denodo
webinar series

DATA VIRTUALIZATION PACKED LUNCH WEBINAR SERIES

Sessions Covering Key Data Integration Challenges
Solved with Data Virtualization

denodo
webinar series

Big Data Fabric: A Necessity For Any Successful Big Data Initiative

Paul Moxon

VP Data Architectures & Chief Evangelist, Denodo

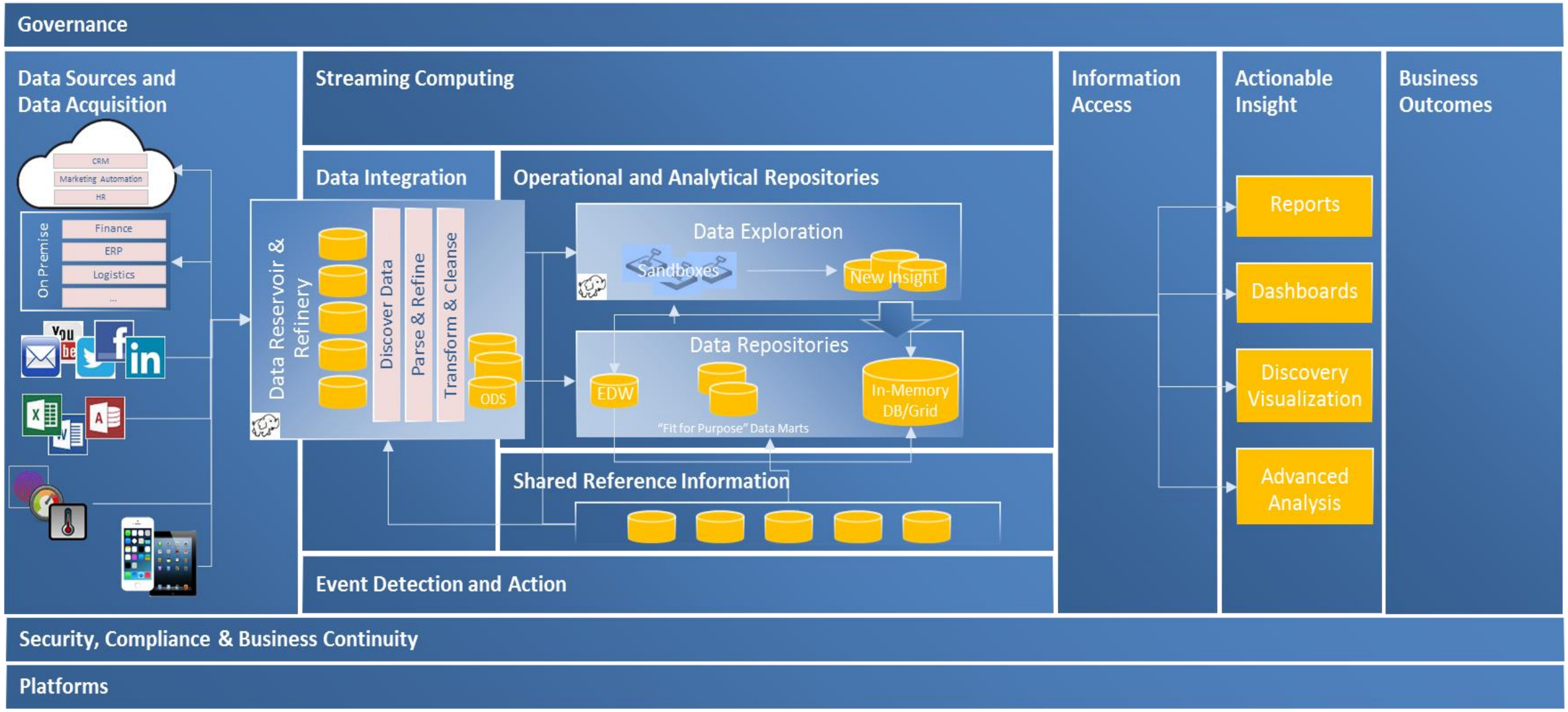
Naren Sankaran

Sales Engineer, Denodo

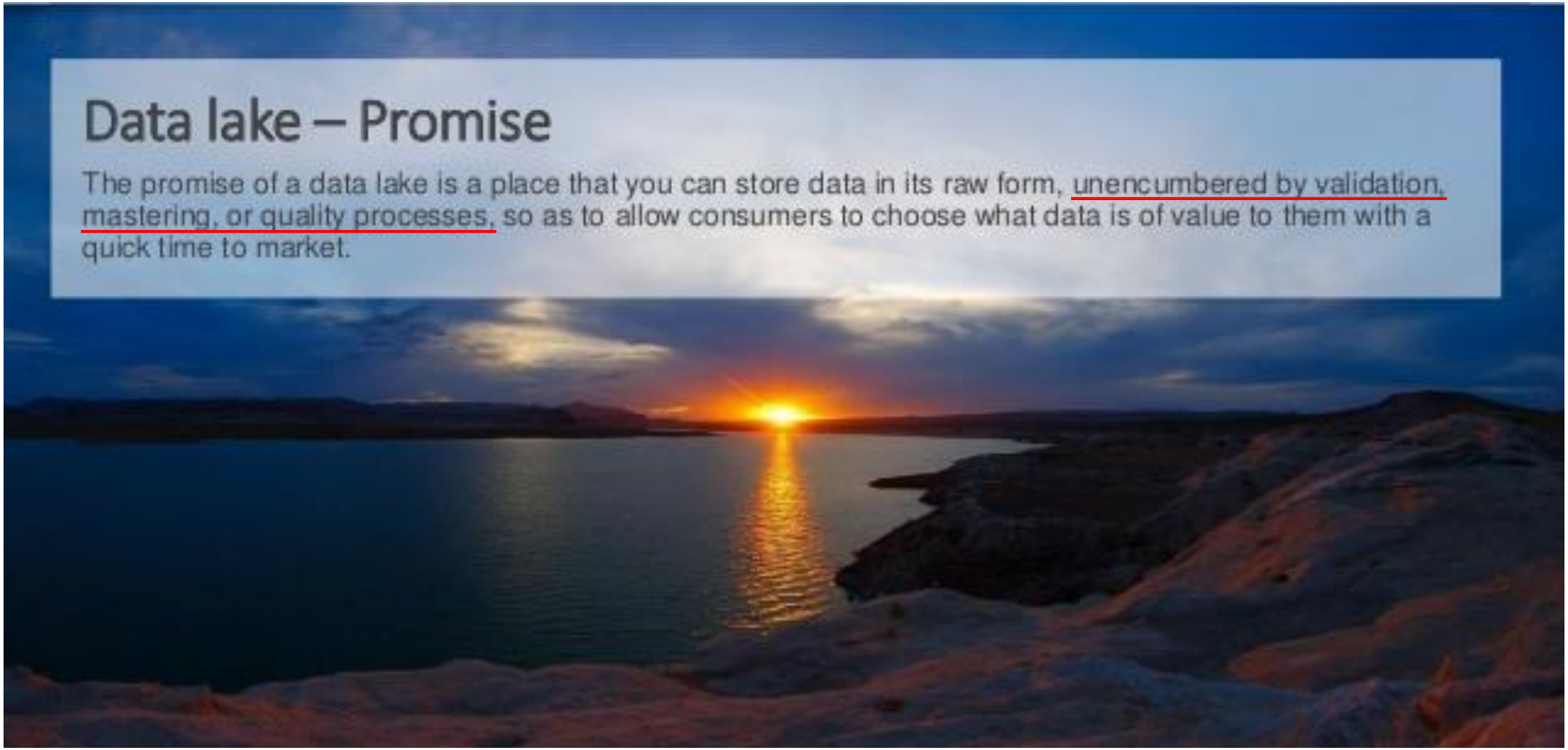
