

Chapter 4: Architecture Framework

- Accidental Architecture
- 4 Architectures
- Information Architecture
- Data Architecture
- Technical Architectures
- Product Architectures
- Terminology



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.



MORGAN KAUFMANN

Key Takeaways

- Inconsistent Data is Key to Limiting BI, Analytics and EDW
 - Information is data in context & Context changes based on purpose
 - Business processes, groups & application work in data silos
- Data Architecture for Enterprise Data Warehouse (EDW) & BI is more sophisticated than assumed
 - Data architecture reflects the stepwise refinement of data from source to its many uses
- Cloud has many benefits, but it does not impact inconsistent data nor simplifies the Analytical Data Architecture
 - Cloud great for infrastructure & keeping software up to date
 - Data content is no better or worse because it is stored in cloud

Chapter 4:

Architecture Framework

- Accidental Architecture
- Terminology
- 4 Architectures
- Information Architecture
- Data Architecture
- Technical Architectures
- Product Architectures



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

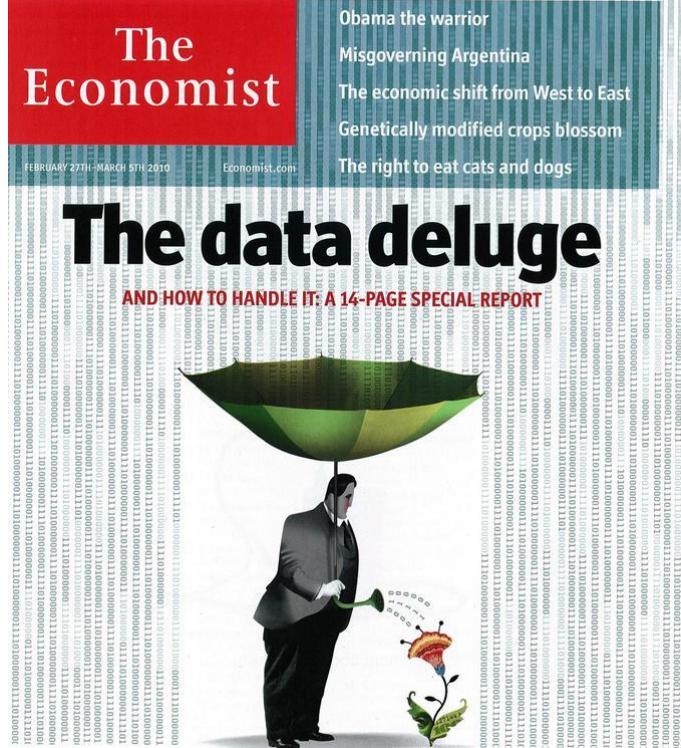
FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.



MORGAN KAUFMANN

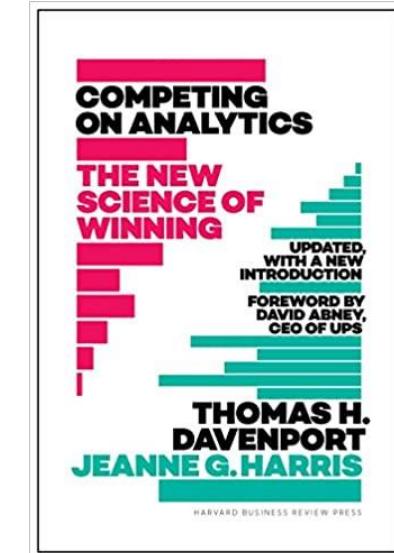
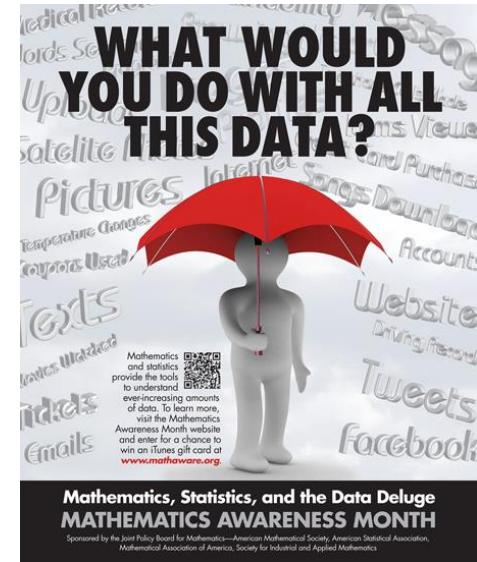
The Data Deluge

- Creating ever increasing amounts of data
 - Society: social media, Internet, news
 - Business
 - Business-to-Business (B2B)
 - Business-to-Consumer (B2C)
 - Internet of Things (IoT): devices, sensors, etc.
- Businesses historically focused on their own data
 - Managed it
 - Exchanged it with others that managed it
 - Mostly business process & events data collected in enterprise applications
- Key differences from past:
 - Unstructured data: big & small
 - External sources: Social media, IoT, Interactions with customers, prospects, suppliers, partners and other stakeholders
 - Many more business processes are being automated in applications (versus being completed manually with spreadsheets) & more data collected on everything



It's the Analytics, not just the Data by itself

- Business demand for analytics
 - Increase sales, manage costs & increase profits
 - Interact with customers, partners & suppliers
 - Respond to competitive pressures
 - Comply with government & industry regulations
 - Examine economic trends
- Need spans industries & enterprise size
- Use spans business & tech roles
- Analytics demand, awareness & hype



Accidental Architecture

Information is data in context

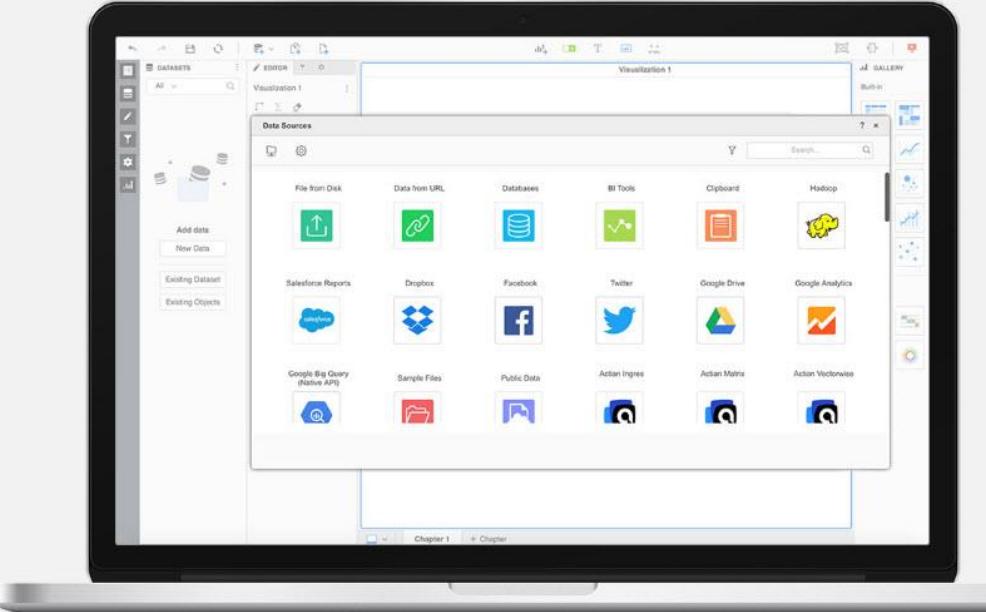


Analytics is NOT just a data connectivity issue

Connect to and blend data from any source.

Using a variety of native connectors, you can tap into nearly any type of data from spreadsheets to Salesforce reports to big data sources like Hadoop. Easily blend data from multiple sources for fast, easy analysis. Quickly define relationships across tables or sources with simple drag-and-drop controls.

[WATCH DEMO >](#)



- Data connectivity is not the inhibitor to obtaining information & insights
 - If you haven't really integrated data before then naïve about real inhibitors
- Overhyped: middleware, microservices, data virtualization, accessing data "as is"
 - Addresses real-time access which is good but not the real problems with respect to information

Information is data in context

According to Peter Drucker, information is “data endowed with relevance and purpose.”

Raw data, such as customer retention rates, sales figures, and supply costs, is of limited value until it has been integrated with other data and transformed into information that can guide decision making.

Sales figures put into a historical or a market context suddenly have meaning—they may be climbing or falling relative to benchmarks or in response to a specific strategy.”

Leandro DalleMule, Thomas H. Davenport, “What’s Your Data Strategy?”, Harvard Business Review, May-June 2017

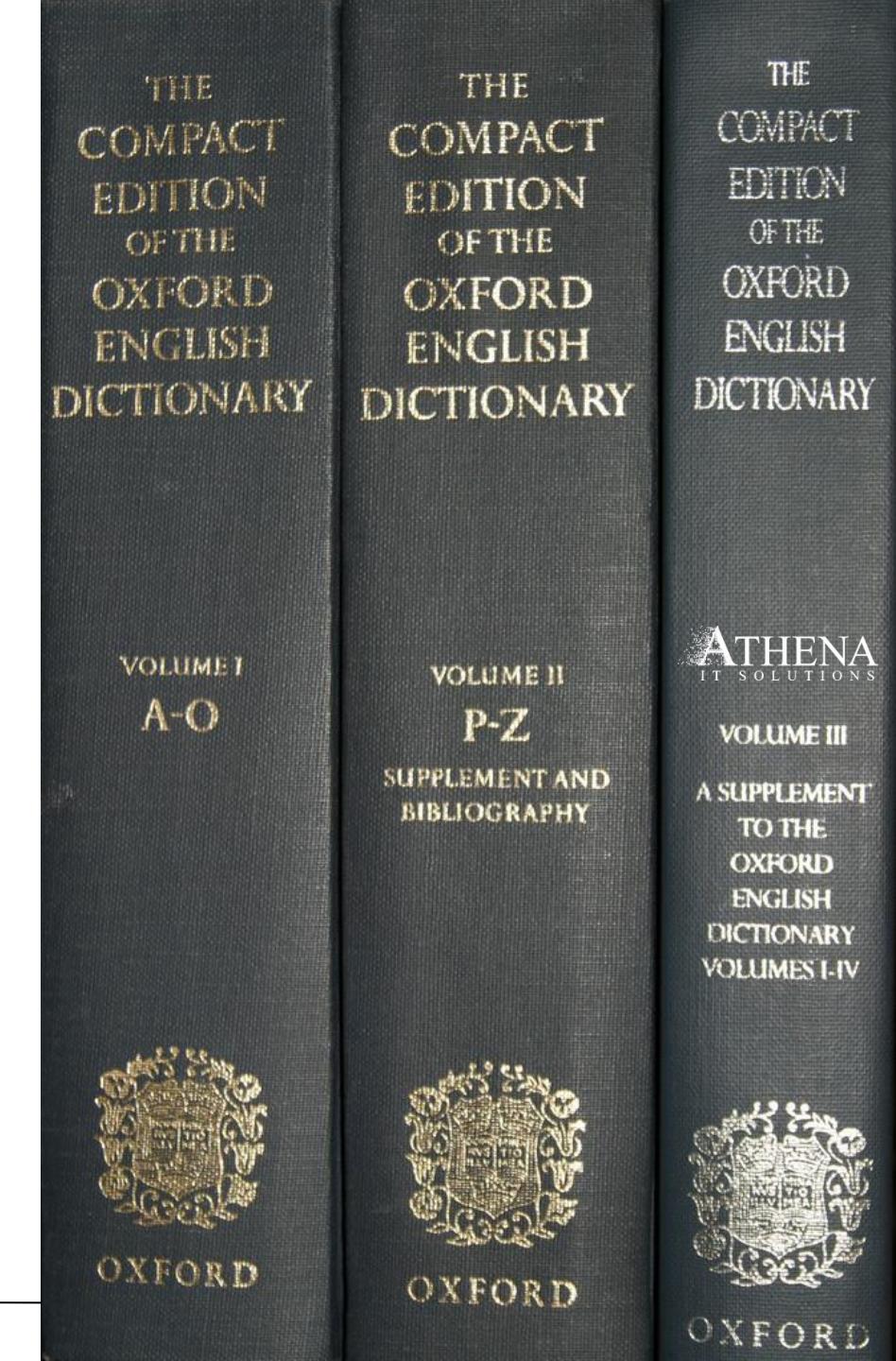
Information is data in context

- Context
 - Why not access the data in place?
 - Simple business questions are not so simple
 - Operational vs enterprise reporting
- Business specific context
 - Business Rules & Filters
 - Metrics
 - Key Performance Indicators (KPIs)
 - Groupings (Hierarchies)
- Data Lifecycle: Data to (various) information contexts
 - No Single Version of the Truth (SVOT)



