

MODERN DATA INTEGRATION

ITMA APRIL 2015

TOM FISH | ALEX JANTZ

TOM FISH

ALEX JANTZ

DATA INTEGRATION CASE STUDIES

THE GOOD! (GOOD?)



THE BAD...

U-VERSE® TV WITH TOTAL HOME DVR®

*More freedom.
More control.*

Watch when you want wherever you are
in your home!

starting at

\$19/mo

[Learn more](#)

With qualifying bundles. For 24 mos. 1-yr term req'd.
Other charges apply. Geo. and svc. restr. apply.

THE UGLY...?

DATA INTEGRATION

“Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. A complete data integration solution encompasses discovery, cleansing, monitoring, transforming and delivery of data from a variety of sources.” - IBM

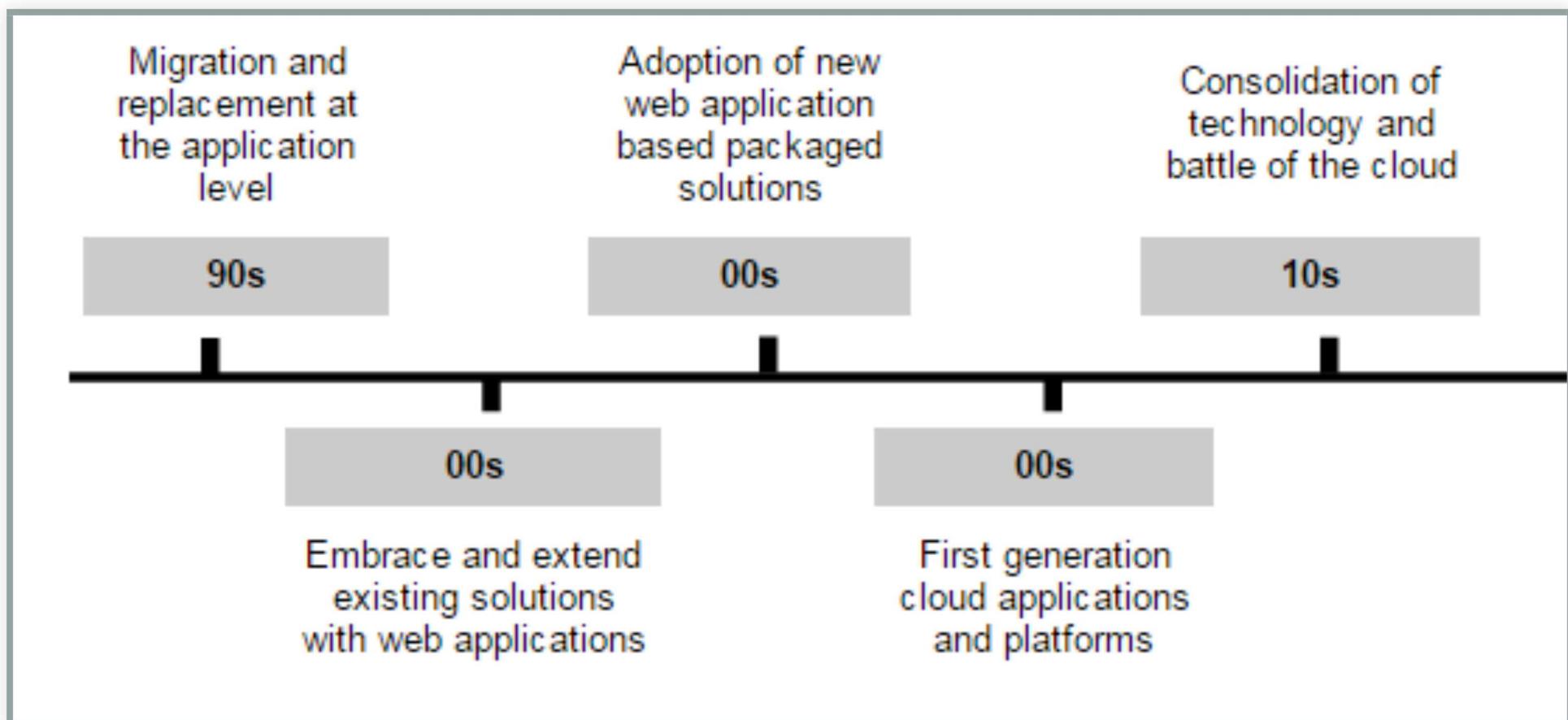
DATA INTEGRATION



TRUE DATA INTEGRATION IS AGNOSTIC OF SOURCE OR TARGET APPLICATION

ETL IS A BRIDGE FOR BI-DIRECTIONAL FLOW

HISTORY



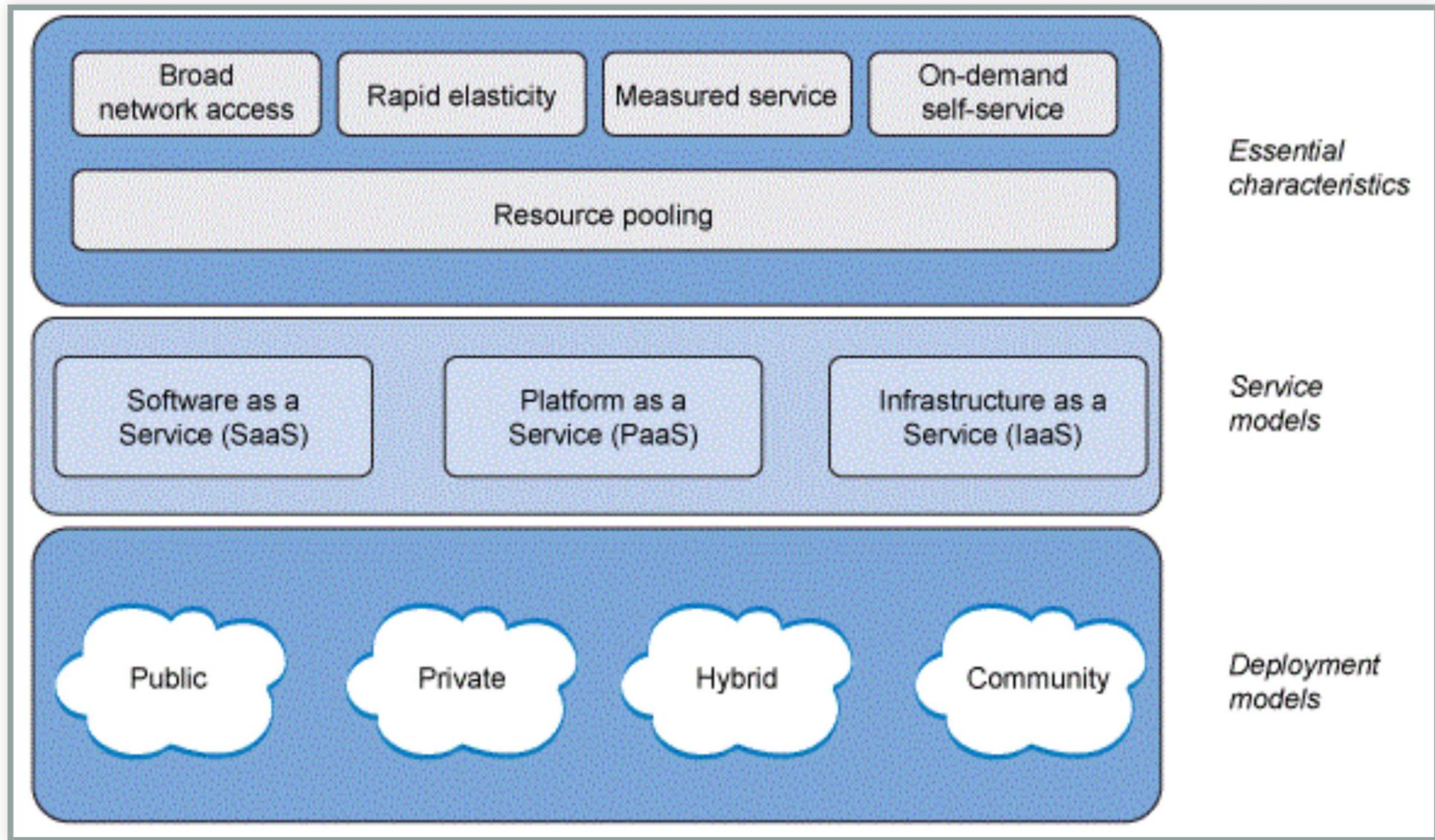
WHERE WE ARE

- Cloud popularity exploding
- “Distributed systems” now routinely global
- “Integration” is area of least satisfaction for cloud users
- Three entirely new scenarios:
 - Ground to cloud
 - Cloud to cloud
 - Cloud to ground

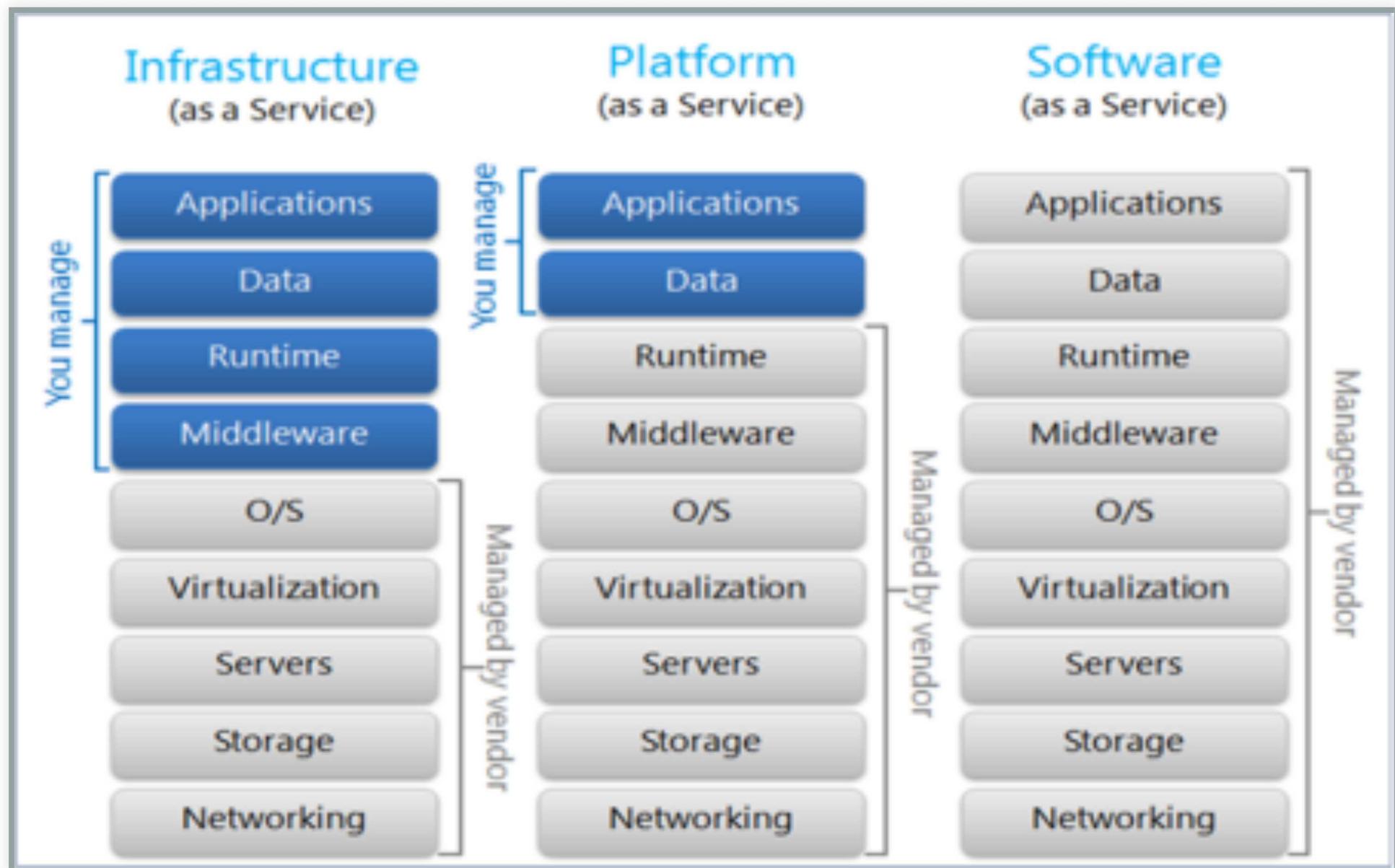
THE CLOUD



NIST DEFINITION



CLOUD DELIVERY MODELS



SAAS

(SOFTWARE AS A SERVICE)

- Examples
 - Salesforce
 - Office 365
 - Google Apps

PaaS

(PLATFORM AS A SERVICE)

- Application framework
- Ability to Scale
- Examples
 - Azure
 - Cloud Foundry
 - Google App Engine

IAAS

(INFRASTRUCTURE AS A SERVICE)

- Security
- Updates
- Examples
 - AWS
 - Rackspace
 - Tier3

CLOUD (STILL) NEEDS

- Data aggregation
- Data replication
- Single source of truth
- Shared business functions
- Distributed, long-running business processes
- Business partner integration
- Integration software!