Agenda

1. The Reality of Data Lakes
2. Big Data Fabric
3. Customer Case Study – Logitech
4. Demo
5. Q&A
6. Next Steps

A Modern Data Ecosystem



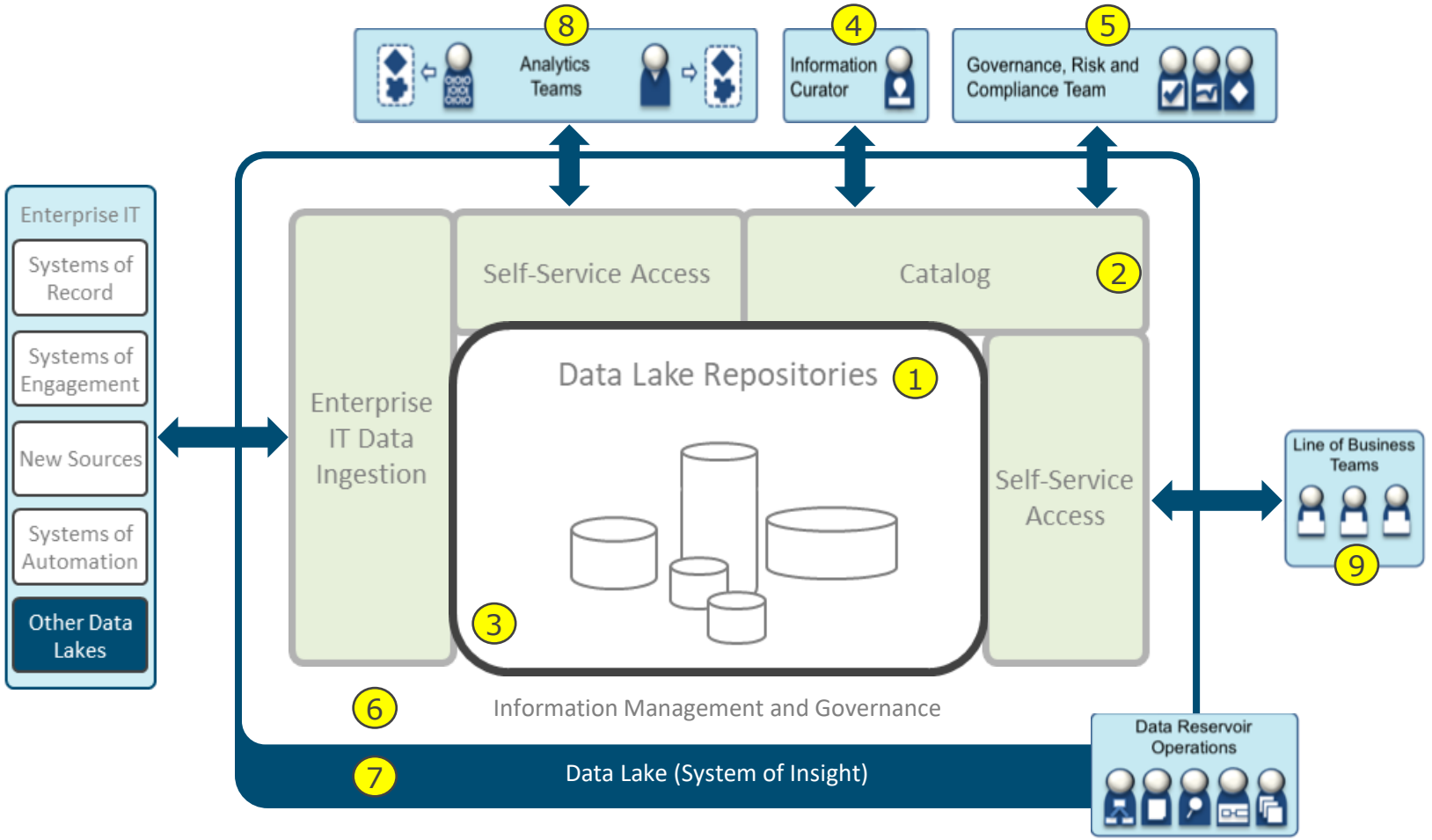
The Promise of the Data Lake



> …Data lakes lack semantic consistency and governed metadata. Meeting the needs of wider audiences require curated repositories with governance, semantic consistency and access controls.”