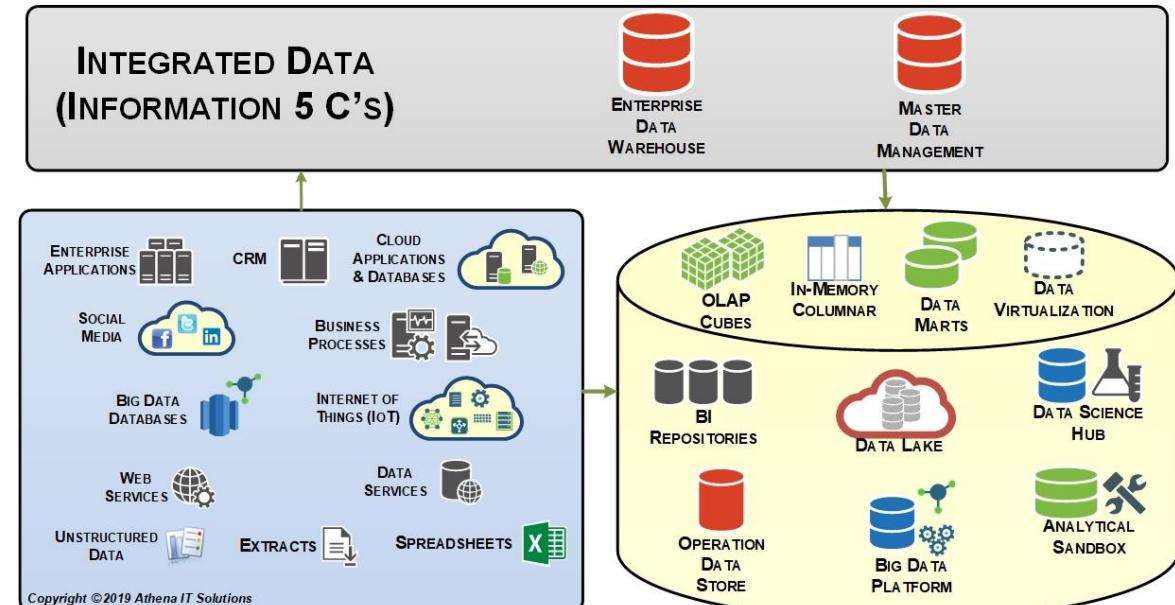
Information 5 C's

- Clean
- Consistent
- Conformed
- Current
- Comprehensive



State of Data: What is Needed for Analysis

- As-Is: Connect to data sources & systems of record (SORs)
 - Direct query to data sources & SORs
 - Operational systems, business applications, IoT, documents & social media
 - Operational reporting, domain specific analysis, real time analysis & alerts
- Needs to be derived: Curated or Integrated Data
 - Step-wise refinement
 - Do once, use many times
 - Business rules – data selection, filters (privacy, security, ease-of-use), metrics, KPIs
 - From As-Is and/or 5C's
- Need 5C's: Data Warehouse, Data Hub, Master Data Management (MDM)



**As-Is: DATA SOURCES
(DATA CAPTURE & PROCESSING)**

- TRANSACTIONAL, OPERATIONAL, MONITORING
- BUSINESS PROCESSES
- SOCIAL MEDIA & COLLABORATION

**CURATED OR INTEGRATED DATA
(NEEDS TO BE DERIVED)**

- ANALYTICAL DATASTORES
- ODS
- DATA LAKES

Myths & Misconceptions

- Single version of the truth (SVOT)
- All that's needed is better access to data (the eternal vendor pitch)
- Everything should be in a data warehouse (DW)
- Everything should be in a data lake
- Schema-on-read is all we need
- You do not need a DW anymore
- The problem is spreadsheets
- Everything is operational (as-is) reporting



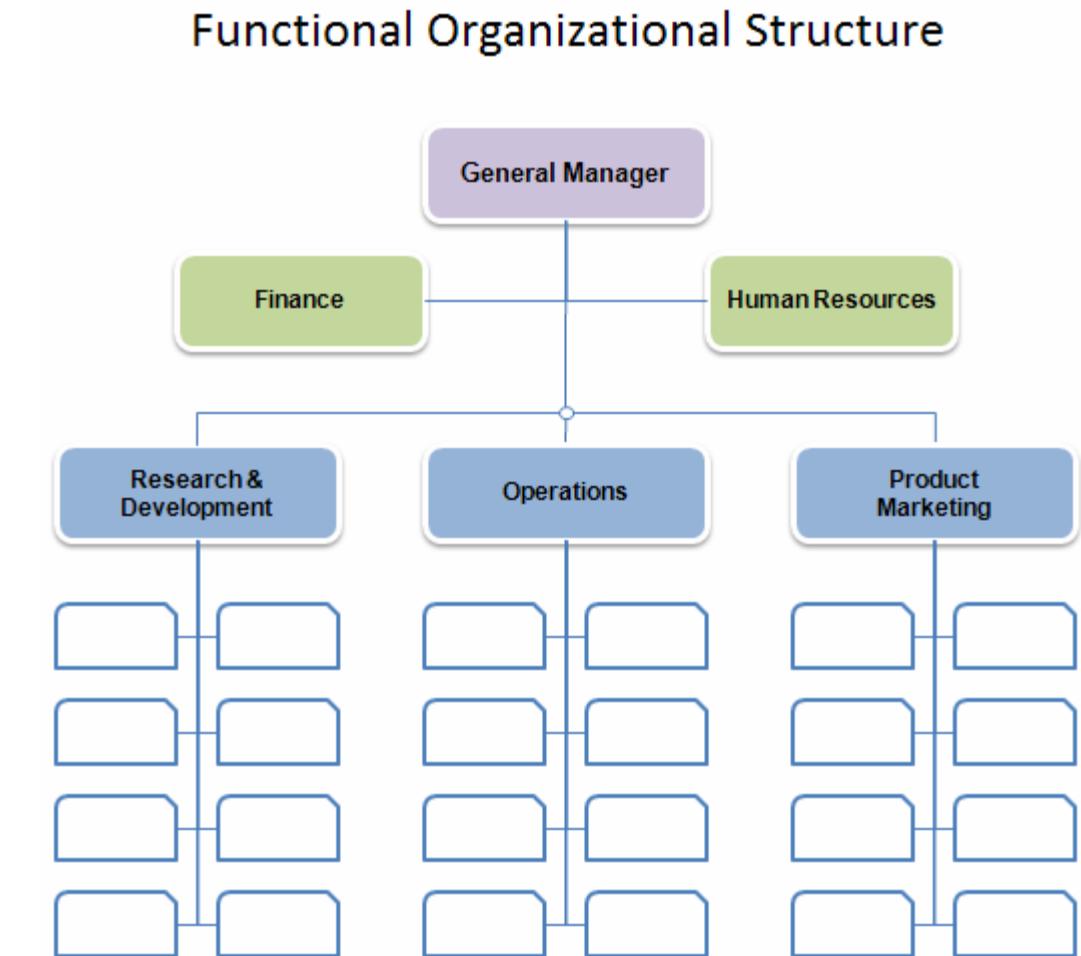


Accidental Architecture

Data silos are rapidly expanding

Business Organizations are setup as Silos

- Organizational Structures
 - Functional
 - Product
 - Geographic
- Enabling Applications
 - Business & operational applications
 - Cloud applications
 - Business application vendors



Business & Operational Applications support Silos

- Business applications
 - Enterprise Resource Planning (ERP)
 - Customer Relationship Management (CRM)
- Automates business functions & processes
 - Sales: Sales Force & Pipeline Management
 - Marketing: Campaigns
 - Finance: Accounting, Payroll, Budgets, etc.
 - Supply Chain Management (SCM)
 - Manufacturing
 - Product Management
 - Customer Support
 - Human Resources: Recruiting, Compensation
- Cloud Applications (3rd wave)
 - Mostly automating “new” processes not replacing existing ERP



BI & Data Silos

- BI projects driven by business group needs
 - BI focused on business processes & groups
 - Initial solutions deliver misleading ROI
- Data integration
 - Much is custom coded, i.e., SQL
 - DW viewed as extract, transform & load (ETL)
 - Data integration versus application integration (silos)
- Not using “Think Globally, Act Locally”
 - Silo-focused integration (oxymoron)
 - Integration often “embedded” in BI
 - Tools selected tactically
 - Infrastructure, resources & applications also tactical



Technology is no Holy Grail

- Machine learning, big data & data scientist hype
 - Everything is big data & big data analytics
 - Technology is too tightly associated with it
- Technology myths
 - Technology solves all
 - Silver bullet solution
 - The “One”
 - Technology – One Size Fits All
 - Architecture – One Way to Do Something
 - Vendor – One Neck to Choke
 - Stereotype



Our Current Mistakes

- Cloud is Magic
- Data Lake avoids thinking & design
- ML (or what people say is ML)
- Data Science w/o science
- Data Engineering == coding
- Parts vs Whole
- Not Understanding...
 - Data
 - Data Uses
 - Personas
 - DB purpose & design
 - Schema purpose & design
- Not learning from previous work

WHEN HYPE DOESN'T MATCH REALITY: DATA AND ANALYTICS FAILURES



Failure rates for analytics, AI, and big data projects = 85% – yikes!

July 23, 2019 by Brian T. O'Neill



[Home](#) > [Big Data and Analytics](#)

Five Reasons Why Your Data Science Project is Likely to Fail

Edited by: Chris Preimesberger, eWEEK | March 27, 2019



eWEEK DATA POINTS: More than 85 percent of big data projects fail. A number of factors contribute to these failures, including human factors, and challenges with time, skill and impact. Here are some precautionary data points of advice.



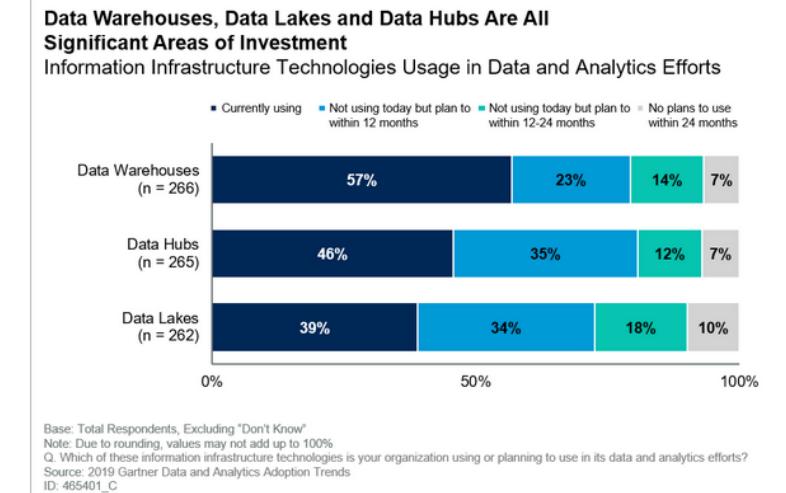
Companies are forging ahead with digital transformation at an unprecedented rate. A recent survey by Gartner Research found that 49 percent of CIOs are reporting that their enterprises have already changed their business models to scale their digital endeavors or are in the process of doing so.

As companies forge ahead with these transformations, they are infusing data science and machine learning into various business functions. This is not an easy task. A typical enterprise data science project is highly complex and requires deployment of an interdisciplinary team that involves assembling data engineers, developers,

data scientists, subject matter experts and individuals with other special skills and knowledge. Moreover, this talent is scarce and costly. In fact, only a small number of companies have succeeded in building an experienced data science practice. And, while building this team takes time and resources, there is an even larger problem faced by many of these companies: more than 85 percent of big data projects fail.

