GridGain

In-Memory Computing Best Practices

Implement Real-Time Analytics, HTAP and Automation

Rob Meyer

Outbound Product Management

August 21, 2019

In-Memory Computing Best Practices Series
Changing the Order Based on Popular Demand

1. Build an in-memory computing foundation and roadmap
2. Add speed and scale to existing applications
3. Develop new applications, APIs and analytics
4. Implement real-time analytics, HTAP and automation
5. ~~Make the bridge to the cloud~~
6. ~~Implement HTAP and real-time data management~~
7. ~~Become event- and stream-driven~~
8. ~~Think different with machine and deep learning~~

GridGain

Digital Business Requires Real-Time Analytics and Automation

And ETL, Hadoop, Even Streaming Are Often Not Real-Time Enough

The Only Real-Time Architecture That Works for Digital Business Is Streaming and In-Memory Computing For HTAP On Top of Your Existing Analytics

HARVARD
BUSINESS SCHOOL

Leading digital companies generate **better gross margins**, **better earnings** and **better net income** than organizations in the bottom quarter of digital adopters.

Performance Metric	Digital Laggards *(Bottom 25% of enterprises)*	Digital Leaders *(Top 25% of enterprises)*
Three Year Average Gross Margin[1]	37%	55%
Three Year Average Earnings before Taxes	11%	16%
Three Year Average Net Income	7%	11%

GridGain

A Great Customer Experience
Doubles Revenues, More than Doubles Profits

Increasing customer retention rates by 5%
increases profits anywhere from 25% to 95%

Loyal Customers

Buy 67-84% more annually

Buy 9-10x more total

Are 20% of the customer base

Are 80% of ALL future revenues

Dissatisfied Customers

25% customers leave annually

67% leave because of poor service

91% leave without feedback

70% stay if you fix their issue

https://www.helpscout.net/75-customer-service-facts-quotes-statistics/
https://blog.accessdevelopment.com/2018-customer-loyalty-statistics

GridGain

ETL Does Not Support Real-Time Analytics or Automation

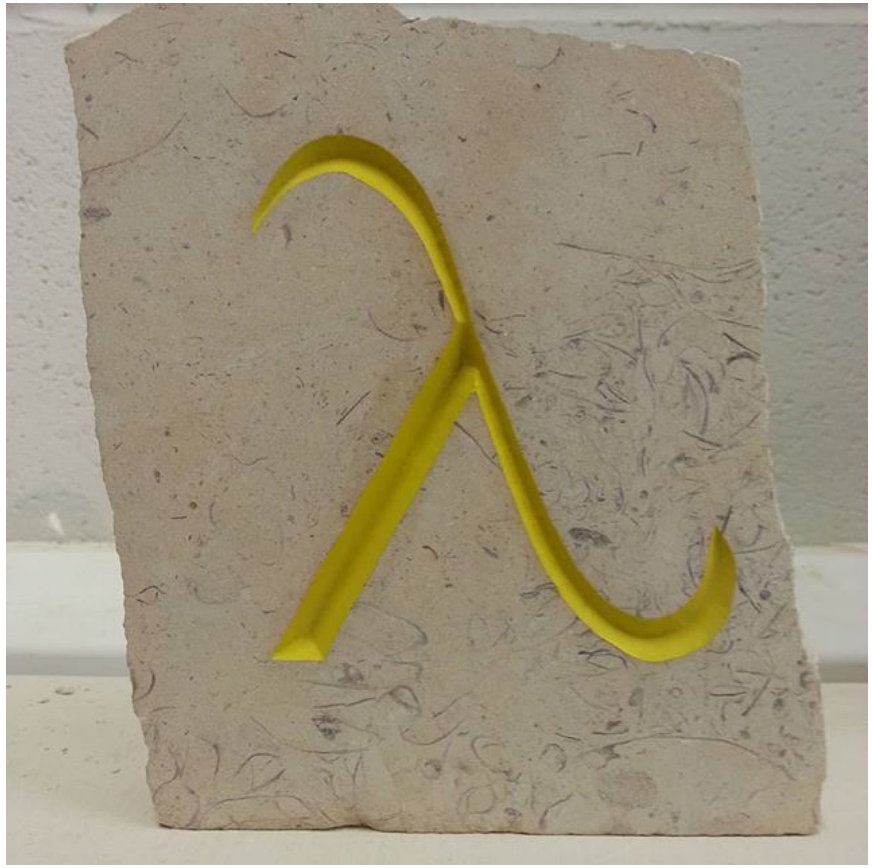
- ETL Creates a Delay
  - Offload analytics from apps
  - Reassemble data across apps
  - Ensure quality and consistency