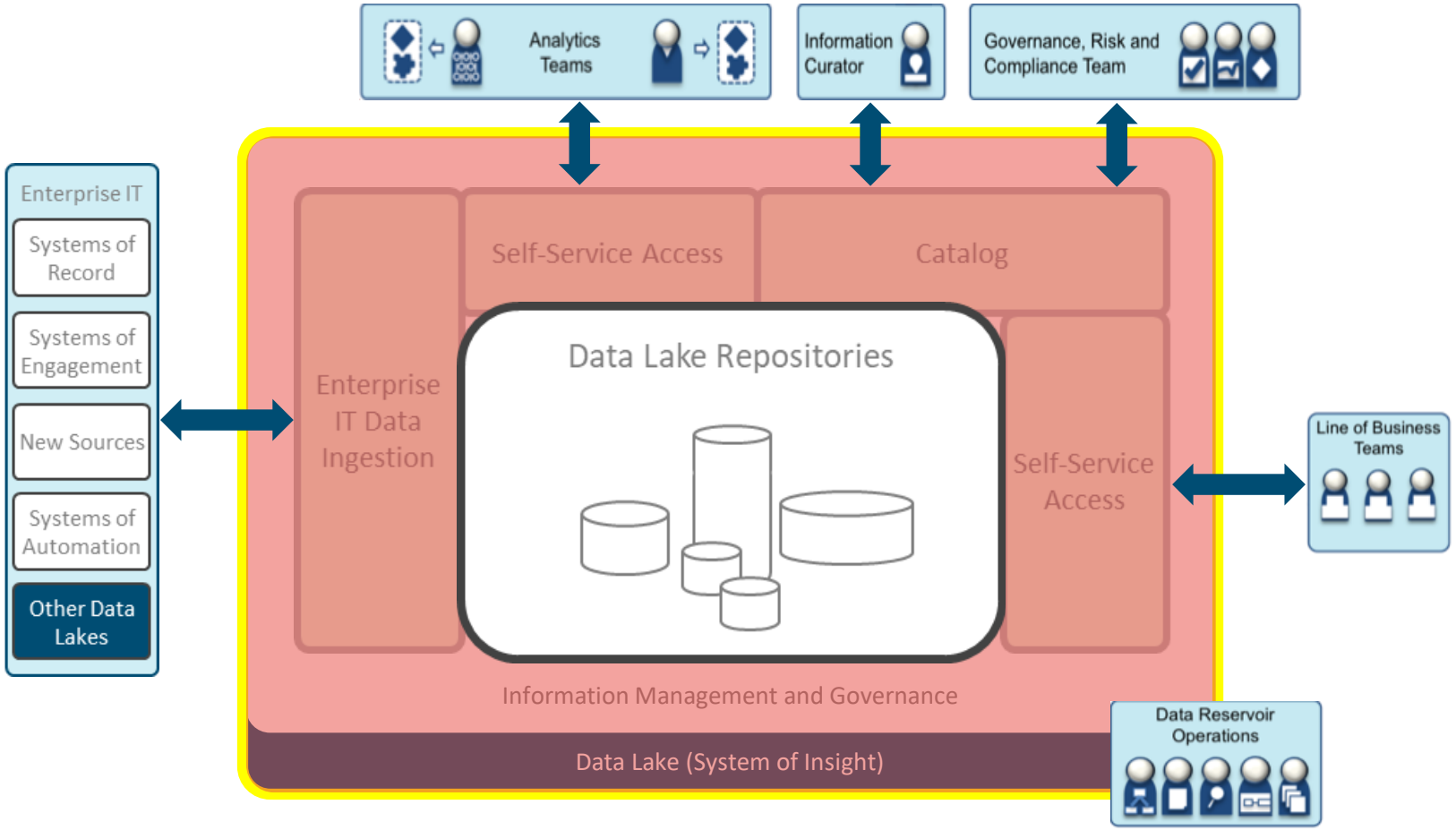
Gartner®

Data Lakes and Big Data Fabric

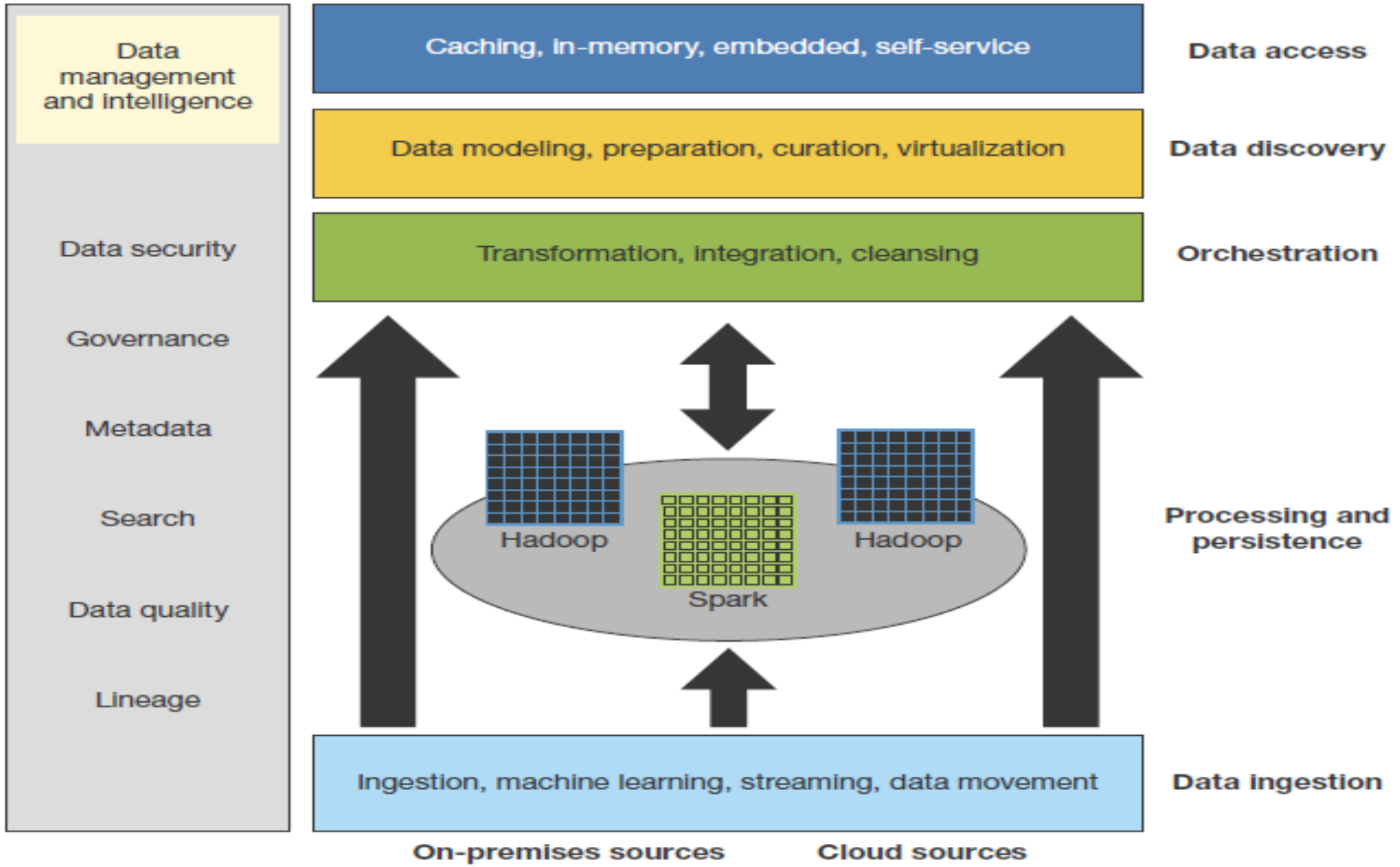


1. Multiple repositories, organized based on source and usage; hosted on appropriate data platform for workload
2. Catalog of data, ownership, meaning and permitted usage
3. No direct access to repositories
4. Curation of all data to define meaning and classifications
5. Business-led information governance and management
6. Active monitoring and management of data
7. Data-centric security
8. Access to raw data to develop new production analytics
9. Moderated, view-based self-service access to data and analytics for Line of Business