NEW CONSIDERATIONS

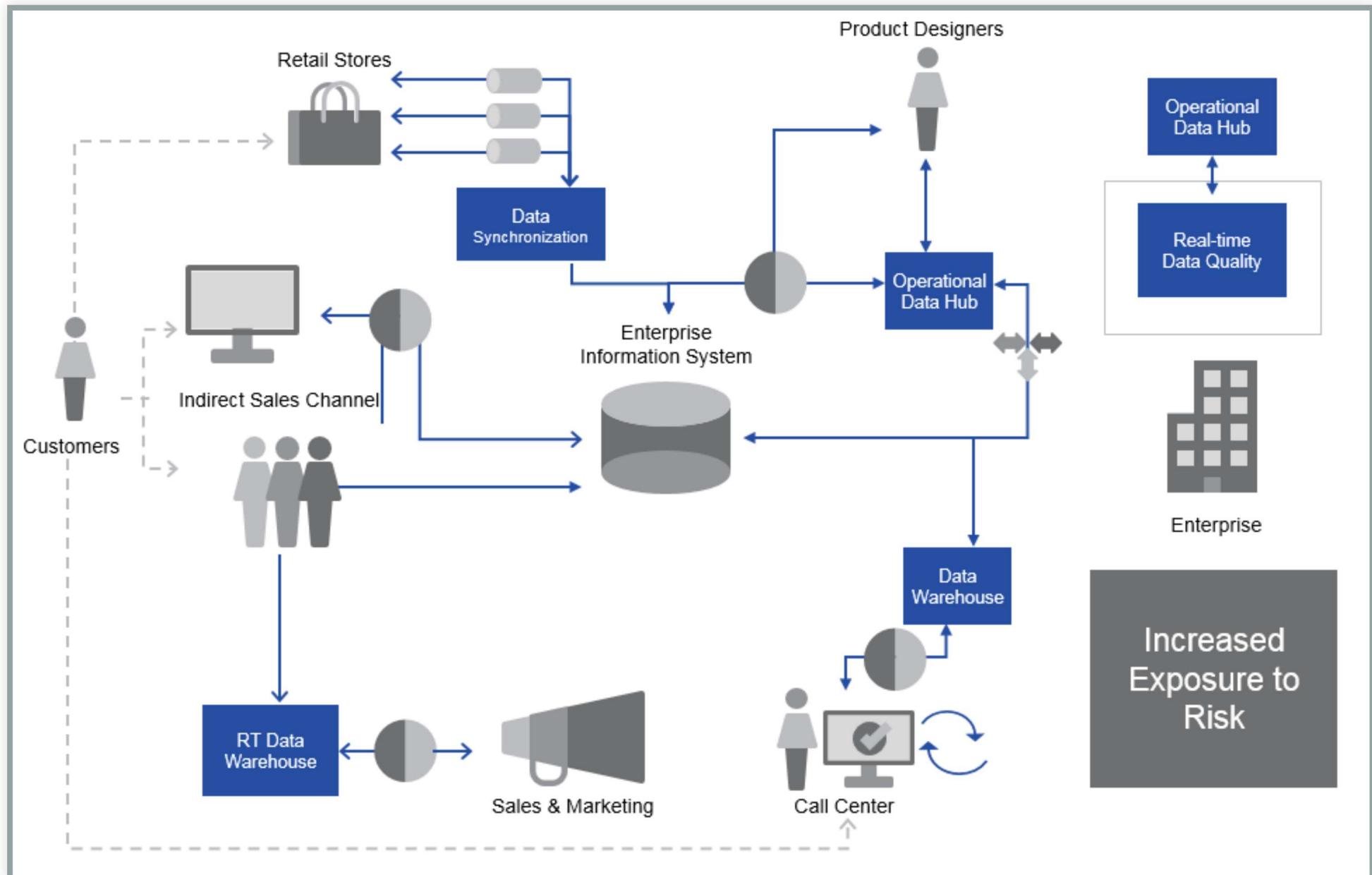
- Network Latency
- Identity Management
- Different SLAs
- Data Security
- Monitoring
- Management
- Interoperability
- Changing Schemas
- Services not Servers
- Connectivity
- REST-first

INTEGRATION

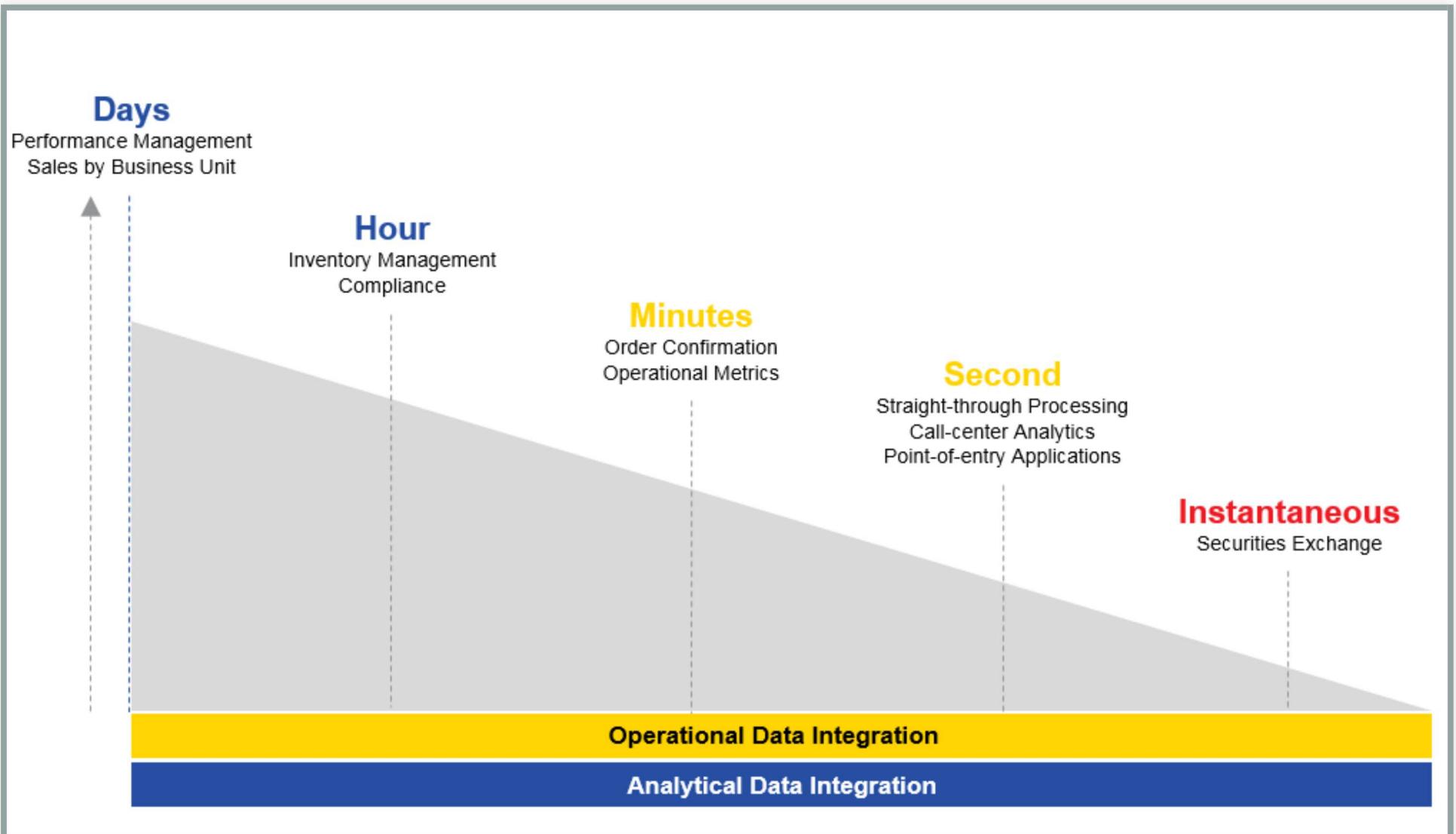
**"APPLICATIONS CREATED IN 2012 USING TRADITIONAL
ARCHITECTURE MODELS WILL BE AN IT-CONSTRAINING LEGACY
BY 2016 THE LEADING BUSINESS APPLICATIONS OF 2016 ARE
DESIGNED TODAY USING NEXUS-ENABLED APPLICATION
ARCHITECTURE PRINCIPLES"**

- GARTNER

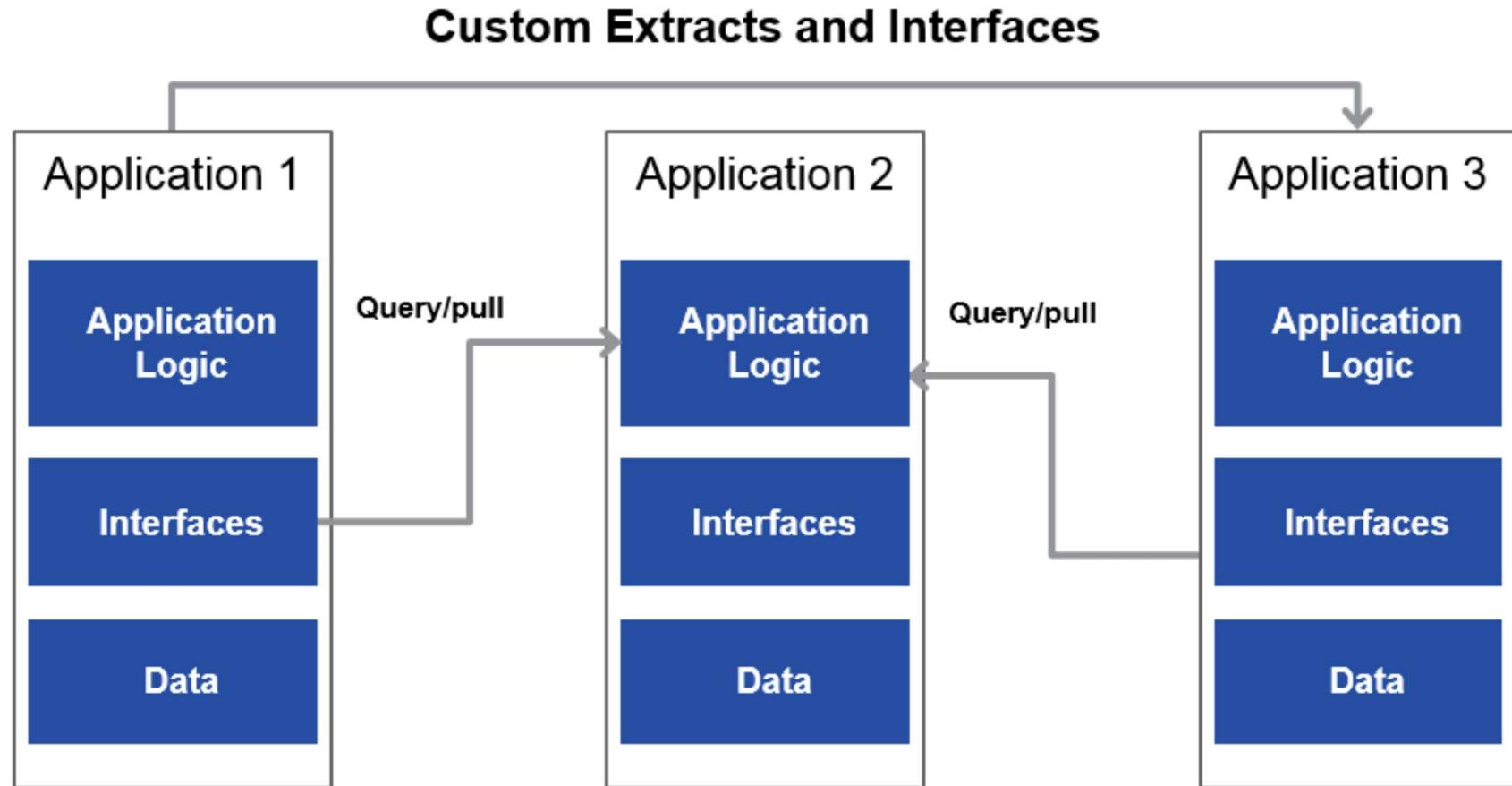
THE NEED FOR TIMELY INFORMATION



WHAT IS REAL-TIME INTEGRATION?