Current State of BI & Analytics

- Significant investment in BI for a couple of decades
- Many BI tools in an enterprise but low adoption
- IT not able to keep up with BI demand
- Business take action in their own hands (next)



- By the end of 2020, the Global revenue in the BI and analytics software market [is forecasted to grow to \\$22.8B, according to Gartner.](#)
- The global BI software market accounted for \$16.3B in 2015, and is expected to reach \$26.5B by 2021, growing at a compound annual growth rate (CAGR) of 8.4 percent between 2016 and 2021, according to [Zion Research](#).
- Big Data and [business analytics software worldwide revenues will grow from nearly \\$122B in 2015 to more than \\$187B in 2019](#), an increase of more than 50 percent over the five-year forecast period.
- BI adoption [is lingering at 30% in the majority of enterprises, according to Gartner](#)

Spreadsheets Fill the Data “Gaps”

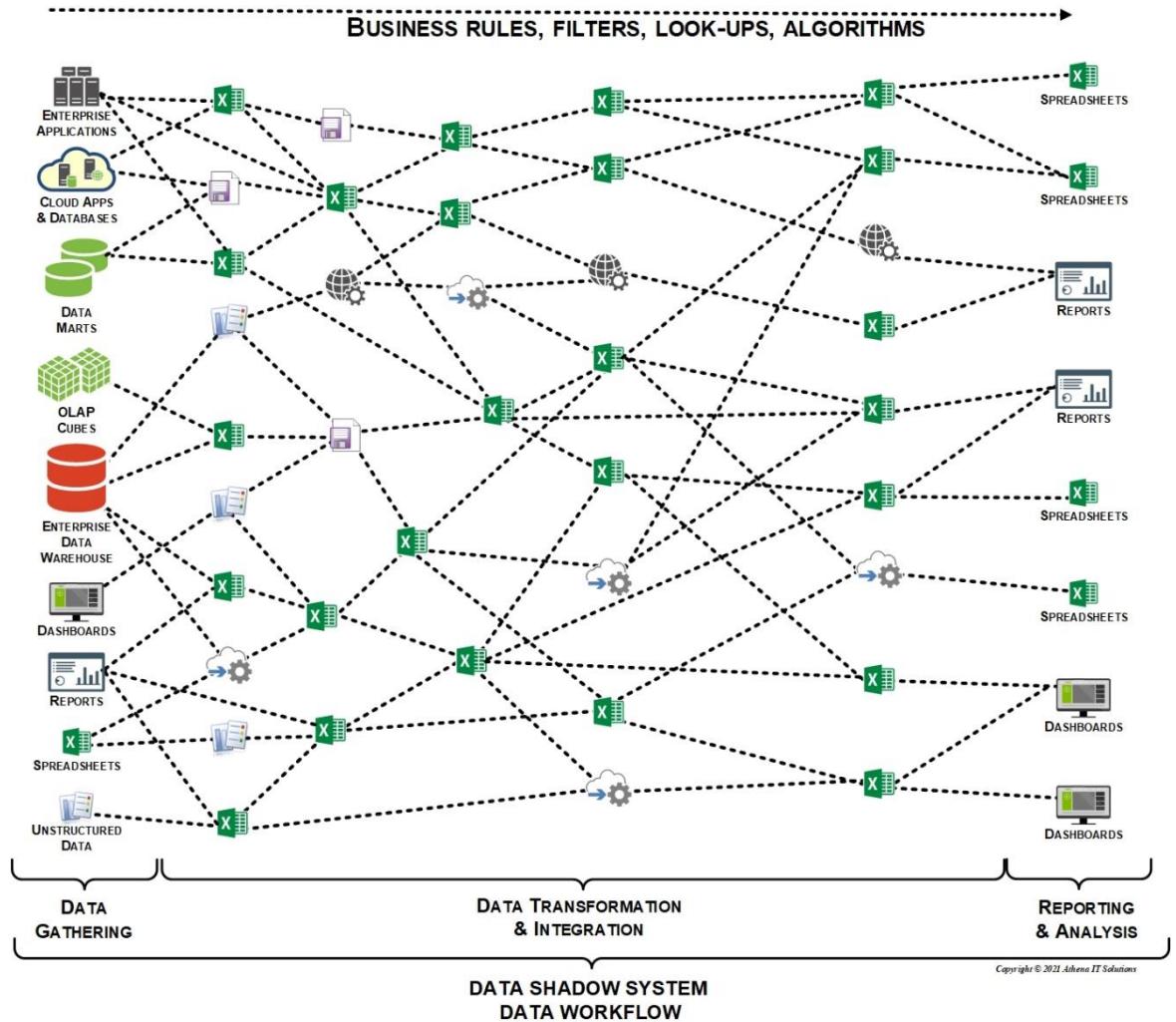
- “50%-80% of all BI/analytics apps are still done in spreadsheets”
(Gartner Research, Boris Evelson)
- Primary tool used by business to present & analyze data
- Only “BI” tool with pervasive business use
- Data superglue
- Often “masks” ineffective data architecture or mismatched BI tools



Spreadsheets evolve to Data Shadow Systems



- Primary tool used for business analysis
- Only BI tool with pervasive business use
- Data superglue
- Often masks ineffective data architecture or mismatched BI tools

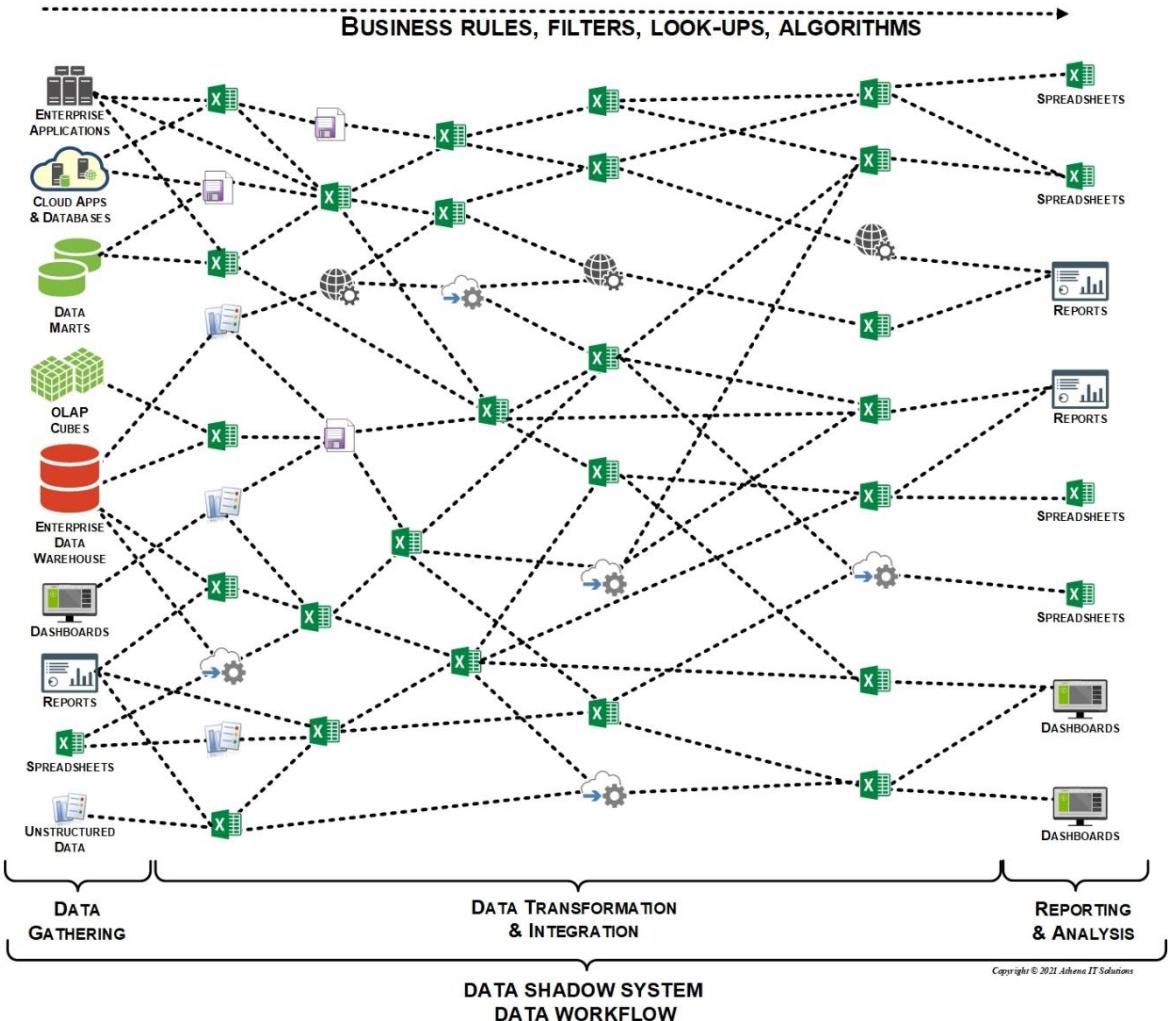


Next Generation of Data Shadow Systems



The new spreadmarts built with:

- Data discovery tools
- Data preparation tools
- Cloud applications
- Big data applications
- ... and spreadsheets

