

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

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

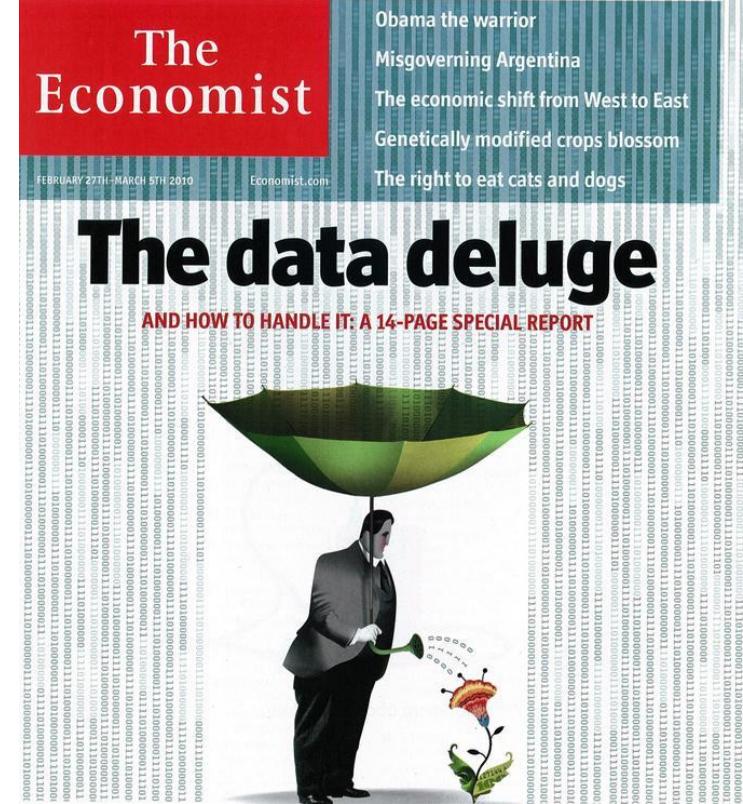
FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.



MORGAN KAUFMANN

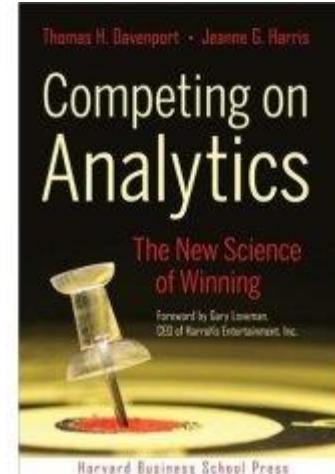
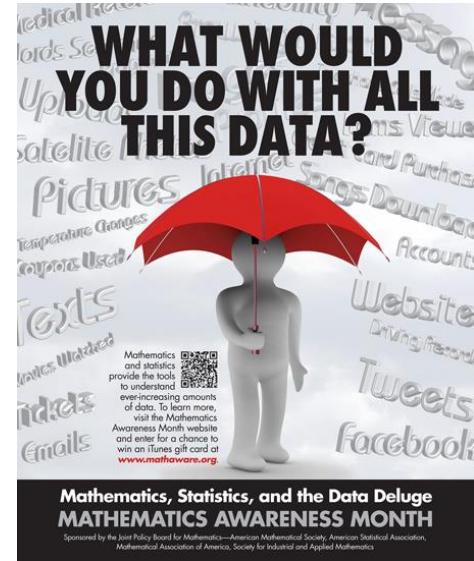
The Data Deluge

- 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 automated (data & applications) & 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



Data Warehousing Market



The global data warehousing market size was valued at \$21.18 billion in 2019, and is projected to reach \$51.18 billion by 2028, growing at a compound annual growth rate (CAGR) of 10.7% from 2020 to 2028.