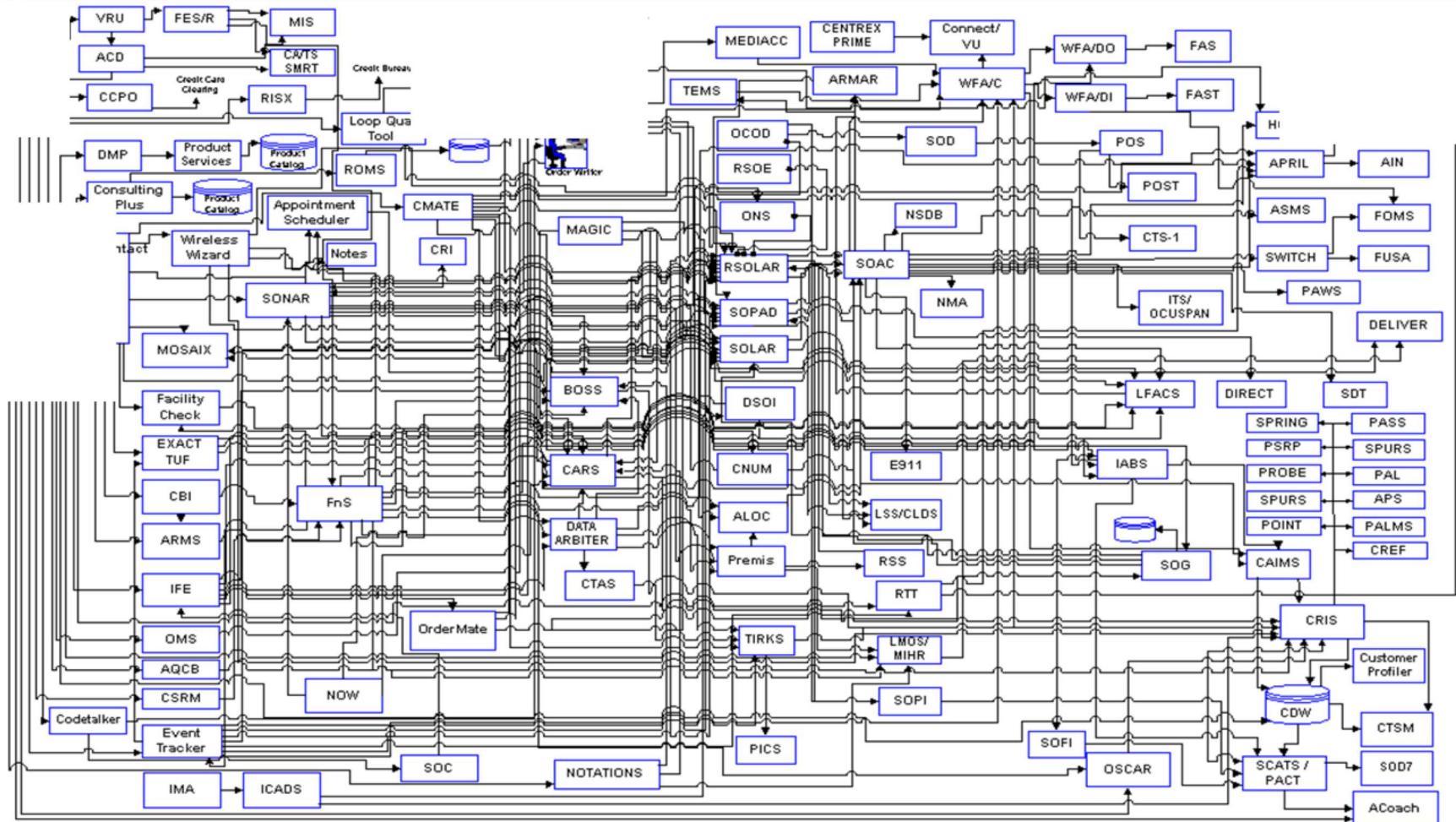


TRADITIONAL INTEGRATION



“Siloed” Applications, Embedded Interfaces and Transformation

TRADITIONAL INTEGRATION: RESULT...



CONSIDERATIONS

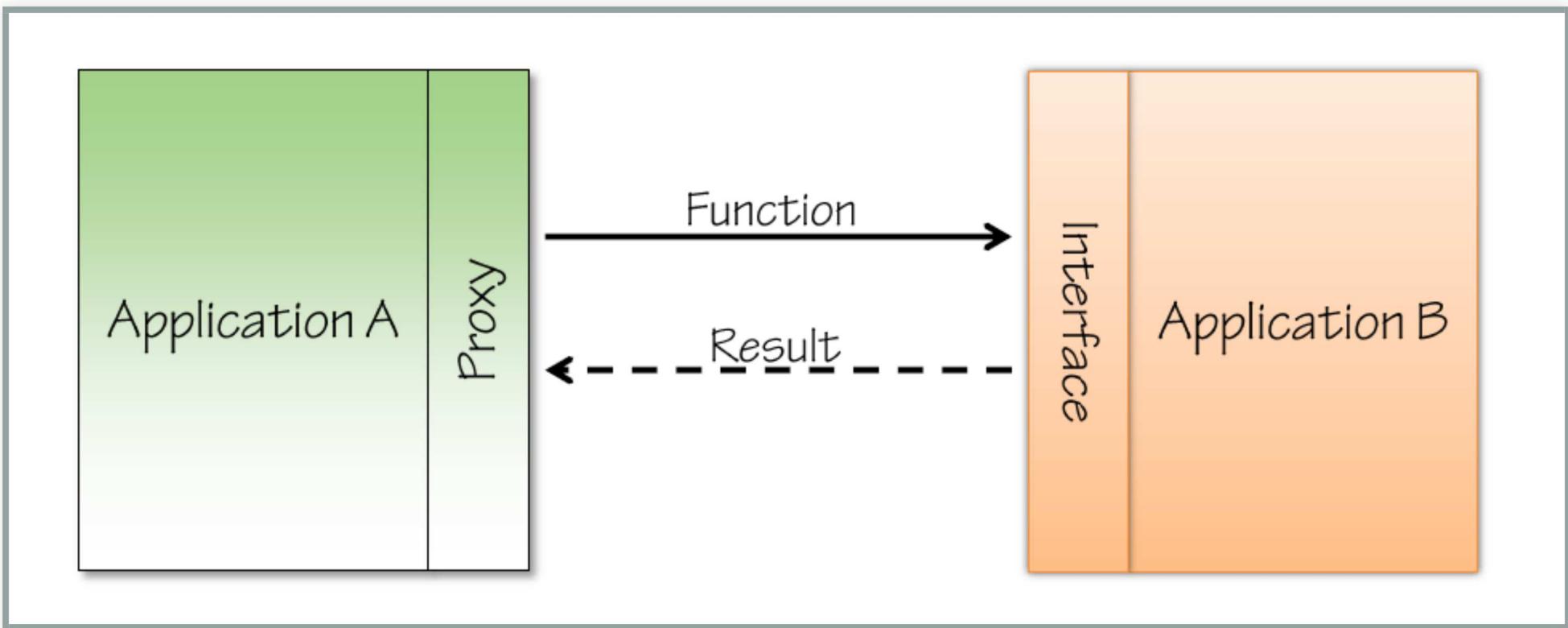
- Complexity of source and targets.
 - Data formats.
 - Level of quality.
 - Availability of source/target interfaces.
- Volume, velocity; and variety.
- Delivery requirements.
 - Real-time, near real-time, batch/scheduled.
- Loose coupling/reusability.
- Performance/Availability.

INTEGRATION PATTERNS

HOW CAN I INTEGRATE MULTIPLE APPLICATIONS SO THAT THEY WORK TOGETHER AND CAN EXCHANGE INFORMATION?

- Remote Procedure Call
- Asynchronous messaging
- Shared database
- File transfer

REMOTE PROCEDURE CALL



REMOTE PROCEDURE CALL

**DEVELOP EACH APPLICATION AS A LARGE-SCALE OBJECT OR COMPONENT WITH ENCAPSULATED DATA.
PROVIDE AN INTERFACE TO ALLOW OTHER APPLICATIONS TO INTERACT WITH THE RUNNING
APPLICATION.**

REMOTE PROCEDURE CALL: WHEN TO USE IT

- Want to share data among independent apps
- Data needs to stay with the source
- Service Oriented
 - Encapsulation
 - Abstraction
 - Interoperable
- Good for mashup applications

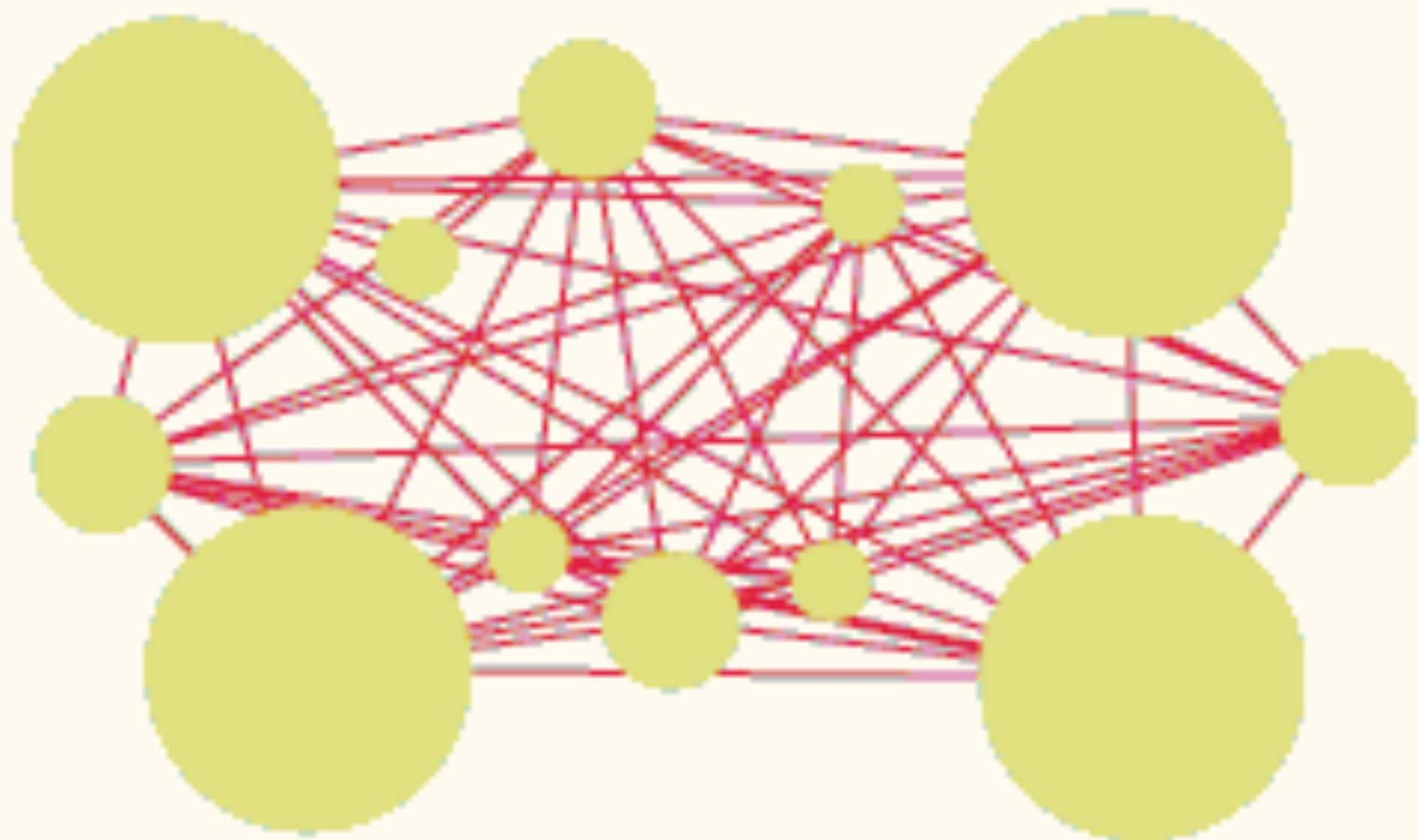
REMOTE PROCEDURE CALL: CONSIDERATIONS

- Coupling / Blocking
- Security
 - lacks uniformity
 - lowest common denominator
- Cloud
 - Http Oriented
 - AVOID Transactions

