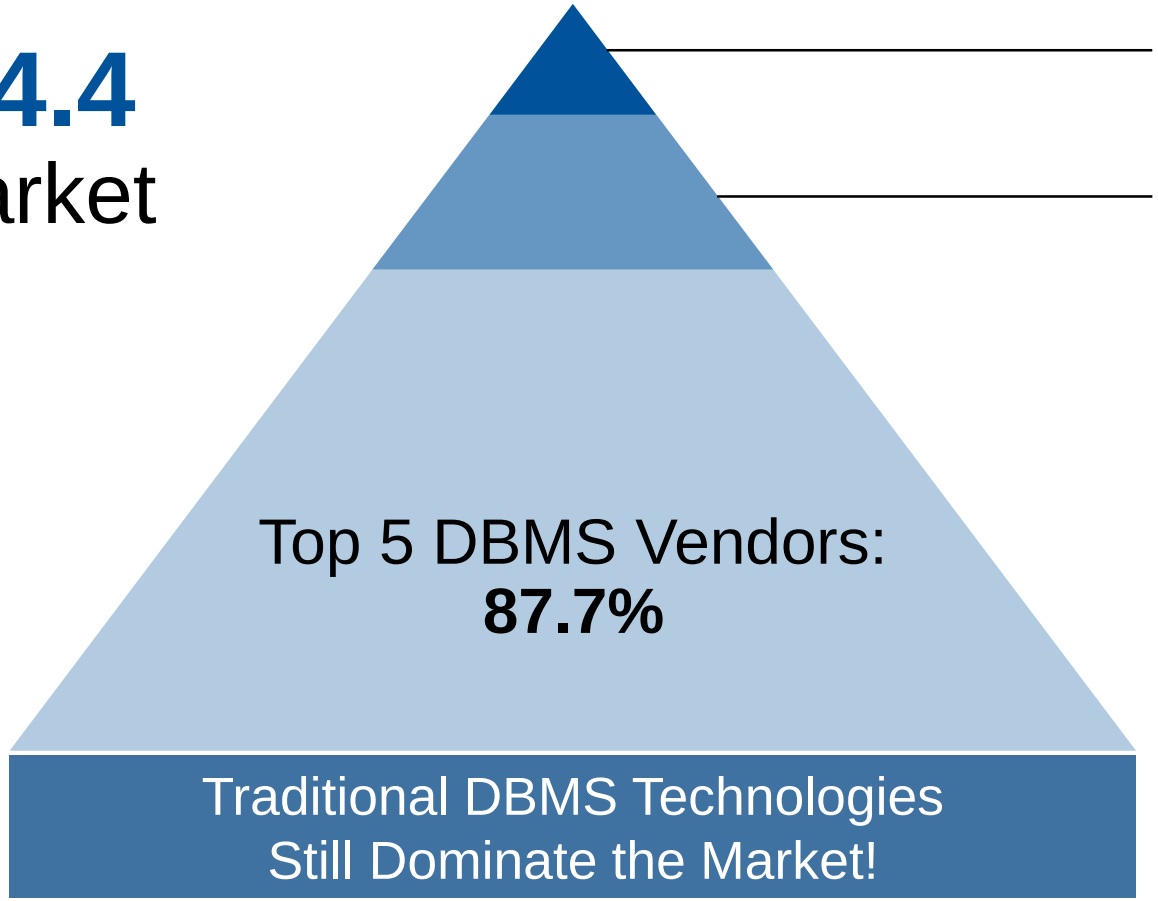
Why It Won't Happen

- Rapid ingest of data into schema-on-read platforms is easier than conforming to a relational model
- Increasing demand for analysis of nonrelational data that does not fit easily (or efficiently) into an RDBMS