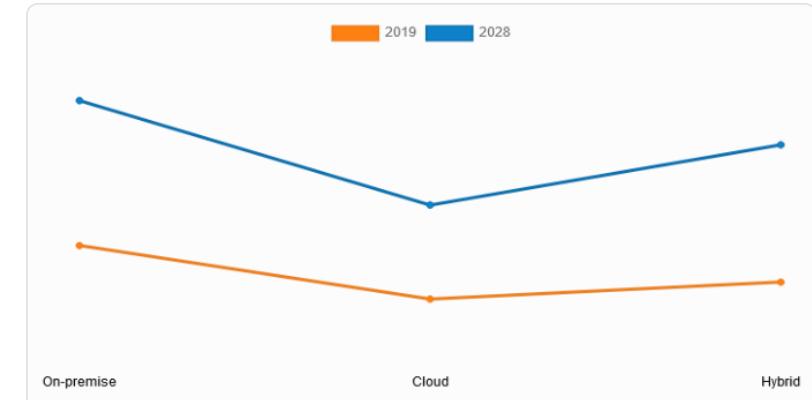
Data Warehousing Market

By Offering Type



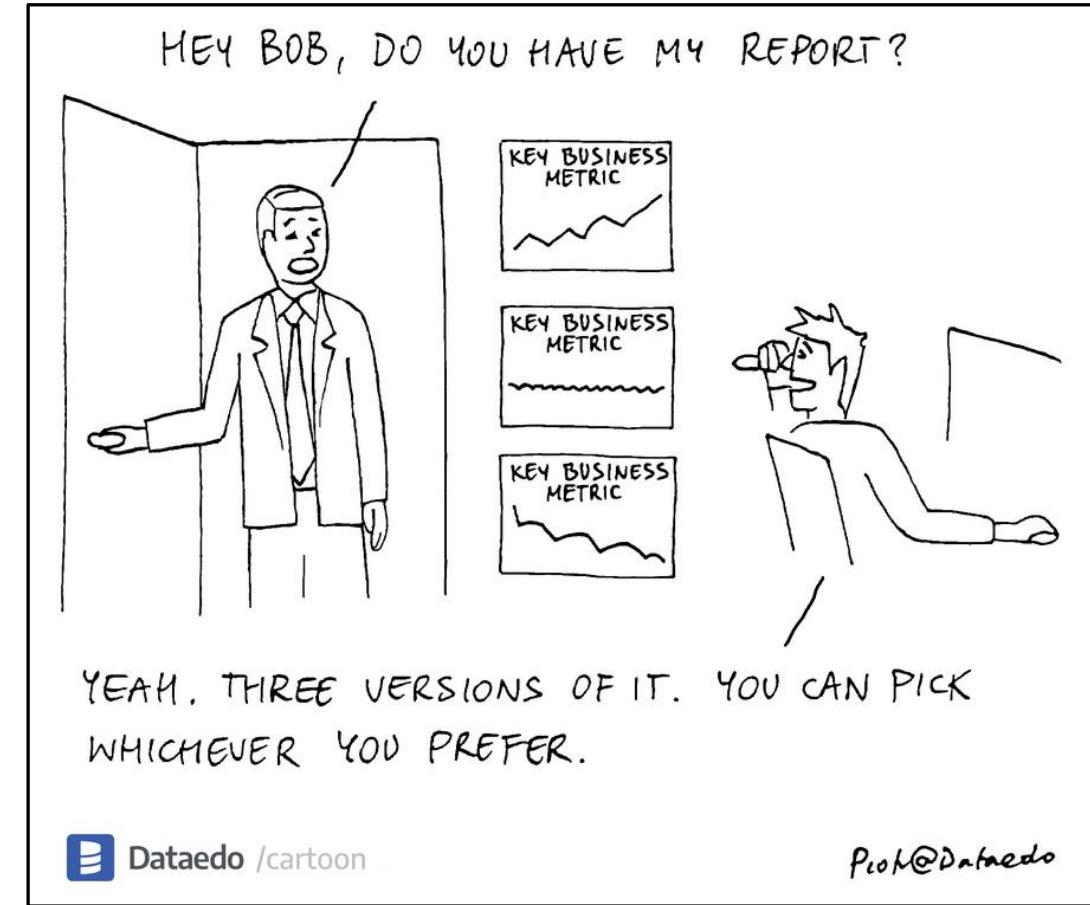
Data Warehousing Market

By Deployment Mode



Accidental Architecture

Information is data in context

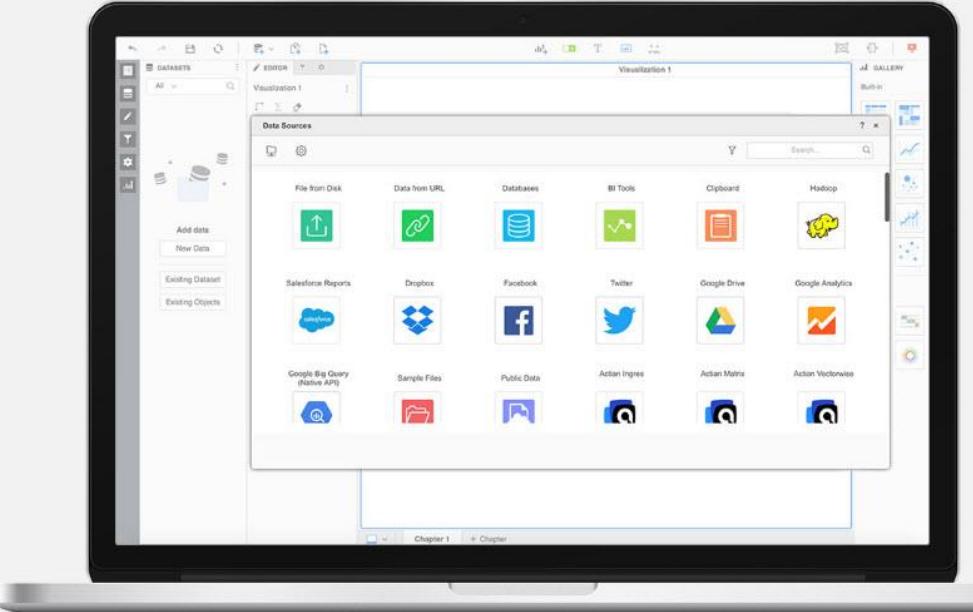


Analytics is NOT just a data connectivity issue

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

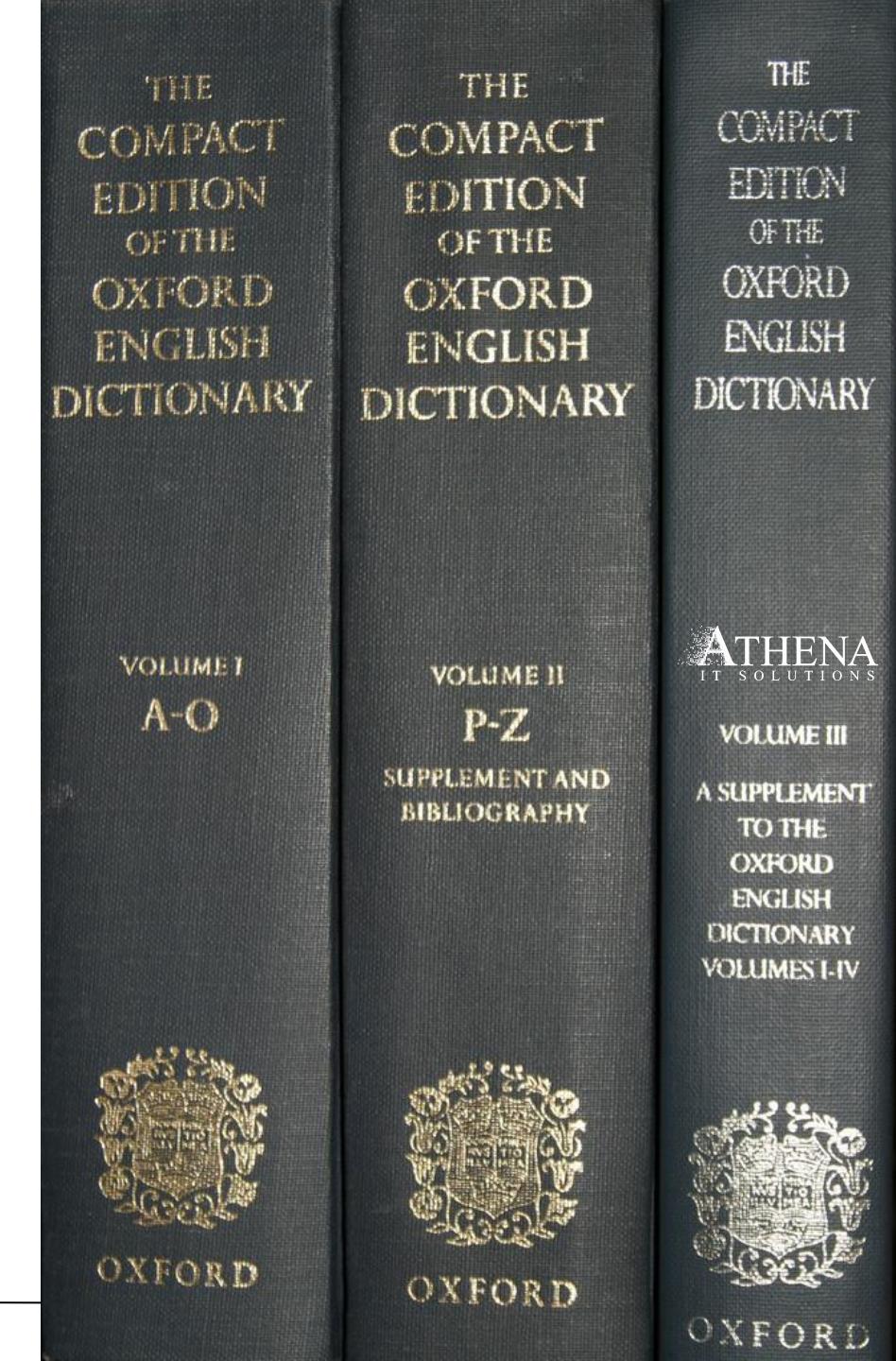
Information is data in context

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



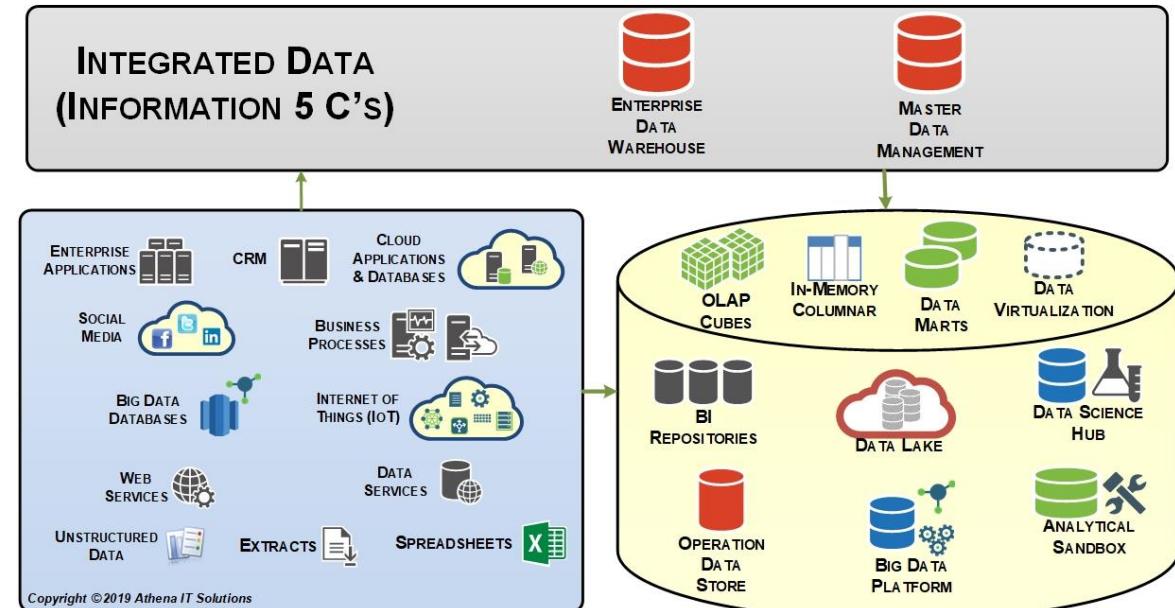
Information 5 C's

- 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



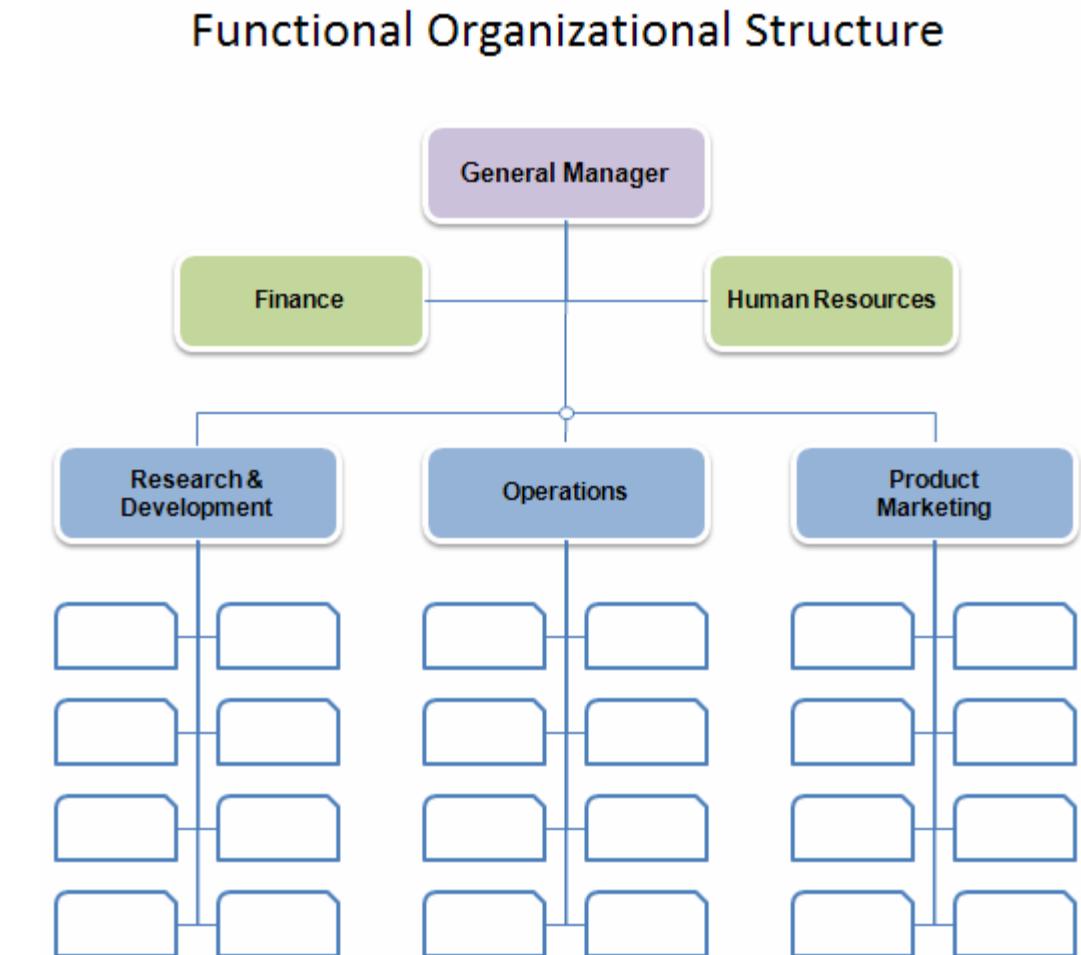


Accidental Architecture

Data silos are rapidly expanding

Business Organizations are Silos

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



Business & Operational Applications

- 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



ORACLE®



Microsoft
Dynamics 365 Sales

Microsoft Dynamics 365
Business Central

sage

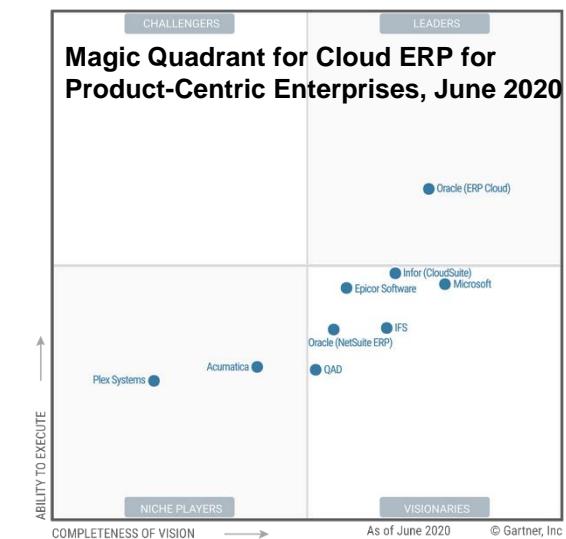
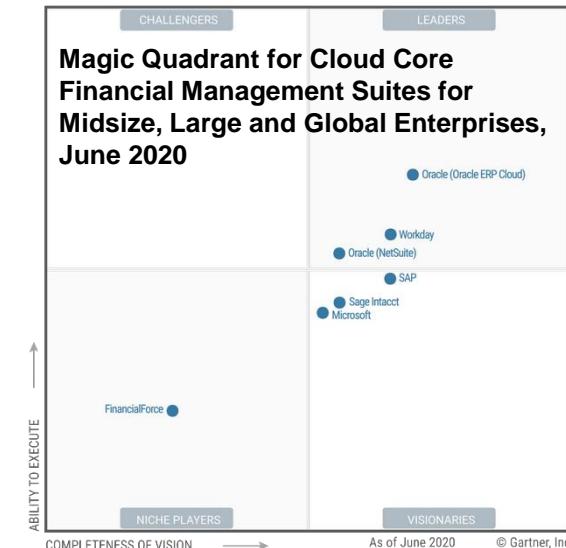


EPICOR.

Acumatica
The Cloud ERP

aptean

ellucian.



Evolution of Business Applications

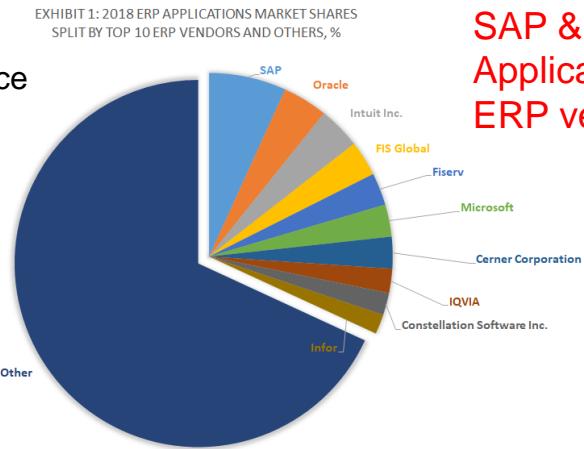
- Custom-built vs off-the-shelf applications
 - 1st wave: Business function applications
 - 2nd wave: ERP (application consolidation)
 - 3rd wave: Cloud applications for above
 - Still many business processes supported by custom applications or by using spreadsheets
- ERP Applications (2nd wave)
 - On-premise initially
 - Private & Hybrid cloud offerings emerge
 - More consolidation (high tech acquisitions)
- Cloud Applications (3rd wave)
 - Mostly automating “new” processes not replacing existing ERP



Top Enterprise Application Vendors (by category)

ERP

(Enterprise Resource Planning)

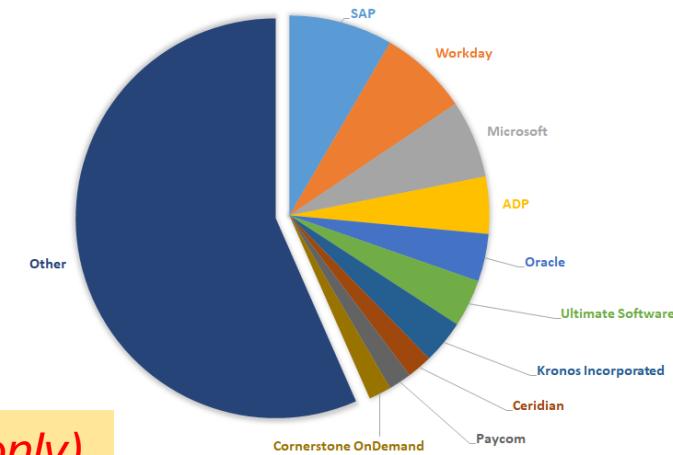


SAP & Oracle
Applications are top
ERP vendors

HCM

(Human Capital Management)

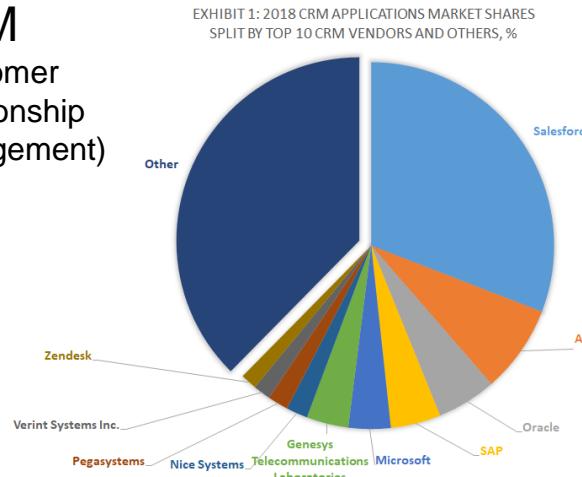
EXHIBIT 1: 2018 HCM APPLICATIONS MARKET SHARES
SPLIT BY TOP 10 HCM VENDORS AND OTHERS, %



fyi: Many different enterprise applications (For illustration purpose only)

CRM

(Customer Relationship Management)

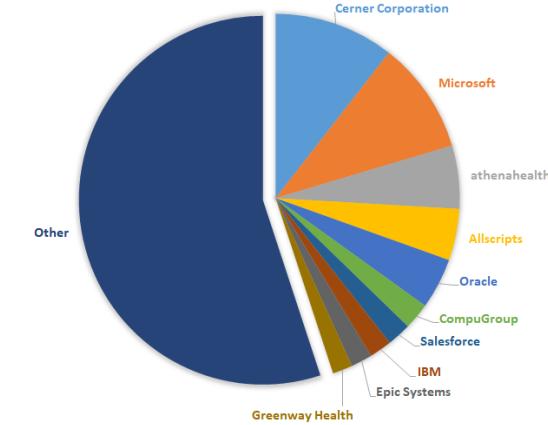


Salesforce & Adobe
are top CRM vendors

SCM

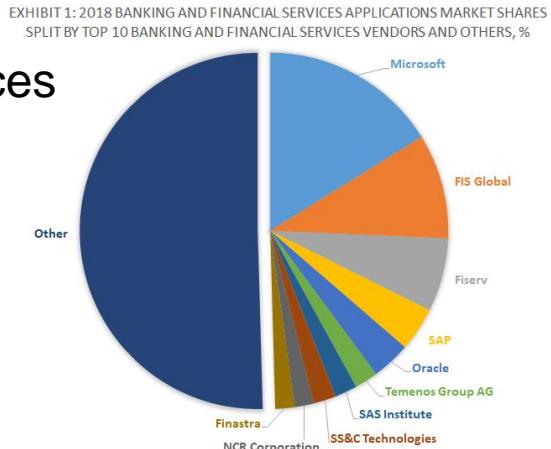
(Supply Chain Management)

EXHIBIT 1: 2018 HEALTHCARE APPLICATIONS MARKET SHARES
SPLIT BY TOP 10 HEALTHCARE VENDORS AND OTHERS, %

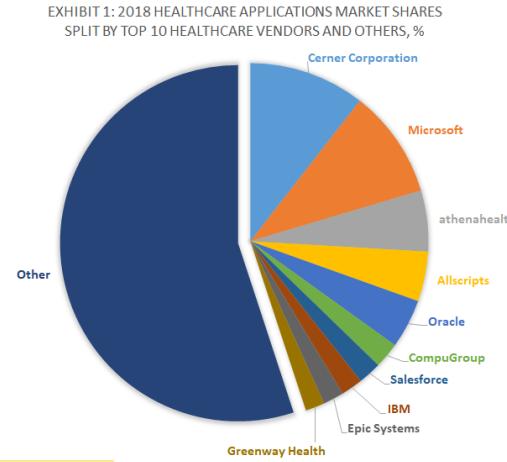


Top Software Vendors by Industry (Sample)

Industry: Financial Services

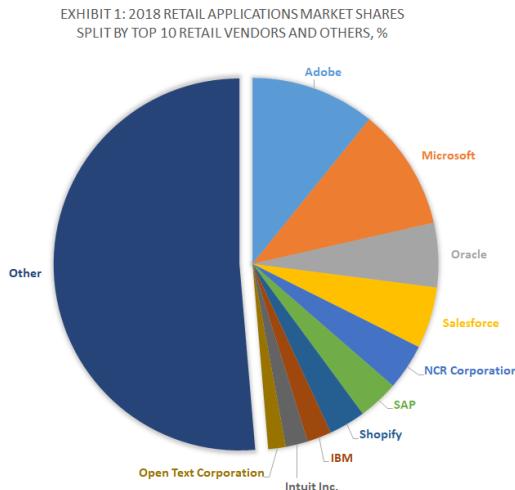


Industry: Healthcare

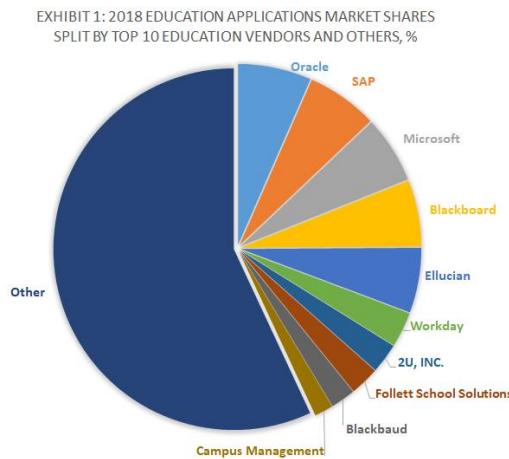


fyi: Many different enterprise applications (For illustration purpose only)

Industry: Retail



Industry: Education



Industry: Education



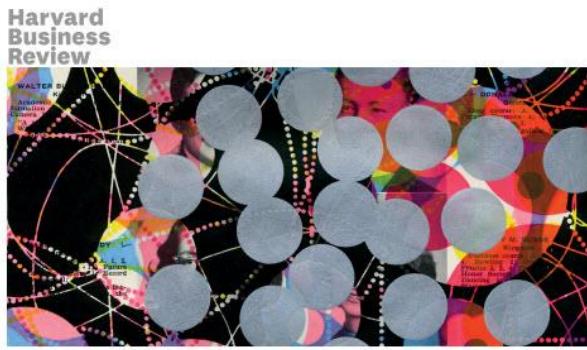
BI & Data Silos

- BI projects driven by departmental 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



Data Scientist: The Sexiest Job of the 21st Century

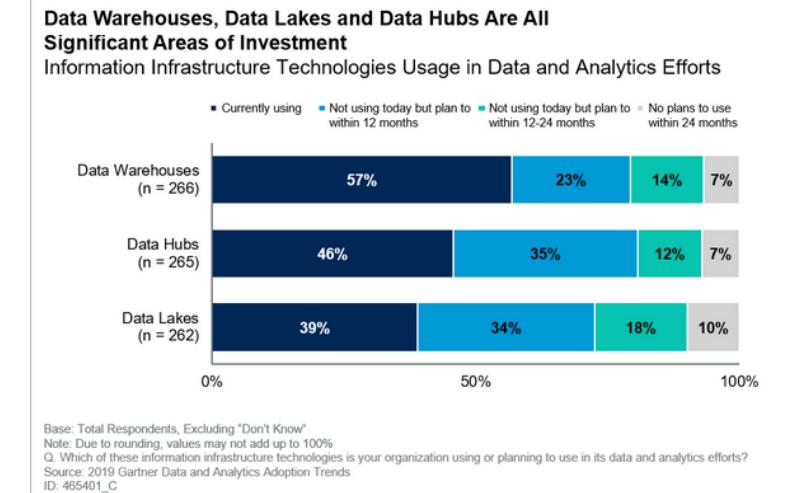
by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE



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”

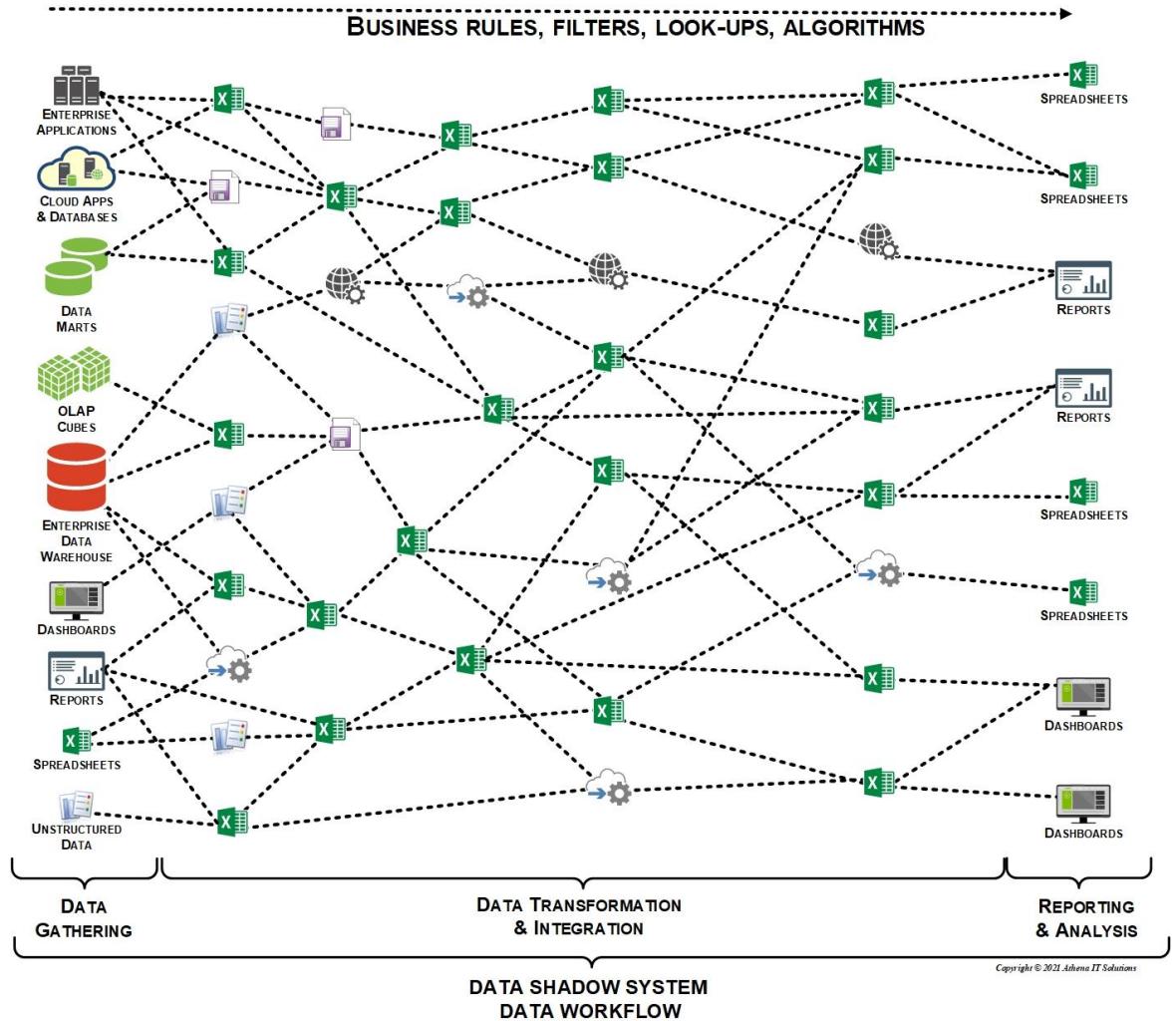
- “50%-80% of all BI/analytics apps are still done in spreadsheets”
(Gartner Research, Boris Evelson)
- Primary tool used by business to present & analyze data
- Only “BI” tool with pervasive business use
- Data superglue
- Often “masks” ineffective data architecture or mismatched BI tools