REMOTE PROCEDURE CALL: TOOLS

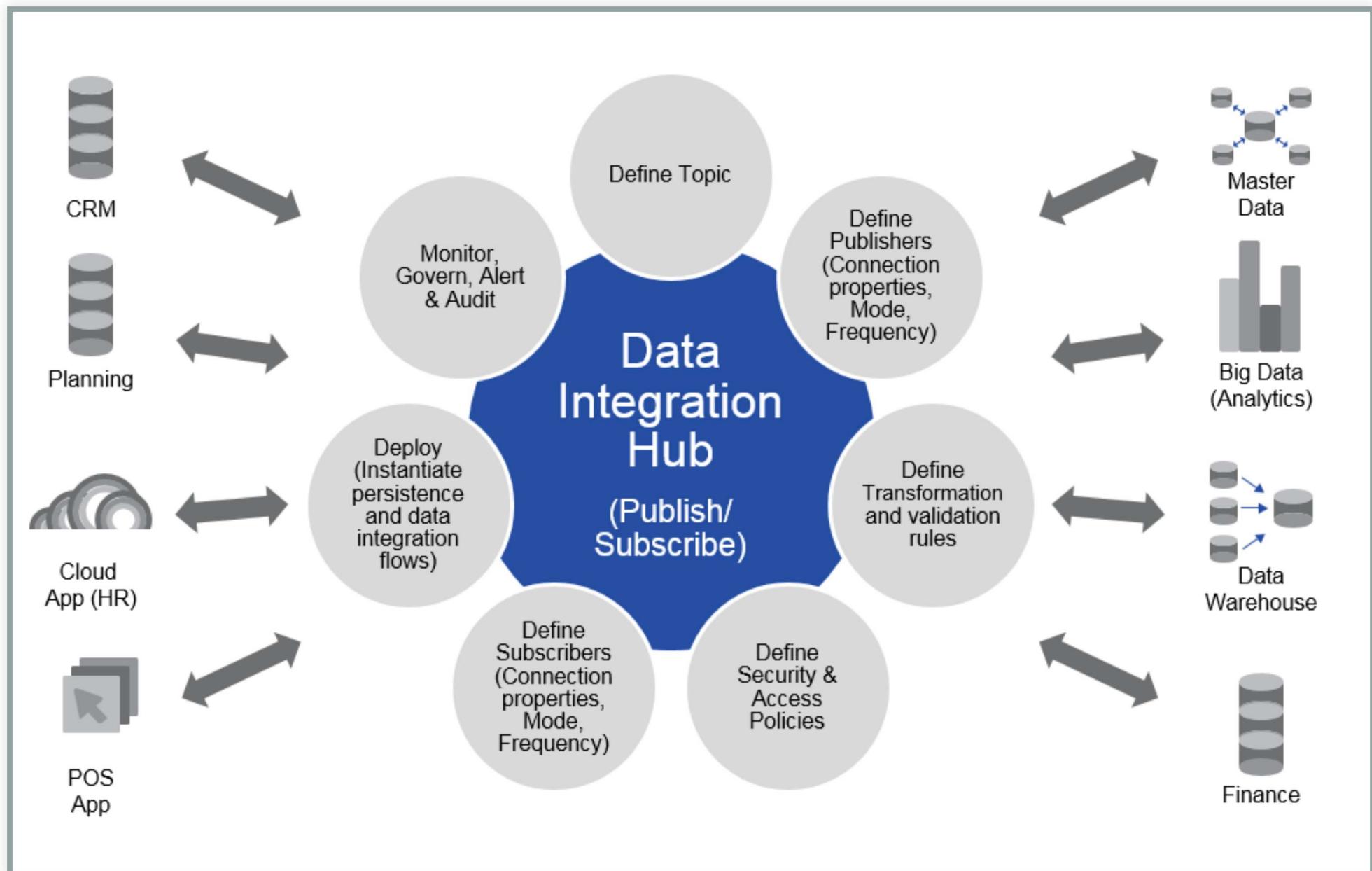
- Biztalk Server
- Mule Cloud Hub
- Node.js, .Net, Java, etc

EXAMPLE

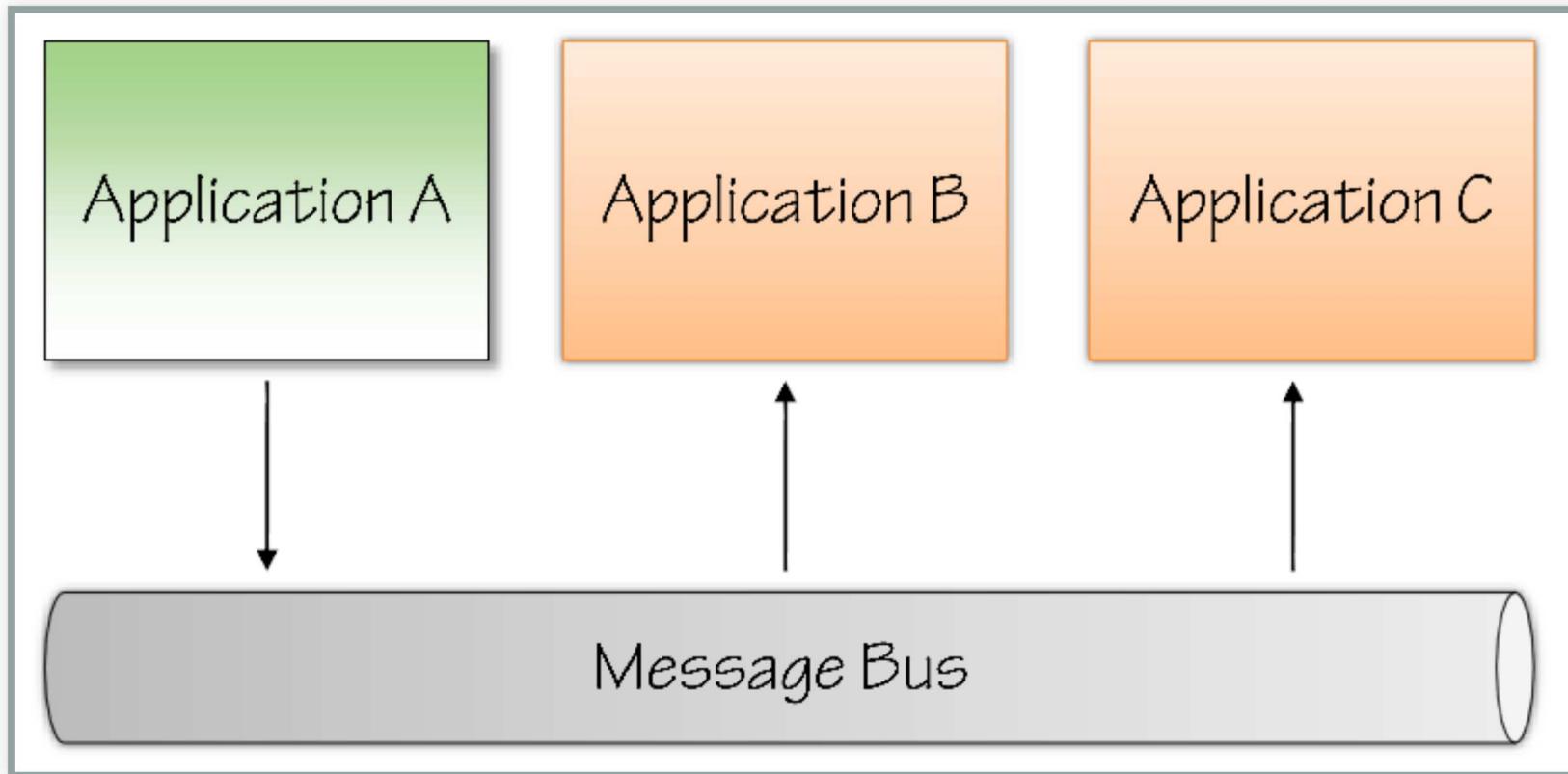


12 SYSTEMS REQUIRE 66 INTERFACES ((12*11)/2=66)

ASYNCHRONOUS MESSAGING



ASYNCHRONOUS MESSAGING



ASYNCHRONOUS MESSAGING

USE MESSAGING TO TRANSFER PACKETS OF DATA FREQUENTLY, IMMEDIATELY, RELIABLY, AND ASYNCHRONOUSLY, USING CUSTOMIZABLE FORMATS.

ASYNCHRONOUS MESSAGING: WHEN TO USE IT

- Fire and Forget
- Share data in responsive and scalable way
- Stateful or Stateless

ASYNCHRONOUS MESSAGING: CONSIDERATIONS

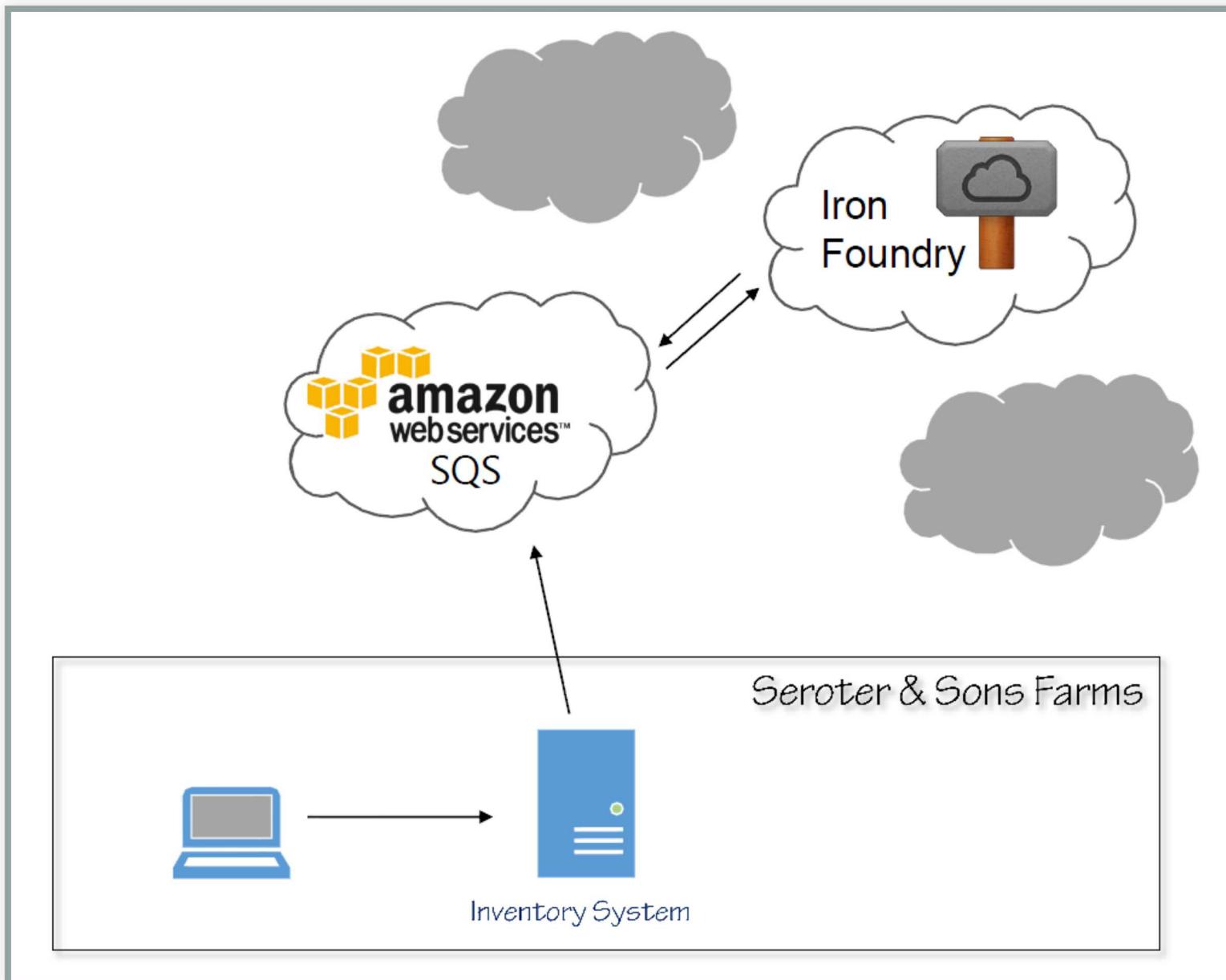
- Not Real Time
- Cloud
 - Poor Application Support
 - Security (firewall challenges)
 - Data limits