Gartner

DBMS Market Dynamics

USD **$34.4**
Billion Market



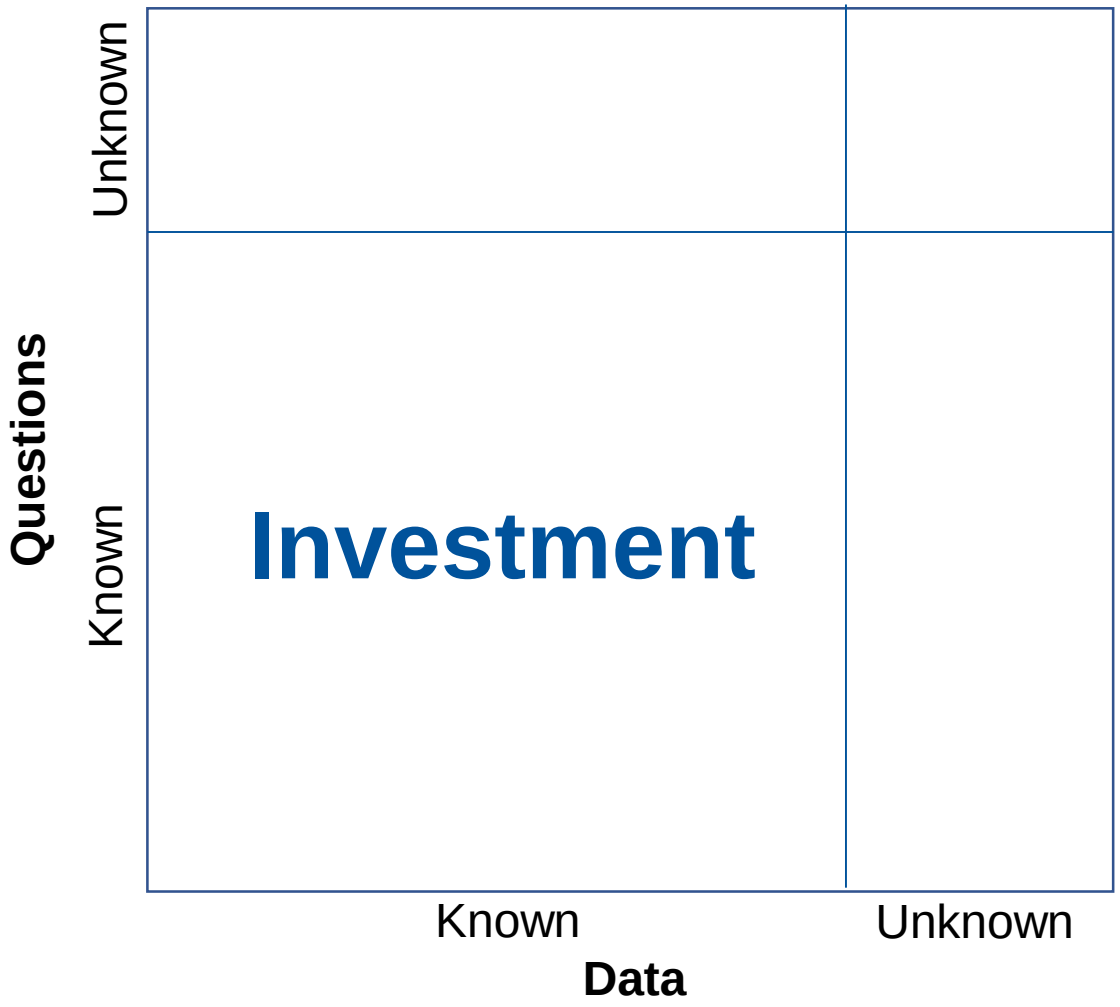
Pure-Play Hadoop: 1.39%

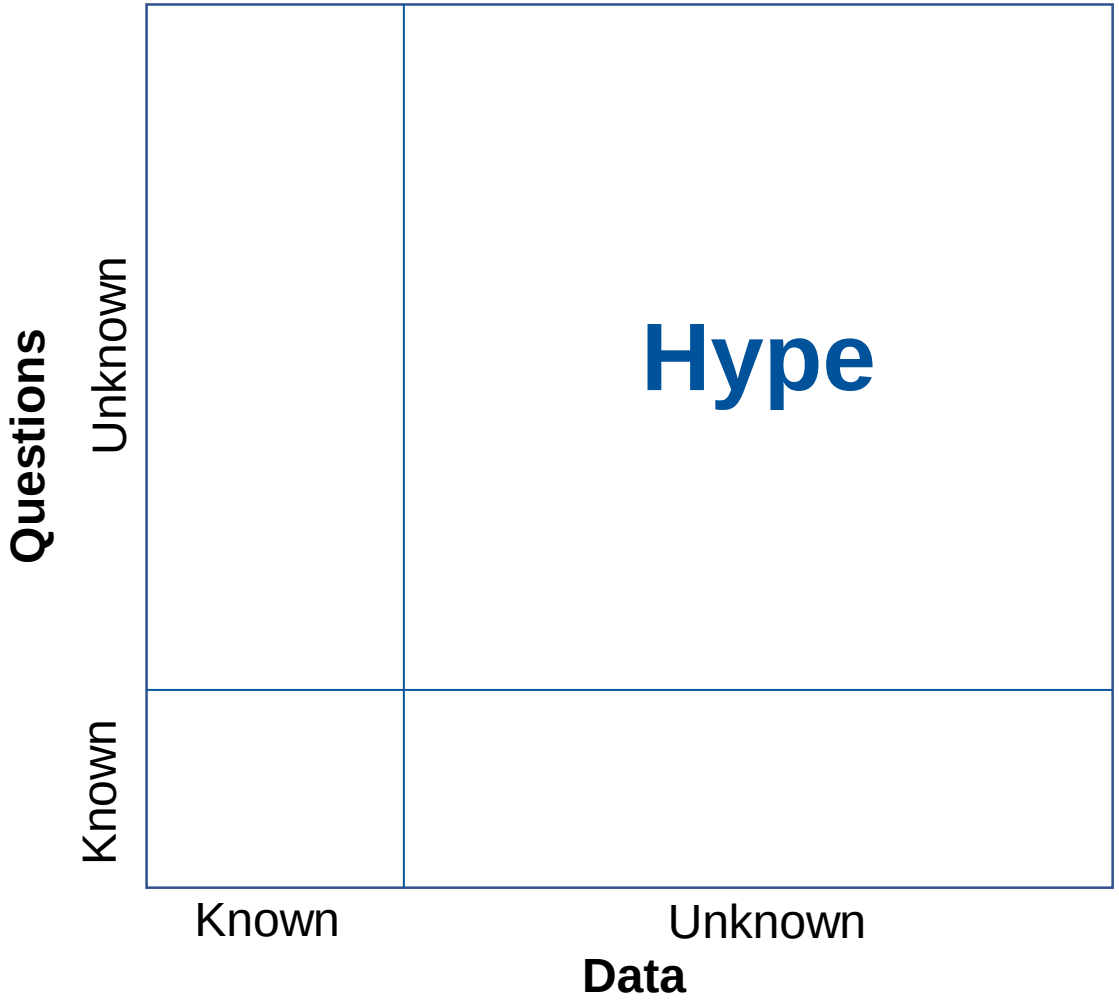
Nonrelational
(Including Hadoop): 5.2%

Source: "Market Share: All Software Markets, Worldwide, 2016," (G00322617)