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

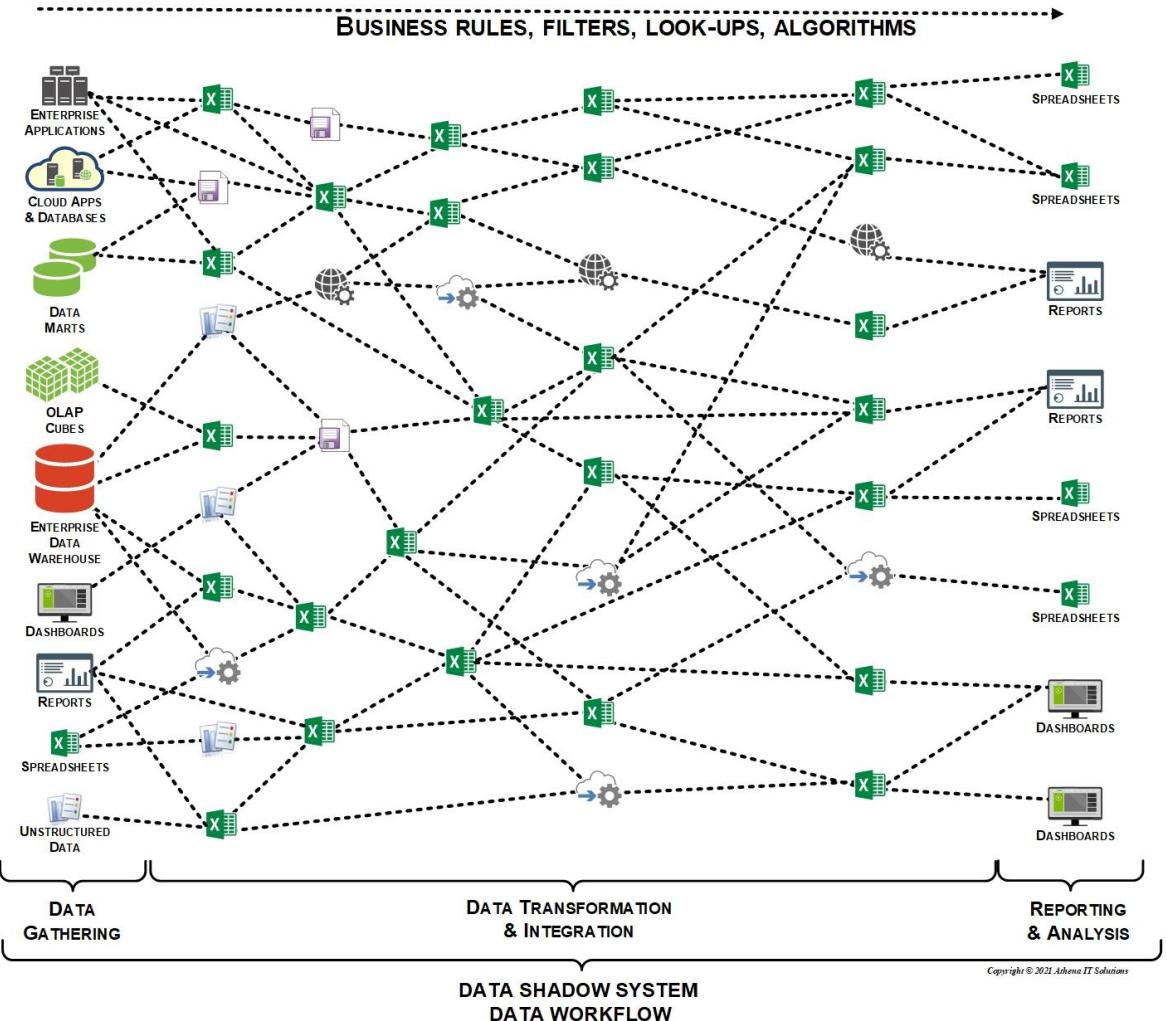


Next Generation of Data Shadow Systems



The new spreadmarts built with:

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

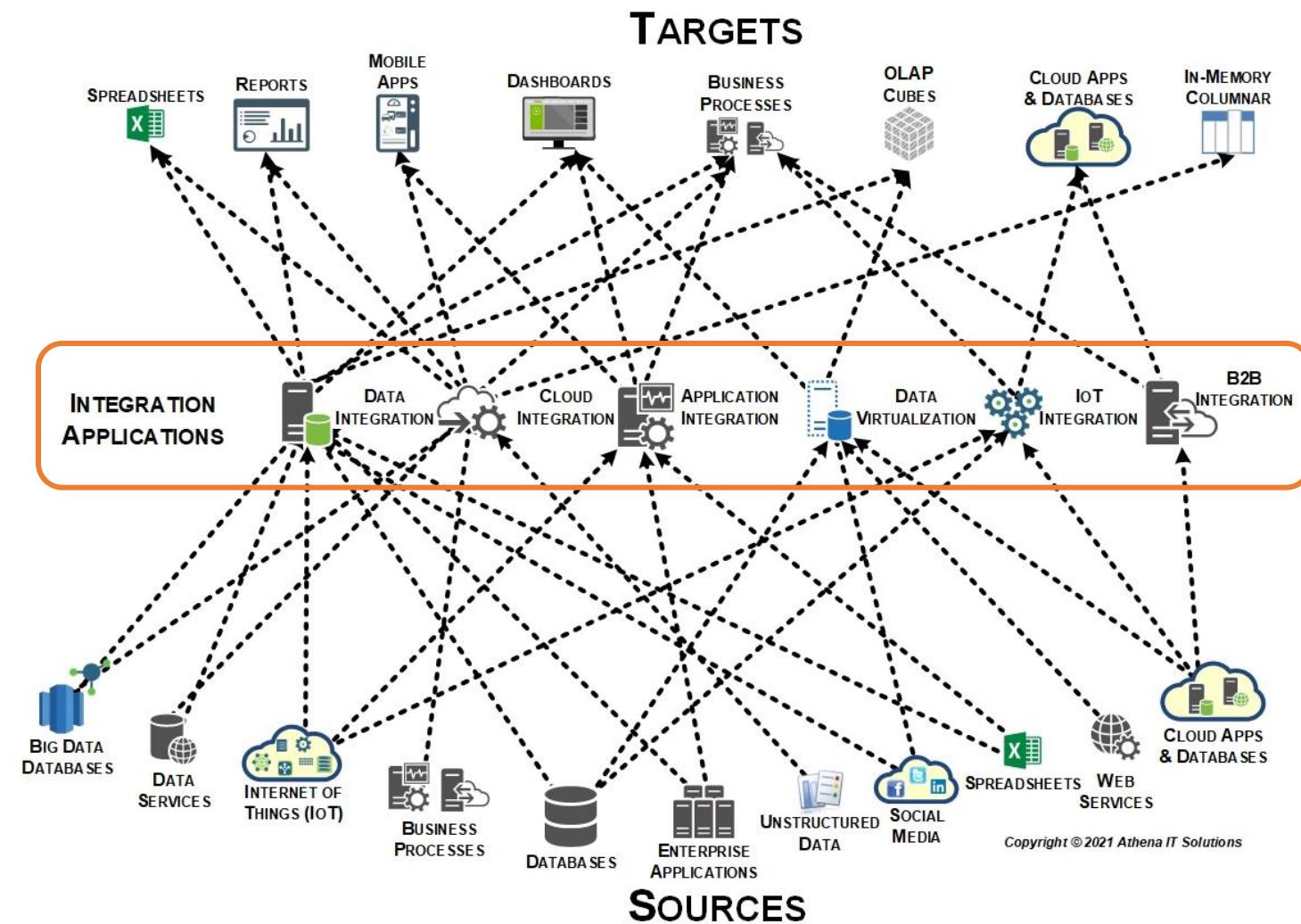


Current State: Integration

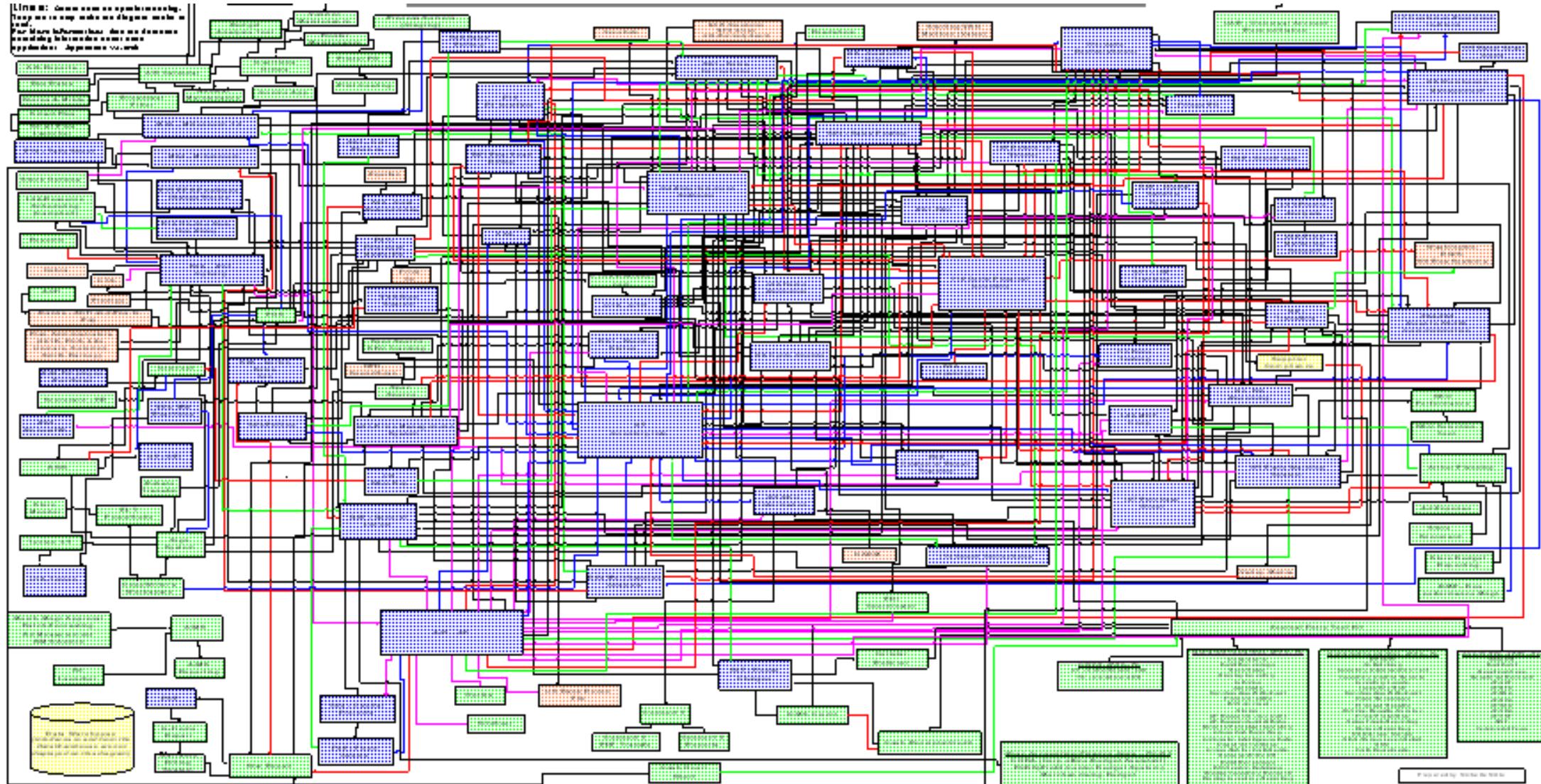
- Primary integration toolsets:
 - Application integration
 - Data integration
- Majority of tool-based work is batch or bulk
- **Much of data integration is custom-coded**
 - SQL
 - Programming language du jour
 - Spreadsheets
- Most data involved in data integration efforts are siloed & not usable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



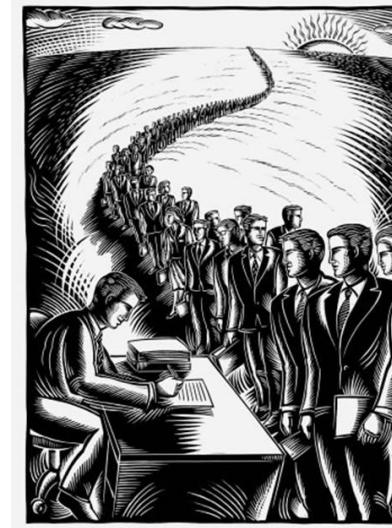
Current State: Integration



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

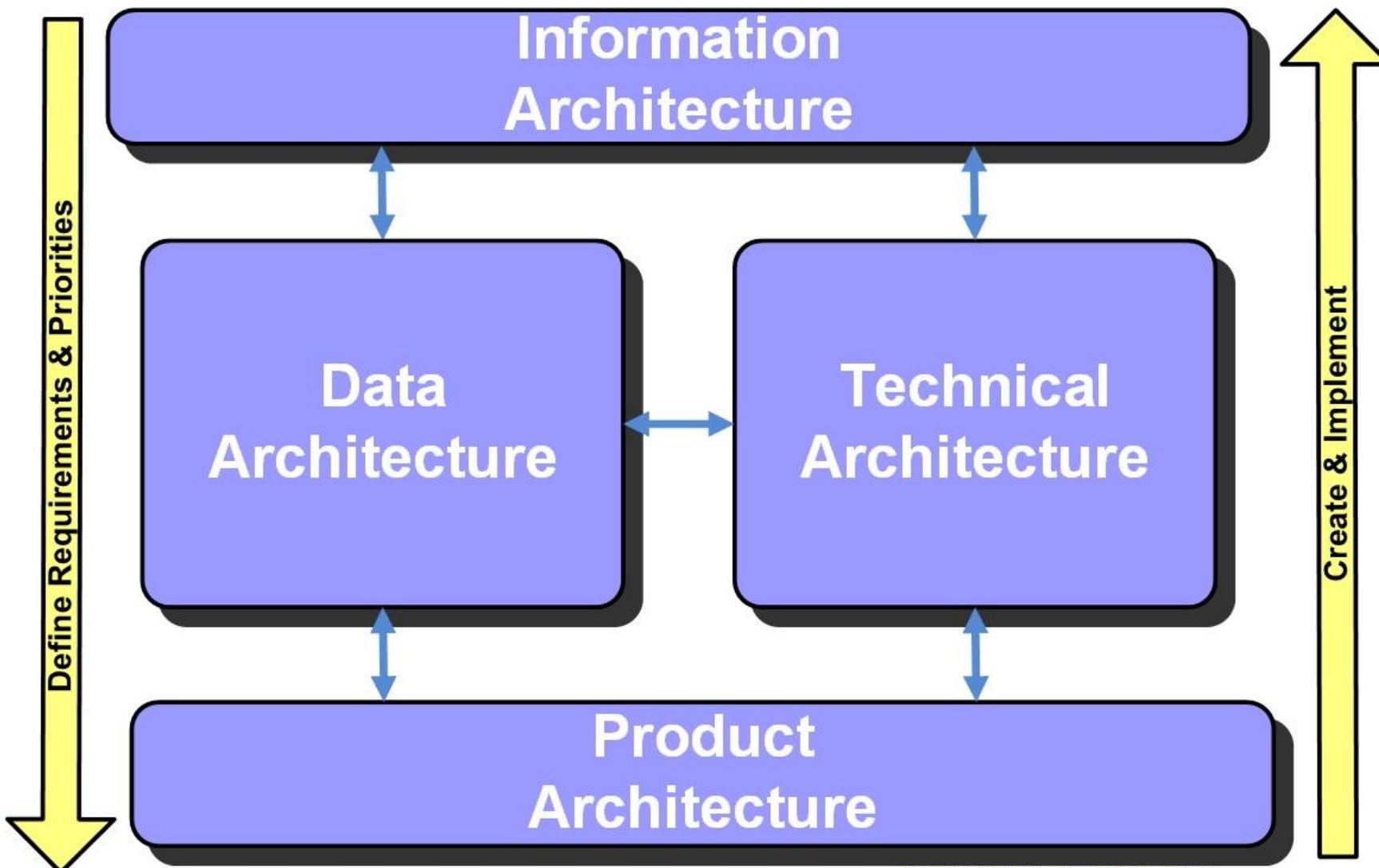
Architectural Blueprints

Analytical Data Architecture

- An enterprise's analytical data architecture (ADA) needs to implement the integration & analytical requirements of an information architecture
- ADA includes:
 - Data Schemas & Models
 - Data Integration & Workflow
 - Policies, Processes & Standards
 - Organization, People, Skills & Politics
 - Technology Architecture
 - Product Architecture

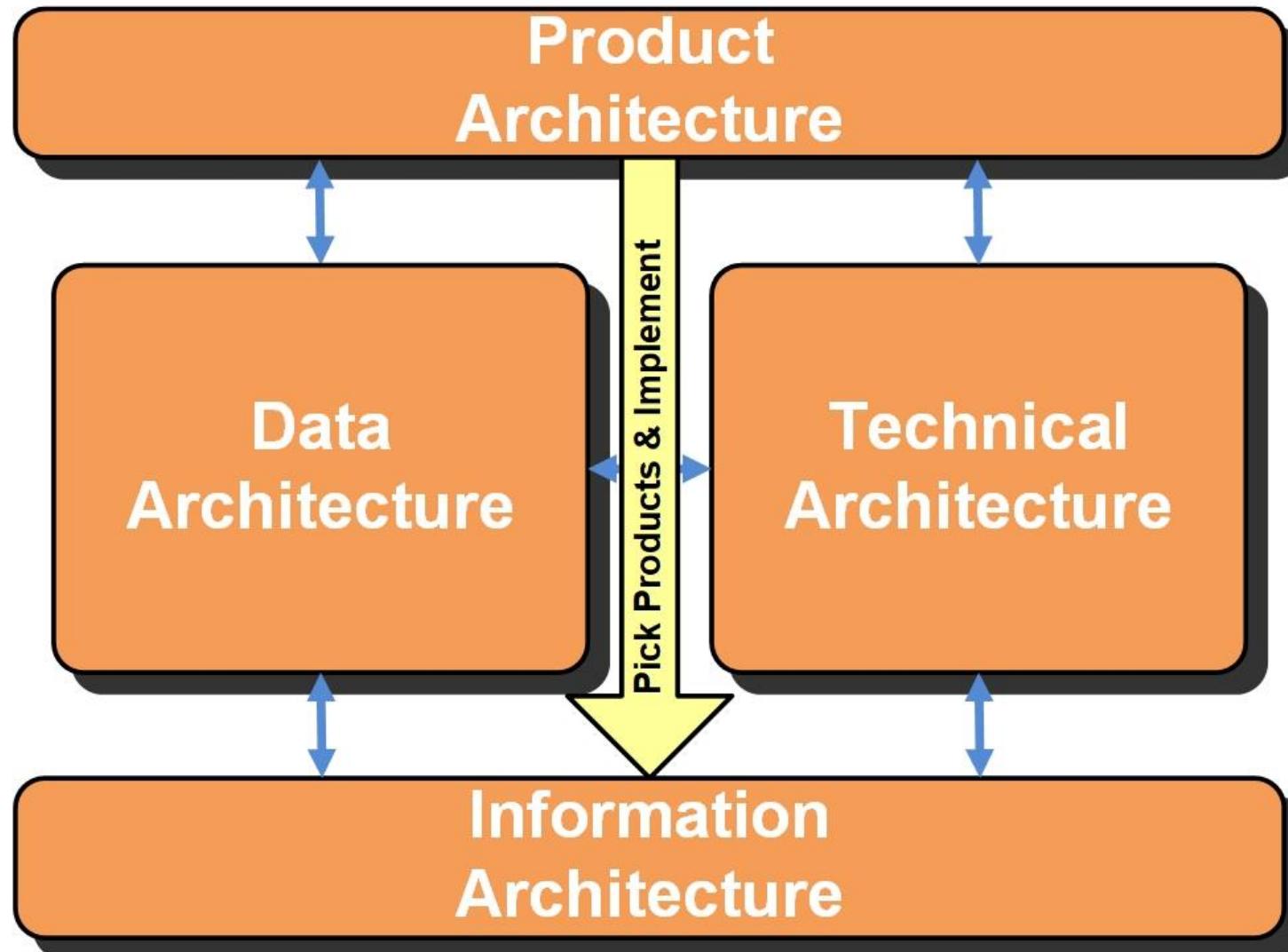


The Four Architectures



Copyright © 2020 Athena IT Solutions

MISTAKE: Architectures – Starting with Product



Copyright © 2020 Athena IT Solutions

MISTAKE: Technology Trap

- Products are our comfort zone
 - Technologists
 - Our society
- Product hype
 - Vendors
 - Analysts & pundits
 - Paid endorsements not transparent
- Case Studies (POCs or Silos?)
- Hype Cycle (Everyone is doing it)