ASYNCHRONOUS MESSAGING: TOOLS

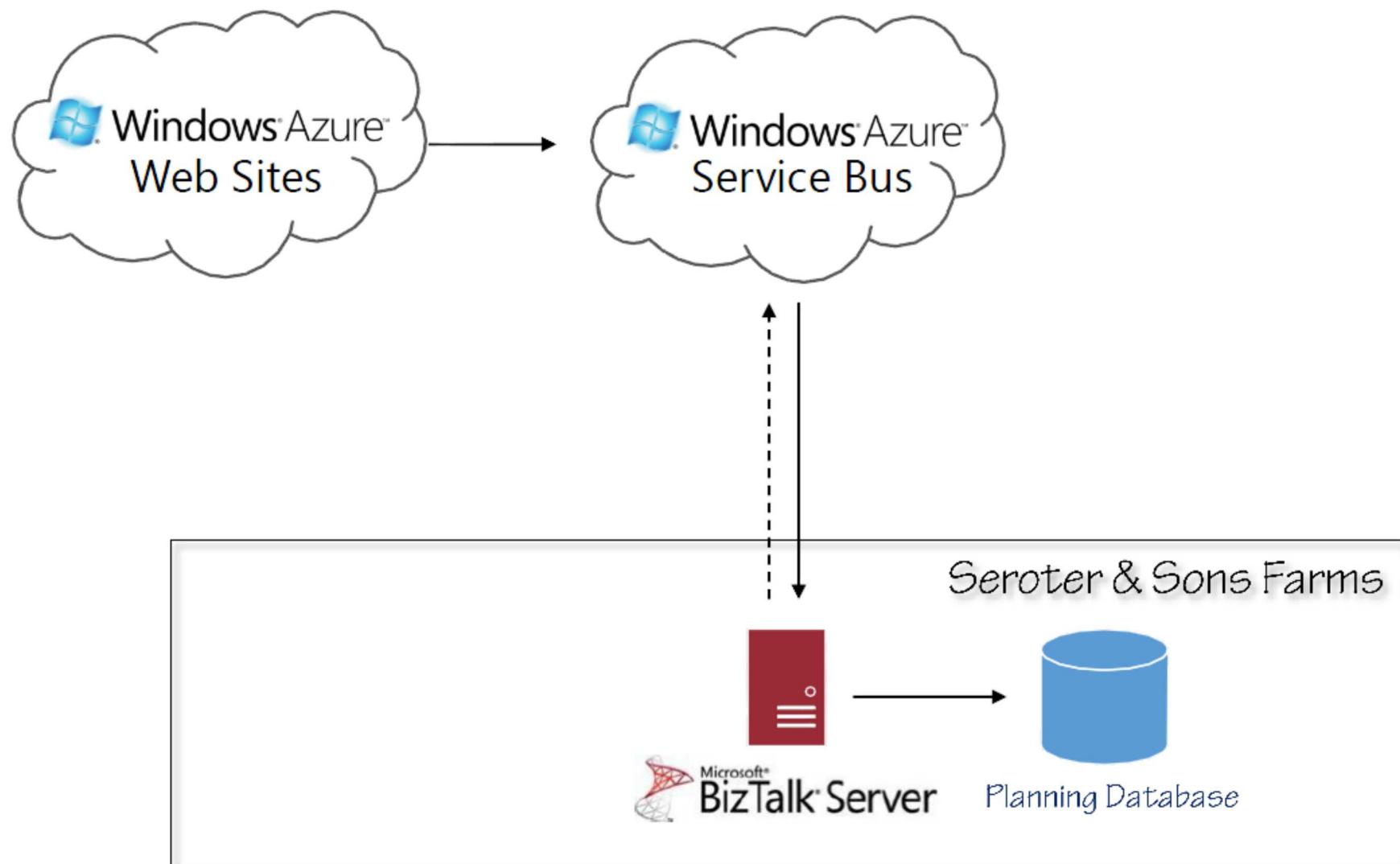
- Biztalk Server
- Mule ESB
- Azure Service Bus

Evaluation Parameter	Hub	Bus Product	ESB
Installation Effort	Less effort	Moderate effort	Moderate effort
Administration	Easy	Complex	Complex
Cost	High	High	Low

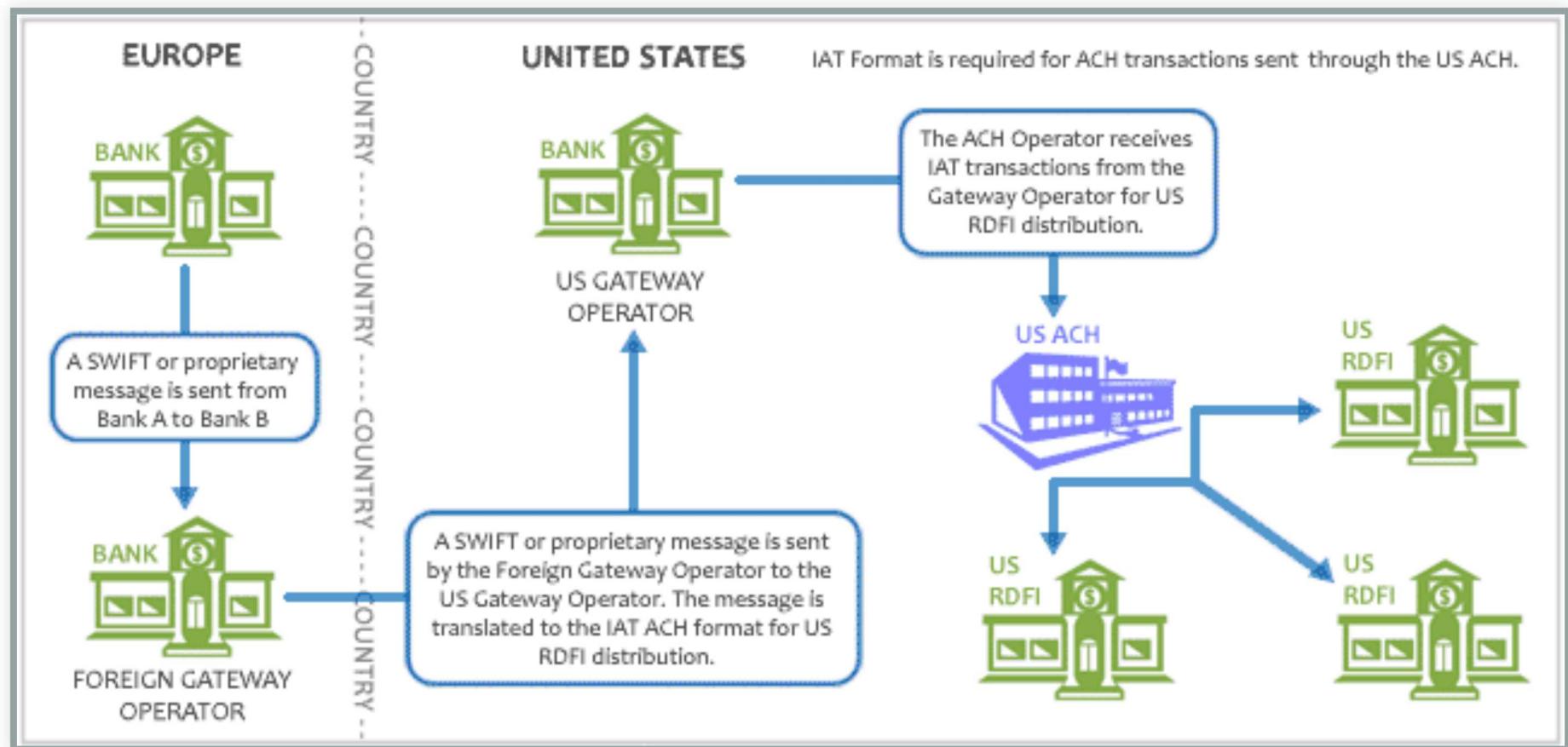
EXAMPLE: GROUND TO CLOUD



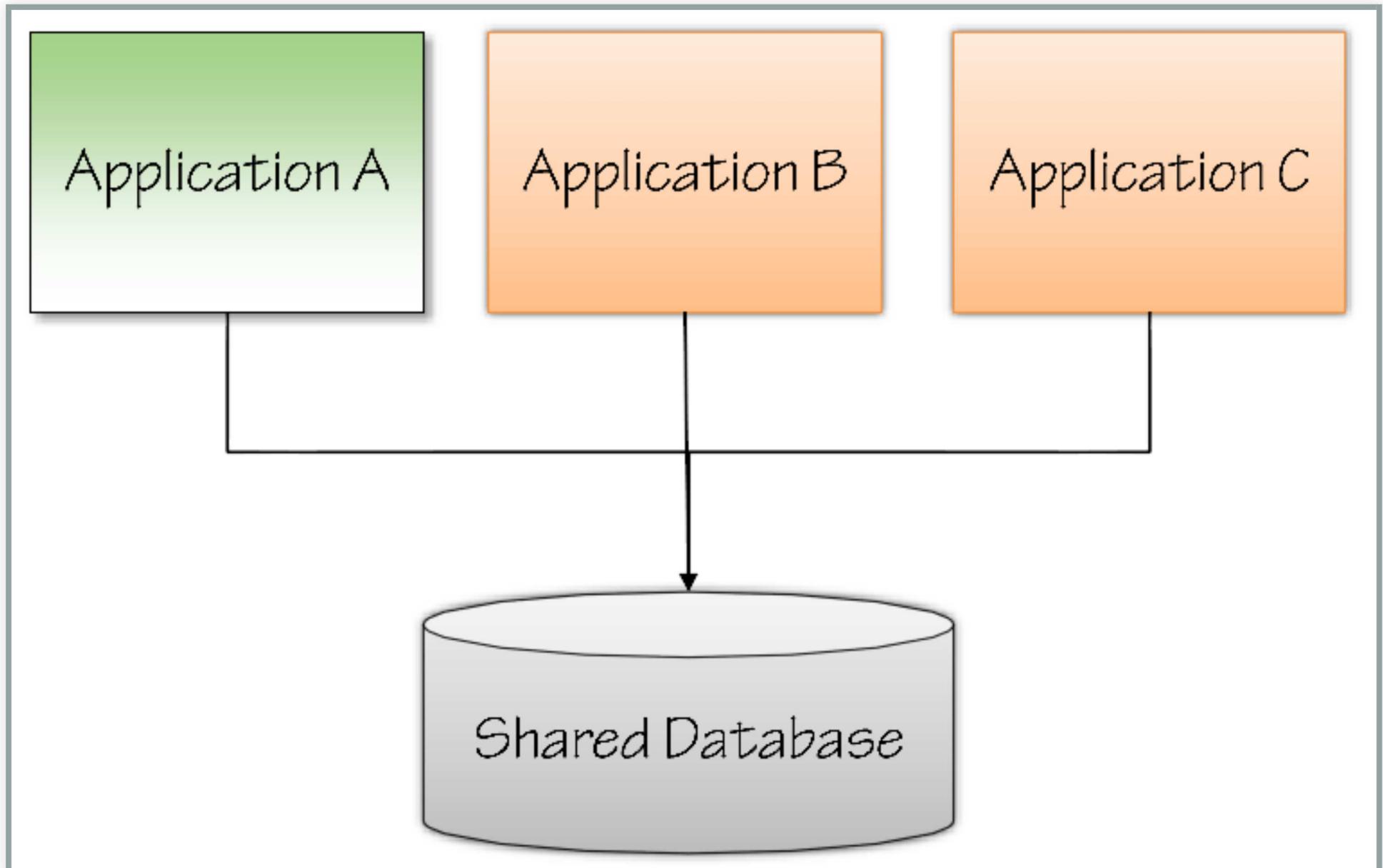
EXAMPLE: CLOUD TO GROUND



AUTOMATED CLEARING HOUSE



SHARED DATABASE



SHARED DATABASE

INTEGRATE APPLICATIONS BY HAVING THEM STORE THEIR DATA IN A SINGLE SHARED DATABASE.

