Decoding the Big Data Deluge…a Virtual Approach

Dan Luongo,

Global Lead, Field Solution Engineering

Data Virtualization Business Unit, Cisco

“High-volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.”

Gartner

What Changed?

- 2011 Internet-connected things:
  - 15+ billion permanent
  - 50+ billion intermittent

- 2020 Internet-connected things:
  - 30+ billion permanent
  - >200 billion intermittent

Content & Services via Connected Products

Biometrics

Sensors & Devices

Building & Infrastructure Management

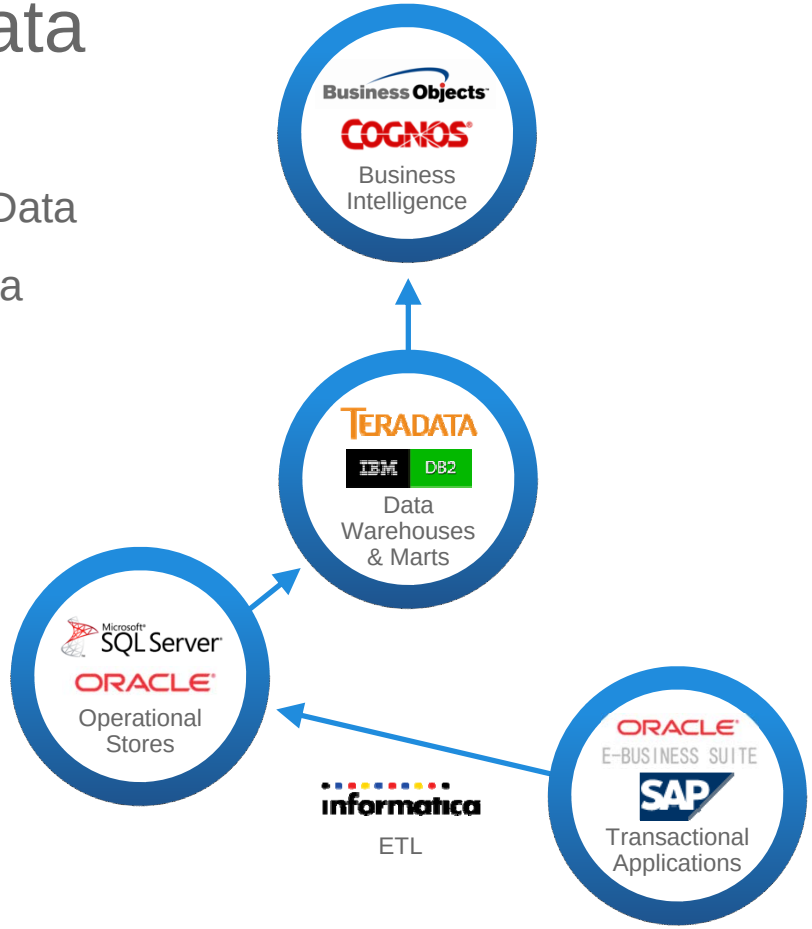
What Are Your Big Data Plans?

- Have Invested
- Investing within next year
- Investing within two years
- No Plans
- Don’t Know