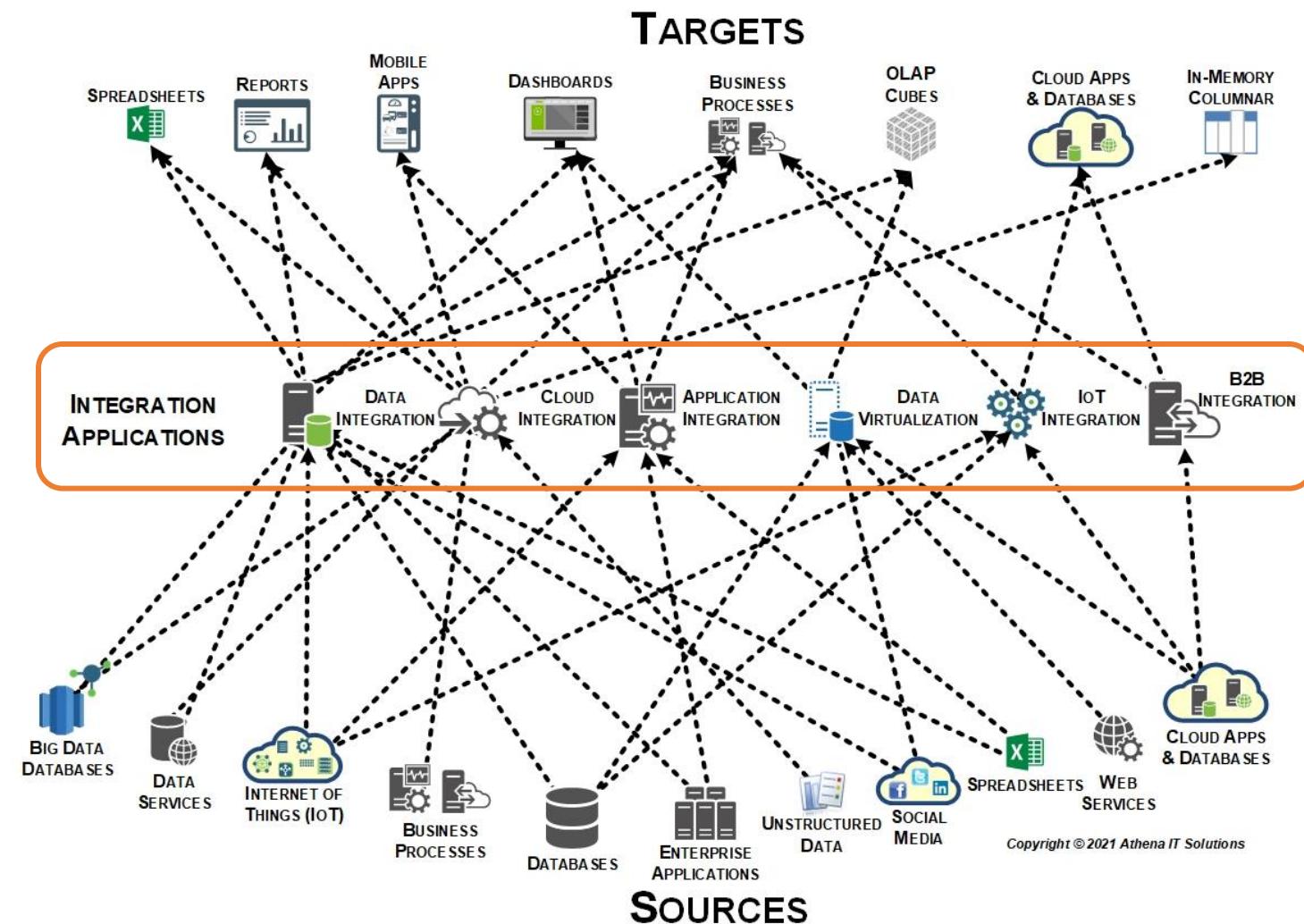


Current State: Integration

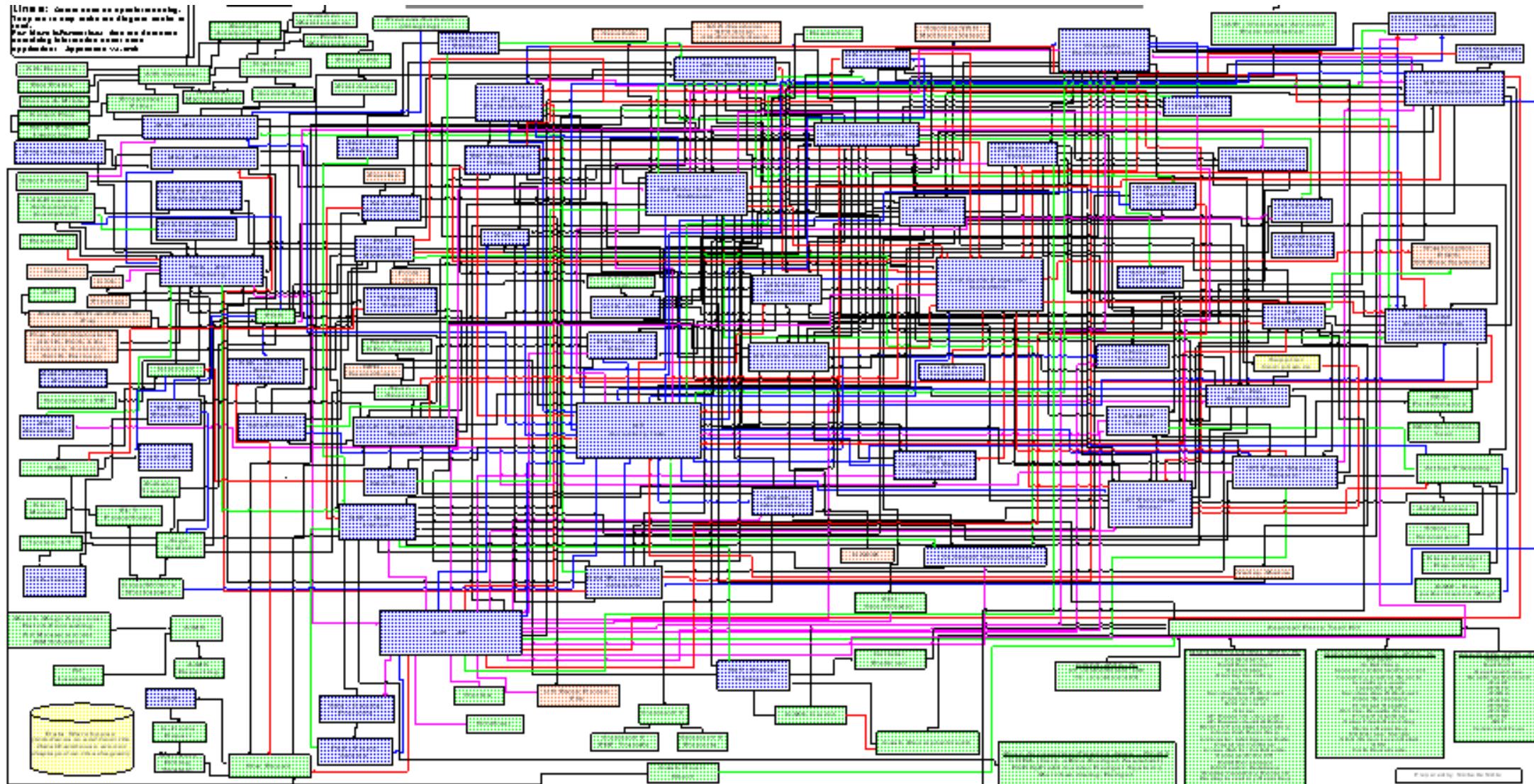
- Primary integration toolsets:
 - Application integration
 - Data integration
- Majority of tool-based work is batch or bulk
- **Much of data integration is custom-coded**
 - SQL
 - Programming language du jour
 - Spreadsheets
- Most data involved in data integration efforts are siloed & not usable by other processes
- Data governance is not pervasive



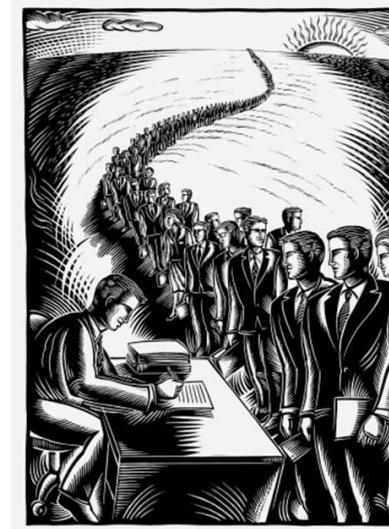
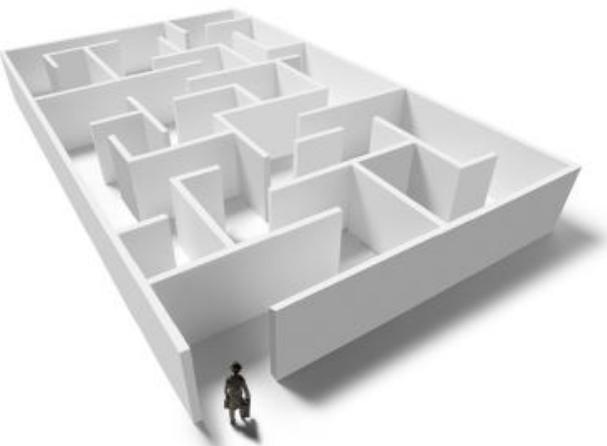
Current State: Integration



Accidental Architecture for Data & BI



Current State of Data & Analytics



Chapter 4: Architecture Framework

- Accidental Architecture
- Terminology
- 4 Architectures
- Information Architecture
- **Data Architecture**
- Technical Architectures
- Product Architectures



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.

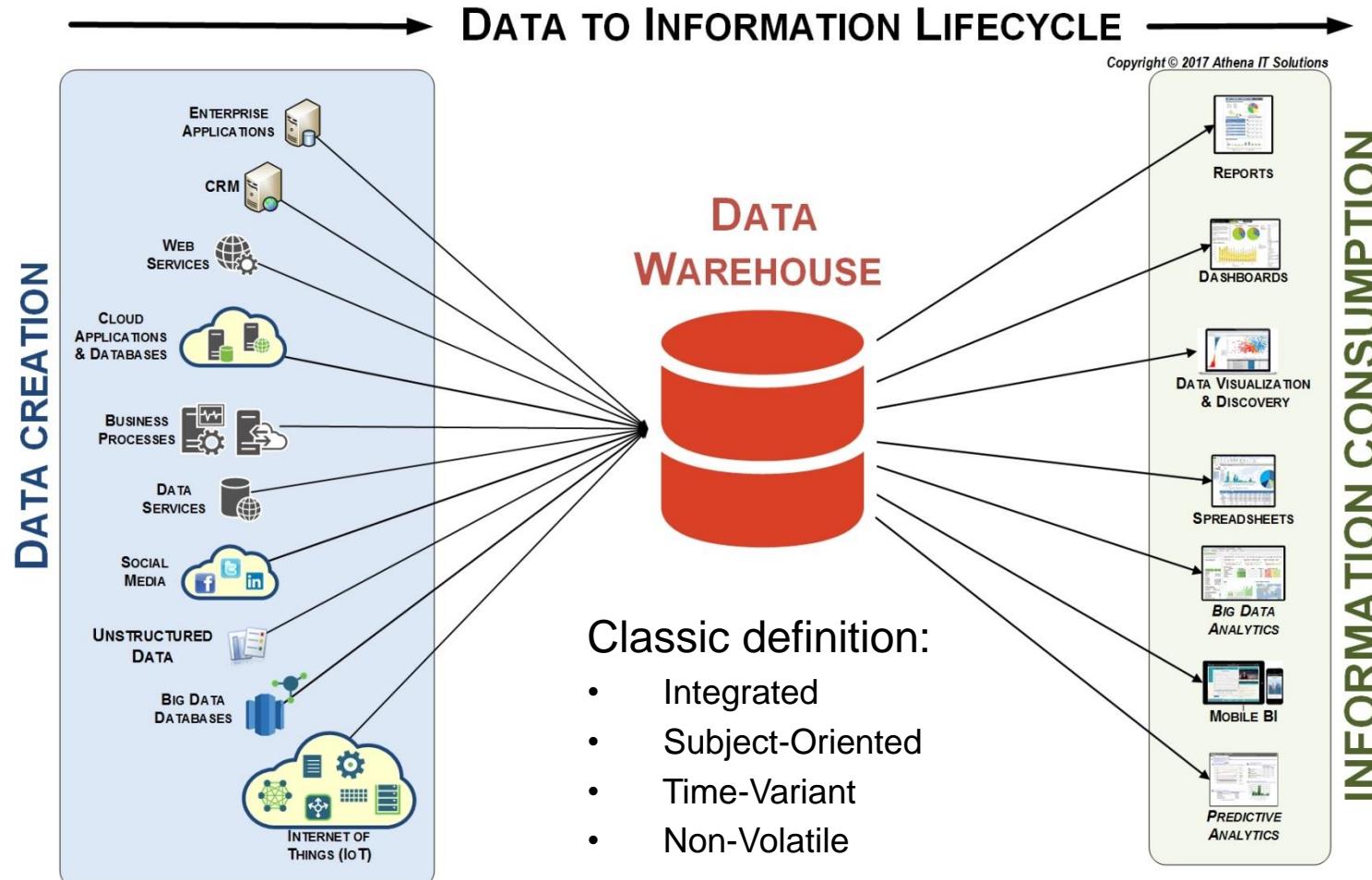


MORGAN KAUFMANN

Data Warehouse (DW)