SHARED DATABASE: WHEN TO USE IT

- Fastest method
- Dependent systems are all consistent
- Want to impose single access strategy

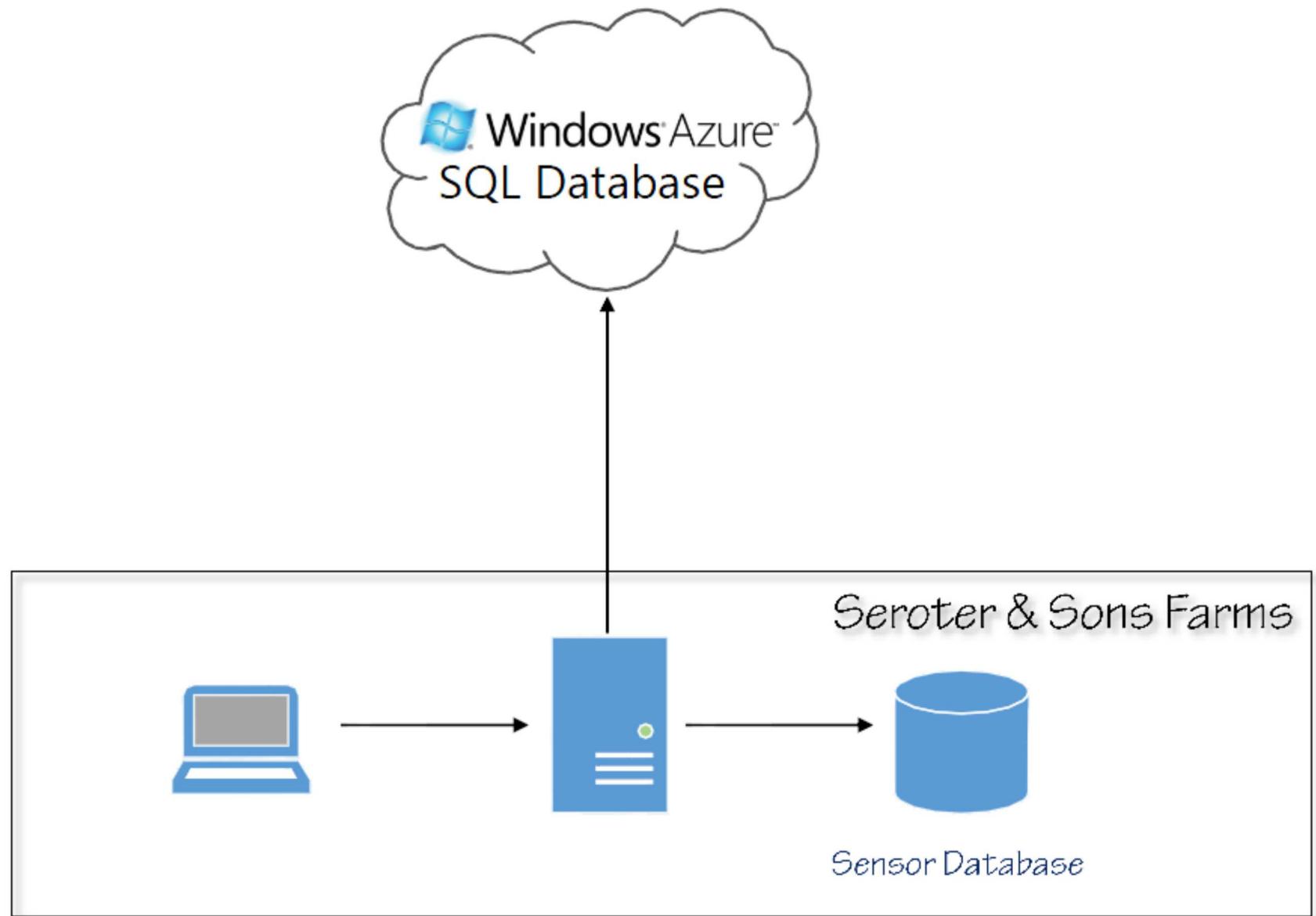
SHARED DATABASE: CONSIDERATIONS

- Difficulty of common format
- Tight Coupling
- Poor Performance over distances
- Scalability / Distribution Concerns
- API or Query based