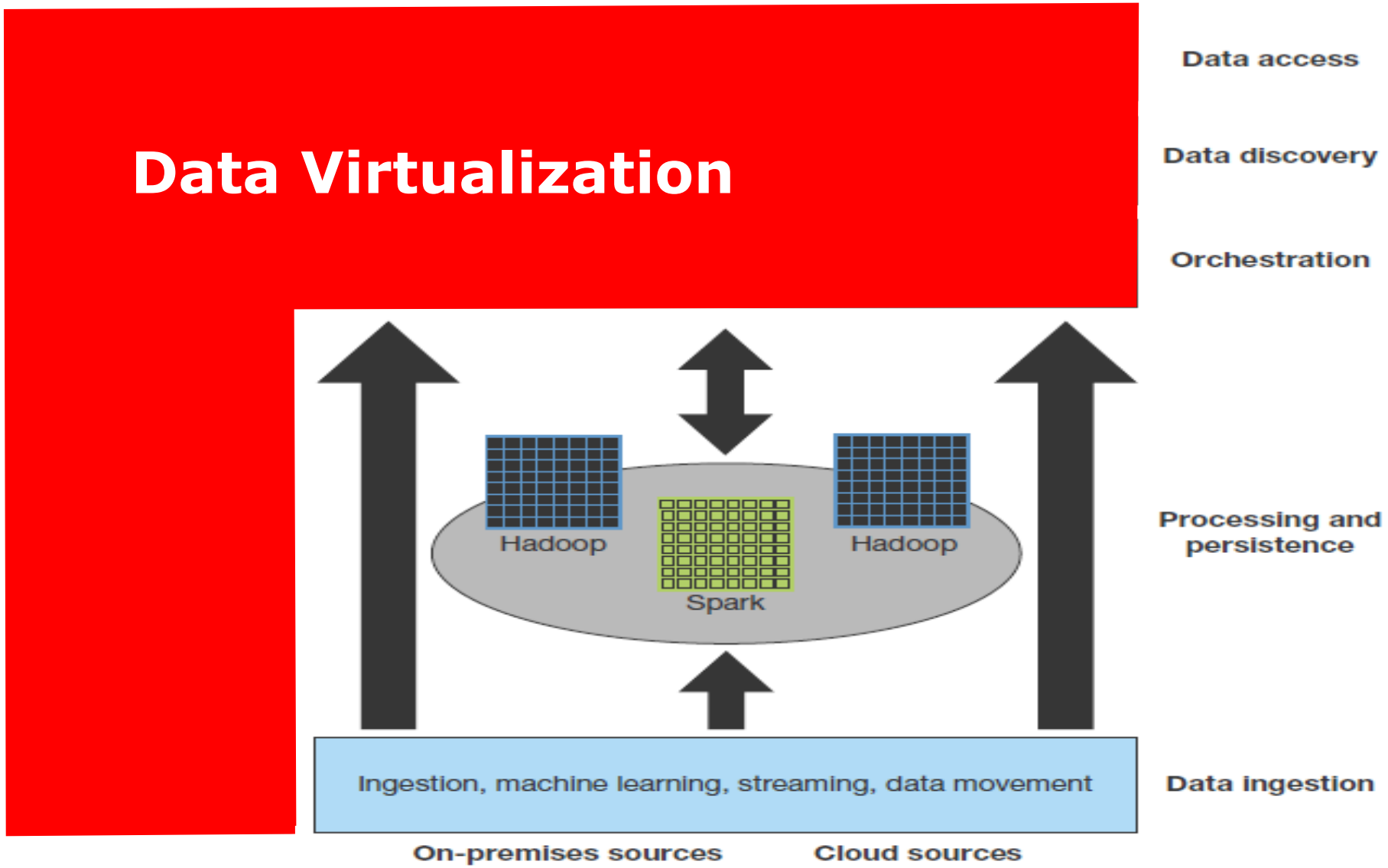
Data Lakes and Big Data Fabric



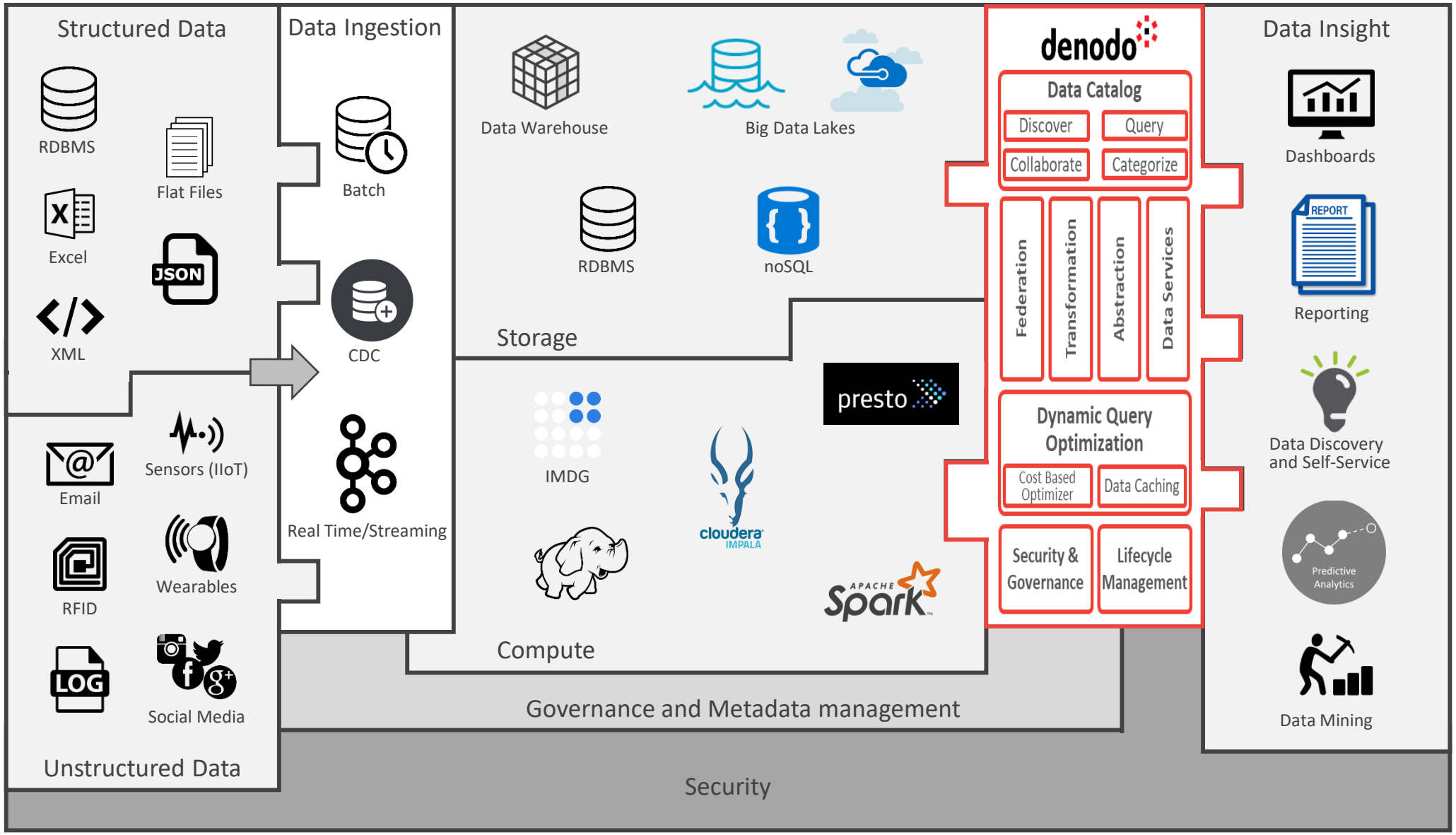
Forrester's Big Data Fabric



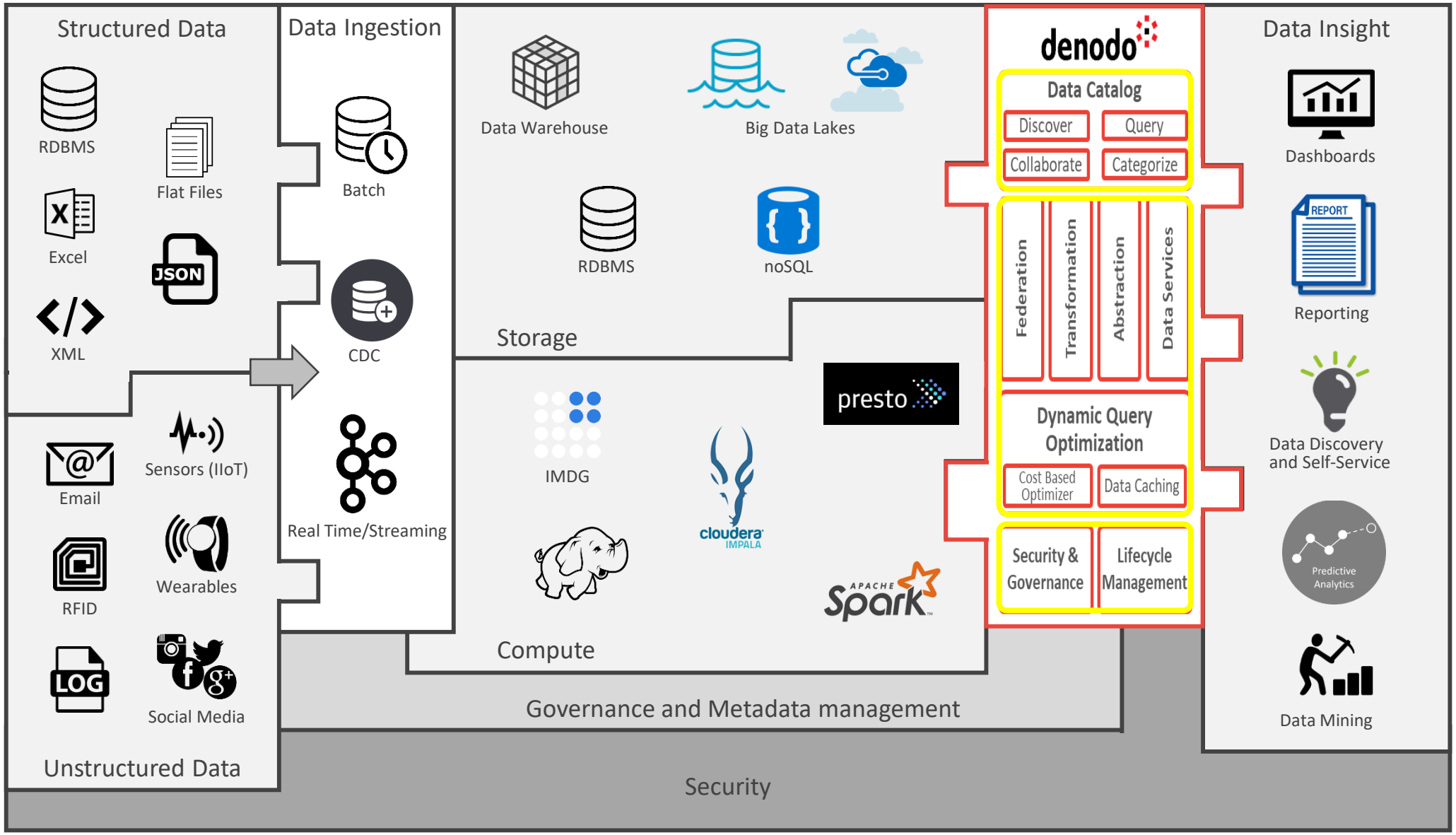
Big Data Fabric and Data Virtualization



Data Virtualization Reference Architecture



Data Virtualization Reference Architecture

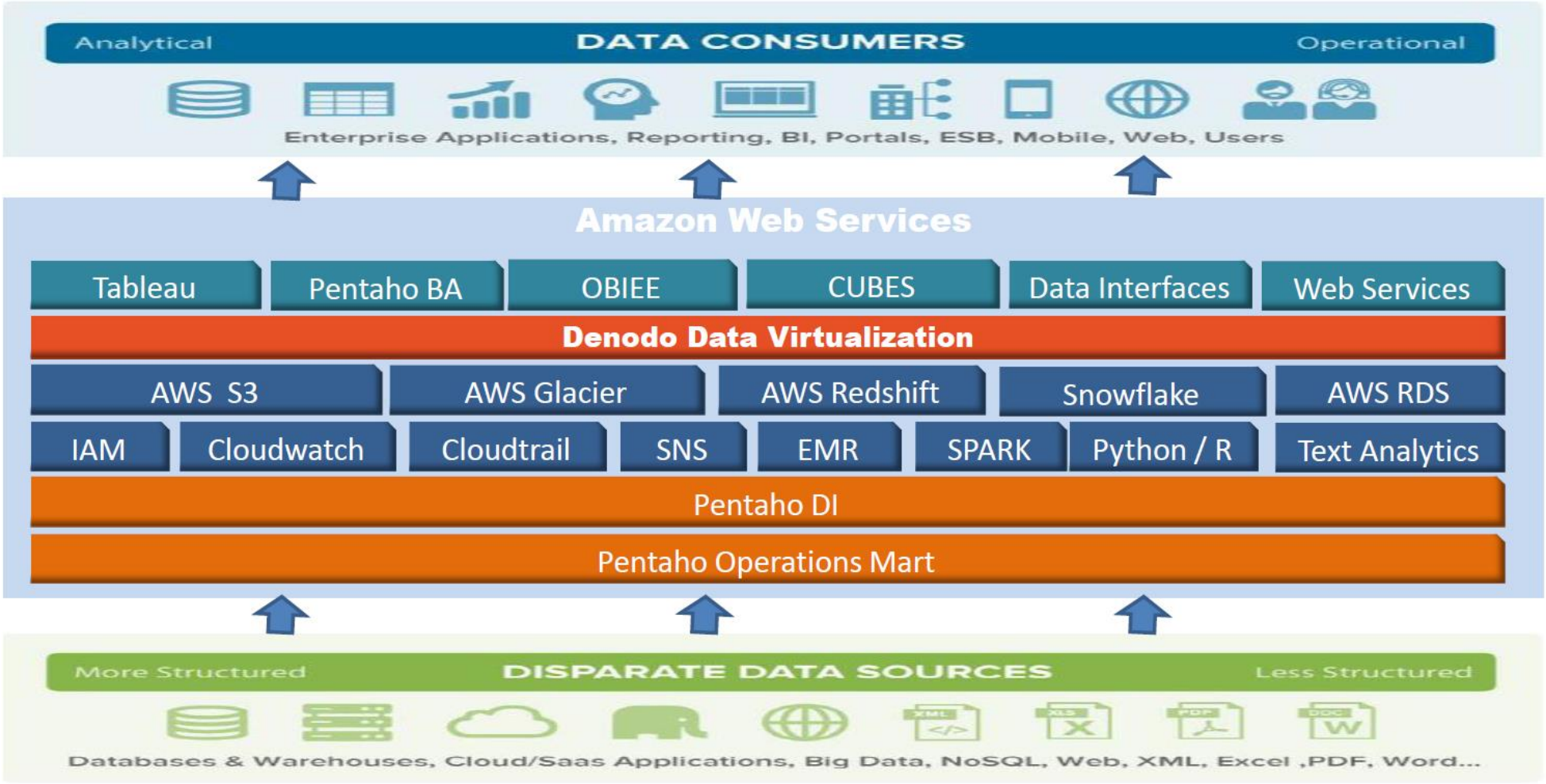


Customer Case Study – Logitech

- Logitech also needed more agile (and cost effective) analytics platform for customer analytics