Big Data Integration Challenge

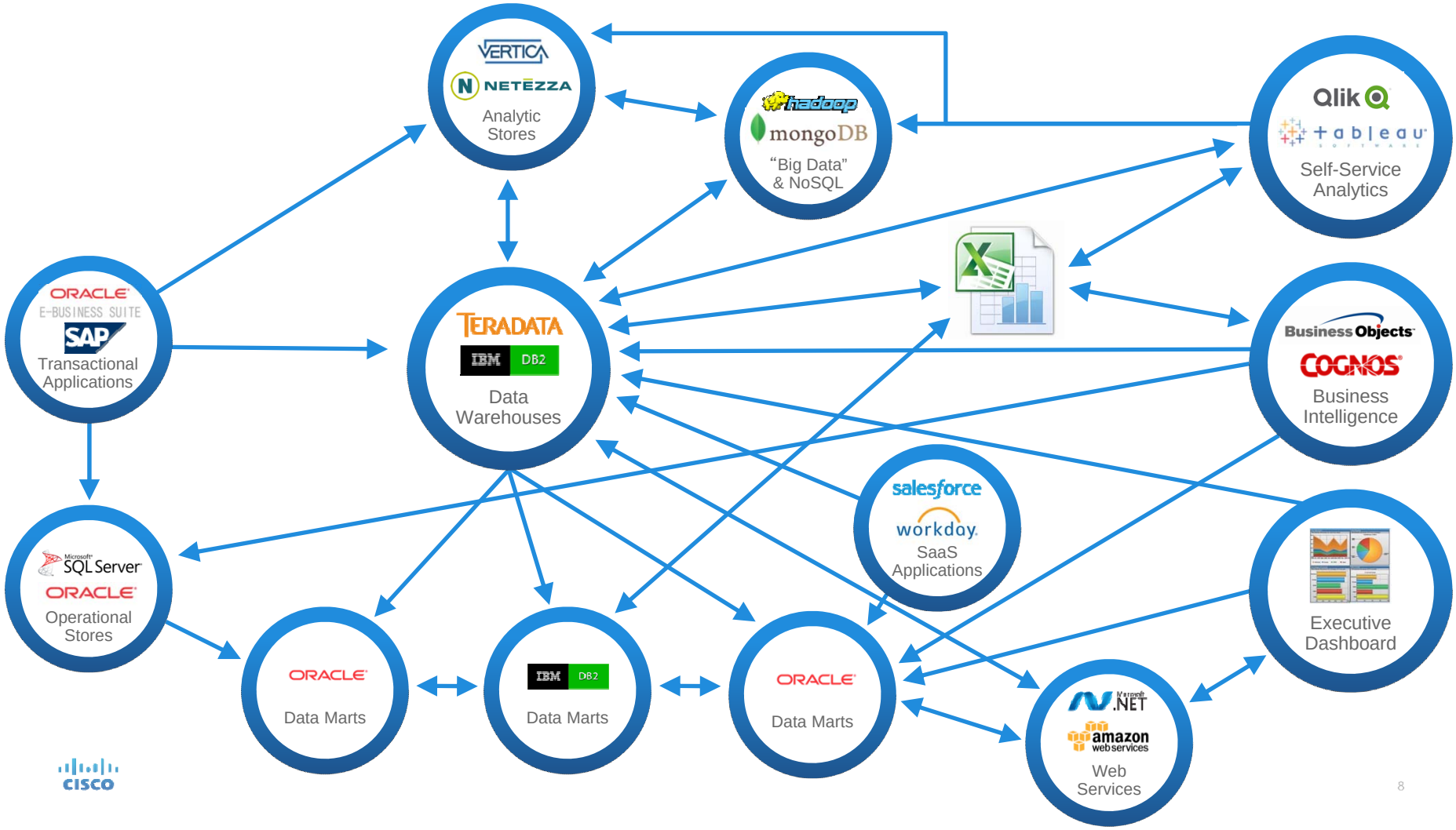
Traditional Data Warehouse

- One Place to Go for Data
- Business View of Data



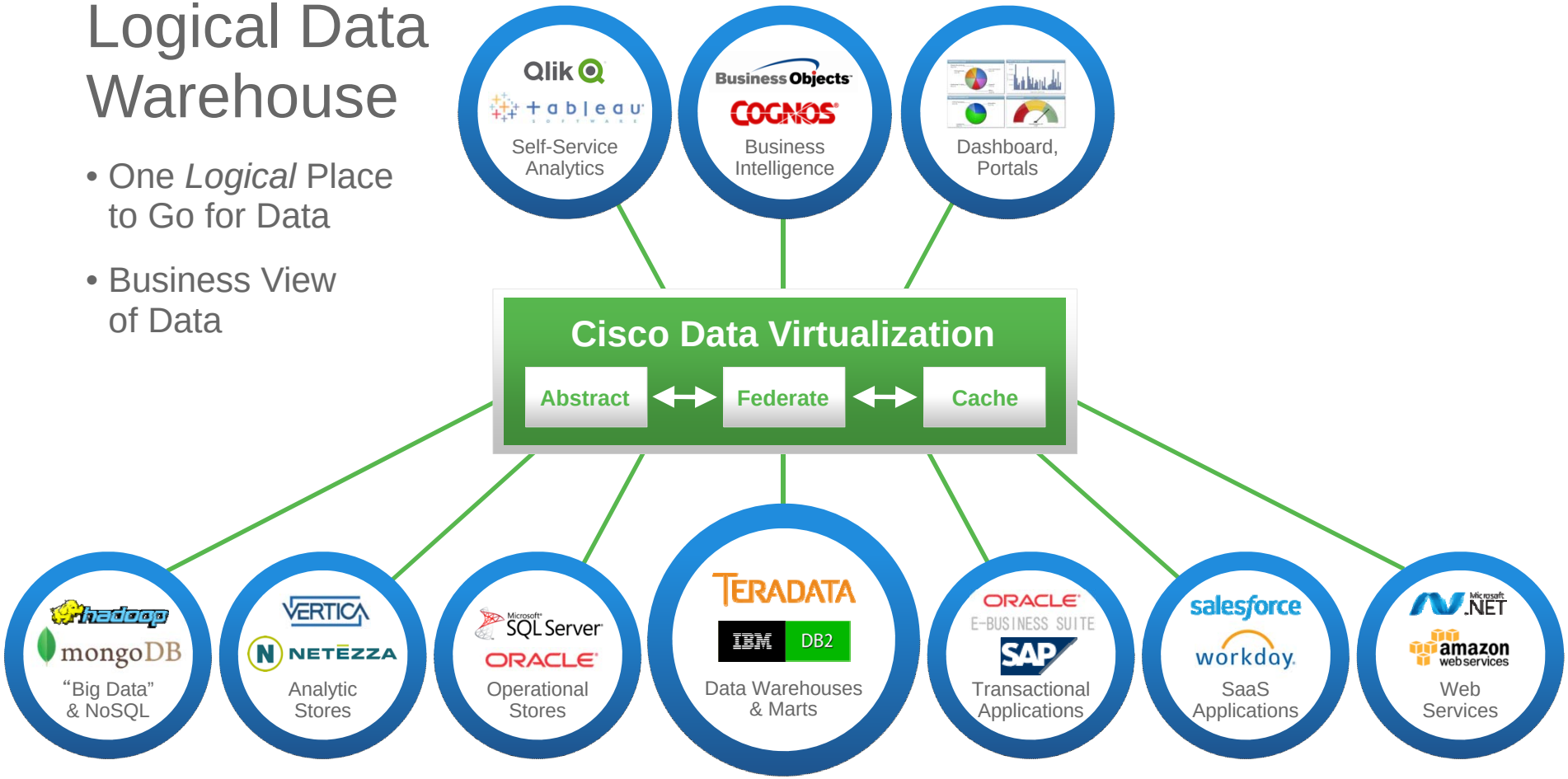
Challenges

- Accelerating Business Demand
- Proliferation of Distributed Silos
- Many Different Access Protocols

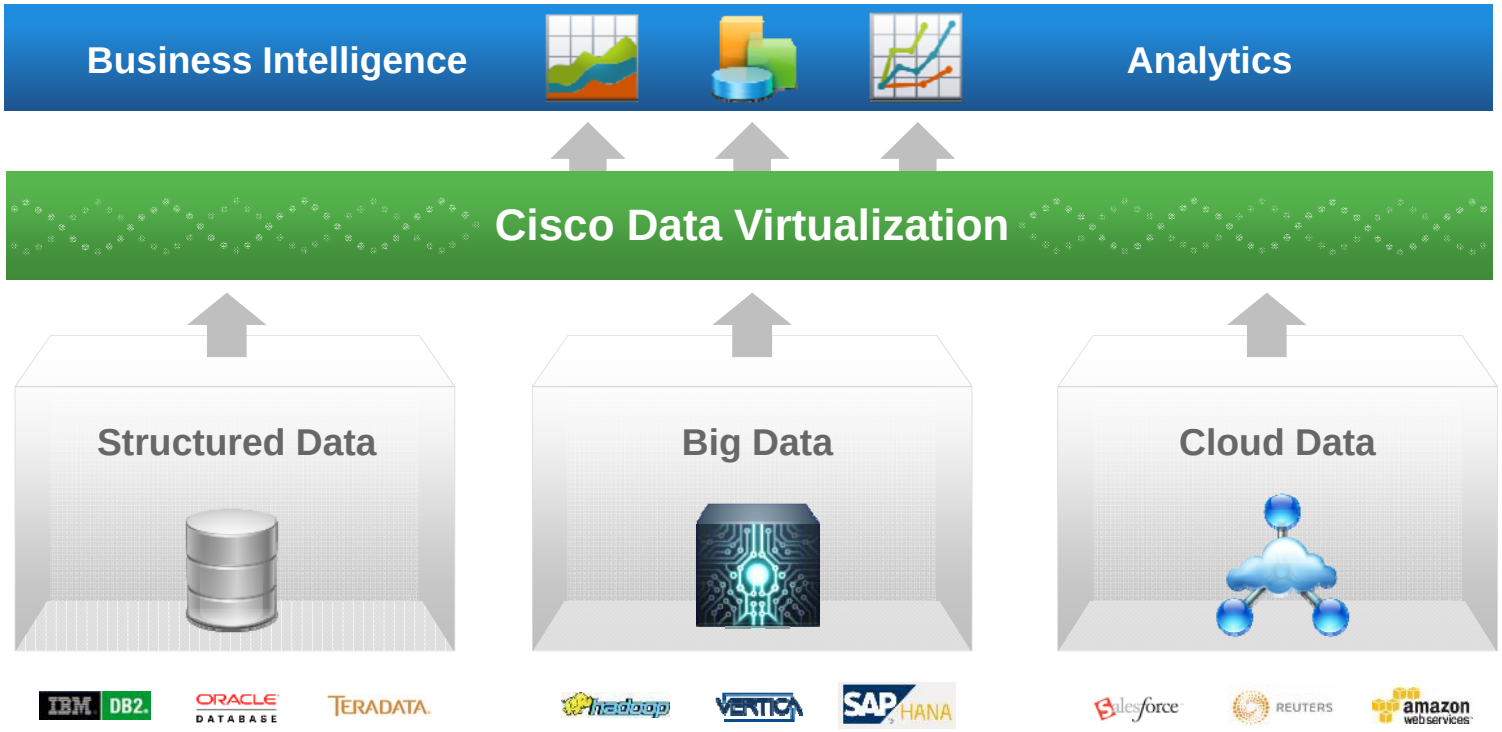

VERTICA
NETEZZA
Analytic Stores
hadoop
mongoDB
"Big Data" & NoSQL
Qlik
tableau
Self-Service Analytics
ORACLE
E-BUSINESS SUITE
SAP
Transactional Applications
TERADATA
IBM DB2
Data Warehouses
Business Objects
COGNOS
Business Intelligence
salesforce
workday.
SaaS Applications
Microsoft SQL Server
ORACLE
Operational Stores
Executive Dashboard
ORACLE
Data Marts
IBM DB2
Data Marts
ORACLE
Data Marts
Microsoft .NET
amazon web services
Web Services
cisco