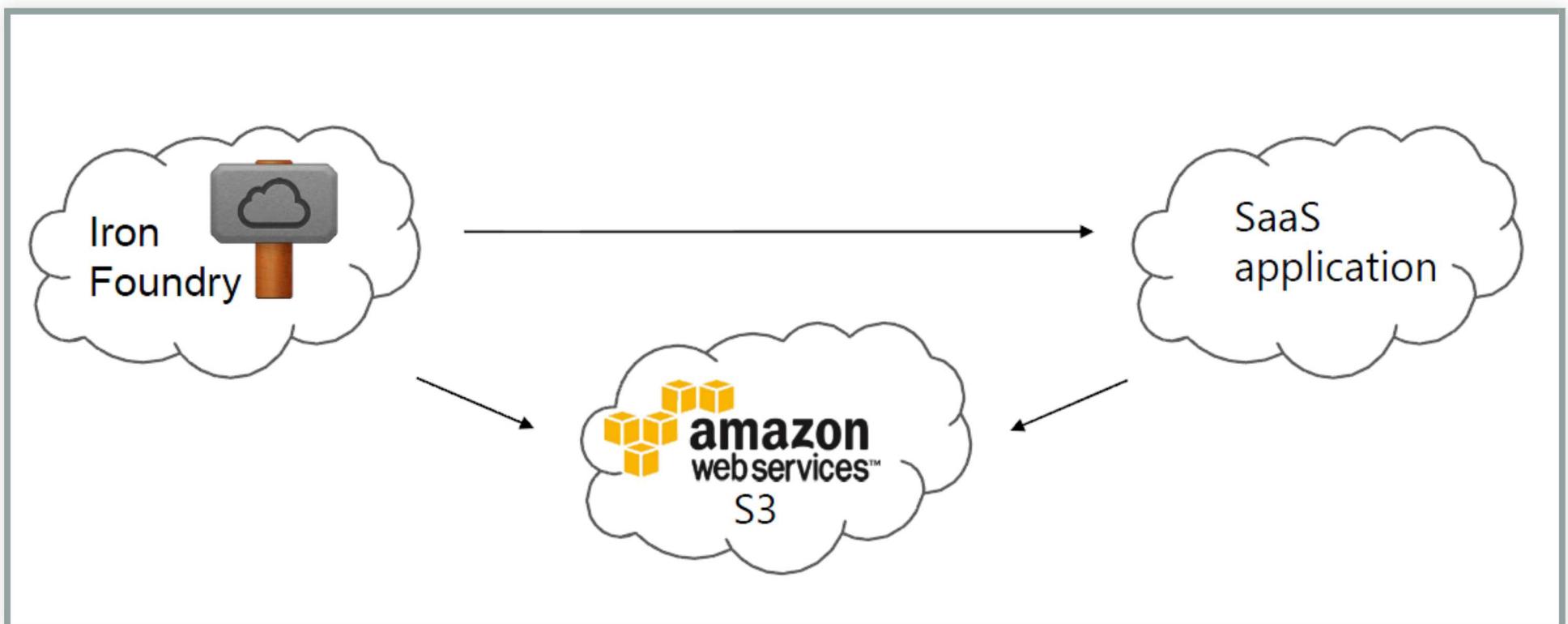
SHARED DATABASE: TOOLS

- Hosted Database
- Vendor based storage products

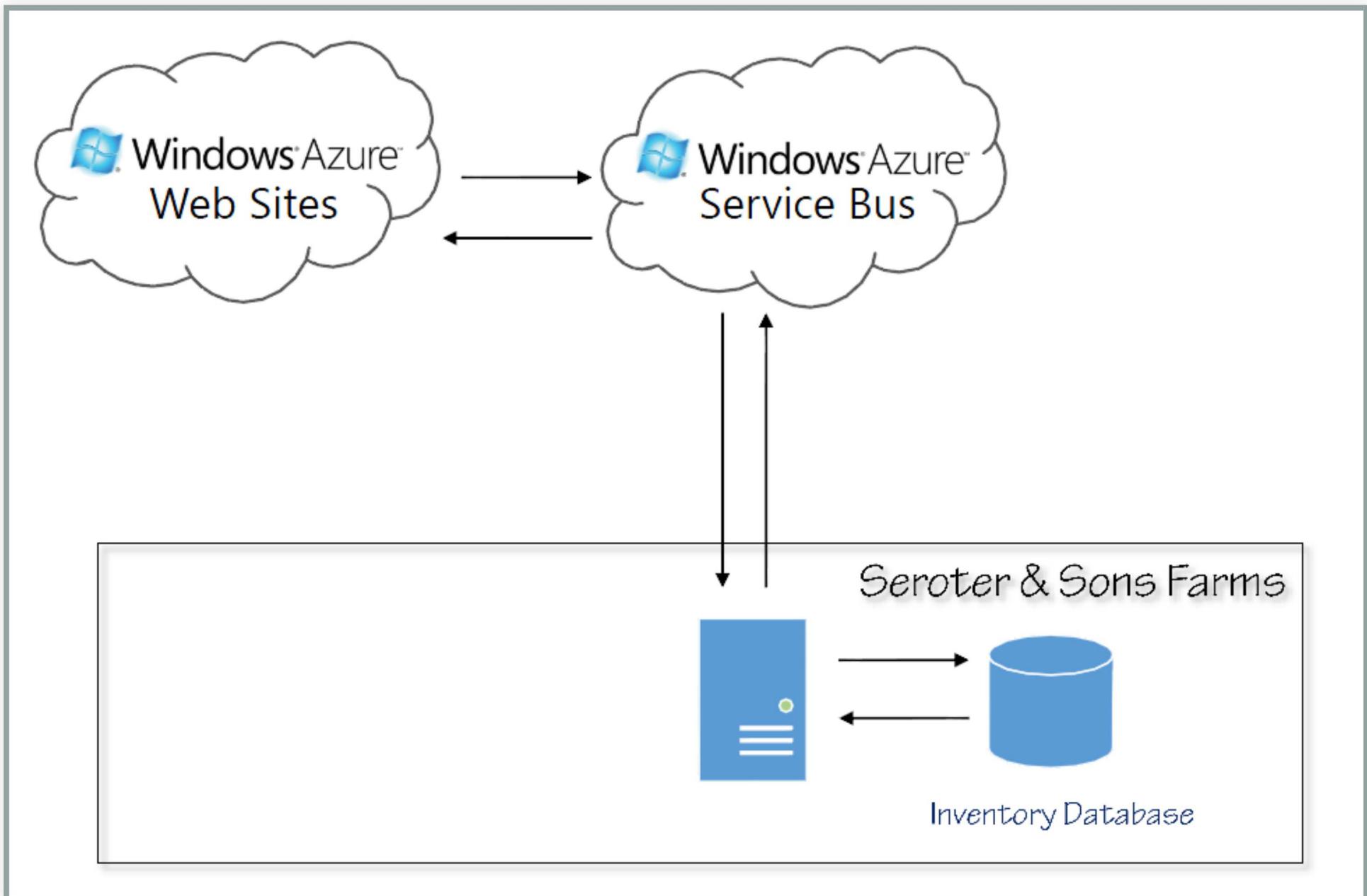
EXAMPLE: GROUND TO CLOUD



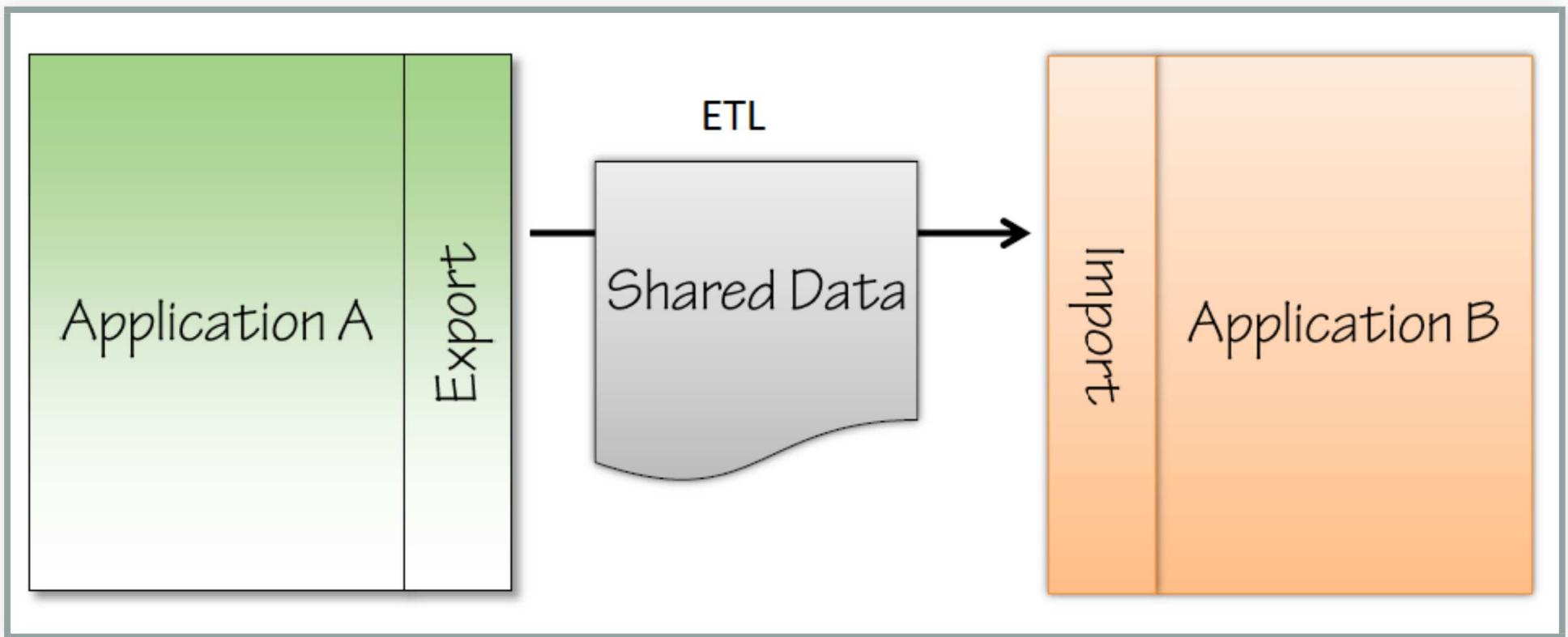
EXAMPLE: CLOUD TO CLOUD



EXAMPLE: CLOUD TO GROUND



FILE TRANSFER



FILE TRANSFER

HAVE EACH APPLICATION PRODUCE FILES CONTAINING INFORMATION THAT OTHER APPLICATIONS NEED TO CONSUME. INTEGRATORS TAKE THE RESPONSIBILITY OF TRANSFORMING FILES INTO DIFFERENT FORMATS. PRODUCE THE FILES AT REGULAR INTERVALS ACCORDING TO THE NATURE OF THE BUSINESS.

FILE TRANSFER: WHEN TO USE IT

- Delayed Data is OK
- Large Datasets
- No / Limited Access to source system
- Data Transforms Needed
- Data Warehouse is destination

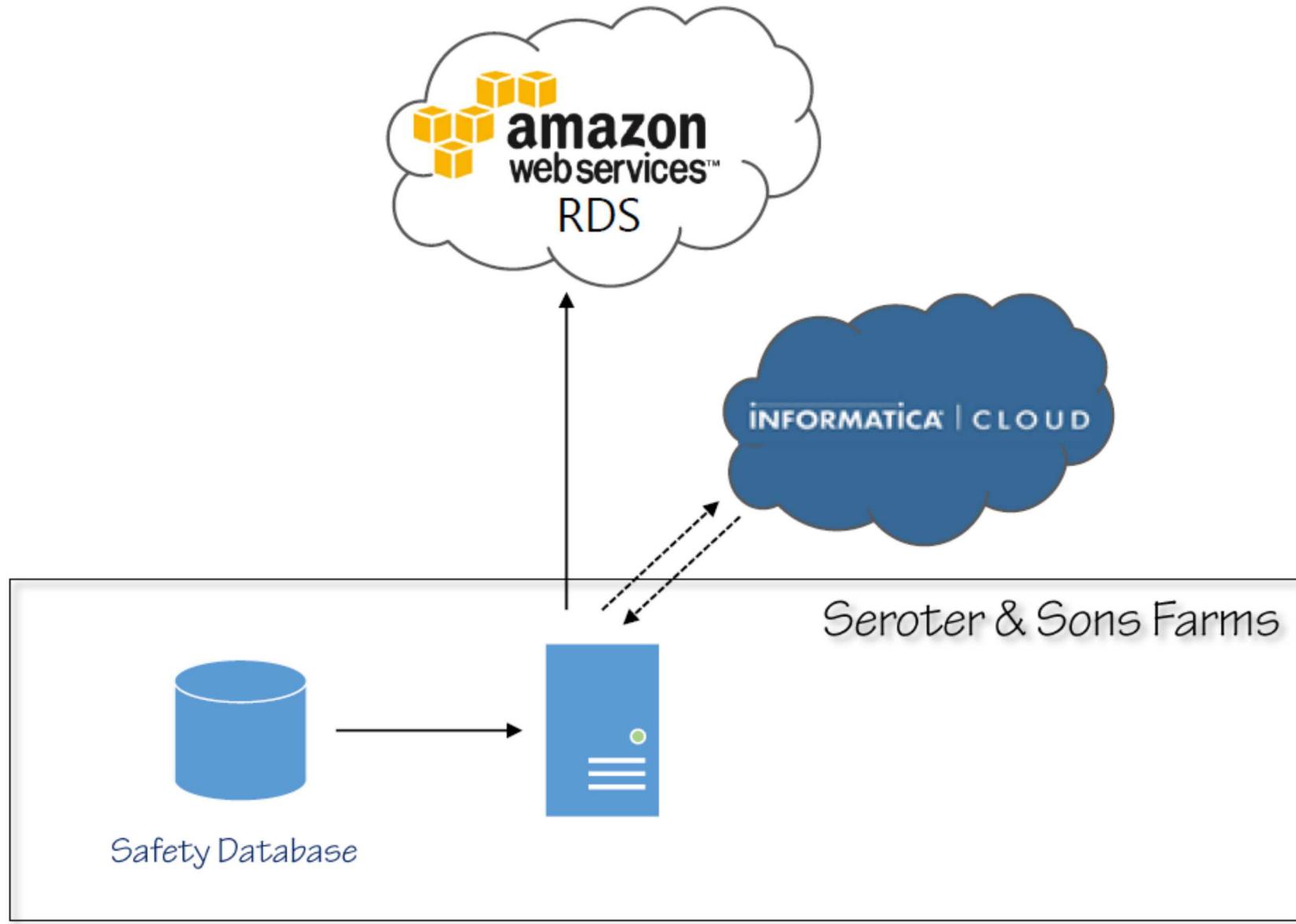
FILE TRANSFER: CONSIDERATIONS

- Time needed to process
- Complex to change / deploy
- Not suited for unstructured data
- Bandwidth constraints

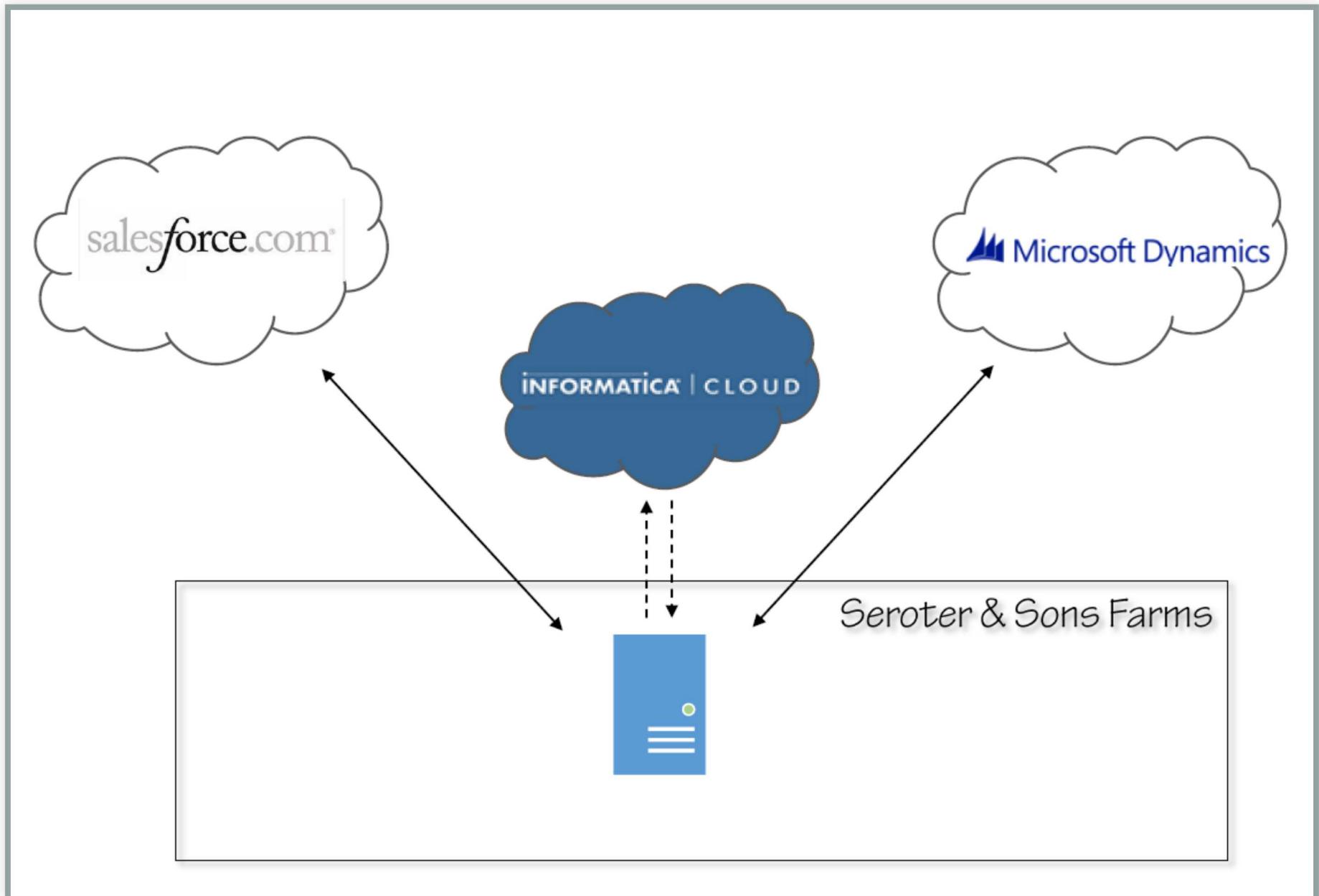
FILE TRANSFER: TOOLS

- SSIS
- Informatica
- Data Stage

EXAMPLE: GROUND TO CLOUD



EXAMPLE: CLOUD TO CLOUD



IPAAS

WHY BUSES DONT FLY IN THE CLOUDS



IPAAS: WHEN TO USE IT

- Integration with SaaS is a requirement
- Hybrid Cloud or Full-Cloud Solution
- Out of box flow / management/ monitoring
- Reusable templates
- Common Adapters