  - Moved analytics to AWS – utilized many different storage and compute capabilities on AWS
  - Redshift, Snowflake for EDW, RDS for relational data, S3 for data ingestion
  - Spark, EMR, NLP for analytics
- Used Data Virtualization as data access layer
  - Users could access data irrespective of storage or processing capability
- Data Virtualization Layer provided data security/access permissions
  - Abstracted security from different data sources
- Data governance also provided by Data Virtualization Layer
  - Data lineage, metadata management, data catalog and discovery, etc.

Customer Case Study – Logitech

SOLUTION ARCHITECTURE



Customer Case Study – Logitech

BENEFITS

- Embraced clou
  and agility, **ena**
- Reusability an
  structure, bala
- More business
  the data they v
- Respond faste
- Array of conne
- Business Laye

LESSONS LEARNT

- Paradigm shift – Traditional RDBMS to Columnar Database
- Denodo - Data Delivery Platform
- Tools will change, Data will stay – Need a platform that keeps all together
- Need of Governance for Modern Analytics
- IT is enabler, not producer!

Product Demonstration

Naren Sankaran

Sales Engineer, Denodo

Inventory Summary

INVENTORY_SUMMARY is a join between:

- REDSHIFT_ITEM
- REDSHIFT_ DATE_DIM
- IMPALA_INVENTORY

Join Conditions:

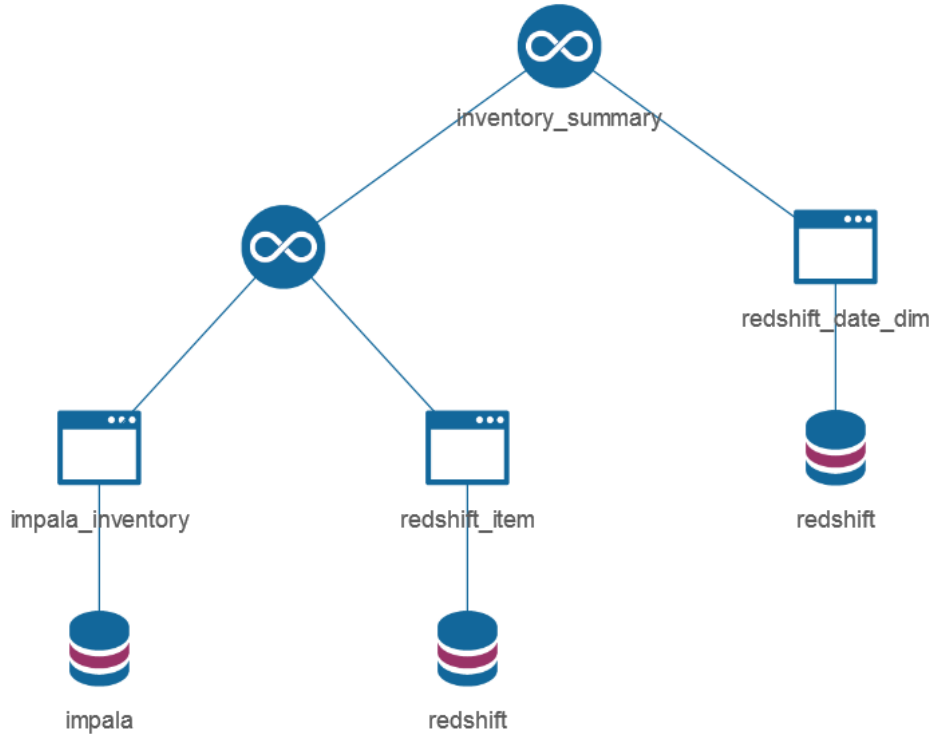
- impala_inventory.inv_item_sk = redshift_item.i_item_sk
- impala_inventory.inv_date_sk = redshift_date_dim.d_date