Logical Data Warehouse

- One *Logical* Place to Go for Data
- Business View of Data



Cisco Data Virtualization

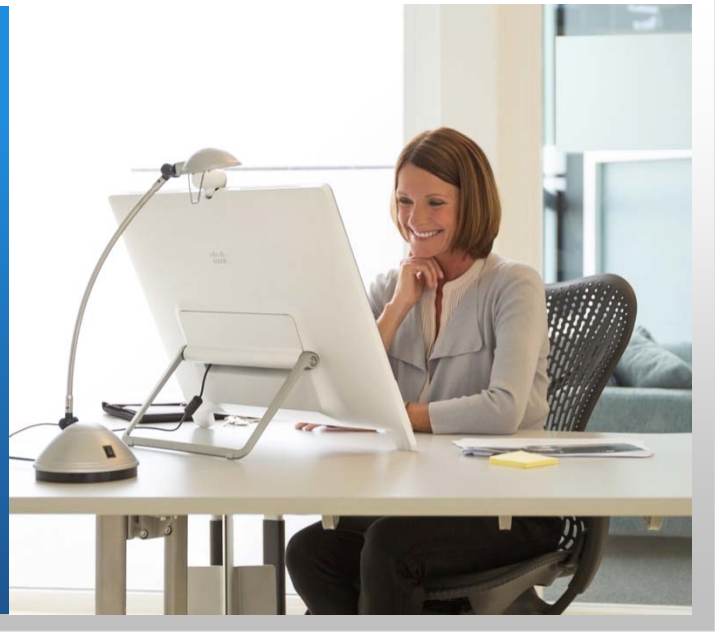


What is Cisco DV, Why is it Unique, Why is it Better?

Cisco DV is agile data integration software that makes it easy to access data, no matter where it resides, and query it across the network as if it is in a single place.

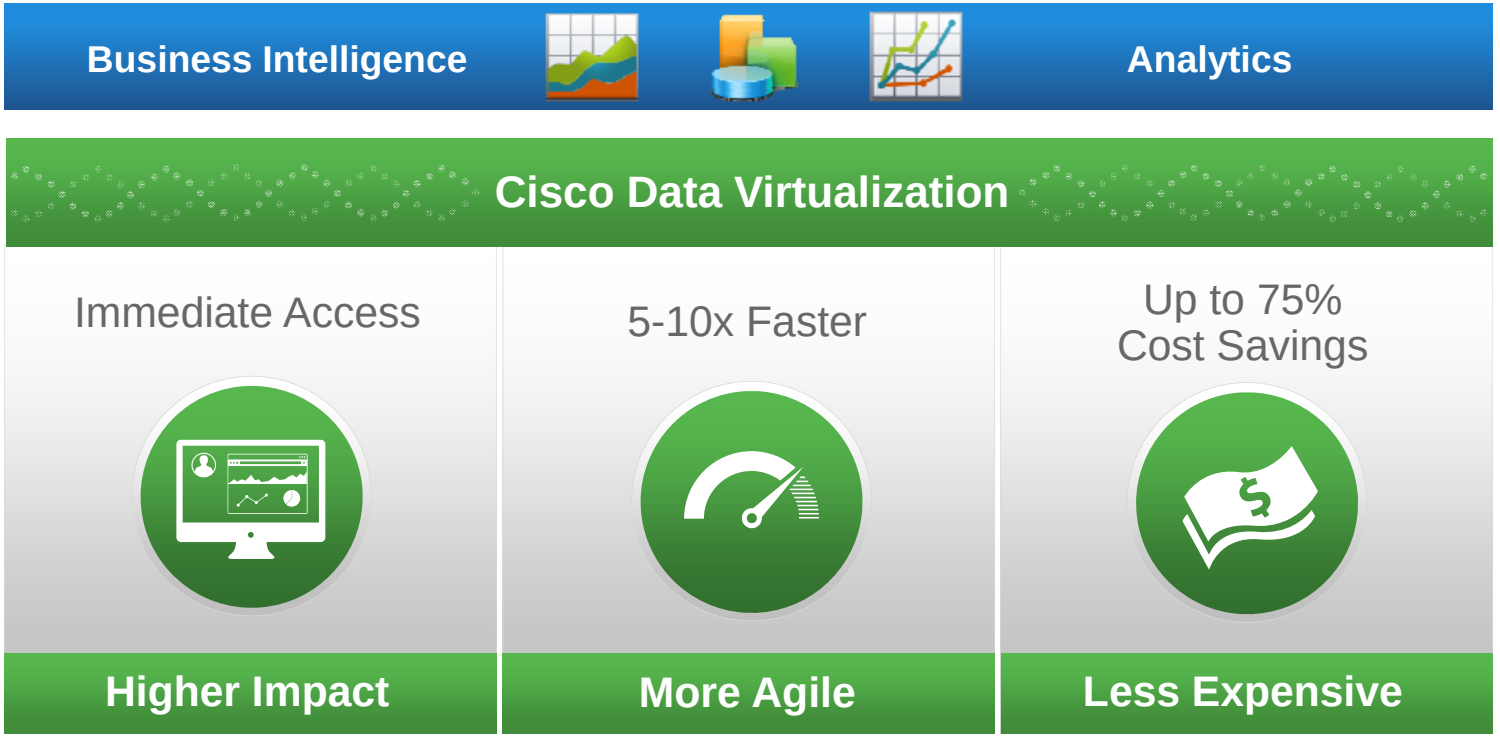
The ability to derive real value from a data virtualization platform is dependent upon