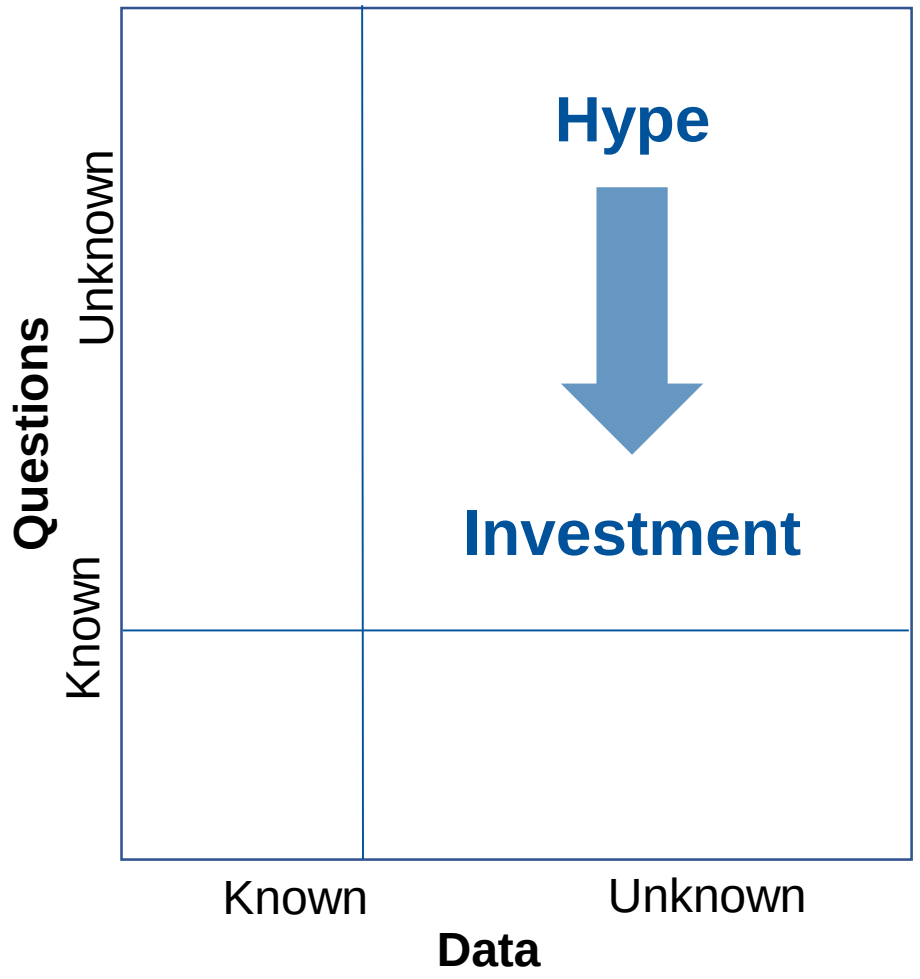
Gartner

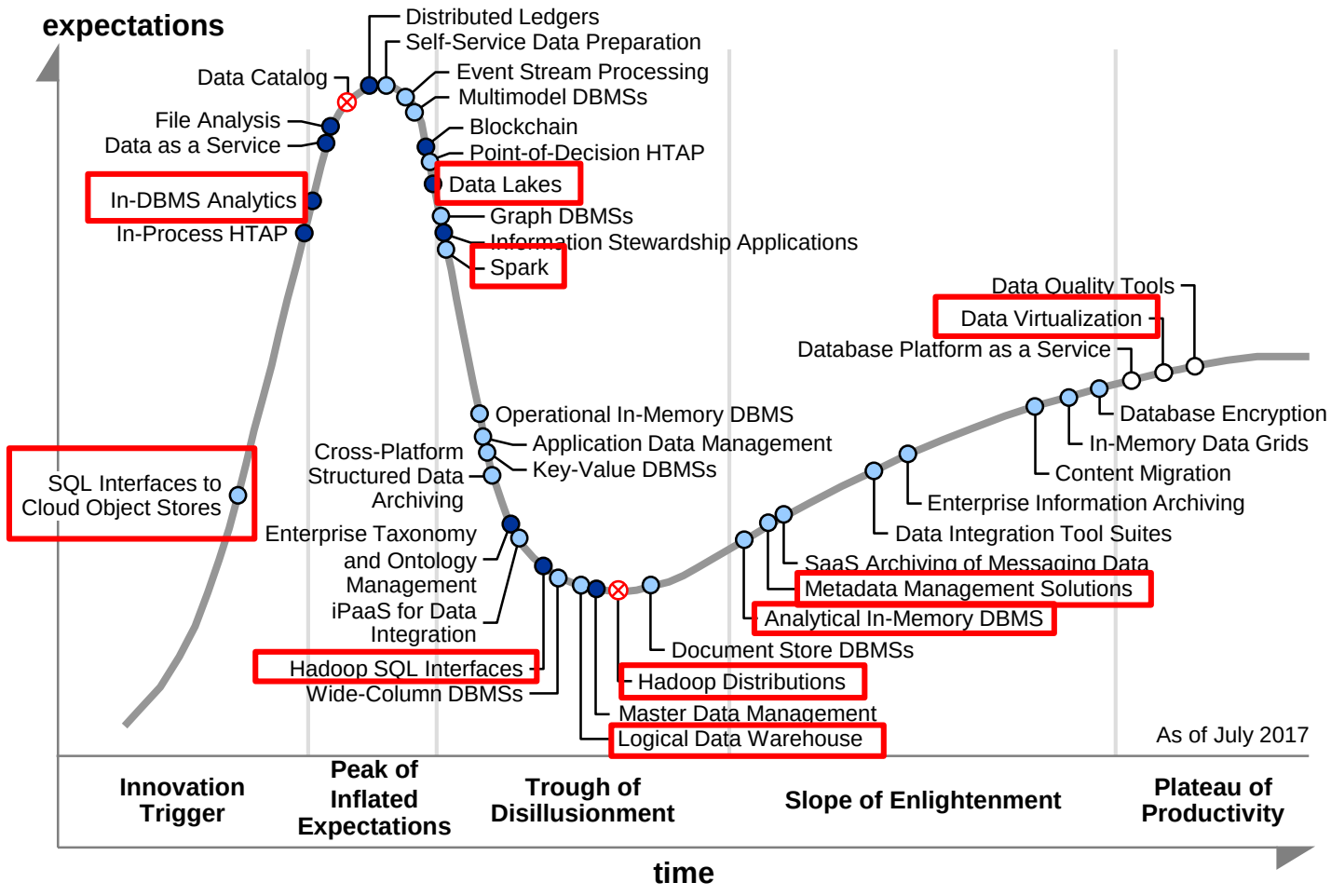
Hype vs. Investment





Hype Exceeds Investment — and Does Not Always Convert!





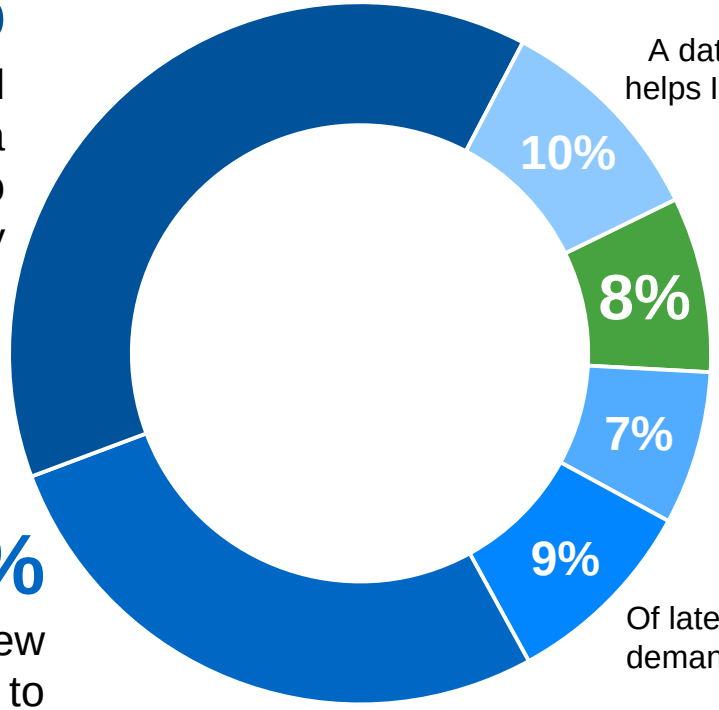
Source: "Hype Cycle for Data Management, 2017," 25 July 2017 (G00313950)

Most Companies Struggle to Meet an Ever-Increasing Demand for Data Analytics, Especially Demand for New Data Types and Having the Ability to Model Independently

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

38%

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

27%

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

Base: n = 175 Gartner Research Circle Members
Q. What scenario best describes how your organization's data capabilities meet current demand?