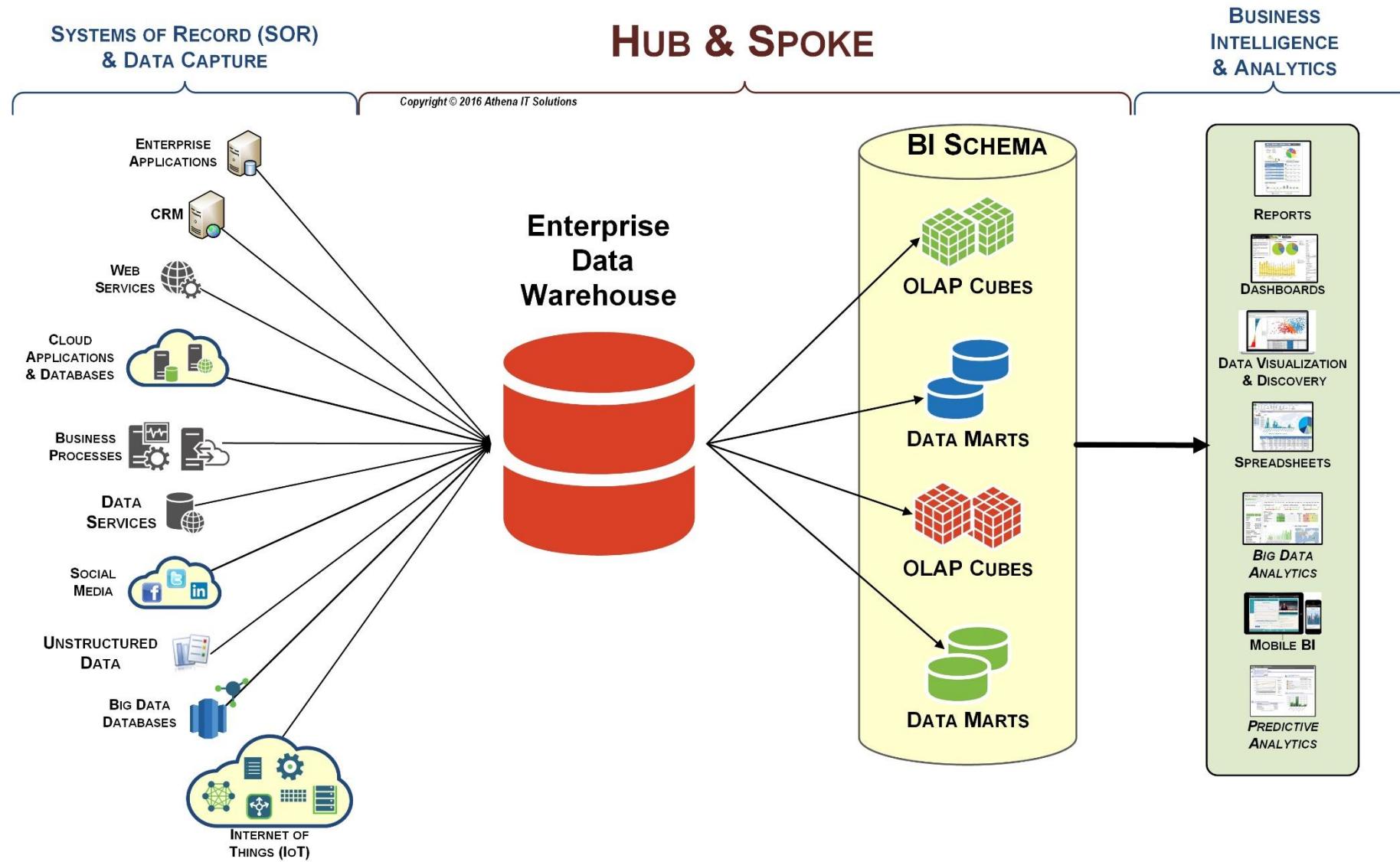


Details to be discussed later in course



In beginning a place to store data for reporting, decision support systems (DSS) & BI in centralized database

Classic DW Evolution – Hub & Spoke

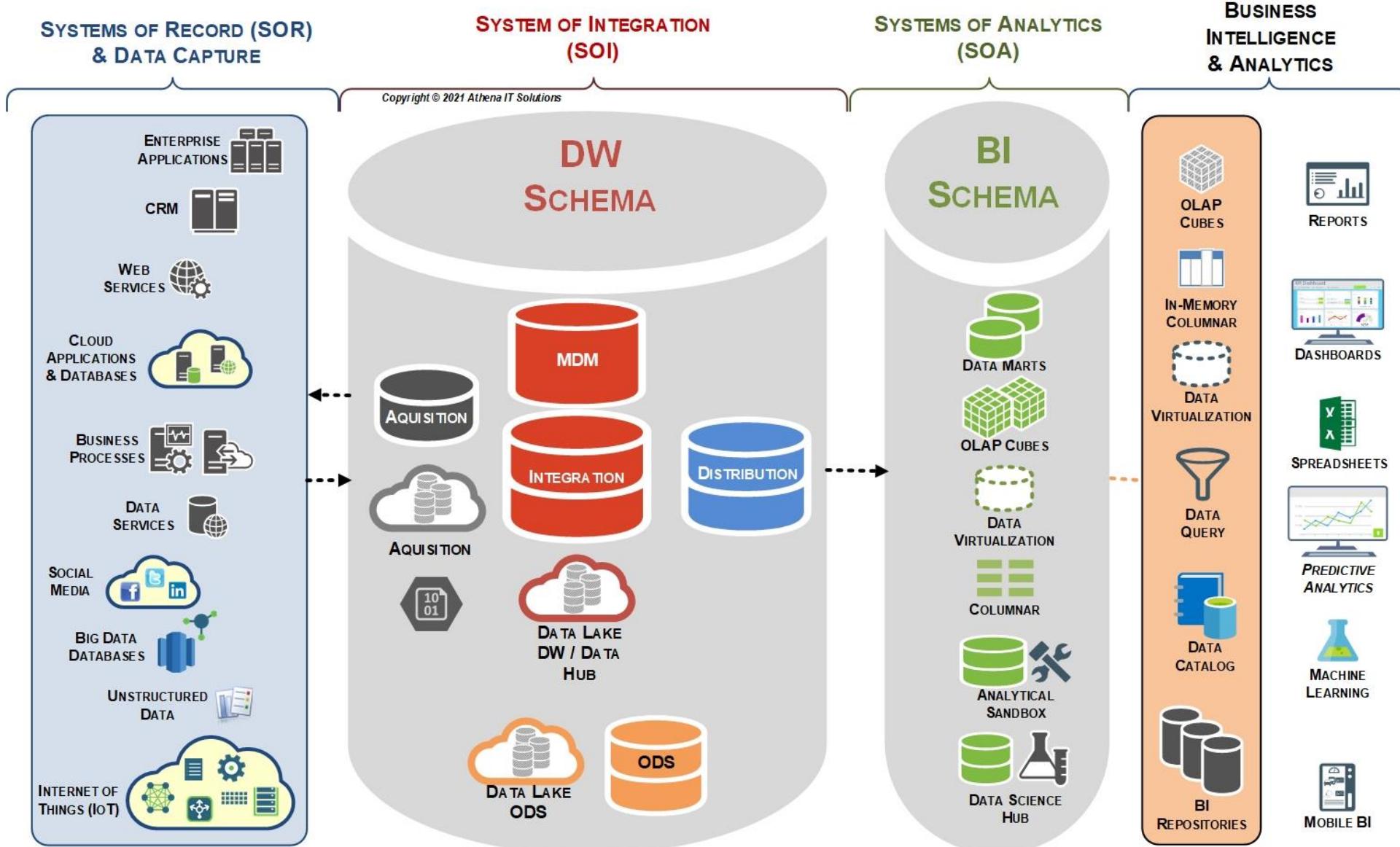


Details to be discussed later in course

Analytical Data Architecture (ADA)