- Federated query optimization
- Breadth of data source systems
- Breath of consuming applications
- Enterprise security, scalability & manageability
- Network level query optimization



Cisco Data Virtualization

Better Business Outcomes, Faster, for Less



Benefits of Data Virtualization: Business Agility

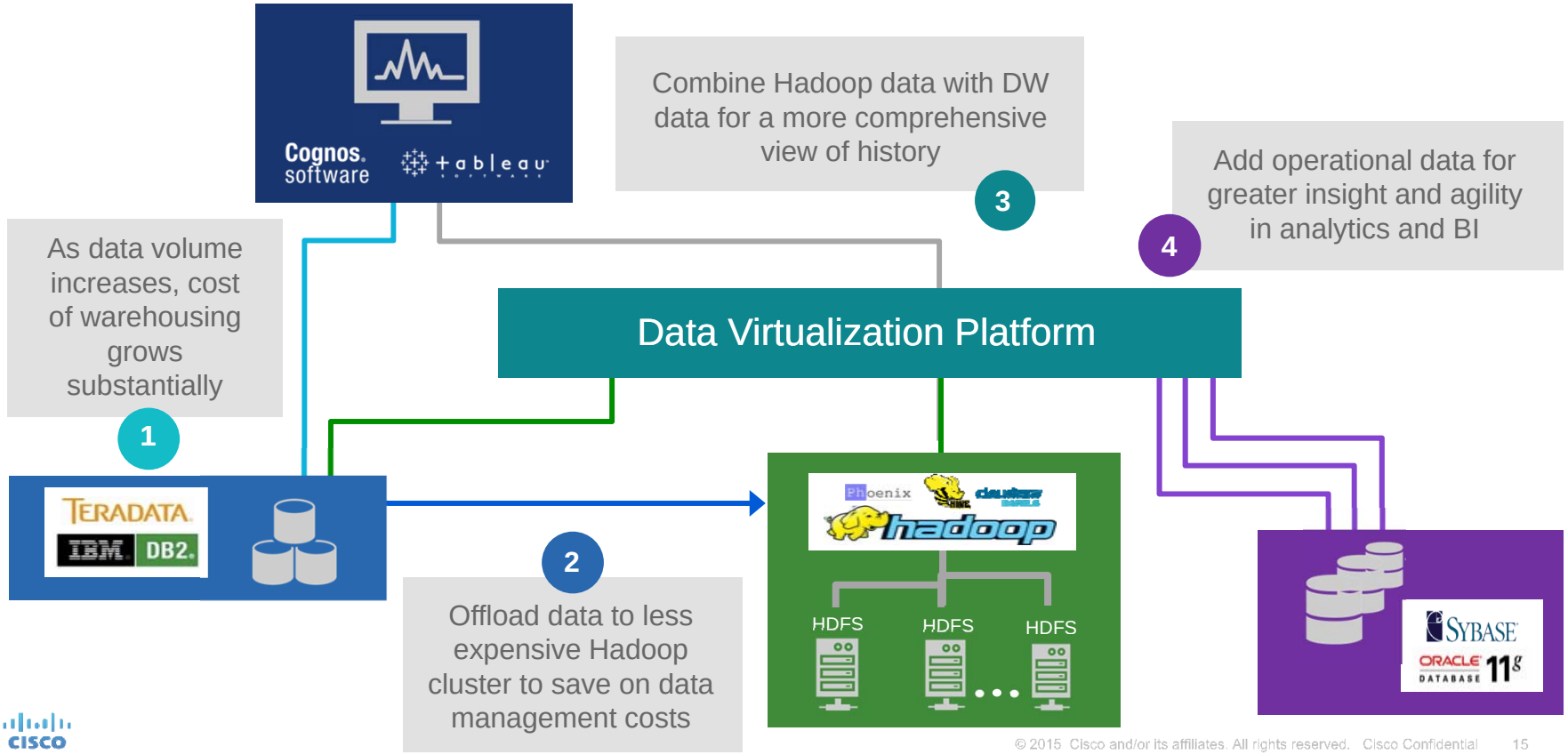
Reduced development effort makes it easier to adapt to changes

- Get to solving the business problem sooner in the project timeline
- Same data does not have to be re-modeled for representation in a different DBMS
- Application programmers don't have to design data movement code; custom data movement code or ETL code doesn't have to be tested and supported
- Application designers can focus on solving the business problem
  - Let data virtualization worry about locality, format, etc.