MISTAKES: Technology Trap

- Each new generation of BI tool is:
 - Easier
 - Faster
 - Requires less technical knowledge
 - Addresses last BI generation's problems
- Each generation has same marketing life cycle
- Assuming product can overcome:
 - People, Politics & Processes (3 P's)
 - Accidental architecture



MISTAKES: The One...Trap

- One customer persona fits all
- One use case fits all
- One size (tool) fits all
- One neck (vendor) to choke

Customer Segmentation is the norm
in most industries but not in IT



MISTAKES: Cultural Trap

- People continue to work in business silos
- People's comfort zone
- People need to get things done
- People love/hate spreadsheets



Our Past Mistakes

- Big Bang, Waterfall Projects
- Centralized, 3NF DW
- Operational Focus
- Parts vs Whole
- Not Understanding...
 - Data
 - Data Uses
 - Personas
 - DB purpose & design
 - Schema purpose & design



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



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

July 23, 2019 by Brian T. O'Neill

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



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,

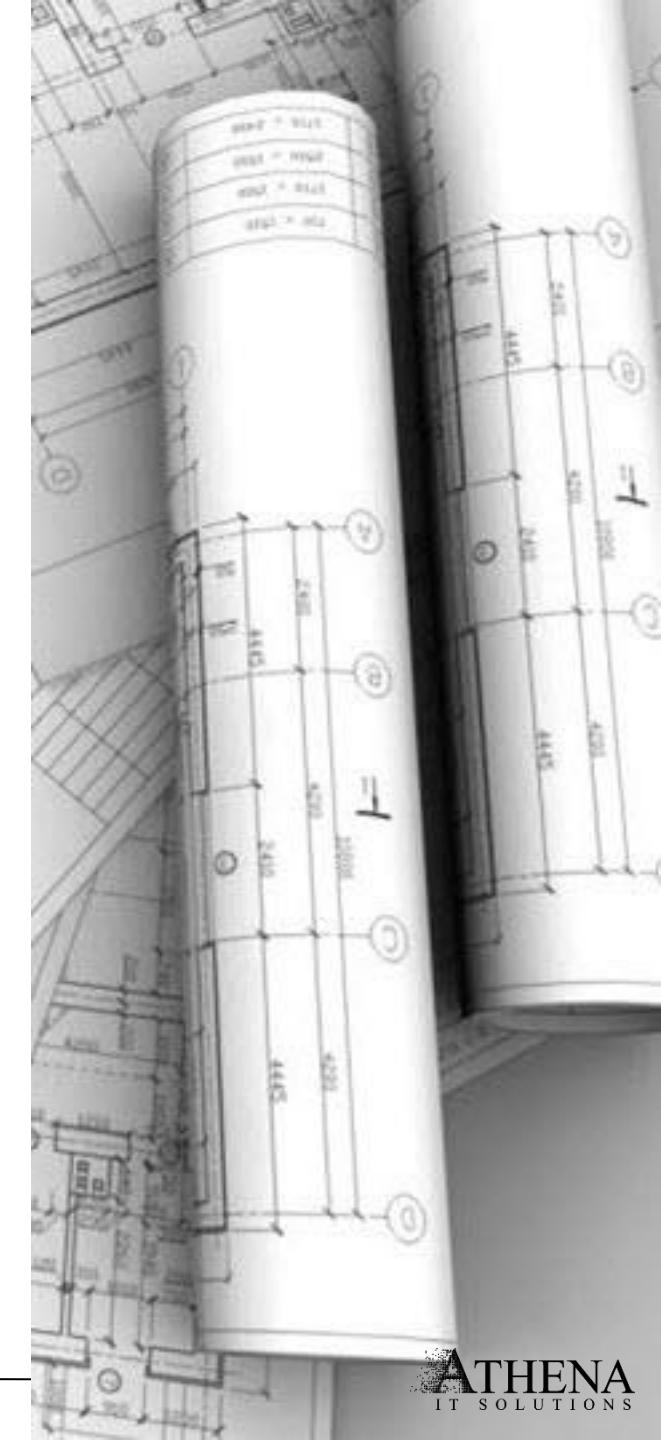
data scientists, subject matter experts and individuals with other special skills and knowledge. Moreover, this talent is scarce and costly. In fact, only a small number of companies have succeeded in building an experienced data science practice. And, while building this team takes time and resources,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

Funny the way it is

- Data pipeline, NoSQL, Hadoop hype – ingest data, no model and then just analyze it
 - That's fine for social media (but language is the schema) but misses
 - (a) the basic fact that data has a schema based on its business context or data processing
 - (b) context changes based on usage by processes, communities (groups) and people
 - Not cost or resource efficient to “lose” schema and then have Python programmers try to define it each time the data is used; And not have any data lineage trail & likely no documentation
 - Most data is NOT going into a data lake, NoSQL databases or Hadoop without a schema or at all
 - ACID versus BASE, would you want your financial systems and your paycheck in a BASE system
 - Data is scattered across many disparate systems that increasing that have different and sometimes conflicting definitions and schemas
- Query vs Integrate
 - Marketing message & hype for decades has been you do not need to integrate data but query it directly or now the phrase is to “use it in place”
 - Both query and integrate are needed
 - Ignoring integration means it happens haphazardly & accidentally

ADA Principles

- Create a framework & design blueprints for purpose
- Design & implement ADA components based on **need, use & skills** - market segmentation rules
- Data & information architectures drive technologies & products
- Create once & use many times
- Realign roles for data integration & BI
- Data & analytical governance is shared responsibility



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

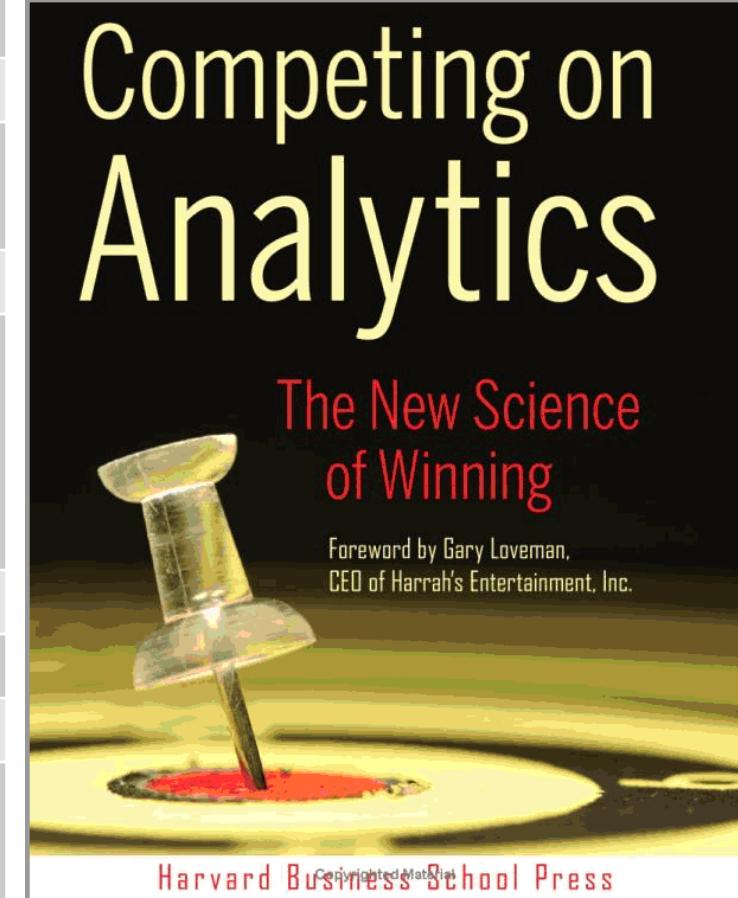
FOREWORD BY CLAUDIA IMHOFF
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.



MORGAN KAUFMANN

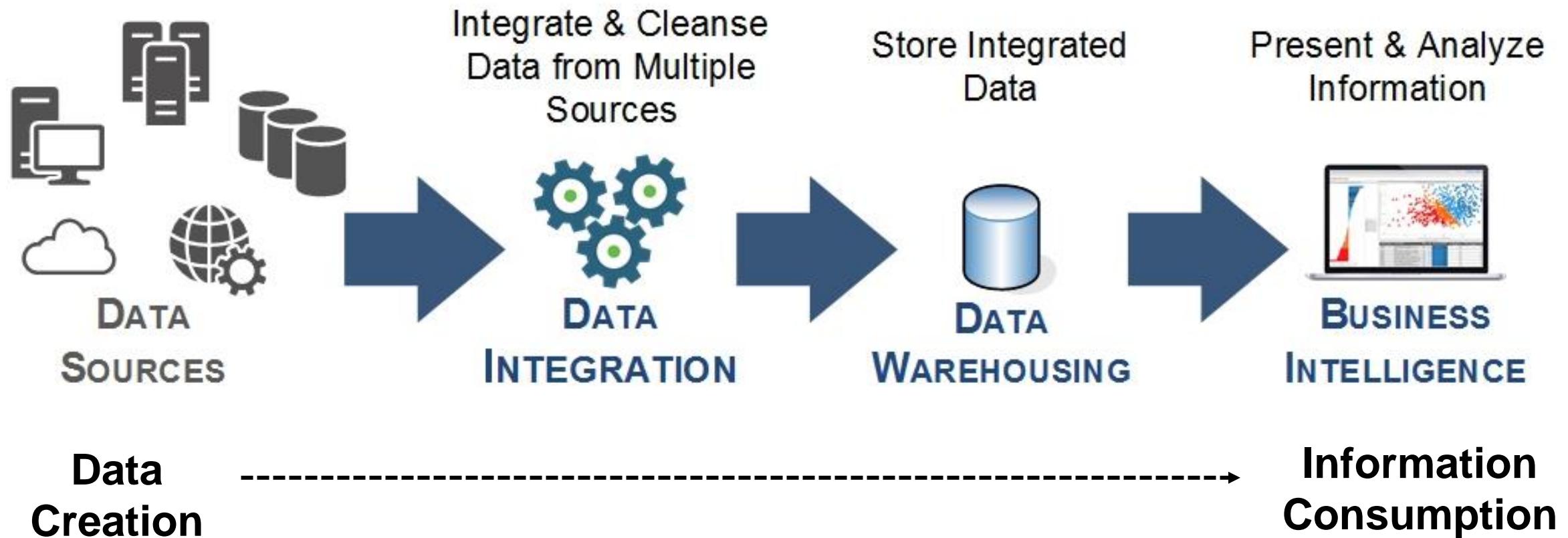
Information Architecture

| Design | Questions |
|--------|--|
| What | <p>What business processes or functions</p> <p>What business decisions or outcomes</p> <p>What are business metrics & Key Performance Indicators (KPIs)</p> <p>What types of analytics</p> |
| Who | <p>Who will have access</p> <ul style="list-style-type: none">Employees, customers, prospects, suppliers or other stakeholdersPrivacy & security |
| Where | <p>Where are the data sources</p> <p>Where it will be integrated</p> <p>Where it will be consumed in analytical applications</p> |
| Why | <p>Why will the BI solution(s) will be built</p> <ul style="list-style-type: none">Business objectivesBusiness & technical requirements |



Data Workflow

- Historical, and many times, current view of Data Warehousing



Why not just access the data where it is

| Operational Data | Analytical Data |
|---|---|
| Structured for efficiently processing & managing business processes, transactions & interactions | Structured for business people to analyze & understand |
| Record the business events or transactions | Need to view data in past, present & future context: AS-WAS, AS-IS, & AS-WILL-BE |
| Keeps latest dimensional context | Tracks changes in dimensions such as products, customers, businesses, geopolitical, account structures & organizational hierarchies |
| Typically contains a relatively short time span | Maintains historical data – supports regulatory requirements, trending |
| Data is spread out over many source systems (business & data silos), making it hard to bring together and analyze | Very enterprise, no matter how large or small, must perform data integration |

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.

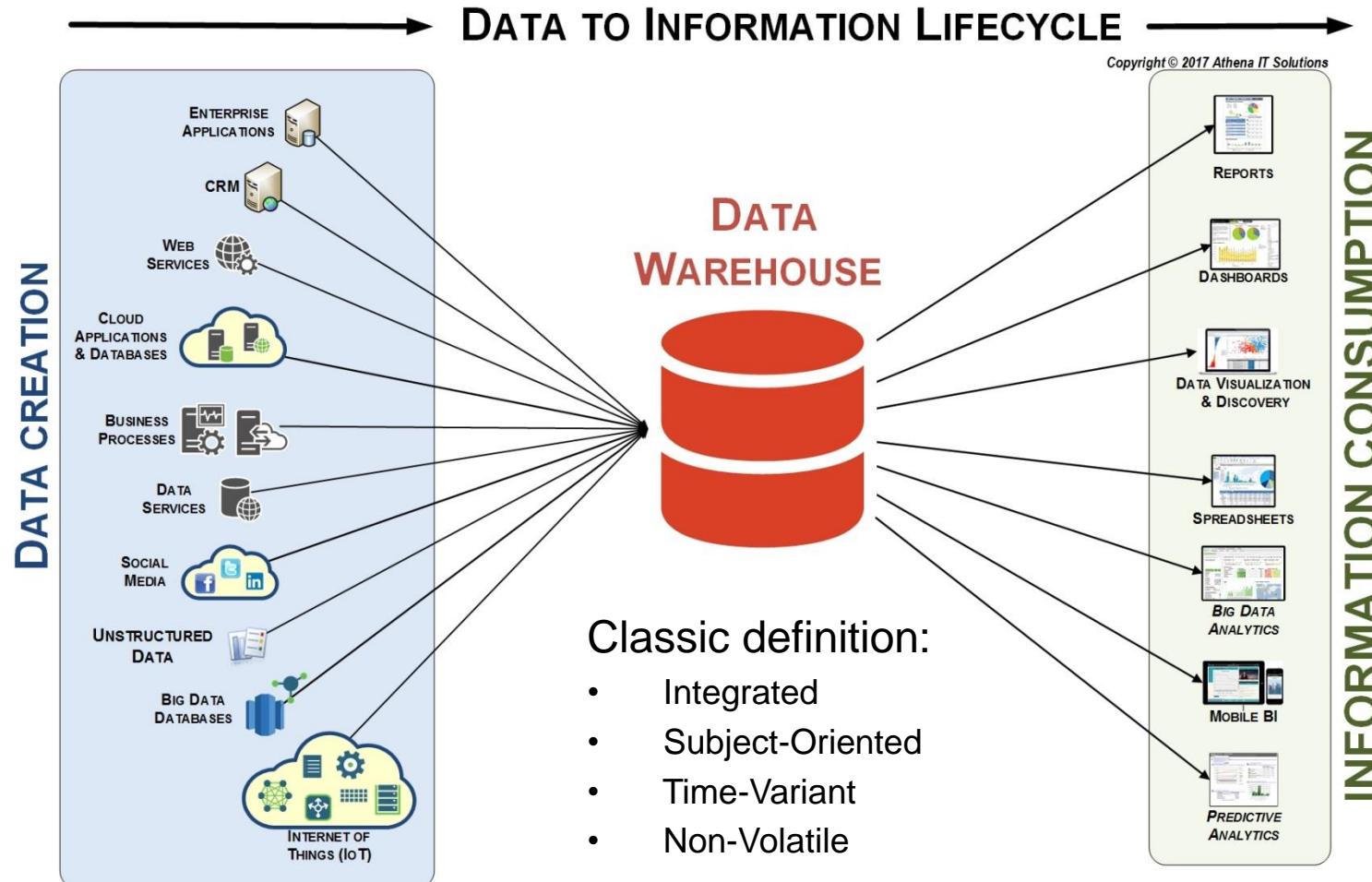


MORGAN KAUFMANN

Data Warehouse (DW)