Details to be discussed later in course



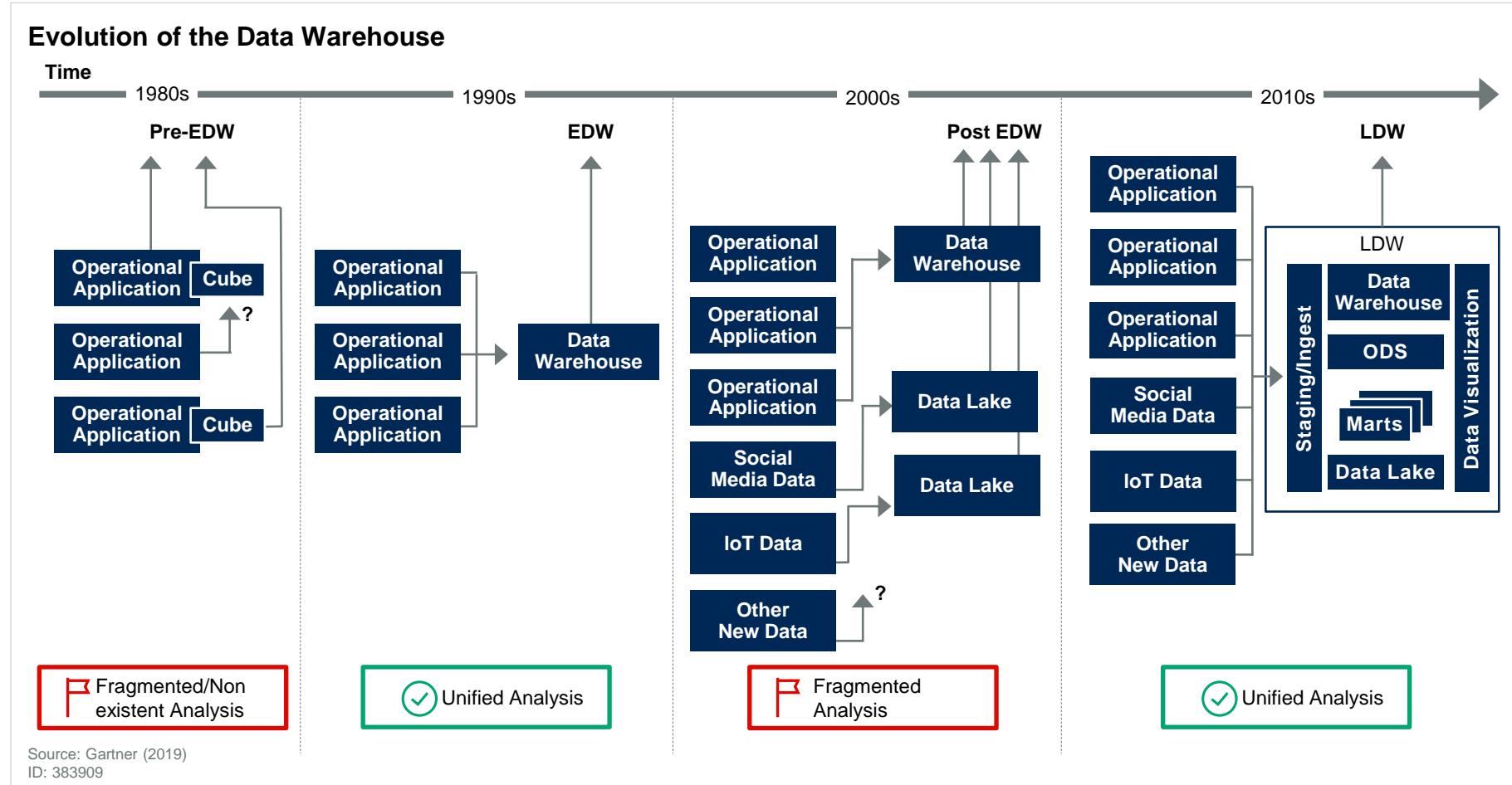
Template

- Use drives design
- Many complementary options, but not all
- Mix & match

Analytical Architecture



Details to be discussed later in course

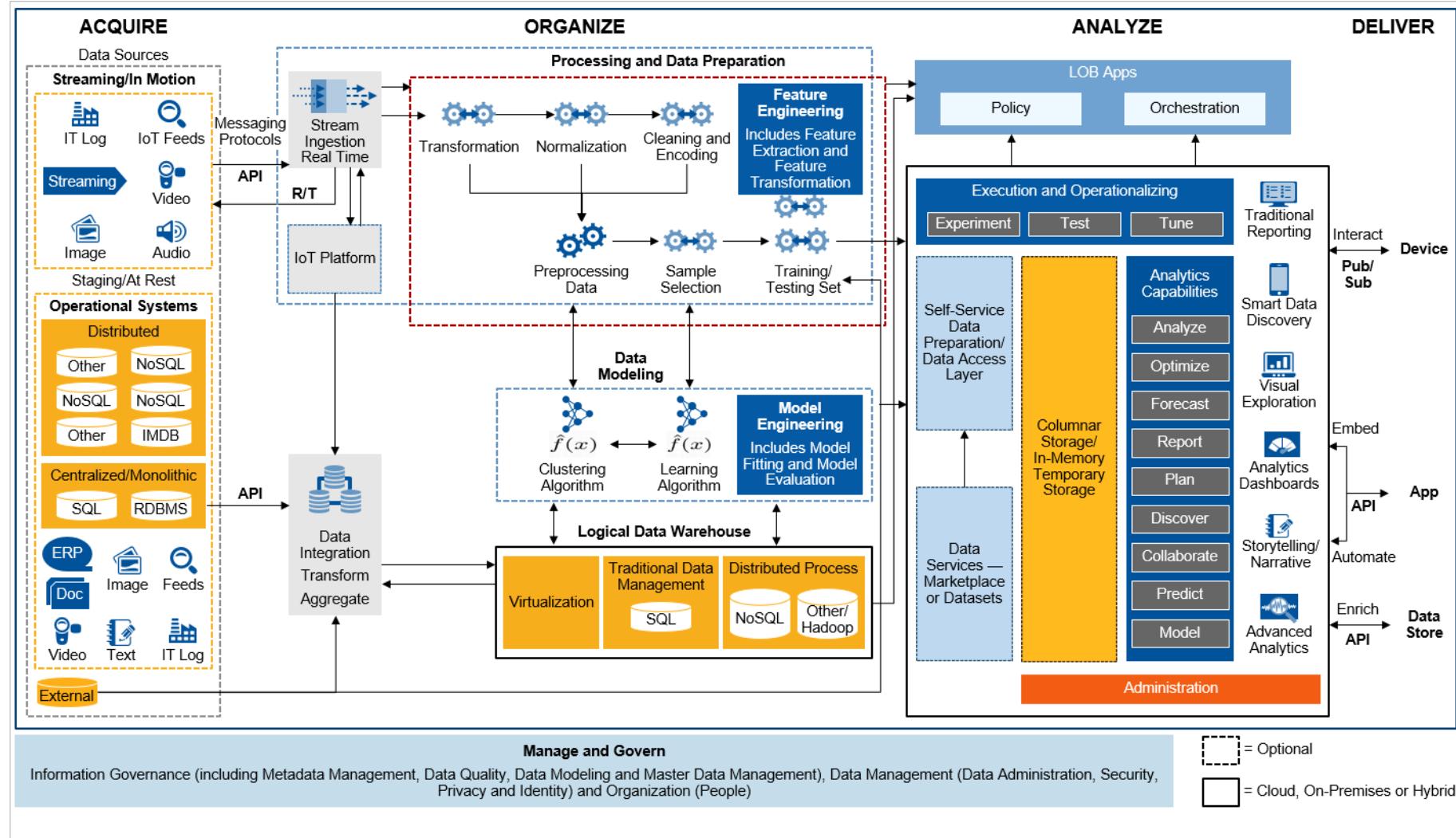


Gartner.

Analytical Architecture



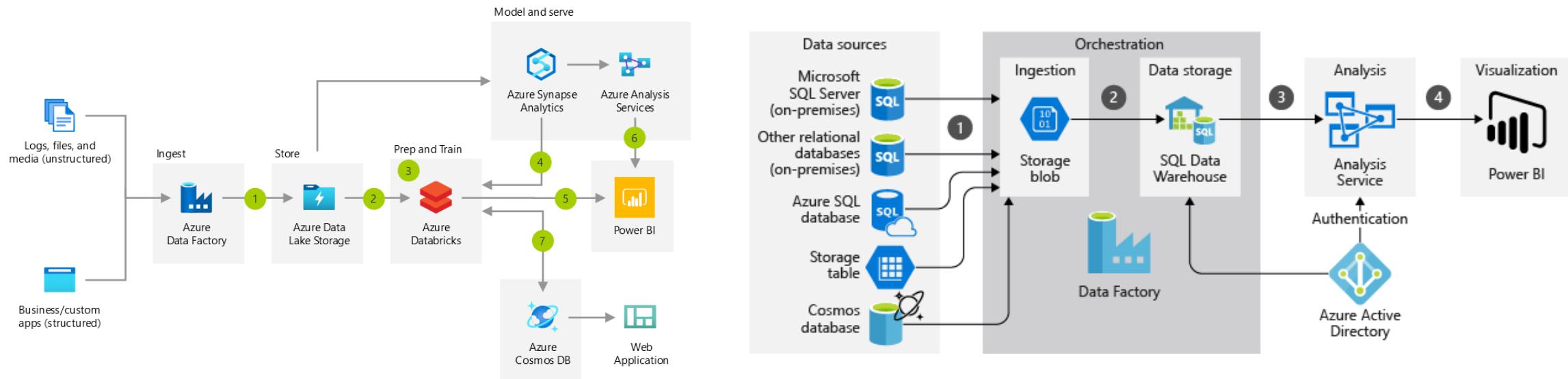
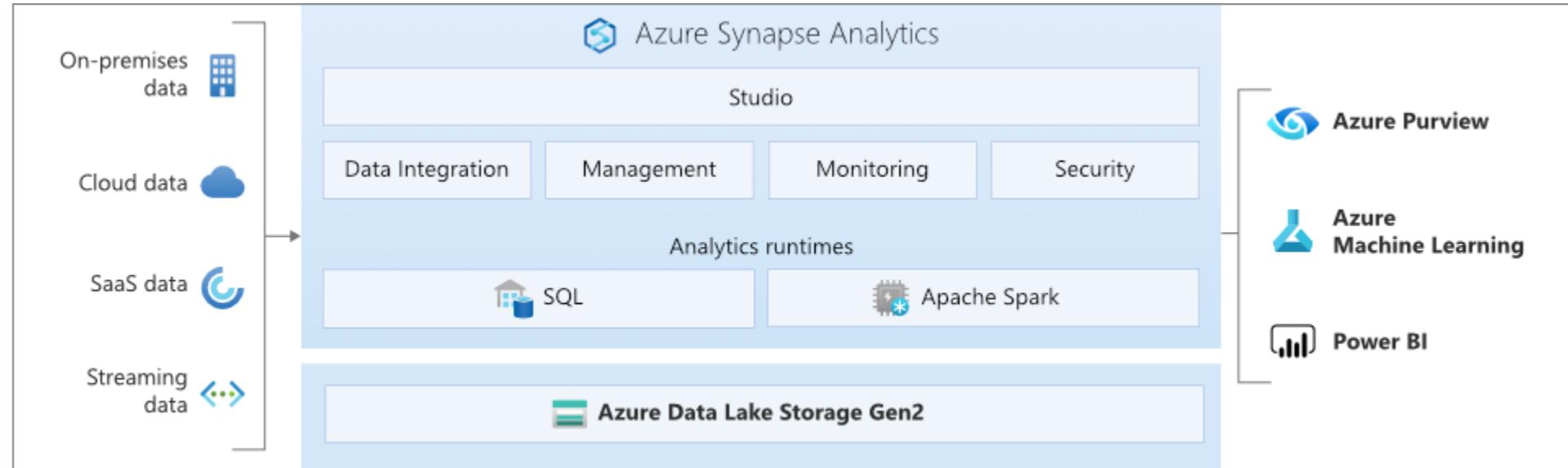
Details to be discussed later in course