Benefits of Data Virtualization: Cost Reduction

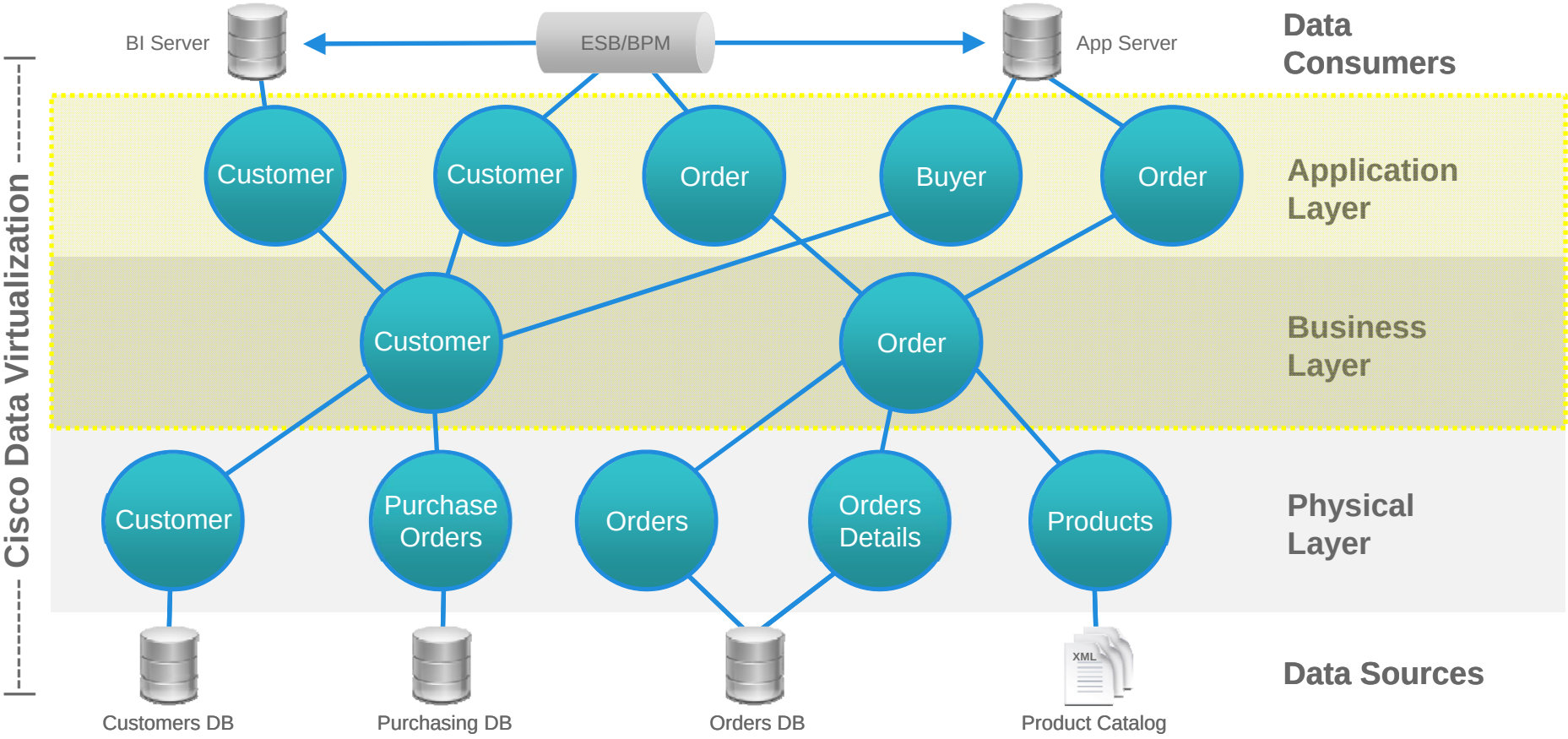
- Duplicate storage eliminated
  - **Duplicate storage *management* also eliminated**
- Data can be moved just when needed; dropped when no longer needed—automatically
- Development effort for application-level data copying eliminated
  - Data access achieved by configuration, **not programming**
- Also, Faster Time to Data – Reduces Op Ex

Data Warehouse Optimization (DWO)