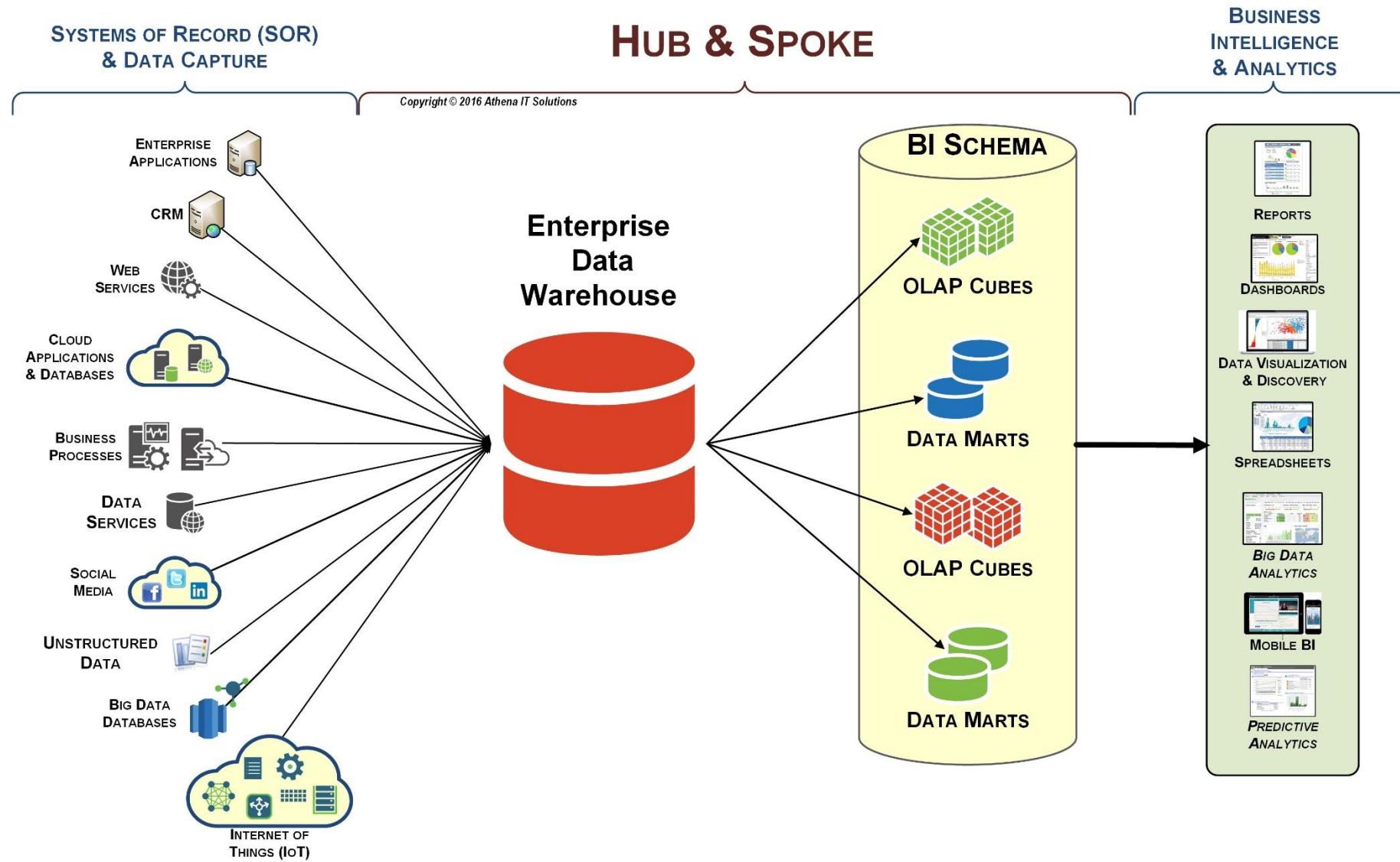


Details to be discussed later in course



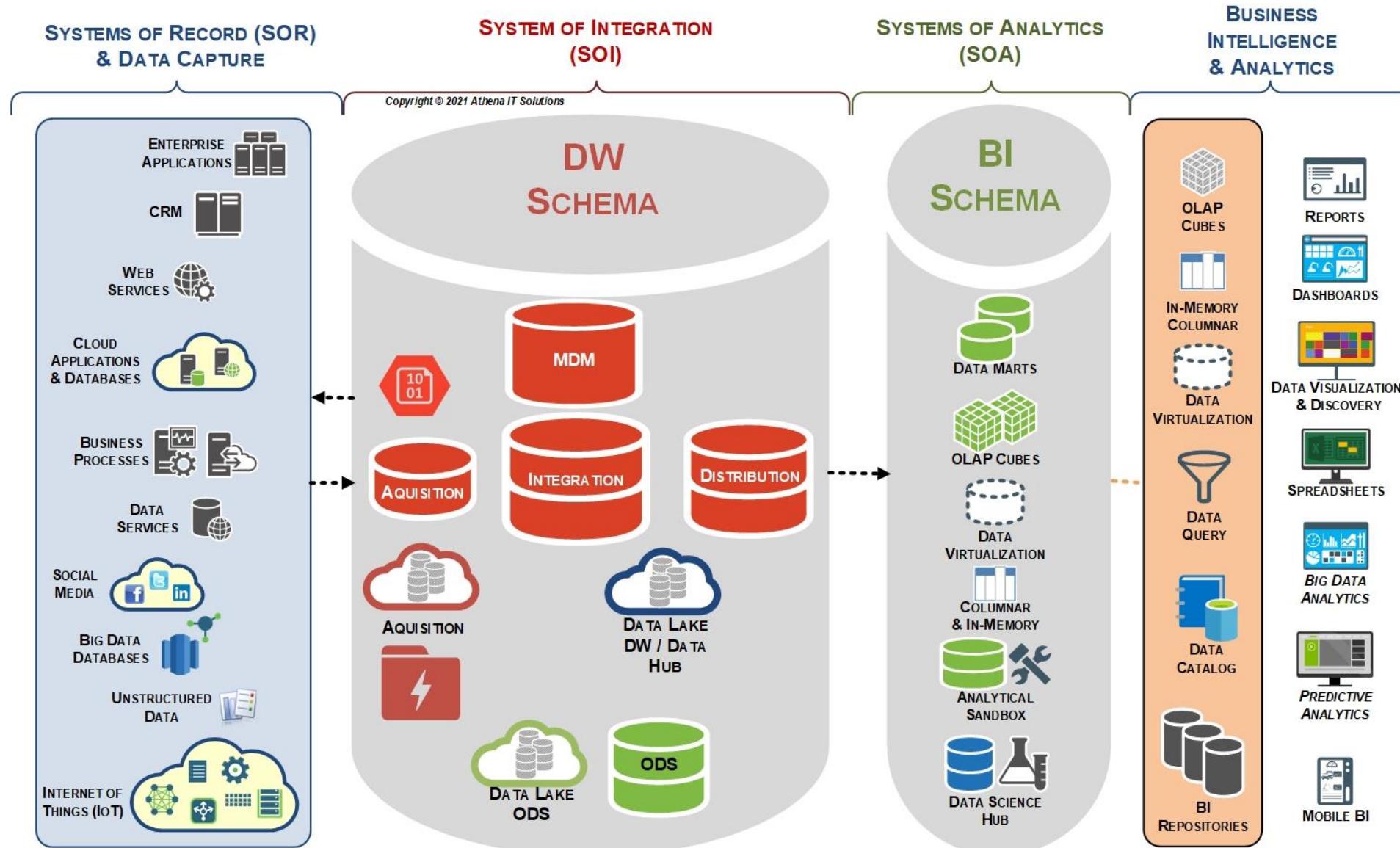
In beginning a place to store data for reporting, decision support systems (DSS) & BI in centralized database

Classic DW Evolution – Hub & Spoke



Details to be discussed later in course

Analytical Data Architecture (ADA)