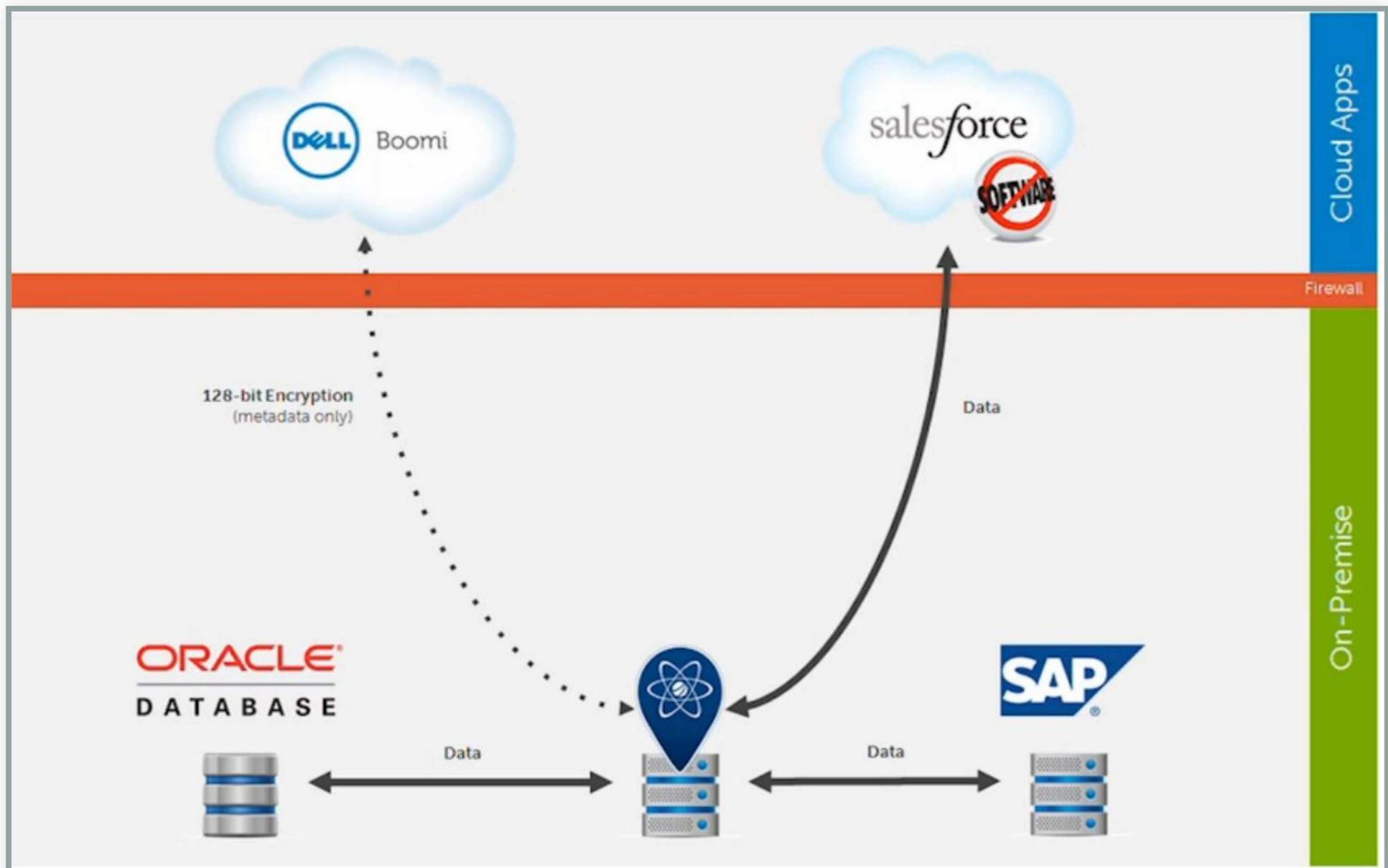
IPAAS: CONSIDERATIONS

- SLA
- Security
- Tool Specific Integrations

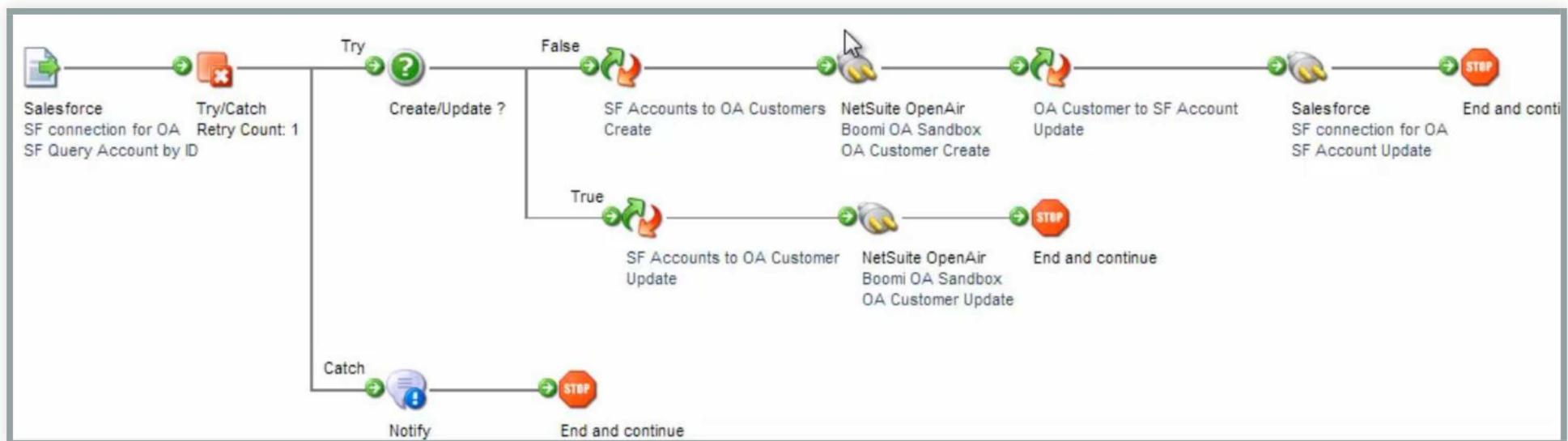
IPAAS: TOOLS

- Dell Boomi
- Snaplogic
- Informatica
- MuleSoft

EXAMPLE DELL BOOMI



EXAMPLE DATA FLOW



ANATOMY OF A TYPICAL ETL PROJECT

PHASES OF AN ETL PROJECT

- Requirements
- Planning
- Analysis
- Implementation

REQUIREMENTS

- Functional
 - Input
 - Output Data
 - Transformations
- Non-functional
 - Time Considerations
 - Availability
 - Frequency and Latency
 - Data Quality

PLANNING

- Biggest risk is unexpected data quality issues
- Iterate

ANALYSIS - EXTRACTION

- **Where** is the data?
- **What** does the data look like?
- **How much** is there?
- **How fast** is the data created or changed?

ANALYSIS - TRANSFORMATION

- Validation
 - Who is the trusted source?
 - What about data that can't be validated?
- Restart Logic
 - Volume dictates methods

ANALYSIS - LOAD

- Everything You Considered in Extract
- Plus Historical Changes
 - SCD Type 1 - Current State Only
 - SCD Type 2 - Full History
 - SCD Type 3 - Previous (or Original) + Current

IMPLEMENTATION

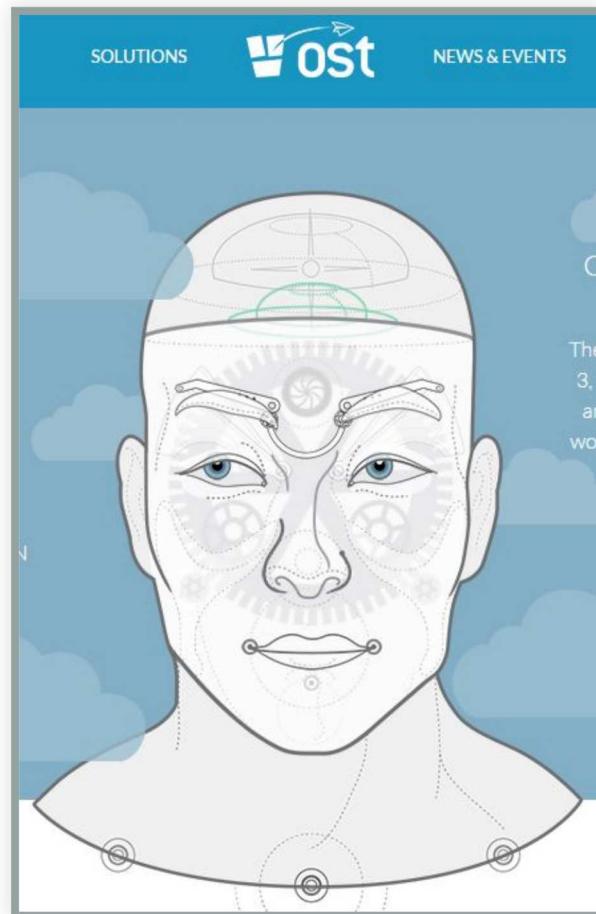
- Enforce a Standard
 - Naming Conventions
 - Consistent Patterns
- Testing
 - Unit Testing
 - Production Data

EXAMPLE OF ETL STANDARDS

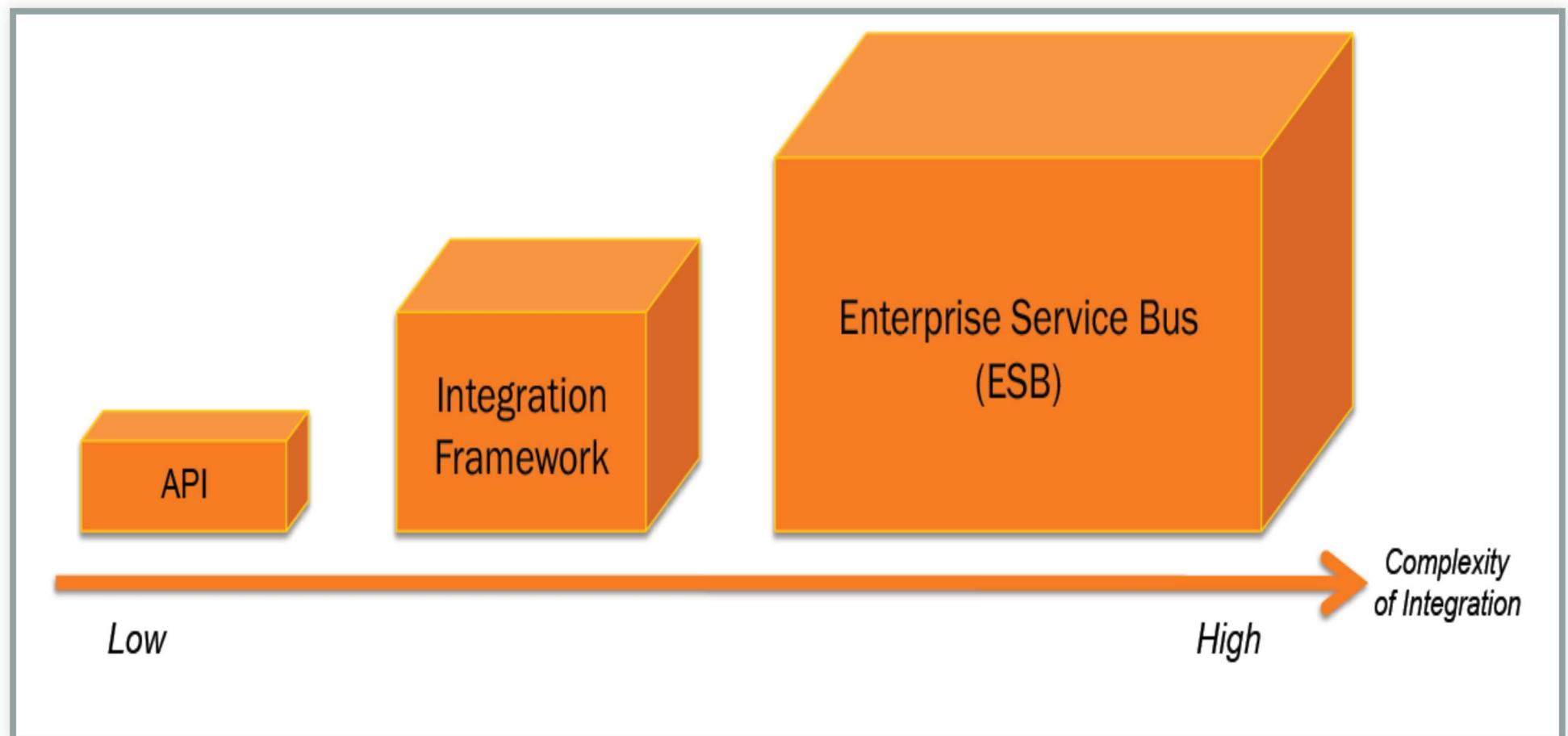
- Extract data needed for business logic locally
- Related: Transformations steps shouldn't need a database connection
- "Land" data often
- Never throw away data invisibly

BRINGING IT HOME

EATING OUR DOGFOOD



KISS



KISS

- Start Small
- Leverage Cloud's Strong Suits
- Different APIs, Frameworks, and Products can help a lot

THANK YOU!