Data Abstraction Reference Architecture

Layered Architecture View



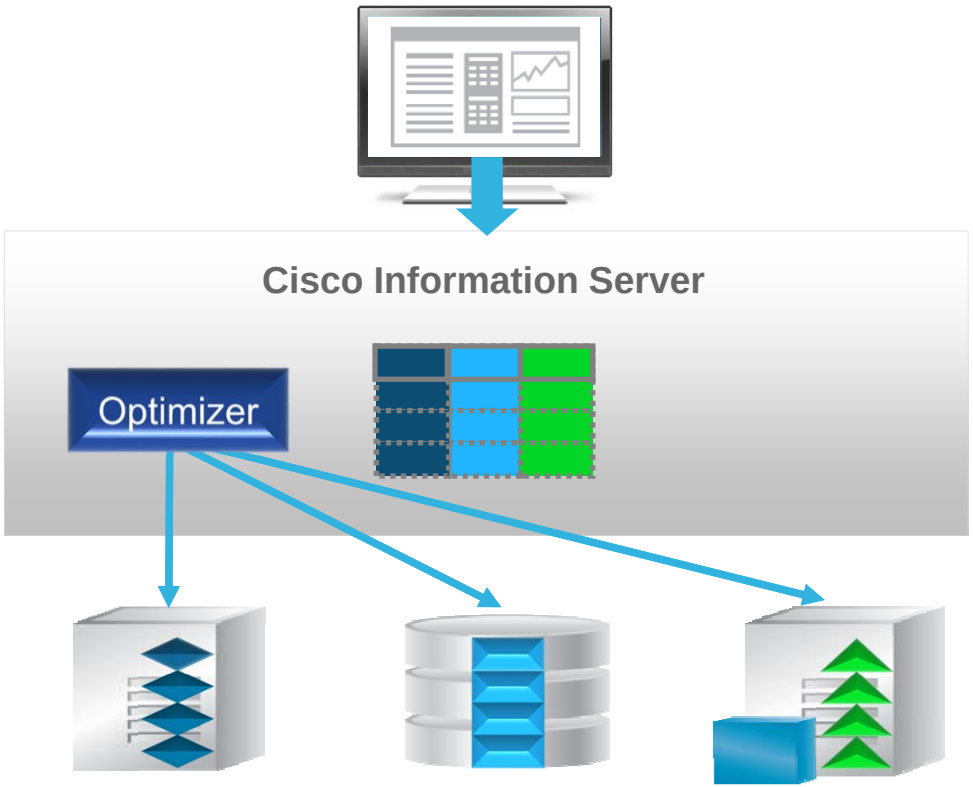
Data Virtualization Runtime

Production Steps

1. Application invokes request
2. Optimized query (single statement) executes
3. Deliver data in proper form

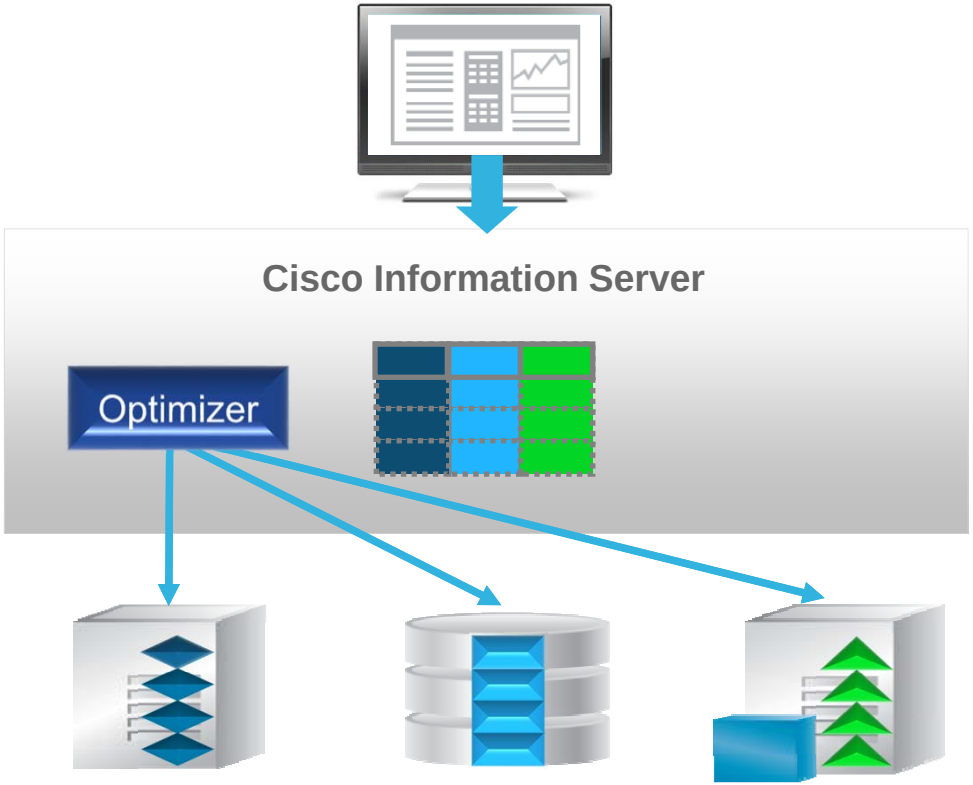
Benefits

- Low latency data
- Optimized performance
- Less replication required

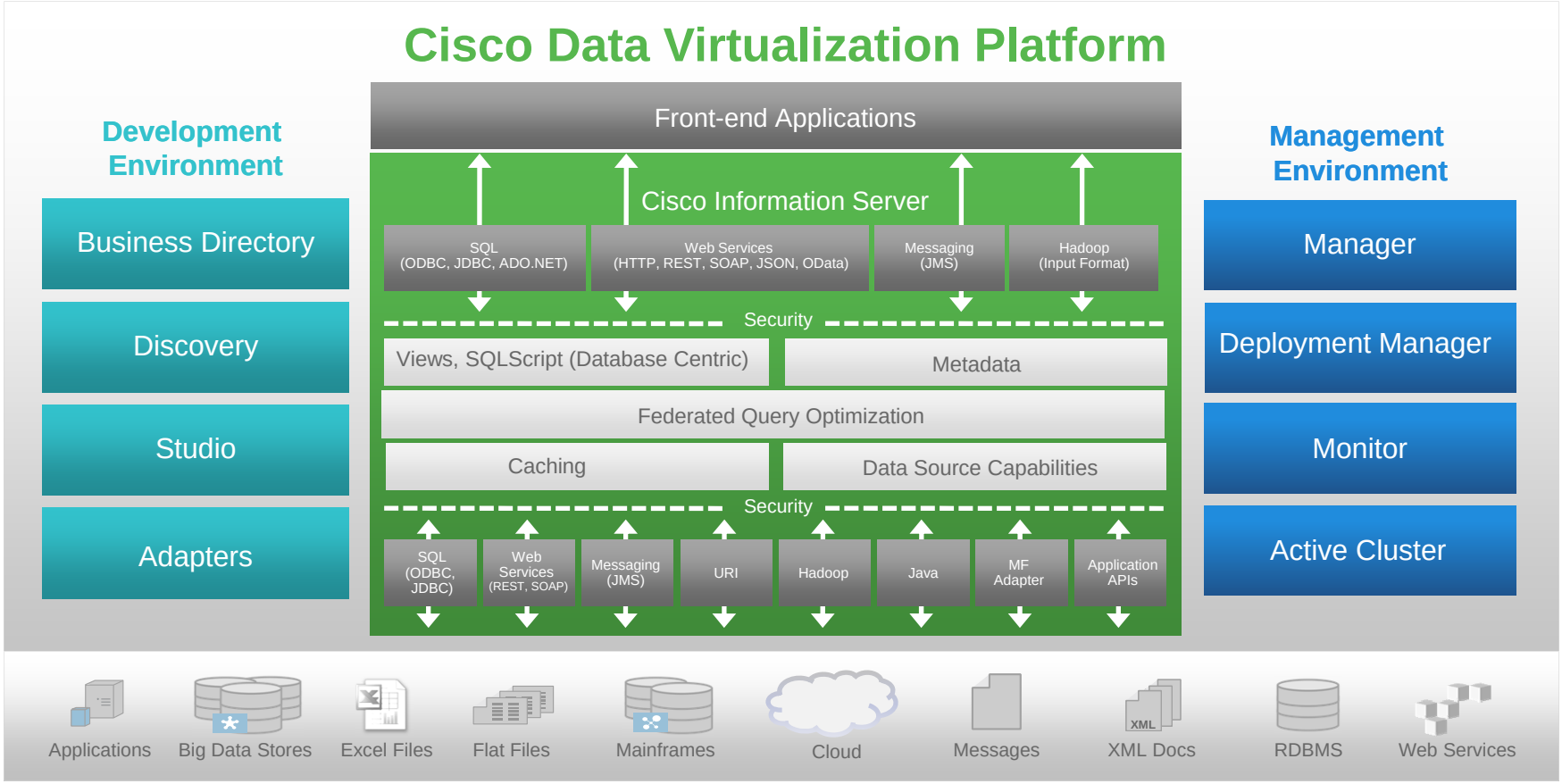


Query Optimization

- Optimal Query Plan (pre-fetch)
  - Auto SQL code generation
  - Cost-based optimizer – using statistics
  - Rule-based optimizer – using user knowledge
- Minimum Network Impact (fetch)
  - On-the-fly plan regeneration
  - Push down joins – leverage source optimizers
  - Minimal data retrieval necessary to complete request
  - Streaming results
  - Streaming-XML technology
- Maximum Join Efficiency (post-fetch)
  - Auto-selection of fastest join operators



Data Virtualization Architectural Overview



Key Takeaways

1. Big is here or its coming
2. Big Data requires new skills
3. Big Data is not a replacement for the traditional DWH
4. Big Data often results in one (or more) new data silos
5. Often times the users wants to augment Big data with enterprise data or vice versa
6. Data Virtualization provides a way to quickly integrate Big Data with enterprise data in a manner that is consistent with your current EIM practices
7. Everyone wins