Details to be discussed later in course

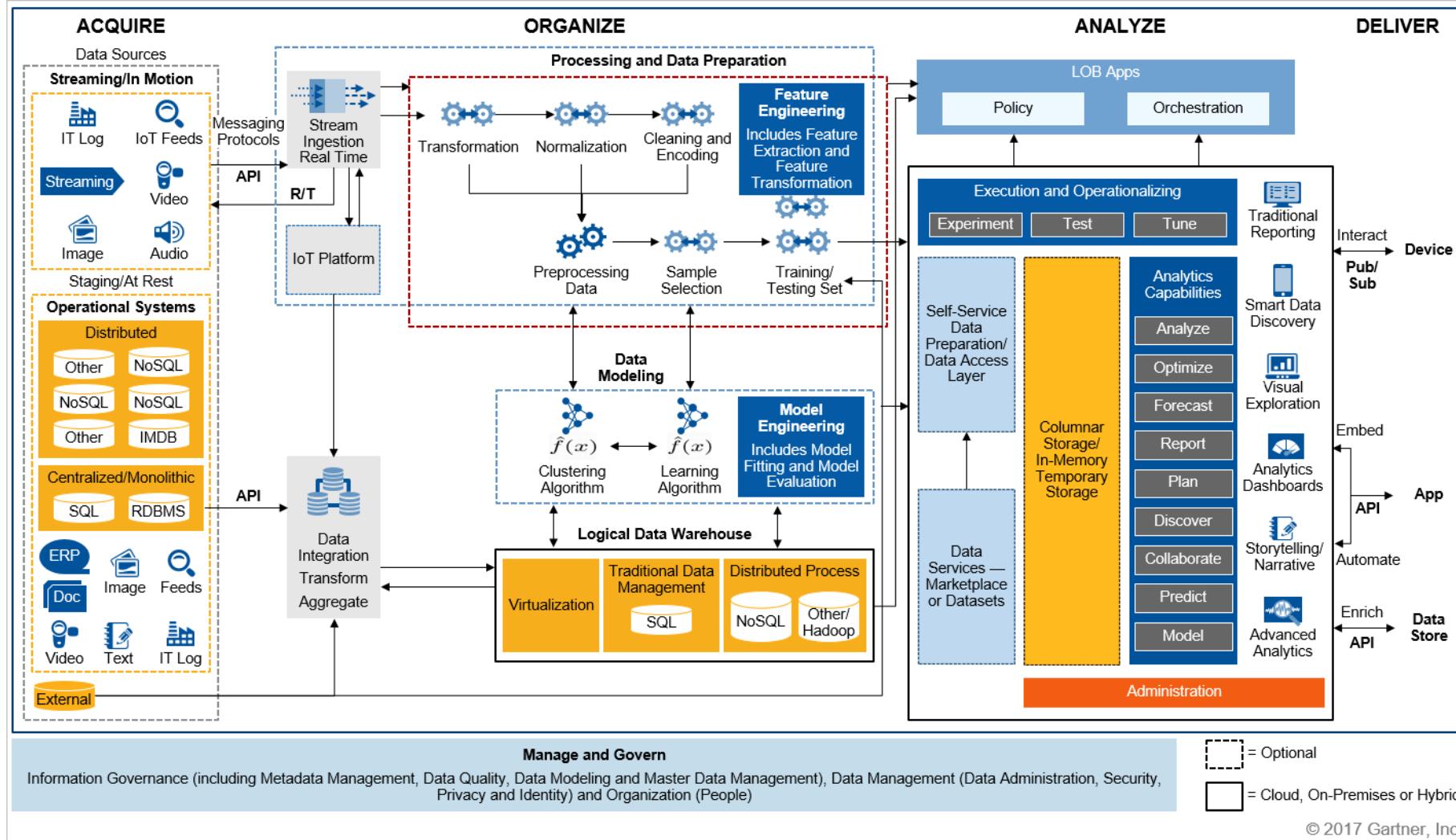
Template

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

Analytical Architecture



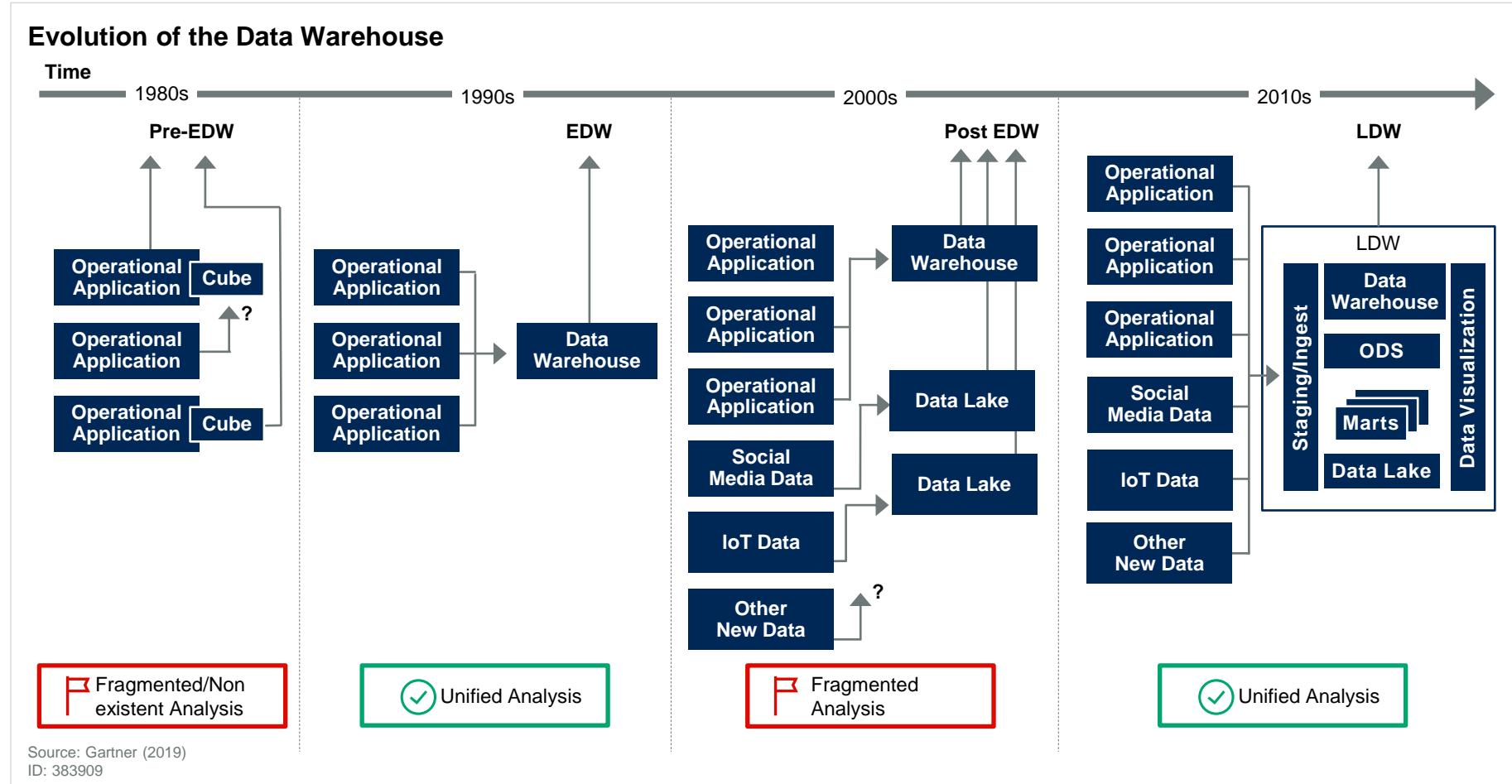
Details to be discussed later in course



Analytical Architecture



Details to be discussed later in course

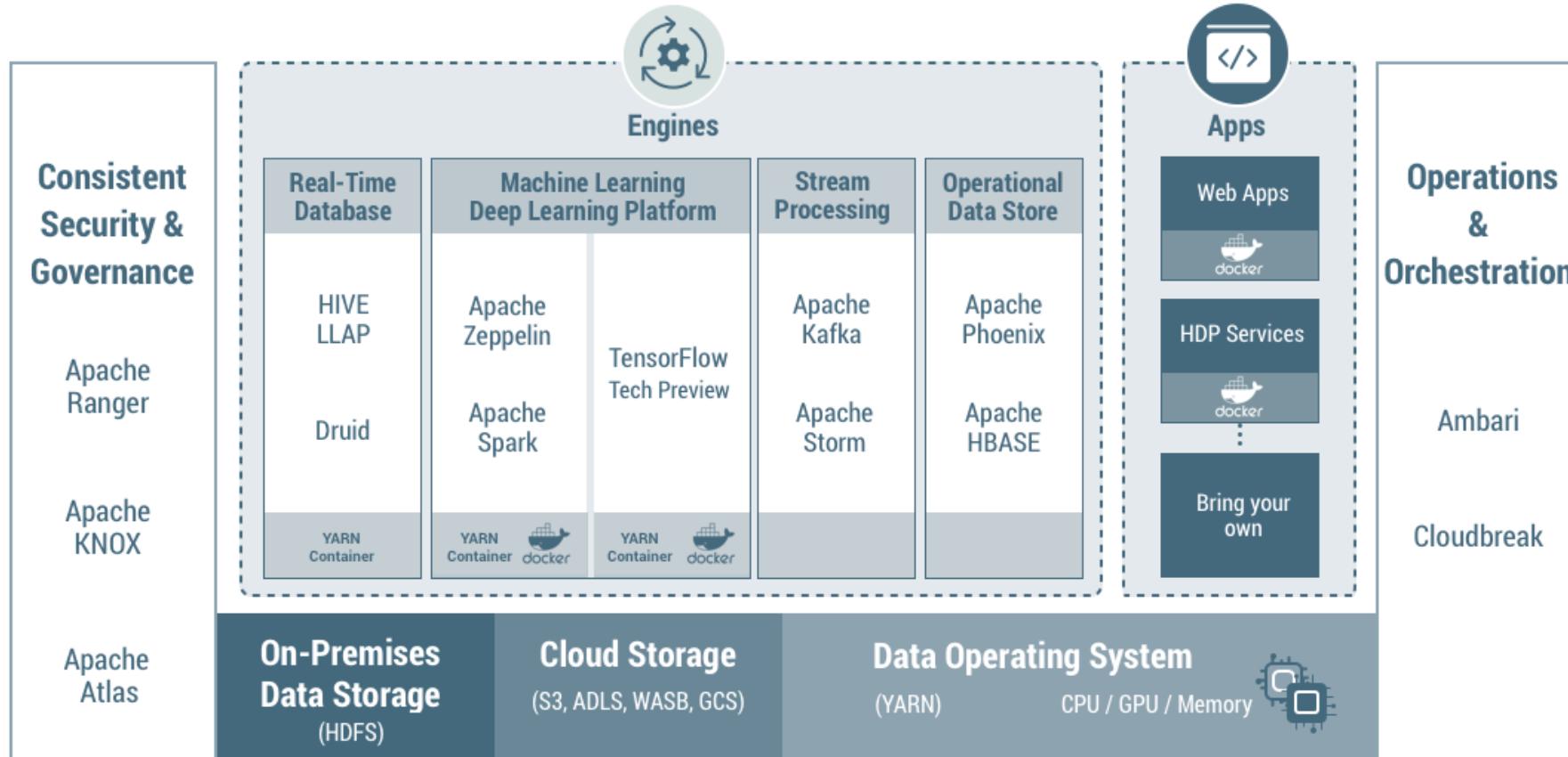


Gartner

Big Data & Analytics Evolves



Details to be discussed later in course



Large, Shared Workloads, Multi-Tenant Clusters



HYBRID



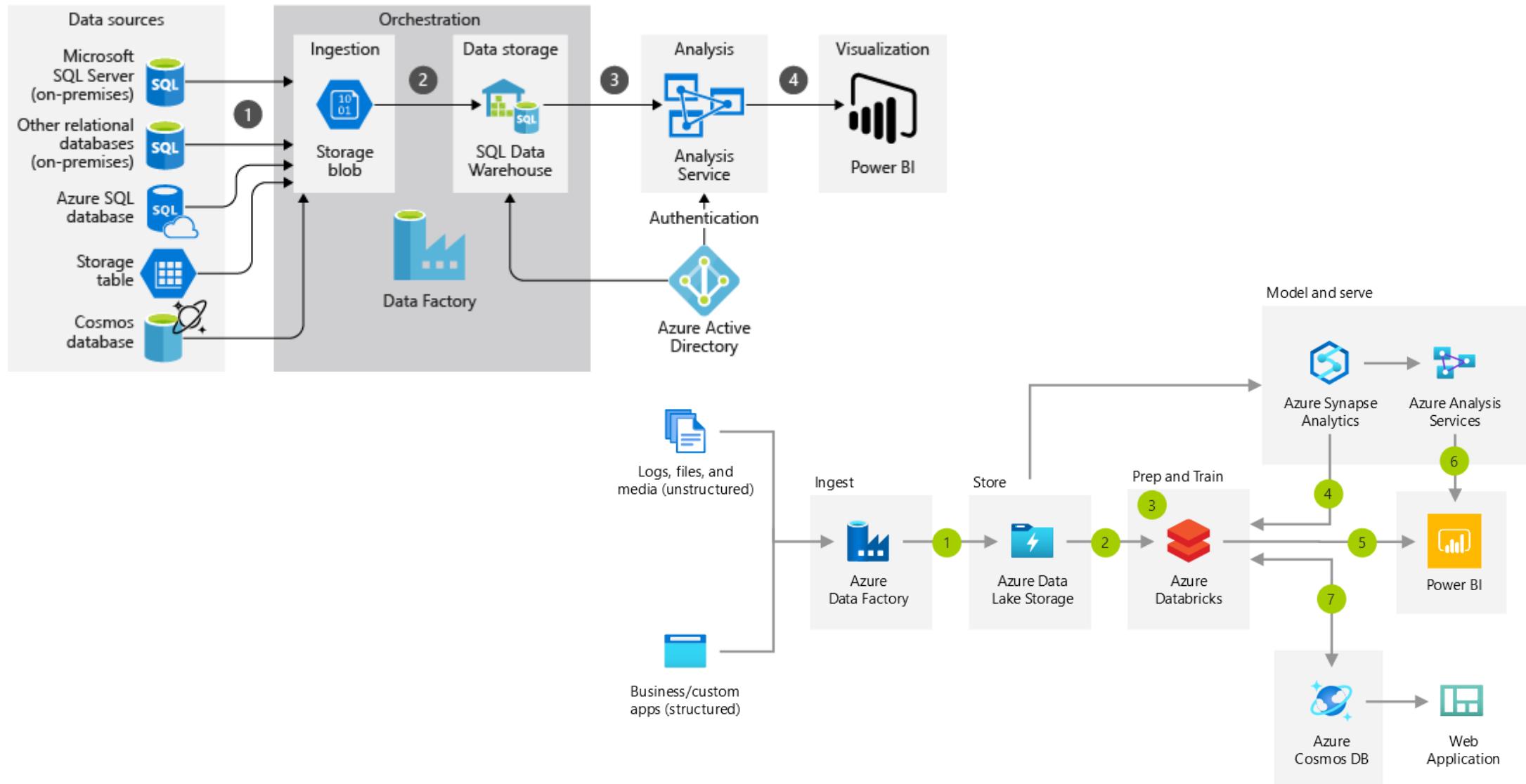
MULTI



DW & Analytics Evolves



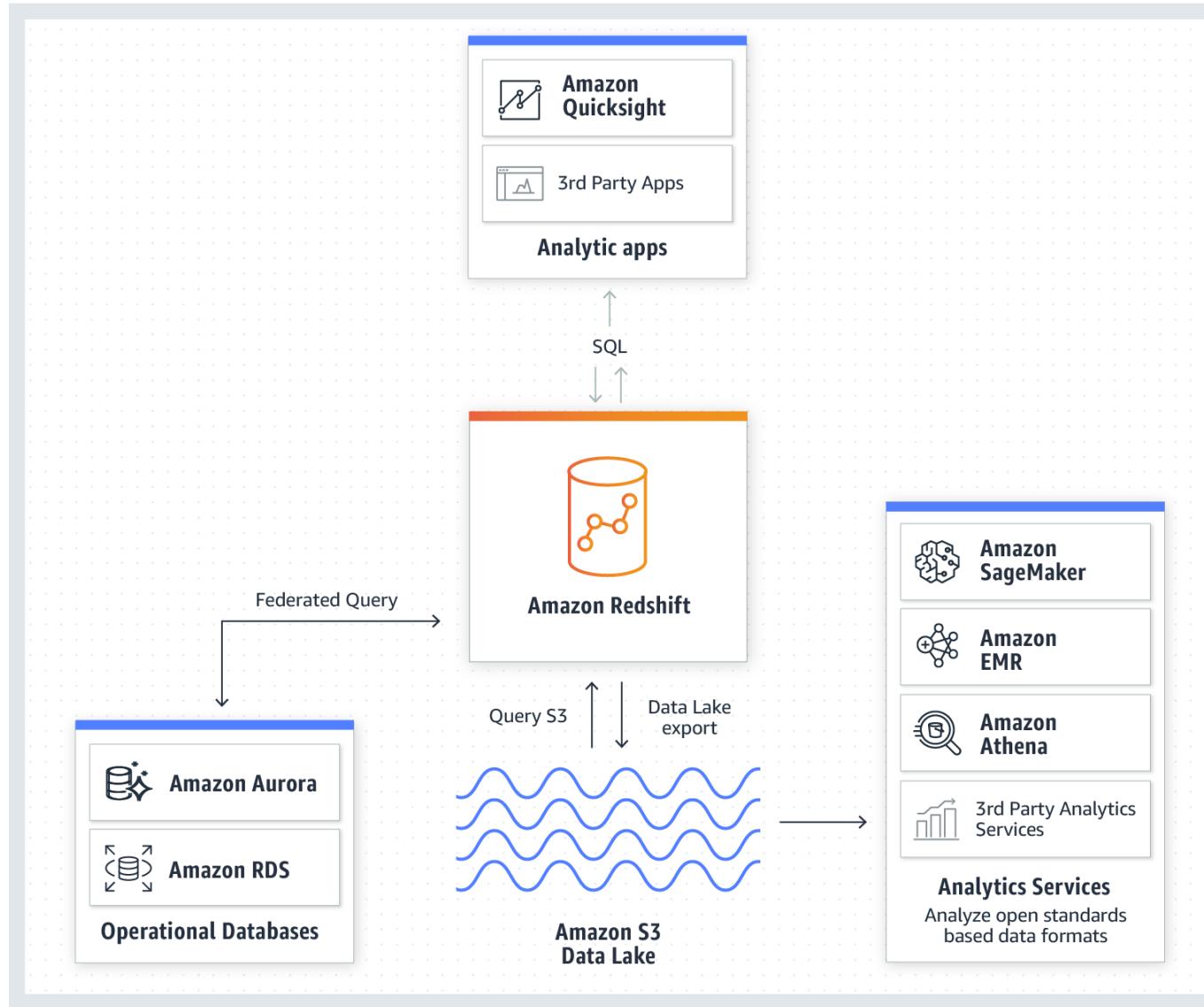
Details to be discussed later in course



DW & Analytics Evolves



Details to be discussed later in course

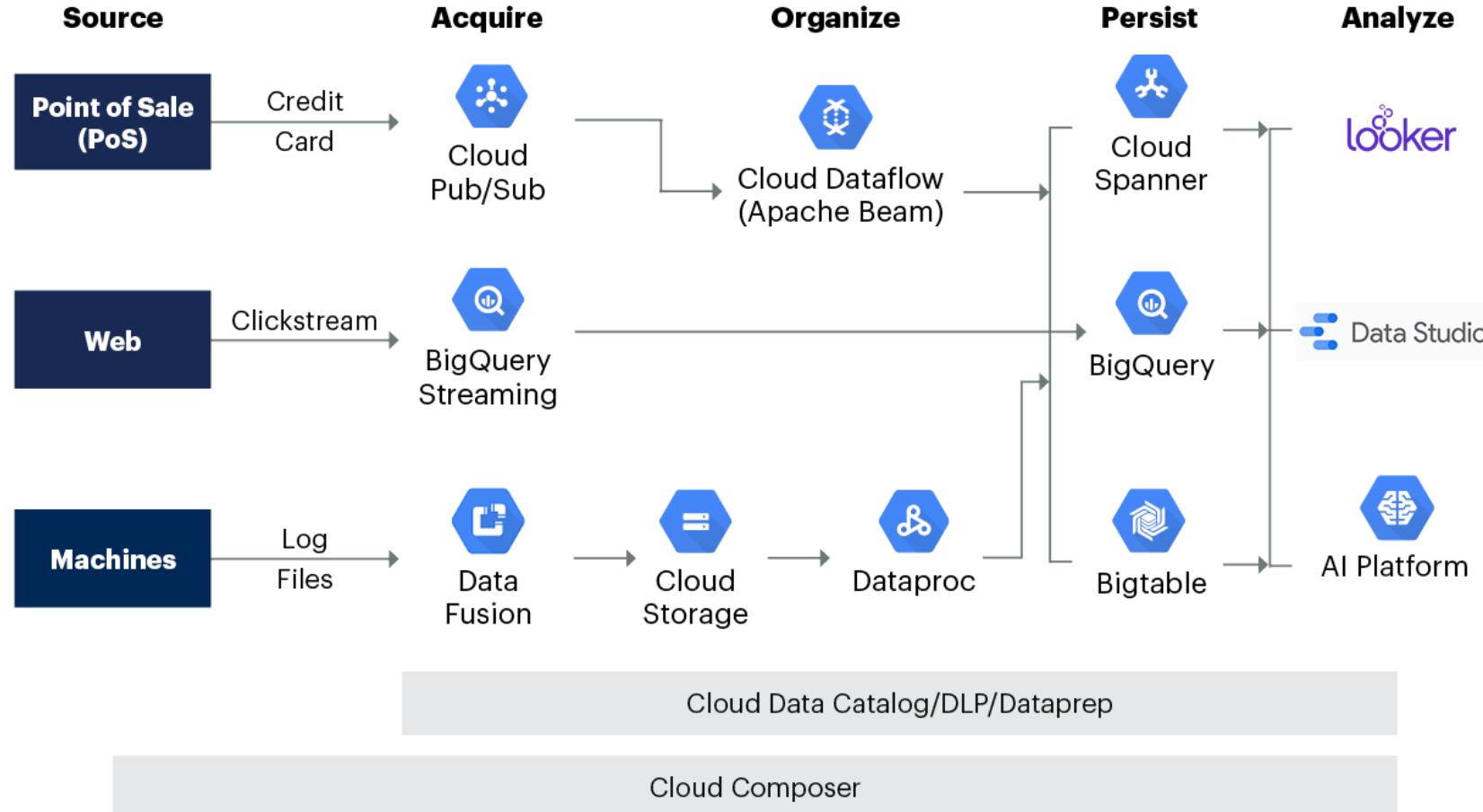


DW & Analytics Evolves



Details to be discussed later in course

GCP End-to-End Sample Data and Analytics Architecture



Source: Gartner

451418_C

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)

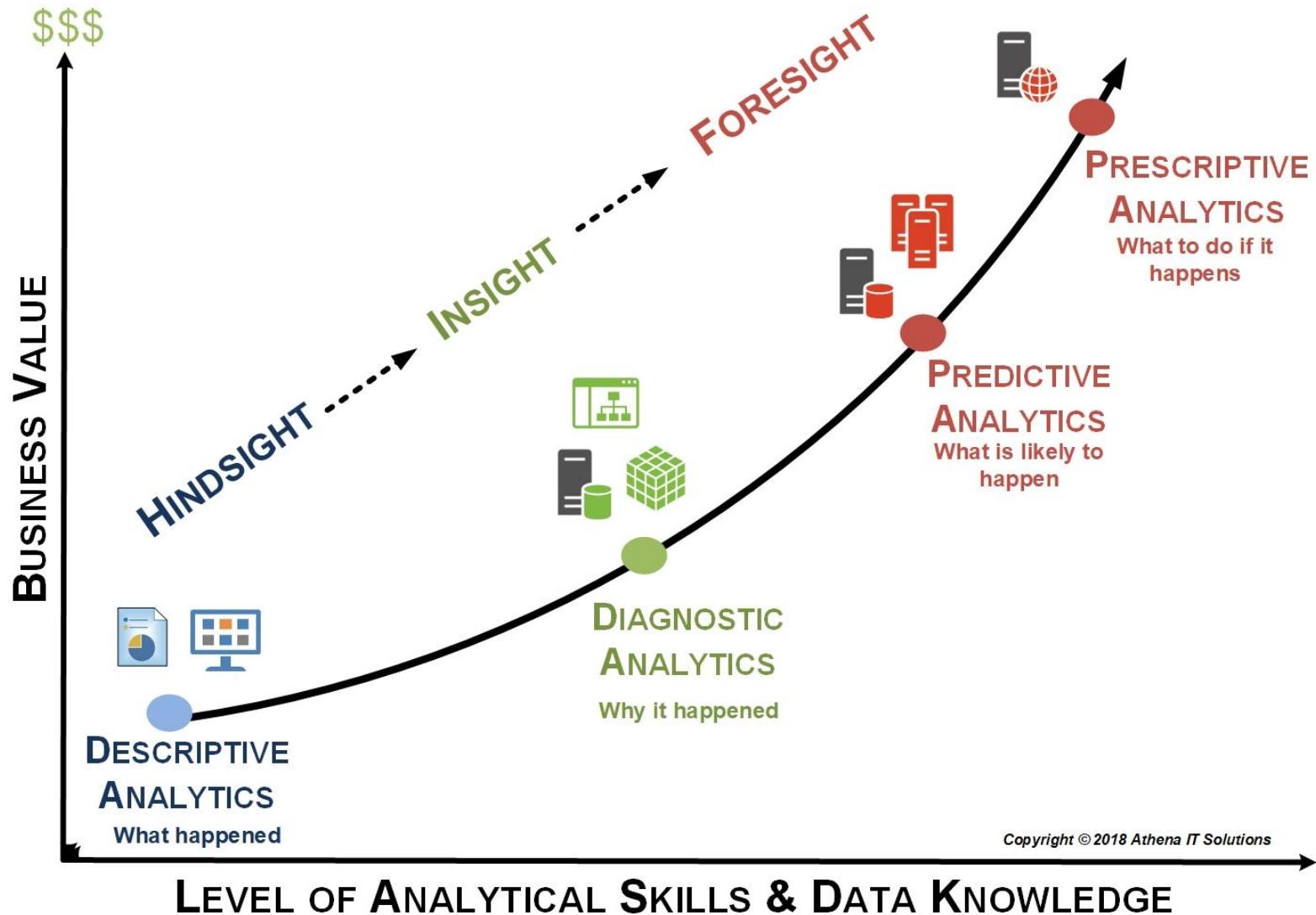
- Enables access and delivery of information to people and processes
- Information presentation layer
- Transform data into business information
 - Goal is to enable business insights
 - Only of business value if actionable
 - ...but both insights & actions depend on people unless model embedded in process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