Key Issues

1. How can you address the increased scope and breadth demanded by DMSA platforms today?
2. Do traditional data warehouse platforms still have a place?
3. How do you build an architecture that leverages the traditional while using innovations?

Magic Quadrant for DMSA 2018

- The Rise of Distributed Data Management Environments
  - LDW Reaches ~15% of Target Market
- Continued Adoption of Cloud and Hybrid Cloud Deployments
- Contracting Use Cases for Hadoop-only DMSA Offerings
- The Rise of China-Based Vendors

Source: "Magic Quadrant for Data Management Solutions for Analytics," 13 February 2018 (G00326691)

Gartner

Invest in Skills

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

92% of organizations report that their current **data management needs** supporting analytics and reporting **remain unmet** and demand new data and data types.

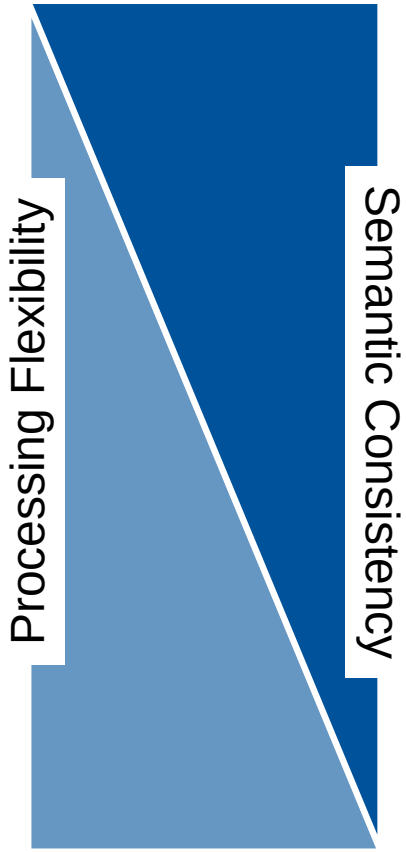
Delivery capacity for prequalified data will almost double by early 2018, **forcing** the **data management architecture** and infrastructure design to **support the rapid conversion** of data science discoveries **into production**.

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

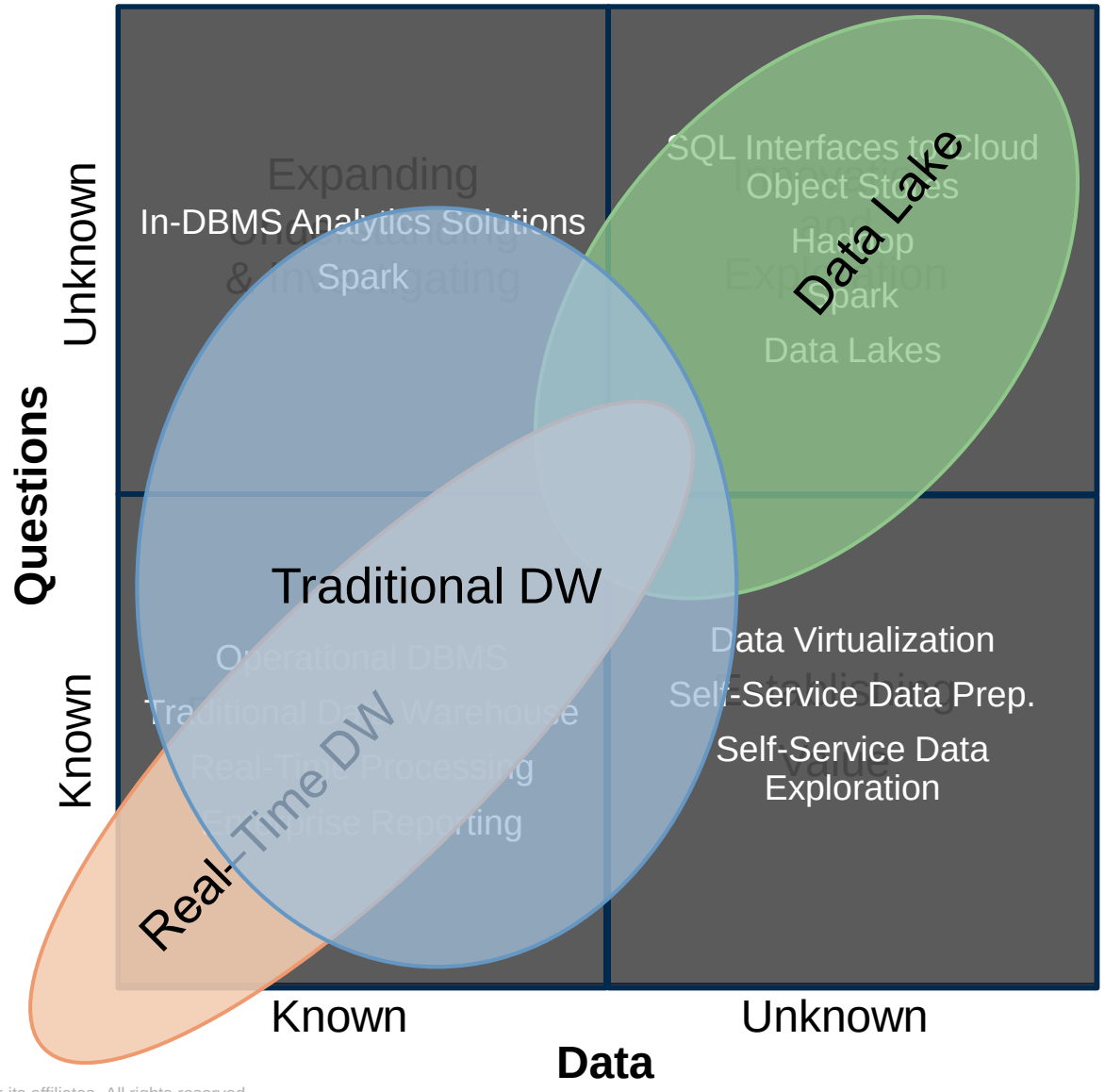
Infrastructures That Balance Optimization Needs With Self-Service Demand