Microsoft Azure Synapse Analytics



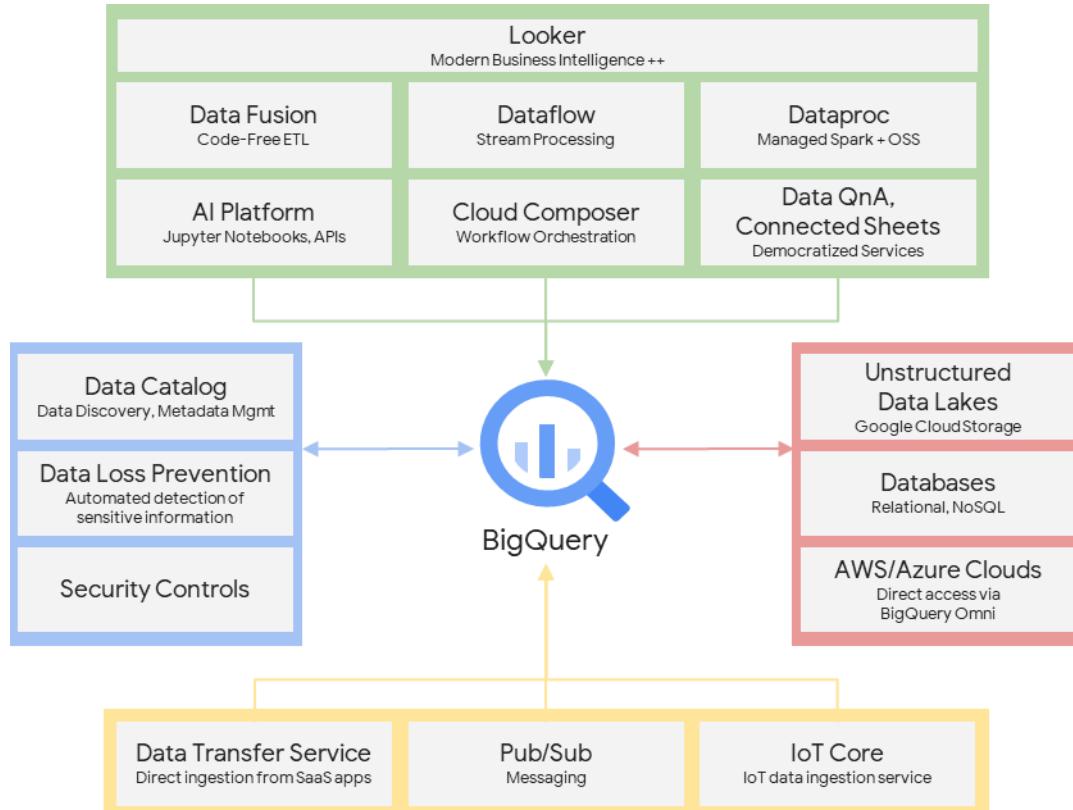
Details to be discussed later in course



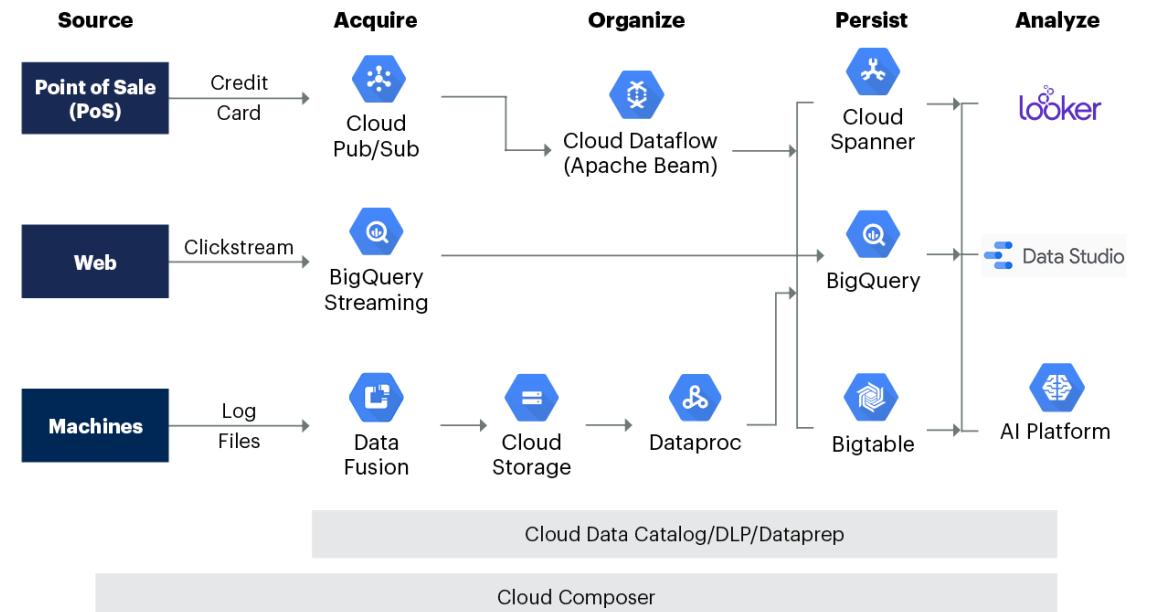
Google BigQuery



Details to be discussed later in course



GCP End-to-End Sample Data and Analytics Architecture

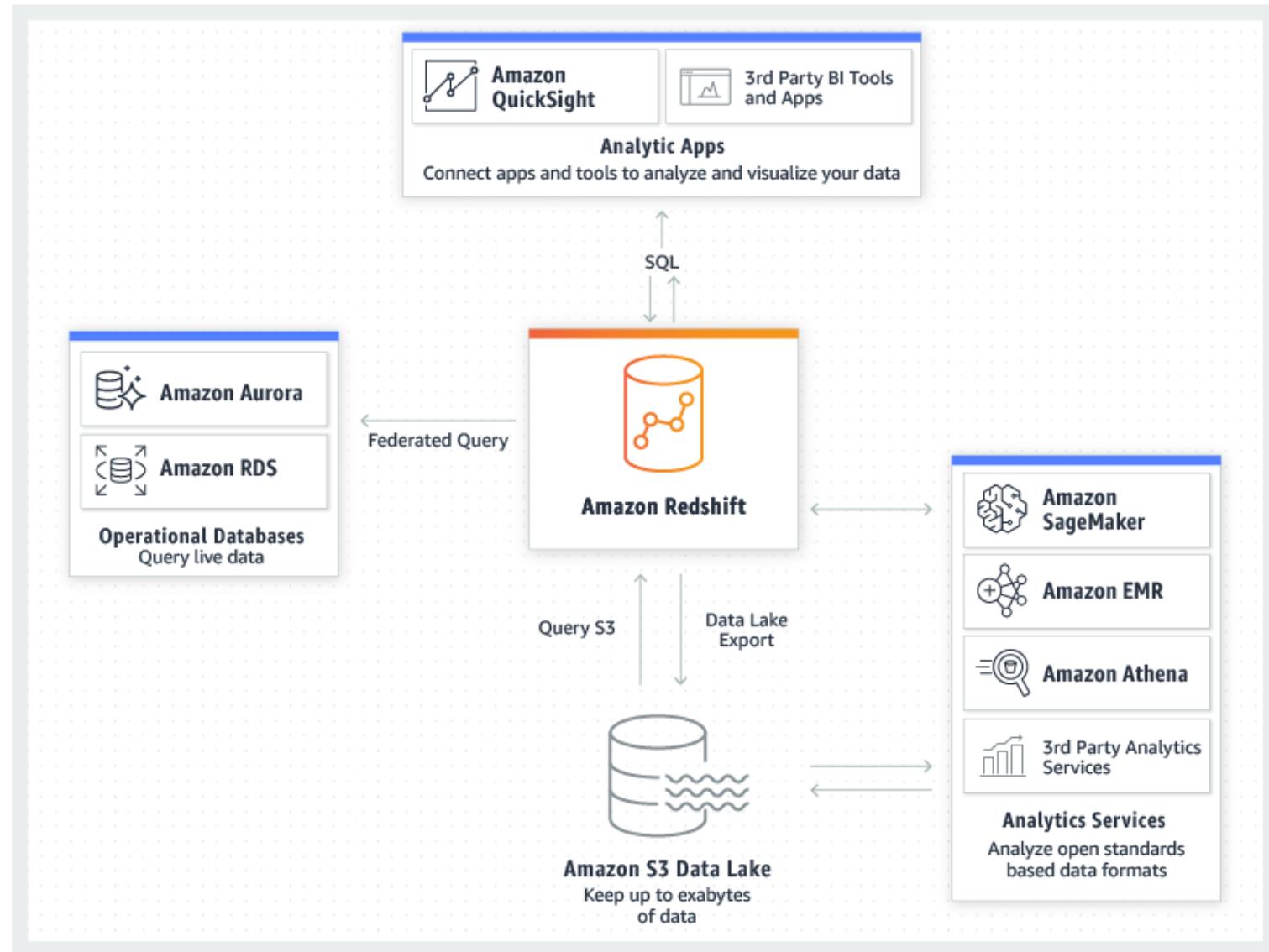


Source: Gartner
451418 C

Amazon Redshift



Details to be discussed later in course



Snowflake Data Cloud



Details to be discussed later in course



Chapter 4: Architecture Framework

- Accidental Architecture
- Terminology
- 4 Architectures
- Information Architecture
- Data Architecture
- **Technical Architectures**
- Product Architectures



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.



MORGAN KAUFMANN

Business Intelligence (BI)

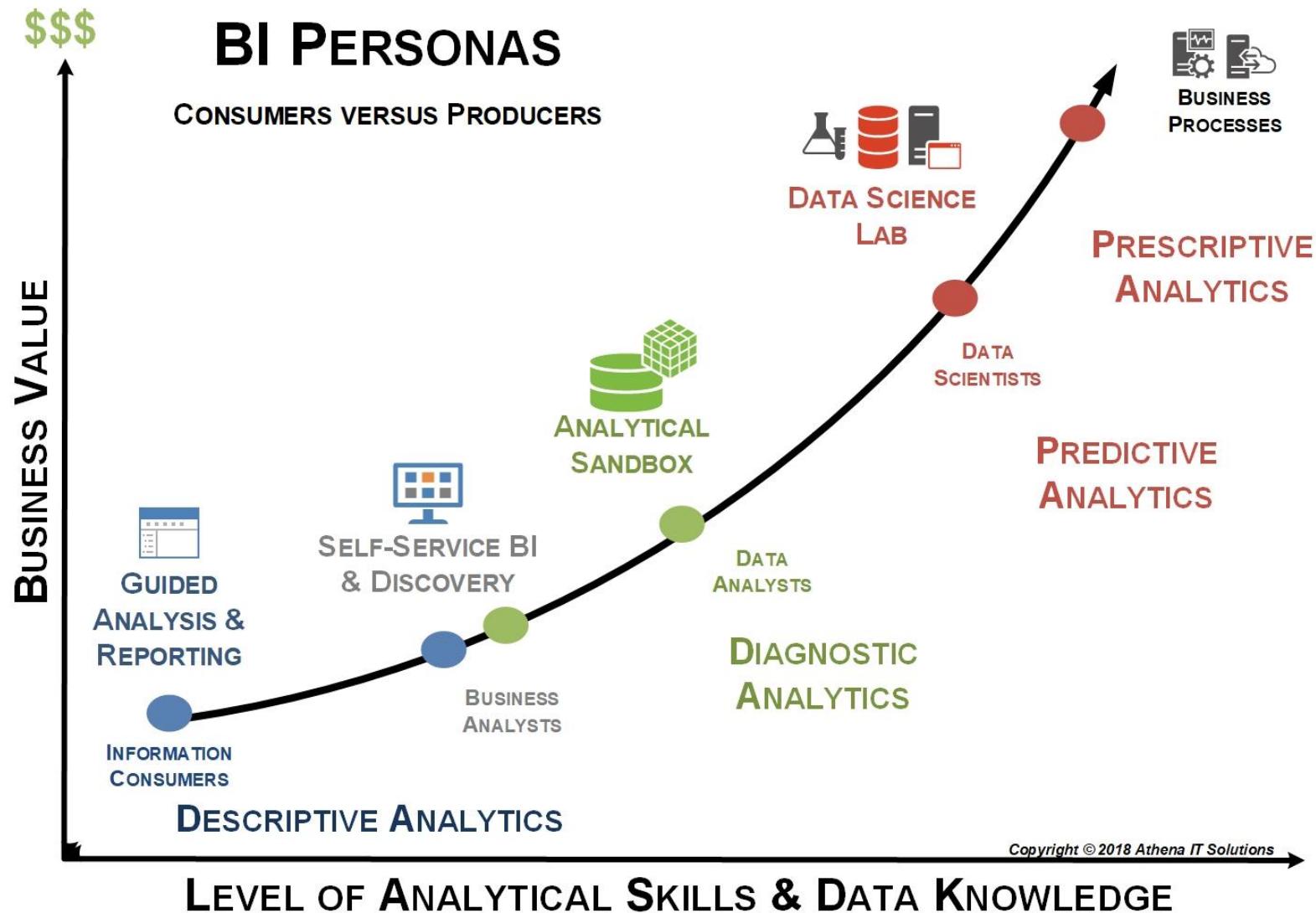
- Enables access and delivery of information to people and processes
- Information presentation layer
- Transform data into business information
 - Goal is to enable business insights
 - Only of business value if actionable
 - ...but both insights & actions depend on people unless model embedded in process (but people developed model initially)



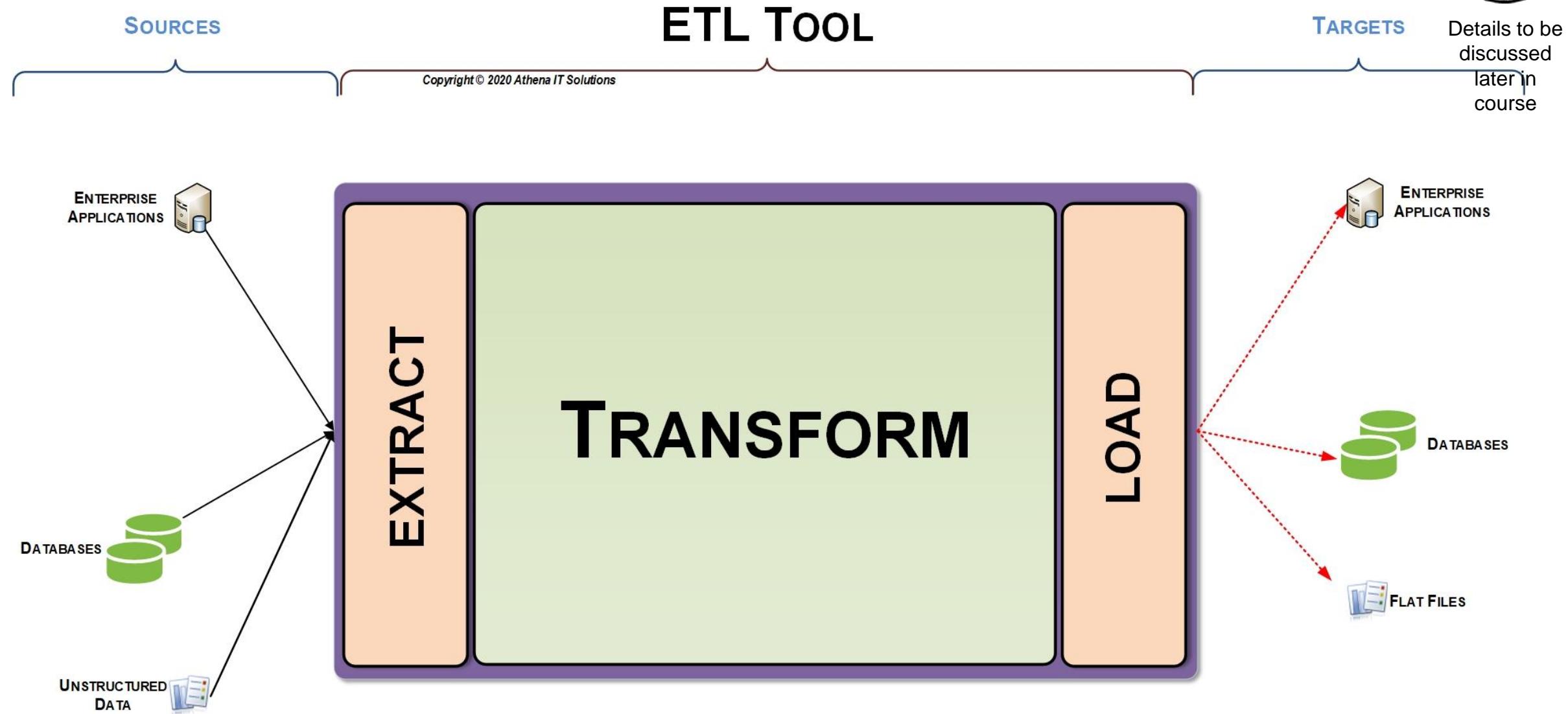
BI Schema – Persona & Analytical Categories