Group By:

- i_item_desc
- i_brand
- d_year

Output:

- Add field “total” which sums the inv_quantity_on_hand



Performance

Examine execution plan of view inventory_summary

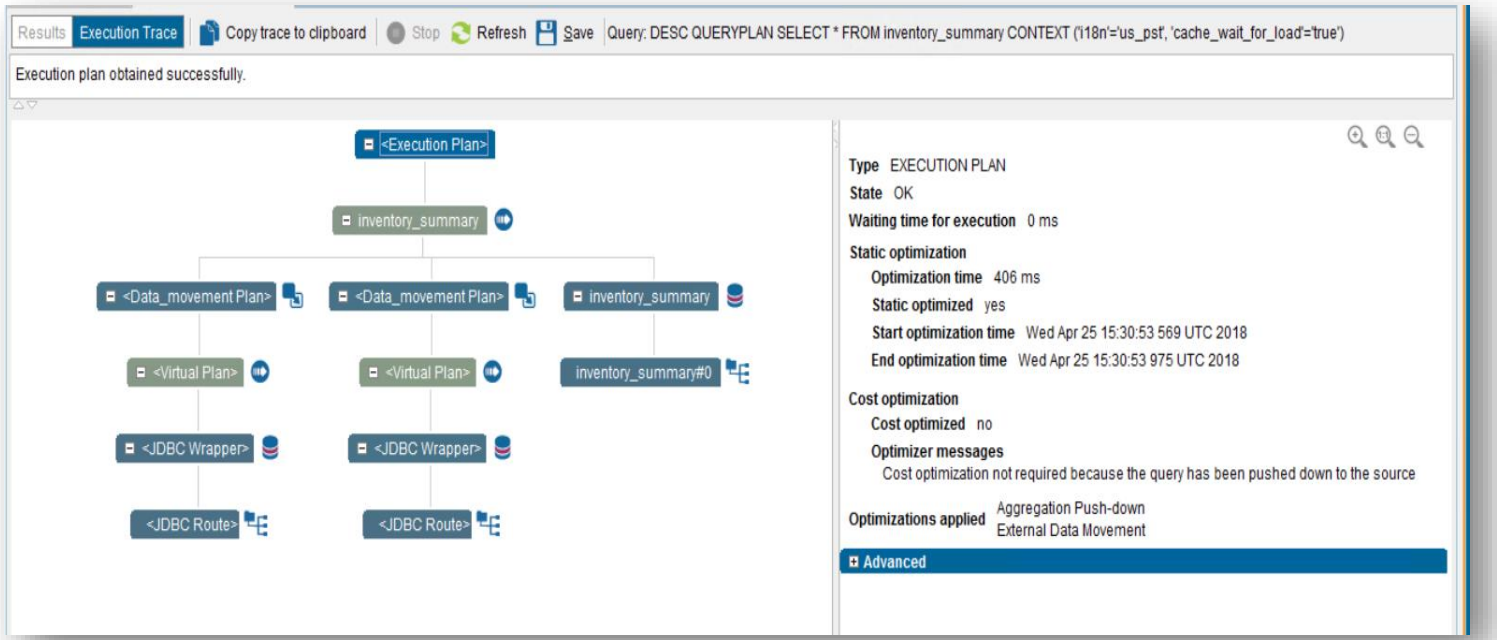
From VQL Shell Execute:

- `DESC QUERYPLAN select * FROM inventory_summary context('datamovement' = 'off')`
- `DESC QUERYPLAN select * FROM inventory_summary`

Compare Execution Plans

- Example of the combination of data movement to an MPP for query acceleration.
- In this case, multiple tables are moved to Hadoop in parallel
- Static optimization also provide an aggregation pushdown to the Hadoop cluster
- Execution time with optimization: 28408 ms
- Execution time w/out optimization: over 900 seconds

These strategies dramatically reduces the data volumes transferred through the network, while pushes down to Impala most of the "heavy lifting"



Demo

Summary

- Big Data initiatives can add deep analytical insight to help organizations find new opportunities or provide more efficient services
- When defining your Big Data initiative, think about how users will find and access the data
- A Big Data Fabric is key to making this easy
  - And making the initiative successful
- Data Virtualization is the right technology for an agile and flexible Big Data Fabric

Q&A

Next steps

Download Denodo Express:
www.denodoexpress.com

Access Denodo Platform in the Cloud!
30 day FREE **trial** available!

Denodo for Azure:
www.denodo.com/TrialAzure/PackedLunch
Denodo for AWS: www.denodo.com/TrialAWS/PackedLunch

denodo
webinar series

Next session

Self-Service Analytics with Guard Rails

Thursday, June 21, 2018 at 11:00am (PST)

Paul Moxon

VP Data Architectures and Chief Evangelist

denodo
webinar series

Thank you!

© Copyright Denodo Technologies. All rights reserved
Unless otherwise specified, no part of this PDF file may be reproduced or utilized in any for or by any means, electronic or mechanical, including photocopying and microfilm, without prior the written authorization from Denodo Technologies.