What Types of Analytics are Needed



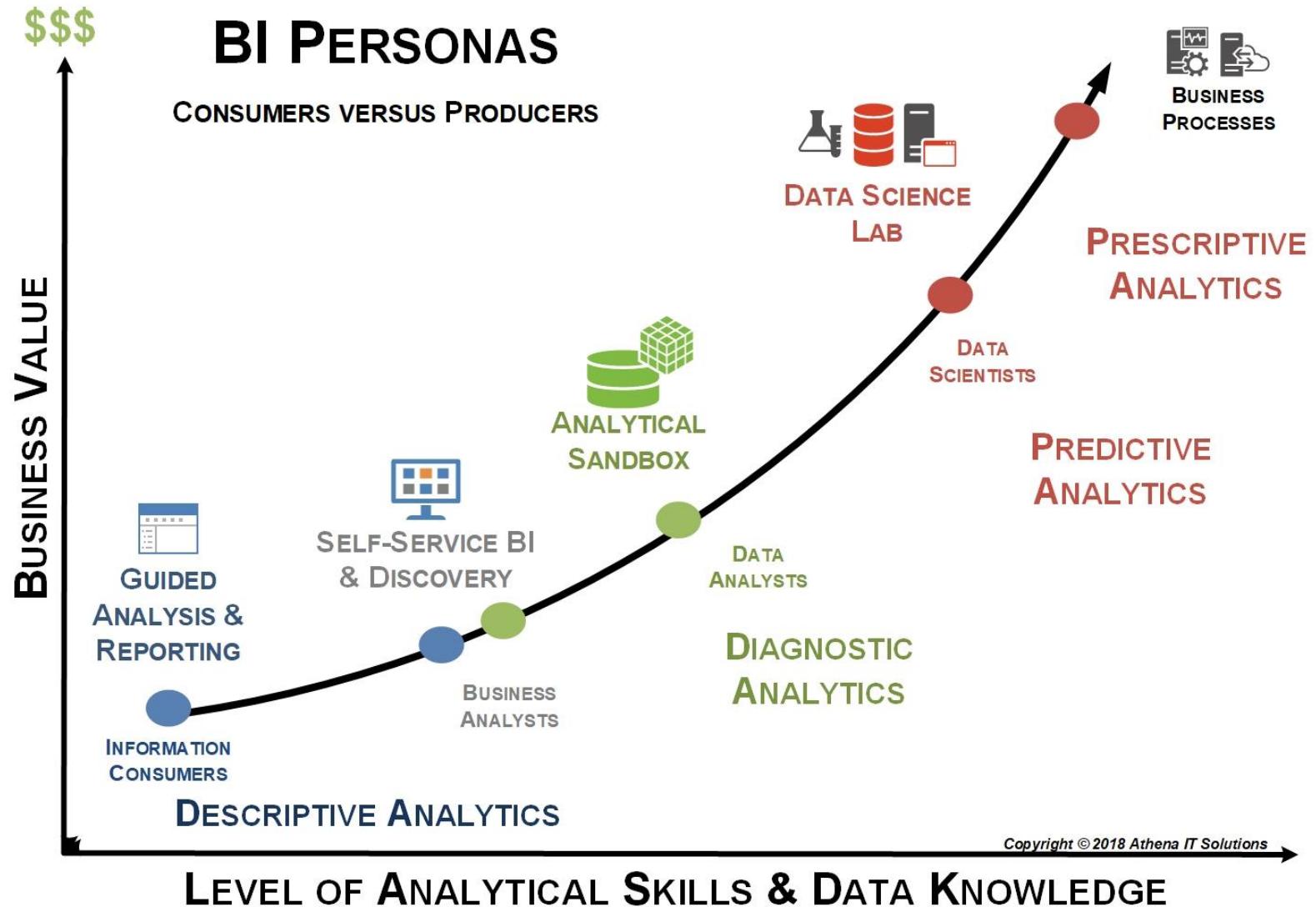
Details to be discussed later in course



BI Schema – Persona & Analytical Categories



Details to be discussed later in course



BI Schema – BI Tool Categories

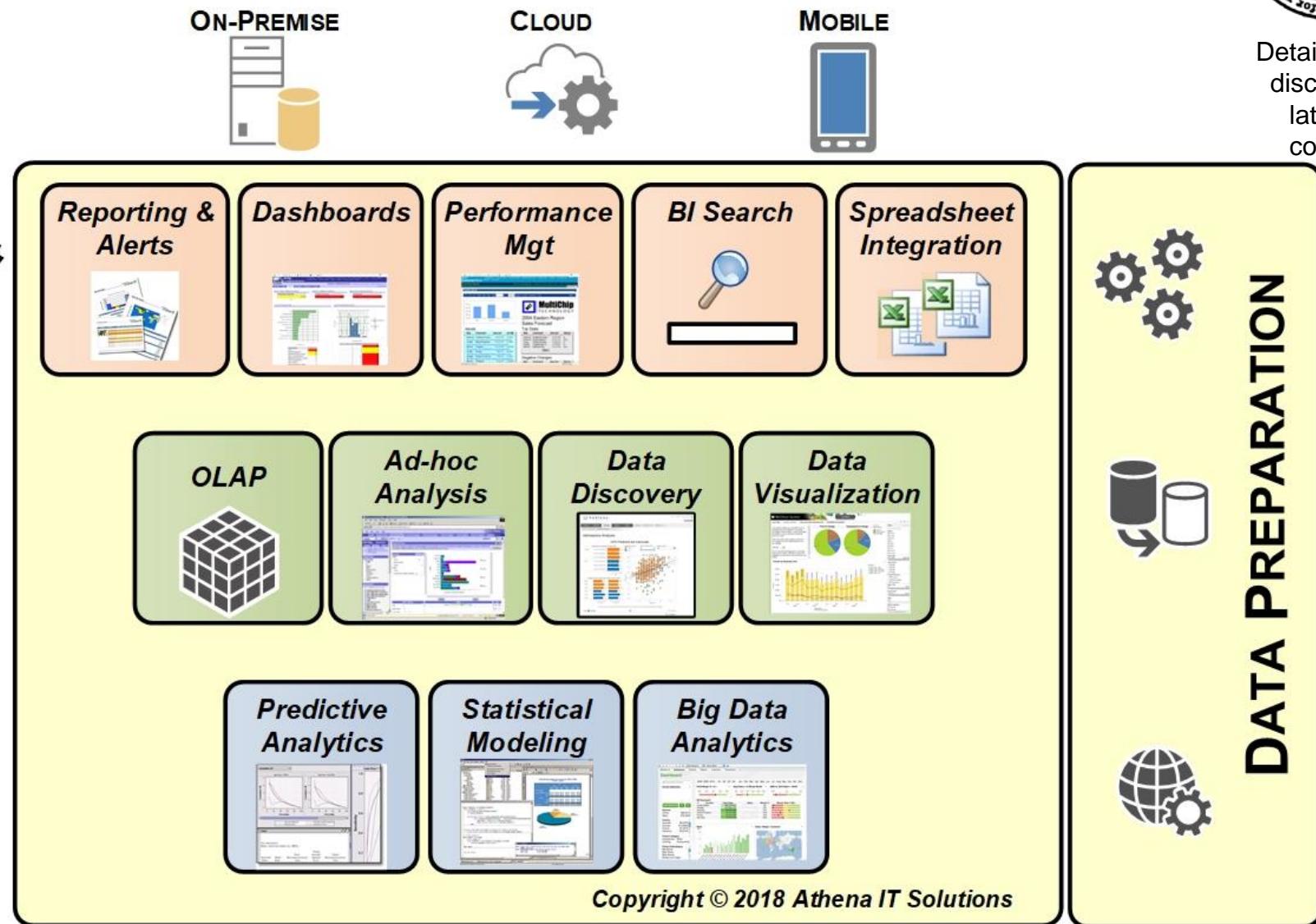


Details to be discussed later in course

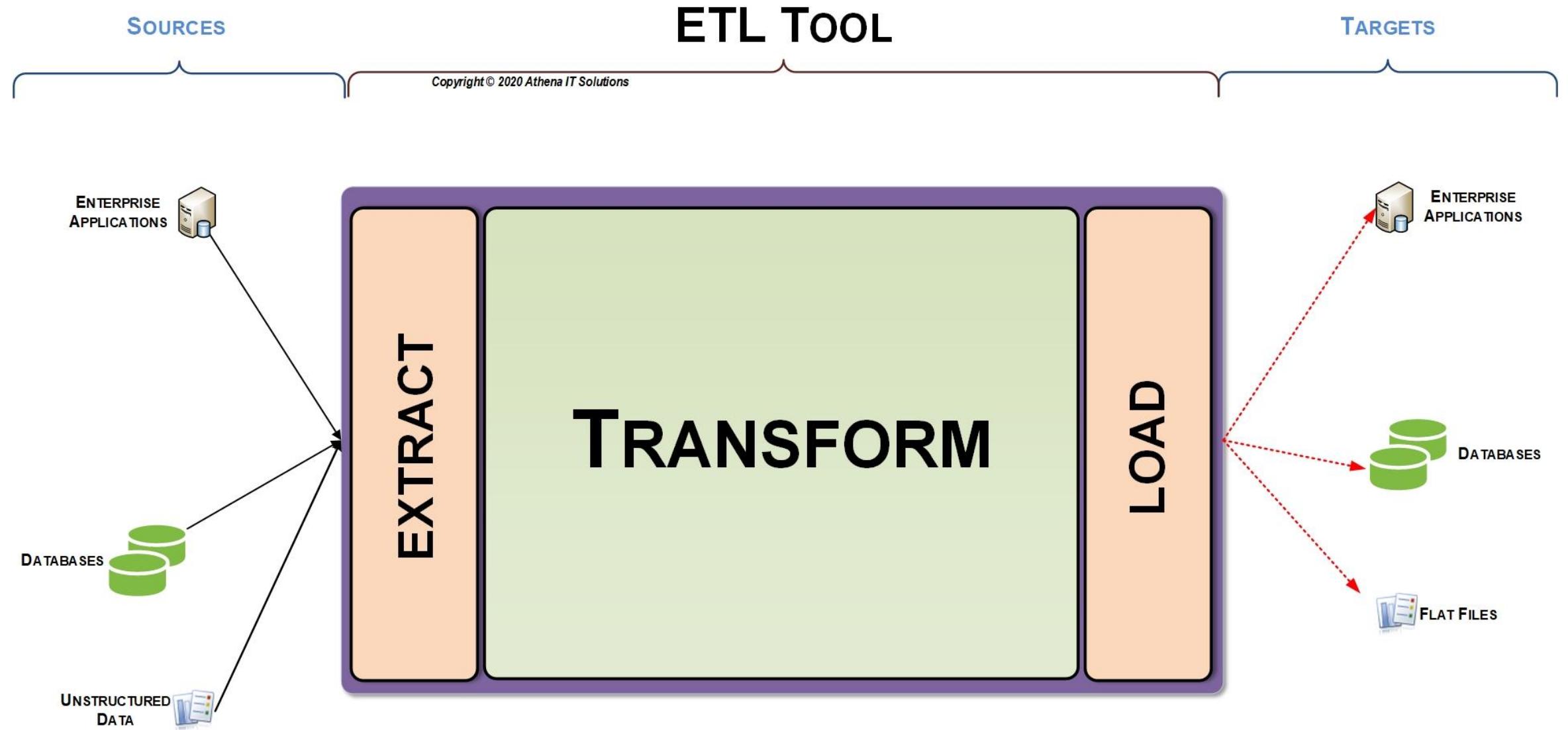
GUIDED ANALYSIS & REPORTING (TRADITIONAL BI)

SELF-SERVICE BI & DATA DISCOVERY

ADVANCED ANALYTICS



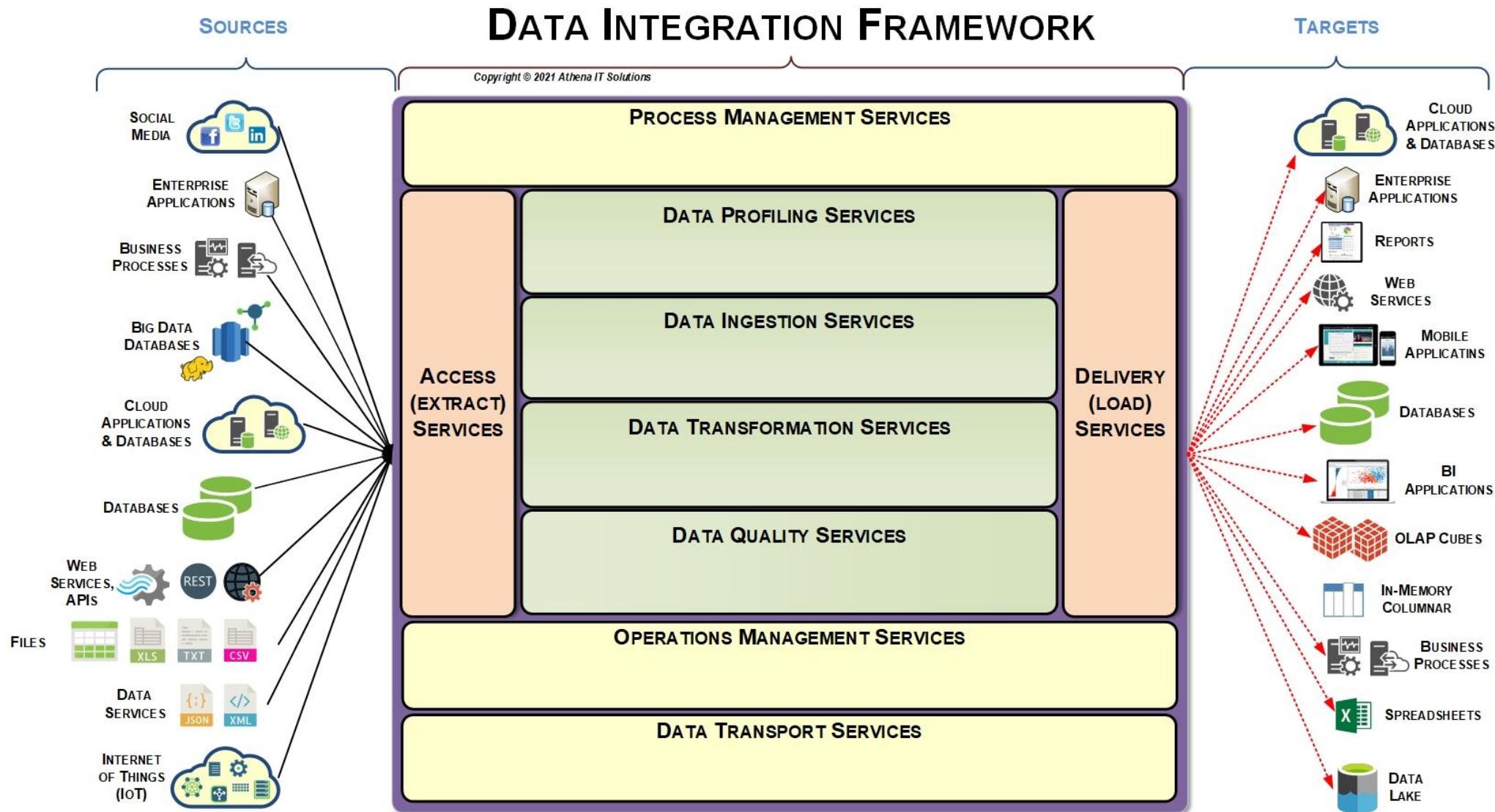
Extract, Transform & Load (ETL) - Beginnings



DI Suite - Services



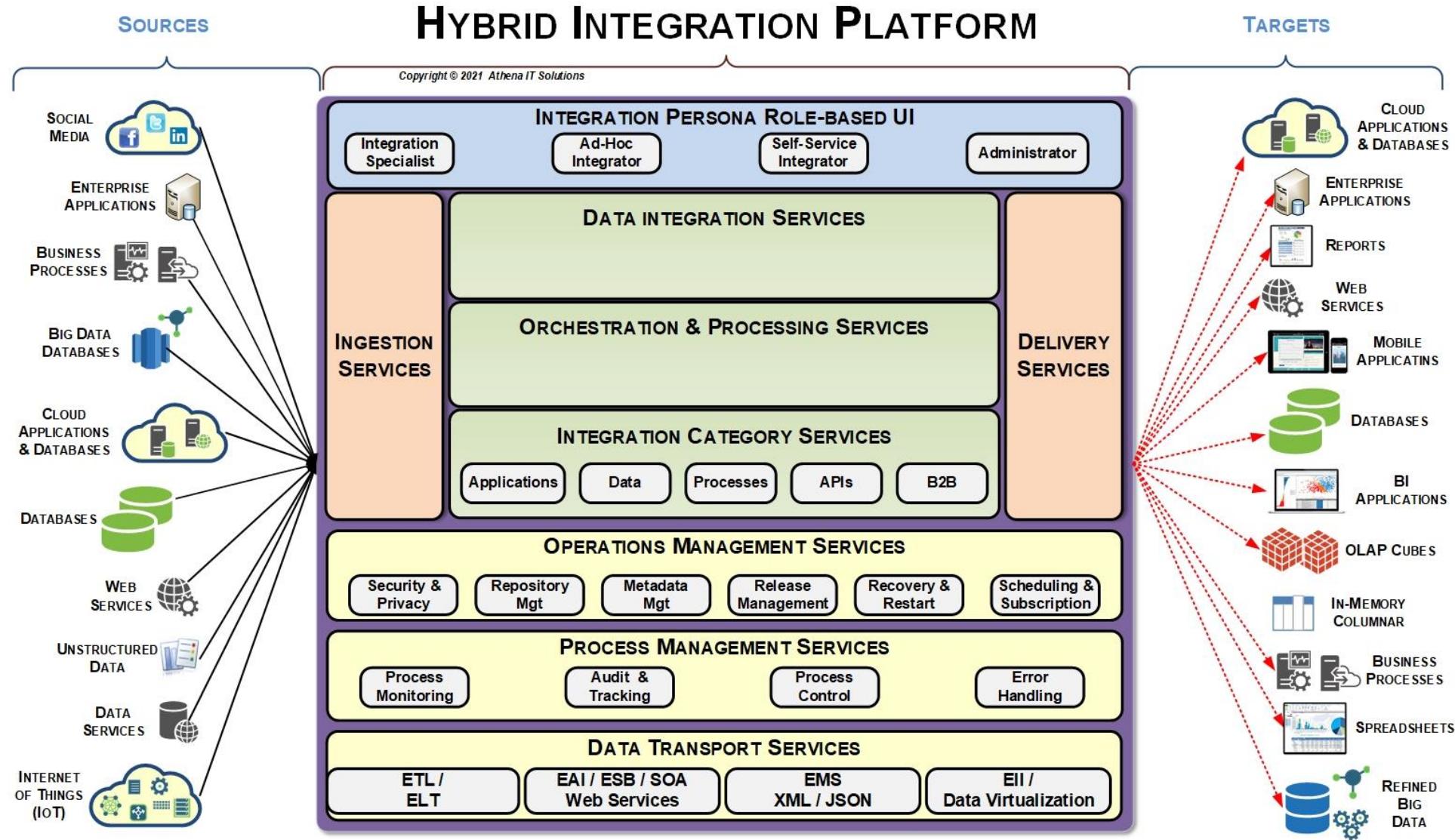
Details to be discussed later in course



DI Suite – Breakdown of Services



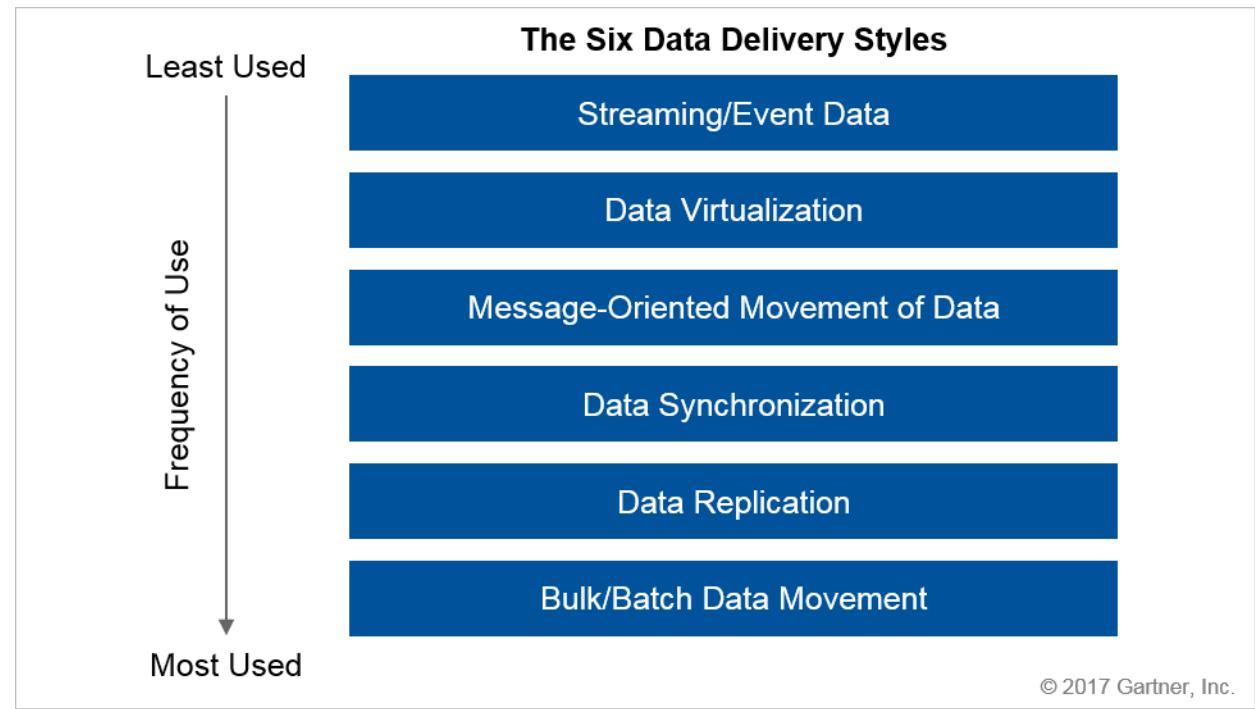
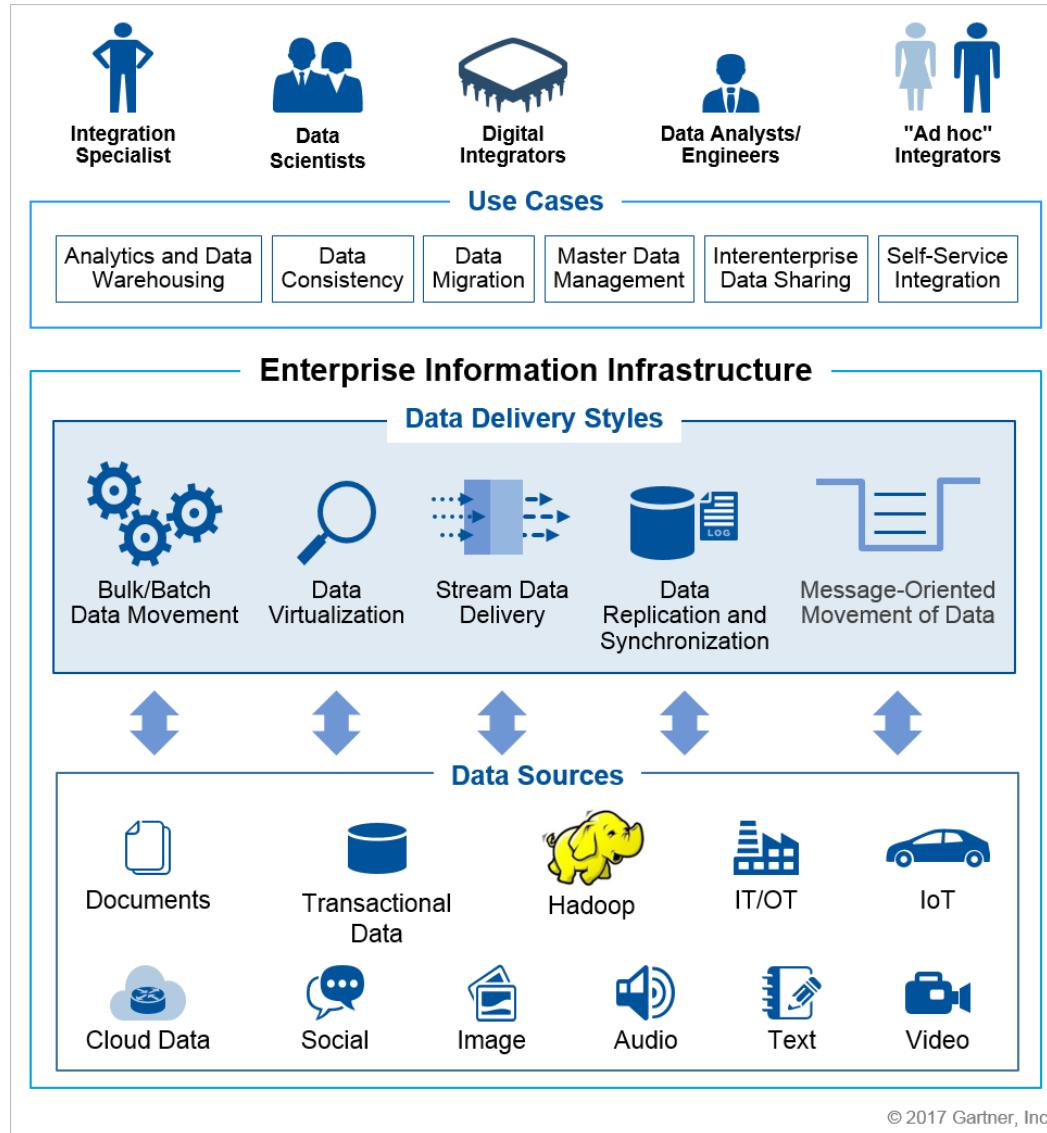
Details to be discussed later in course