	Optimized	Self-Service
Embedded	API/Bots	None
Casual User	Dashboards	Dimensions/Cubes
Business Experts	BI Platform	Analytics Platform
Citizen Scientists	Analytics Platform	Self-Service Data Prep.
Data Engineers	Analytics Platform	Processing Languages
Data Scientists	Processing Languages	Unobstructed Data



Gartner

Technologies for the Known and Unknown Are Complementary …

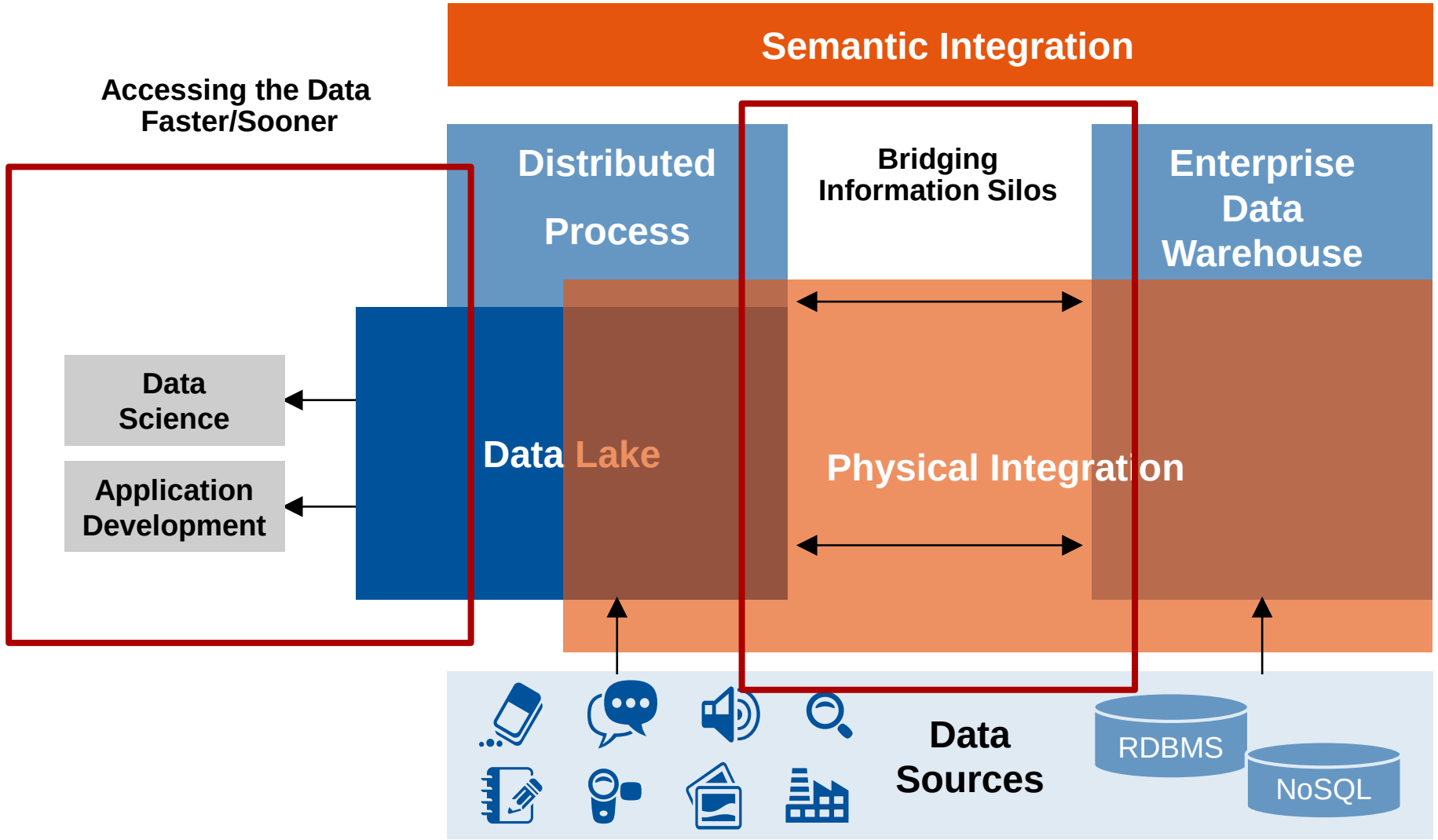


Context Independent Data Warehouse

… and the Lines Are Blurry

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.
- *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.