Details to be discussed later in course



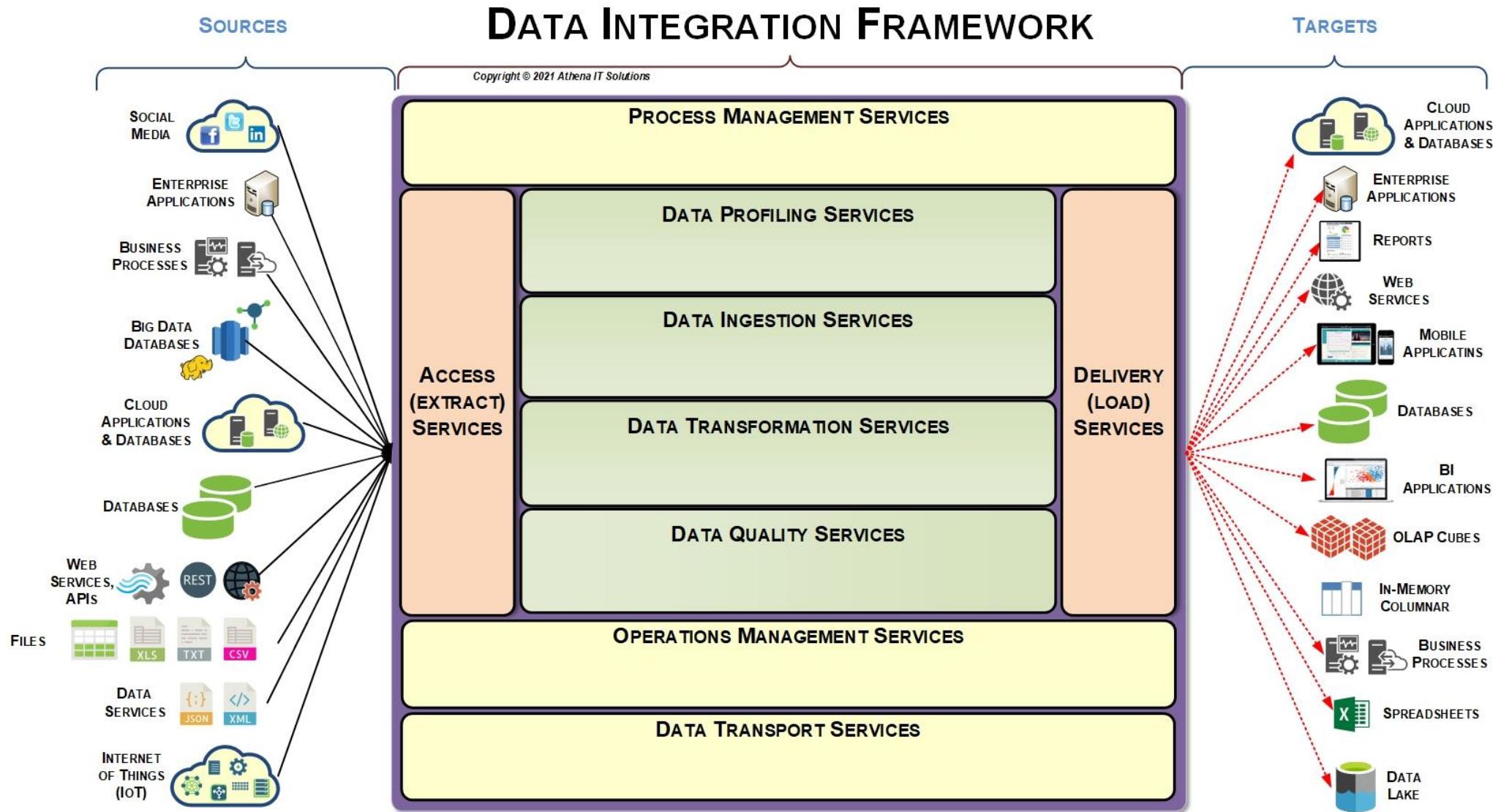
Extract, Transform & Load (ETL) - Beginnings



DI Suite - Services

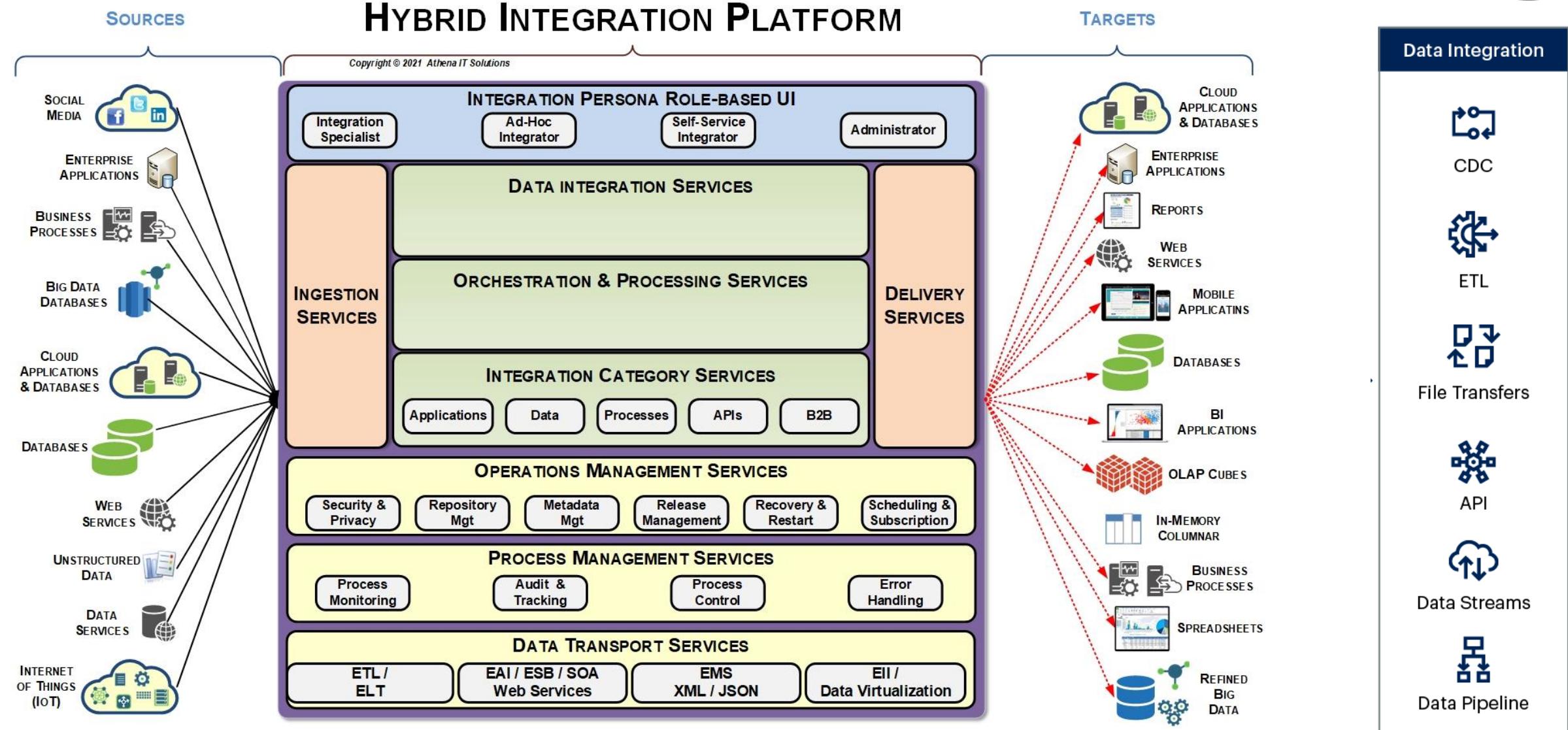


Details to be discussed later in course



DI Suite – Breakdown of Services

Details to be discussed later in course



Chapter 4: Architecture Framework

- Accidental Architecture
- Terminology
- 4 Architectures
- Information Architecture
- Data Architecture
- Technical Architectures
- **Product Architectures**
- Architecture Planning



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics



RICK SHERMAN

FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.

Cloud Database Management Systems

Magic Quadrant for Cloud Database Management Systems

Published 23 November 2020 - ID G00441439 - 56 min read



© Gartner, Inc
Gartner

160678

Source: Forrester Research, Inc. Unauthorized reproduction, citation, or distribution prohibited.



teradata.

databricks

SAP HANA

cloudera

Alibaba Cloud

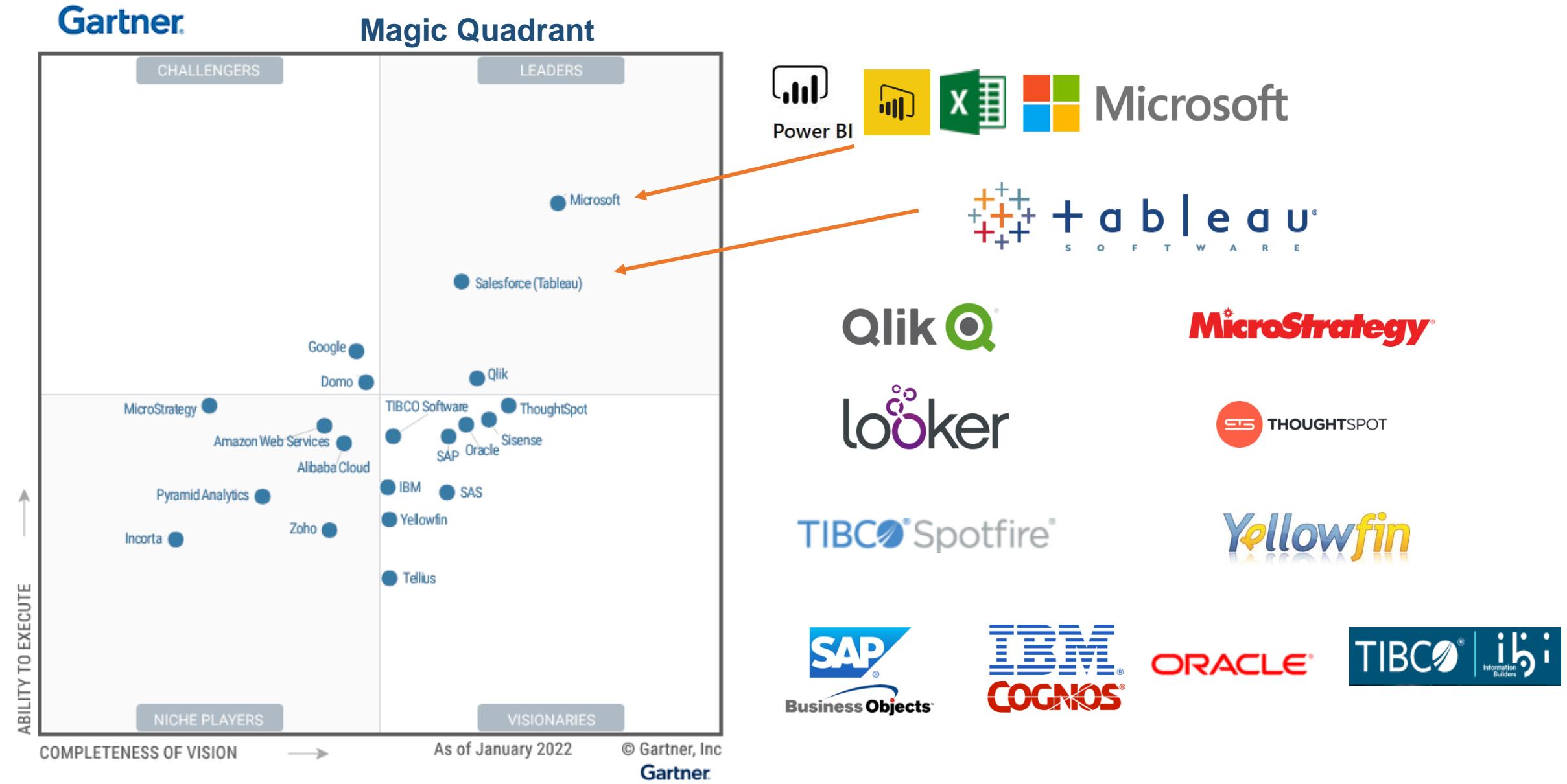
IBM

MarkLogic

InterSystems®
Creative data technology

redis

Analytics and Business Intelligence Platforms



Data Integration



ORACLE®



talend



snapLogic®

Fivetran



Amazon Redshift



Details to be discussed later in course