Data Integration evolves to Hybrid Integration



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.

Relational Era

- Relational emerged in 1980s
- Mainstream in 1990s
- Relational keep adding features for Operational, DW, BI & Analytics
- Relational best practice for DW
- Top 3 vendors the same for decades
- Open-source relational DBMS not top
- Database was constraining factor

ORACLE

teradata.

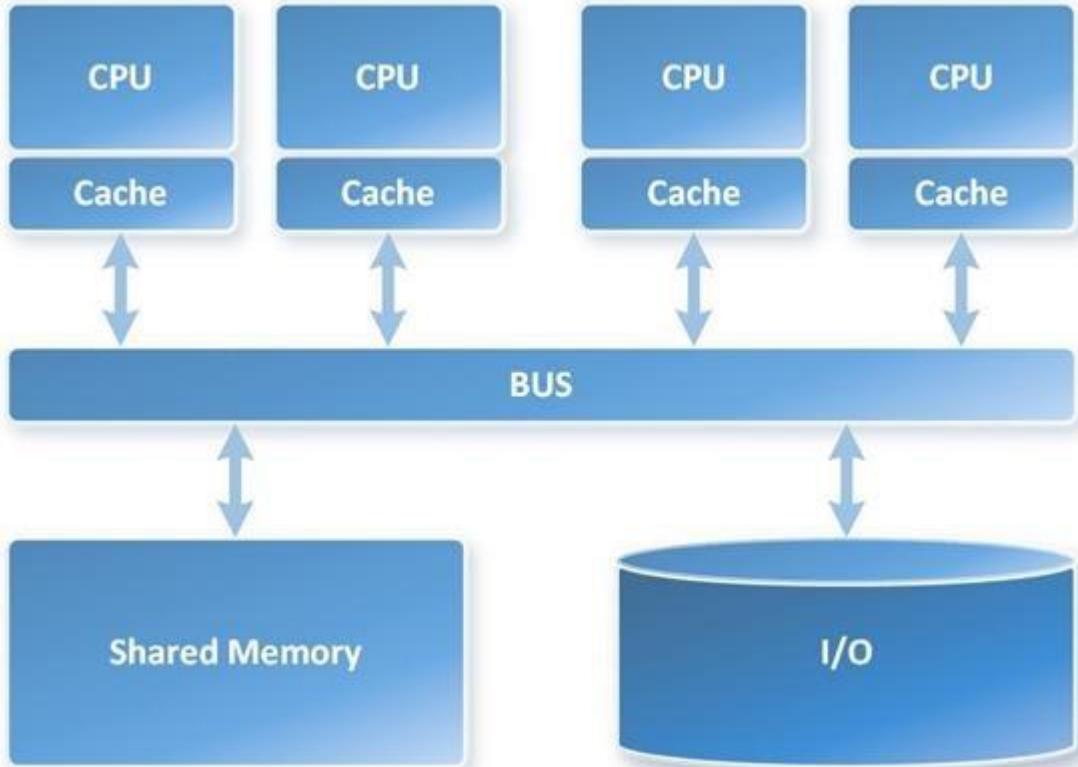


IBM



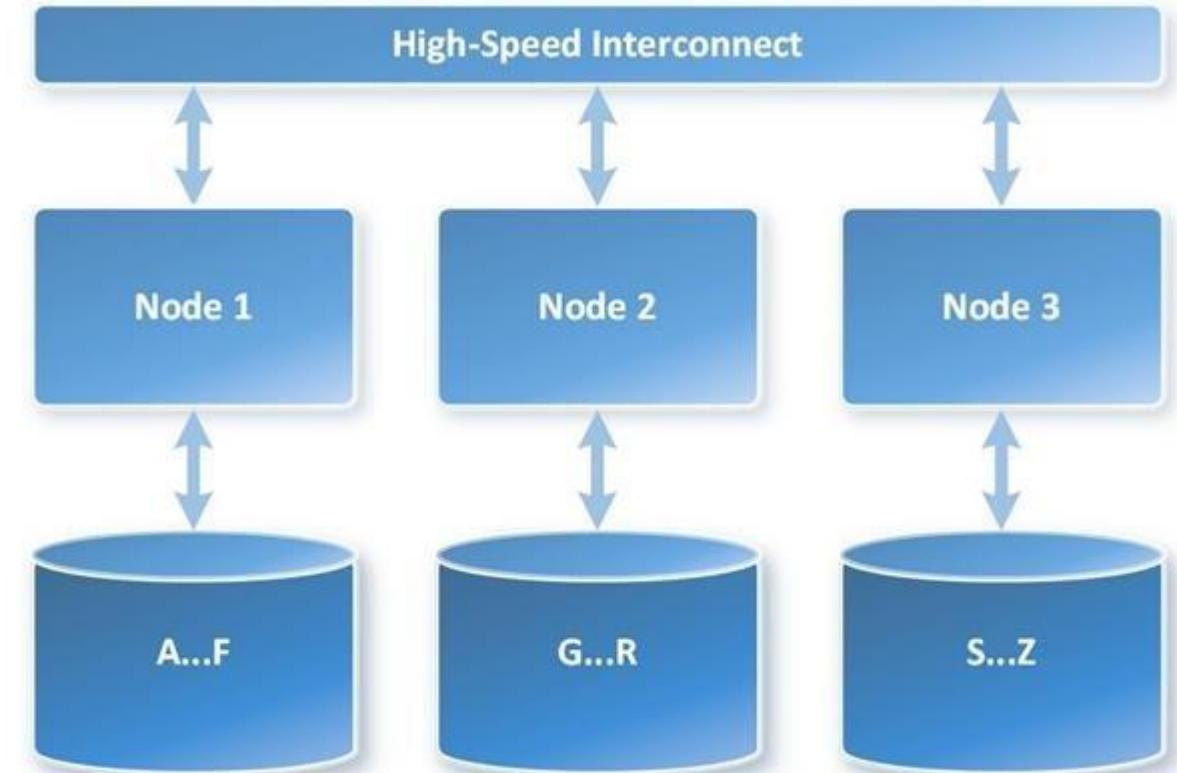
SMP vs MPP

Initially databases built this way



Symmetric Multi-Processing (SMP)

Current best practice. 1st Hadoop & NoSQL then relational migrated to this architecture



Massively Parallel Processing (MPP)

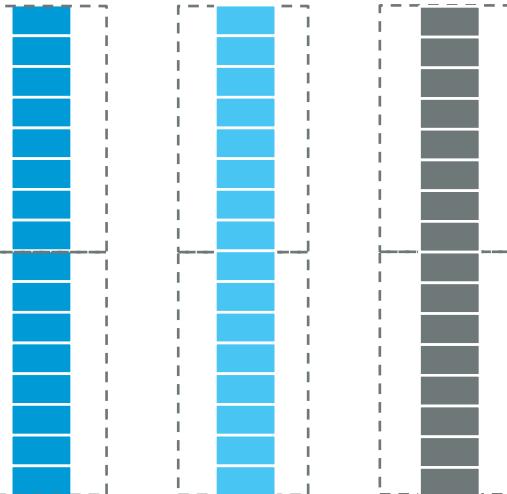
Columnar Databases

Column Stores vs. Row Stores

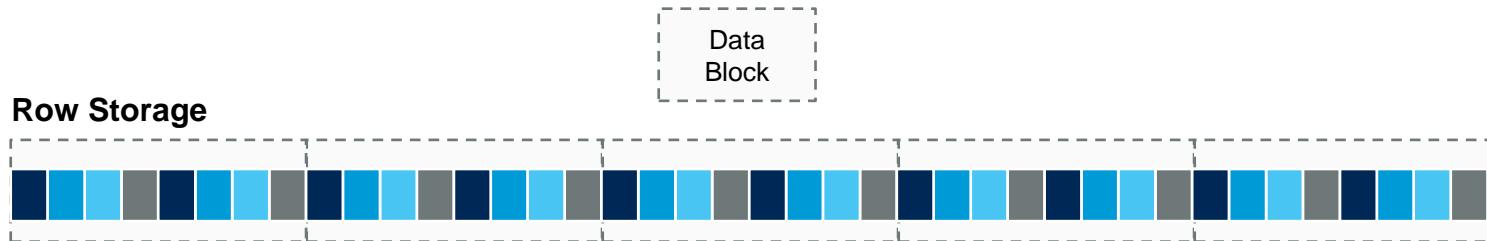
Table

| SKU | Name | Size | Price |
|--------|----------------|------|-------|
| 182701 | Boat | S | 5.99 |
| 526887 | Plane | M | 10.50 |
| 824569 | Spaceship | M | 12.99 |
| 548219 | Car | L | 34.99 |
| 635687 | Motorbike | M | 19.99 |
| 236124 | Sailboat | L | 31.99 |
| 787833 | UFO | S | 5.95 |
| 787834 | UFO Collection | M | 28.50 |

Column Storage

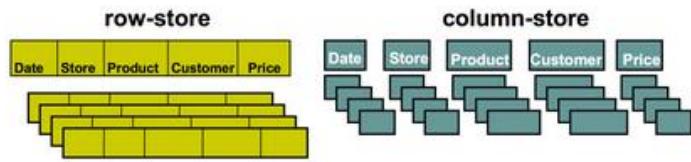


Row Storage



ID: 351983

© 2018 Gartner, Inc.



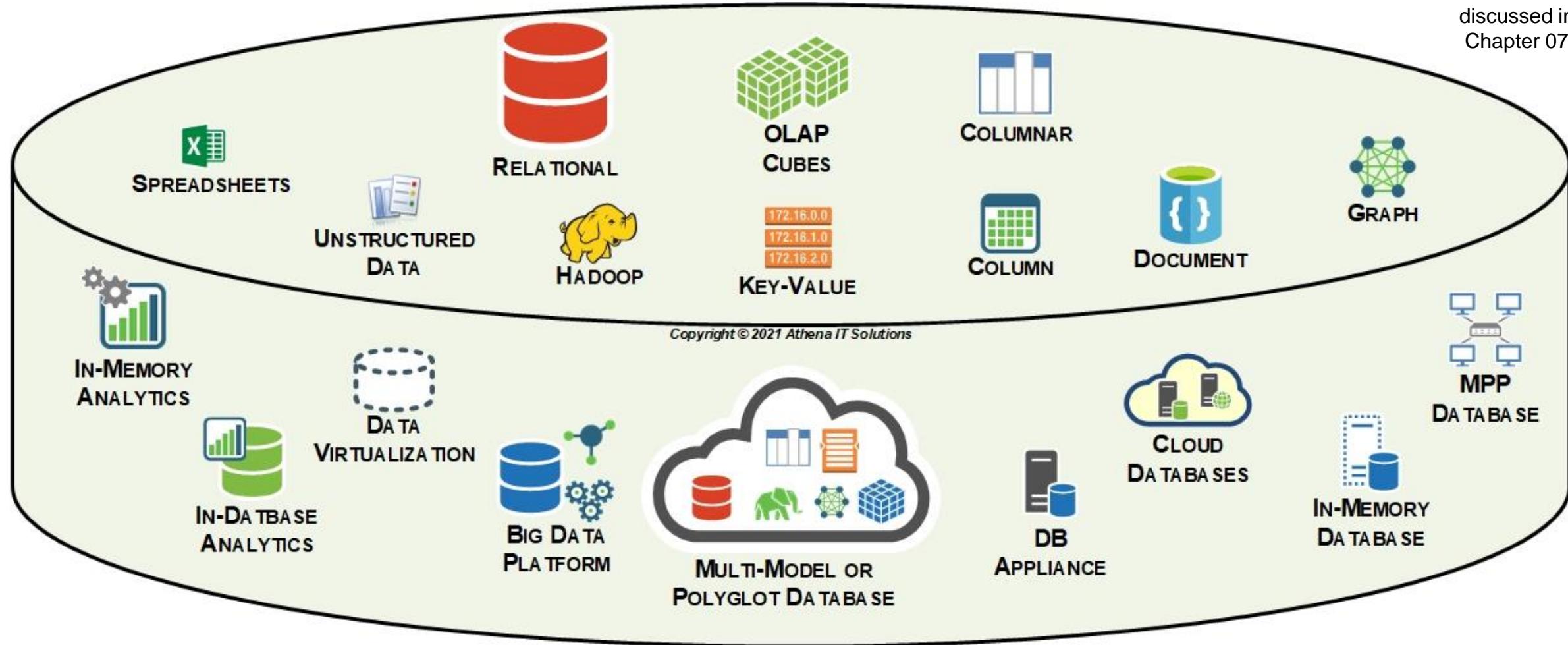
Columnar Databases:

- Column stores == arrays of same field
- Everything is an index
- Compression
- SQL used
- Looks like a relational database to a developer or access method

Database Technologies



Details to be
discussed in
Chapter 07



Cloud Database Management Systems

Magic Quadrant for Cloud Database Management Systems

Published 23 November 2020 - ID G00441439 - 56 min read



THE FORRESTER WAVE™ Cloud Data Warehouse Q1 2021



Analytics and Business Intelligence Platforms



MicroStrategy®

Qlik

looker

THOUGHTSPOT

TIBCO® Spotfire®

Yellowfin

SAP®
Business Objects™

IBM COGNOS®

ORACLE®

TIBCO® | ibi
Information Builders

Data Integration



ORACLE®



Qlik



IBM

talend

denodo



snapLogic®

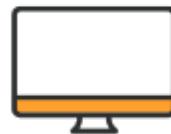
Fivetran



Data Virtualization



Details to be discussed in Chapter 07



Analytics



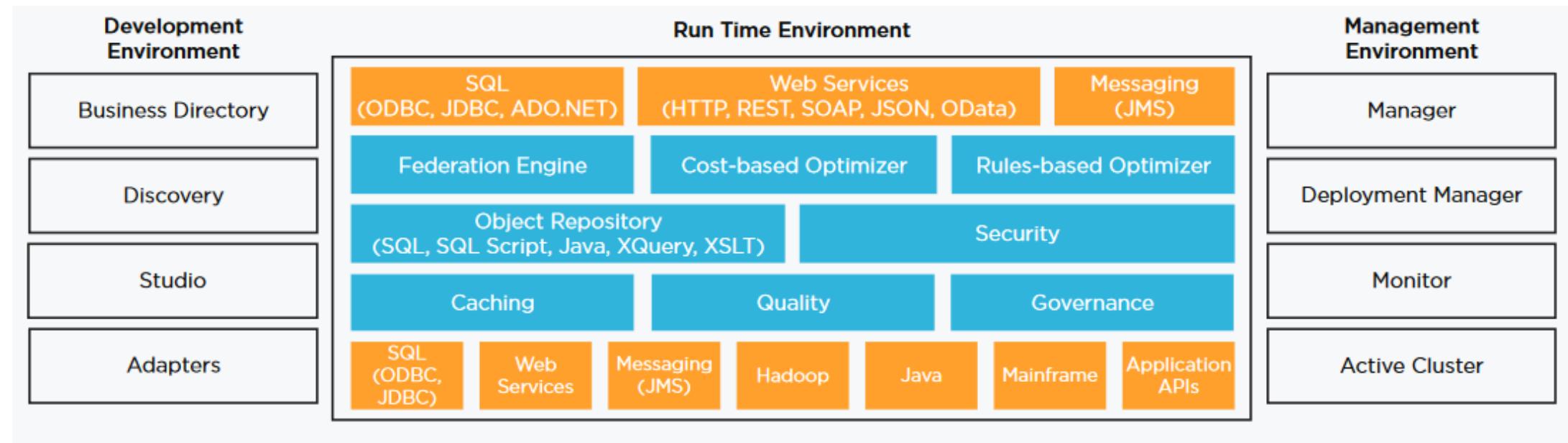
Self-Service



Business Intelligence



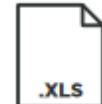
Transactional Apps



Packaged Apps



RDBMS



Excel Files



Data Warehouse



OLAP Cubes



Hadoop/Big Data



XML Docs



Flat Files



Web Services