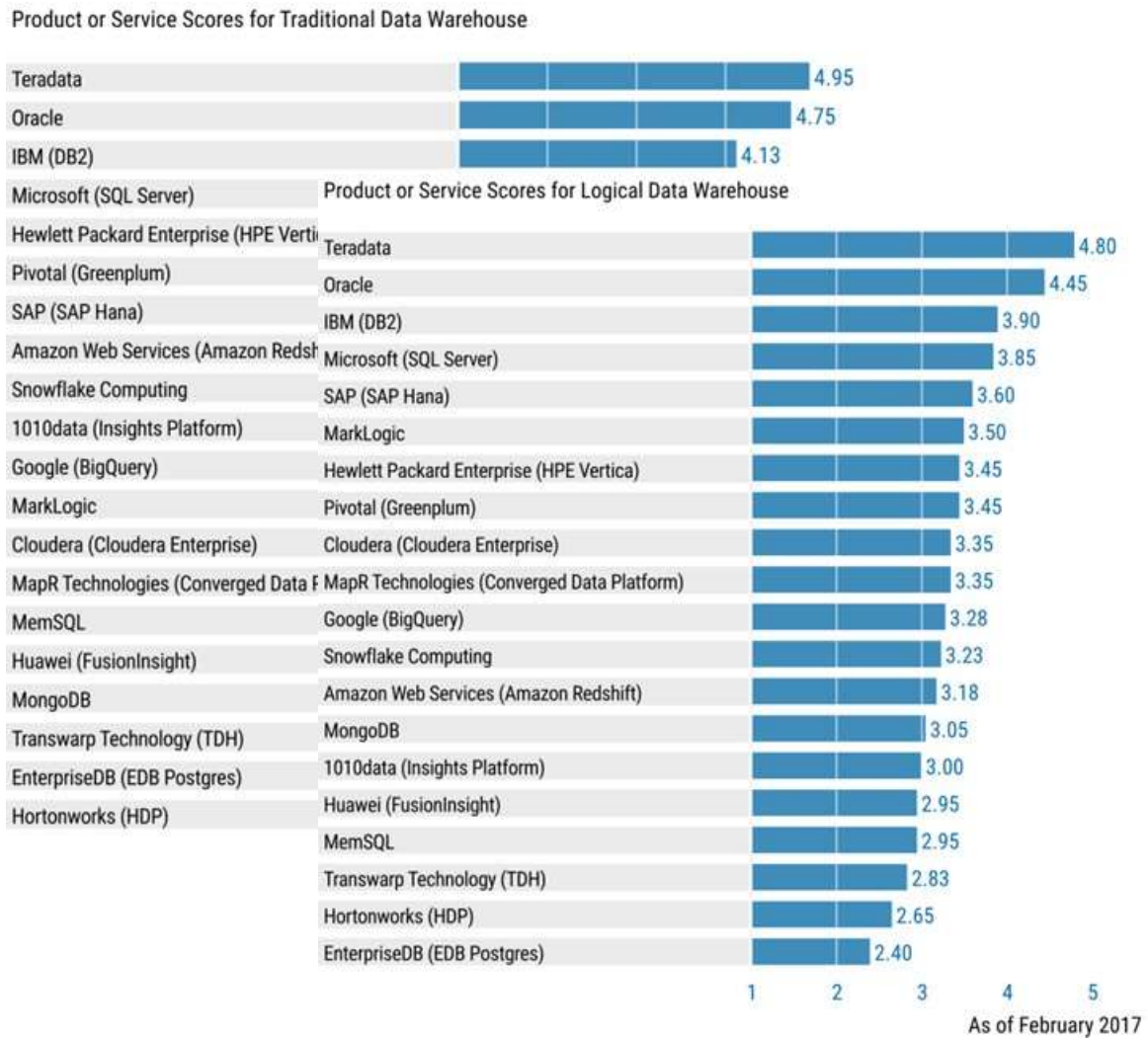
Recommended Gartner Research

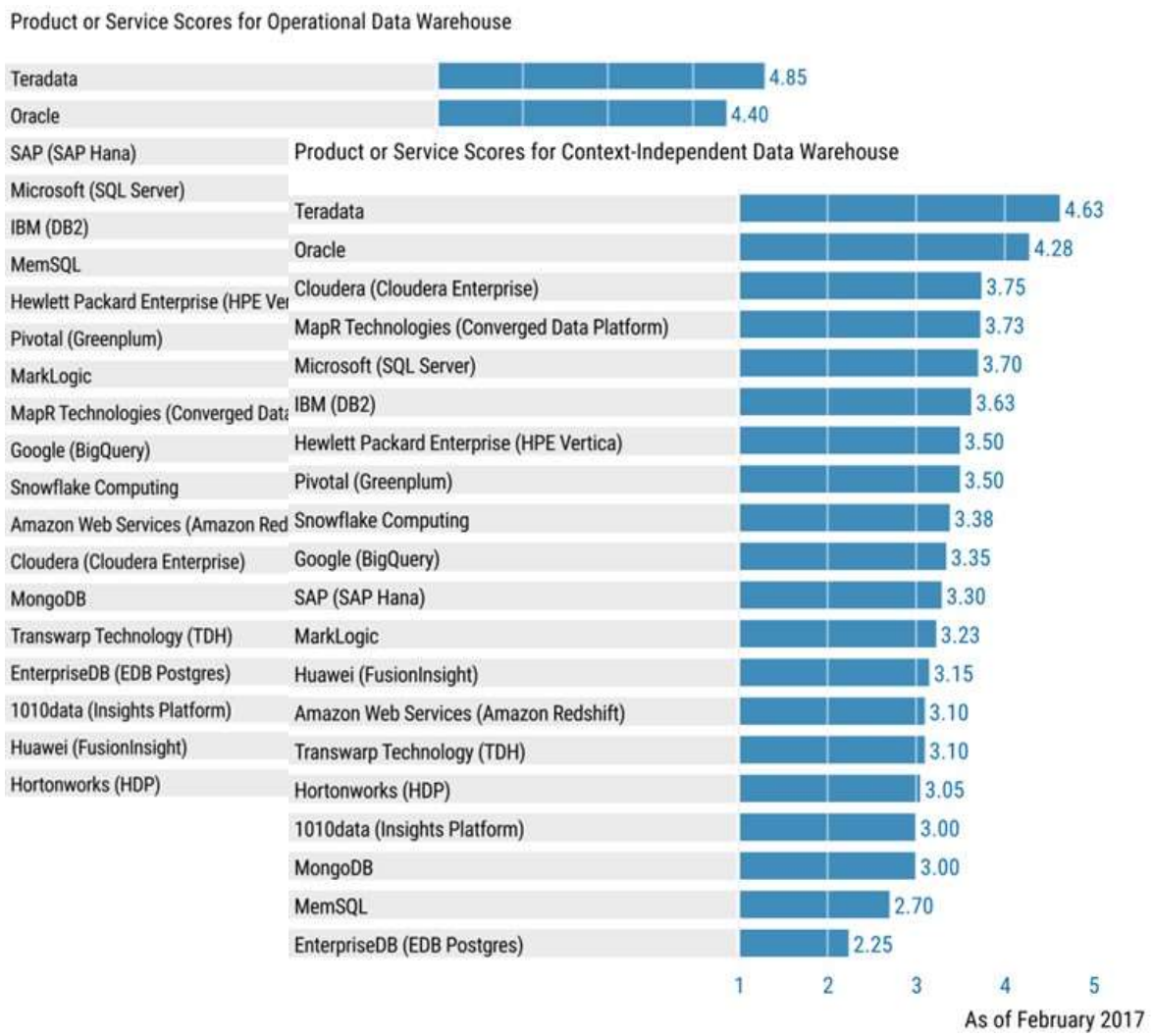
- **Solve Your Data Challenges with the Data Management Infrastructure Model**
  Adam M. Ronthal and Nick Heudecker (G00336474)
- **Efficiently Evolving Data from the Data Lake to the Data Warehouse**
  Rick Greenwald and Ehtisham Zaidi (G00336042)
- **Modern Data Management Requires a Balance Between Collecting Data and Connecting to Data**
  Roxane Edjlali and Ted Friedman (G00334541)
- **Data Management Solutions for Analytics: Current and Future States, 2017**
  Rick Greenwald and Adam M. Ronthal (G00336273)
- **Survey Analysis: New Data and New Analytics Are All Mythology Unless You Add Skills**
  Mark A. Beyer and Adam M. Ronthal (G00331689)

For information, please contact your Gartner representative.

Gartner

Traditional Technologies Do Well in All DMSA Use Cases





Source: "Critical Capabilities for Data Management Solutions for Analytics," (G00314186)

Gartner