- But Customer Issues Must Be Identified (Analyzed) and Fixed (Automated)
  The Moment They Happen to Improve The Experience

Batch and Real-Time Together Is A Necessity

- Batch and Real-Time Transactions and Analytics **Will Both Exist,** and **Will Both Feed Each Other**

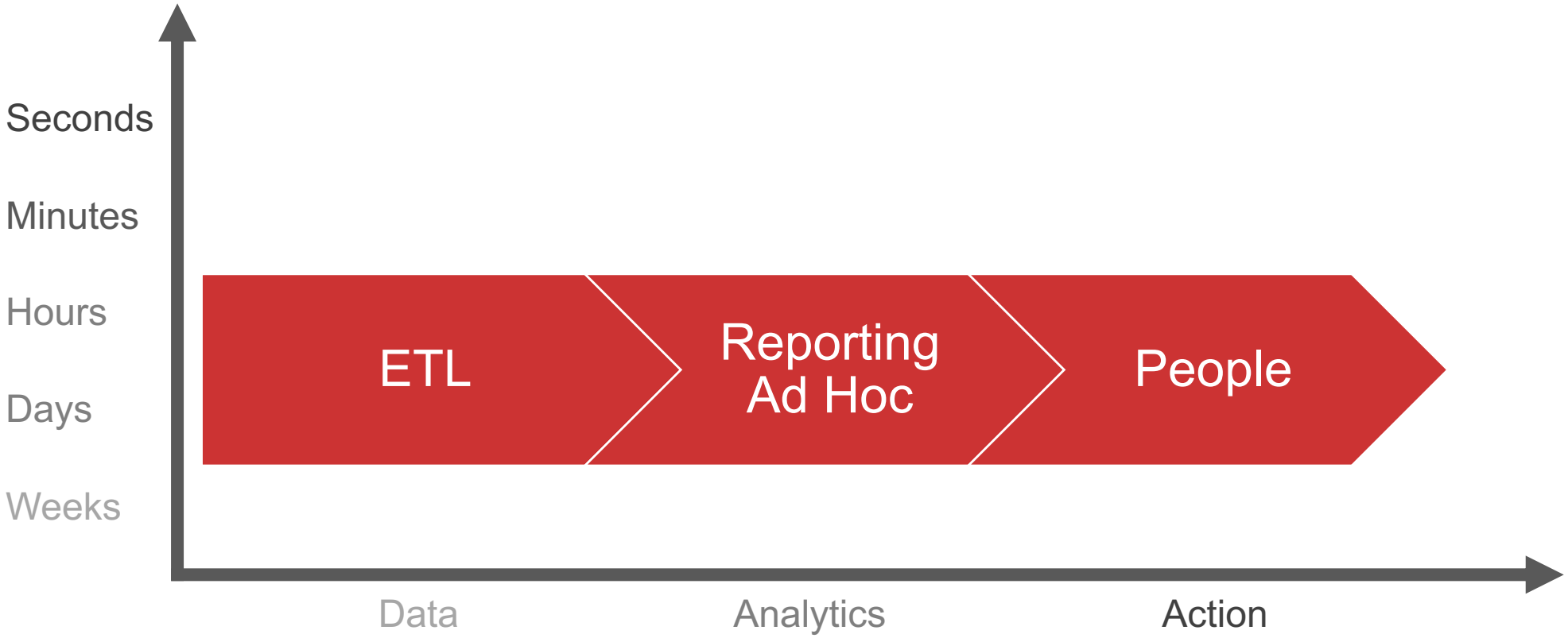
- **They Must be Integrated**
  - Real-time/historical thresholds for monitoring
  - Reporting comparisons
  - Real-time feeds



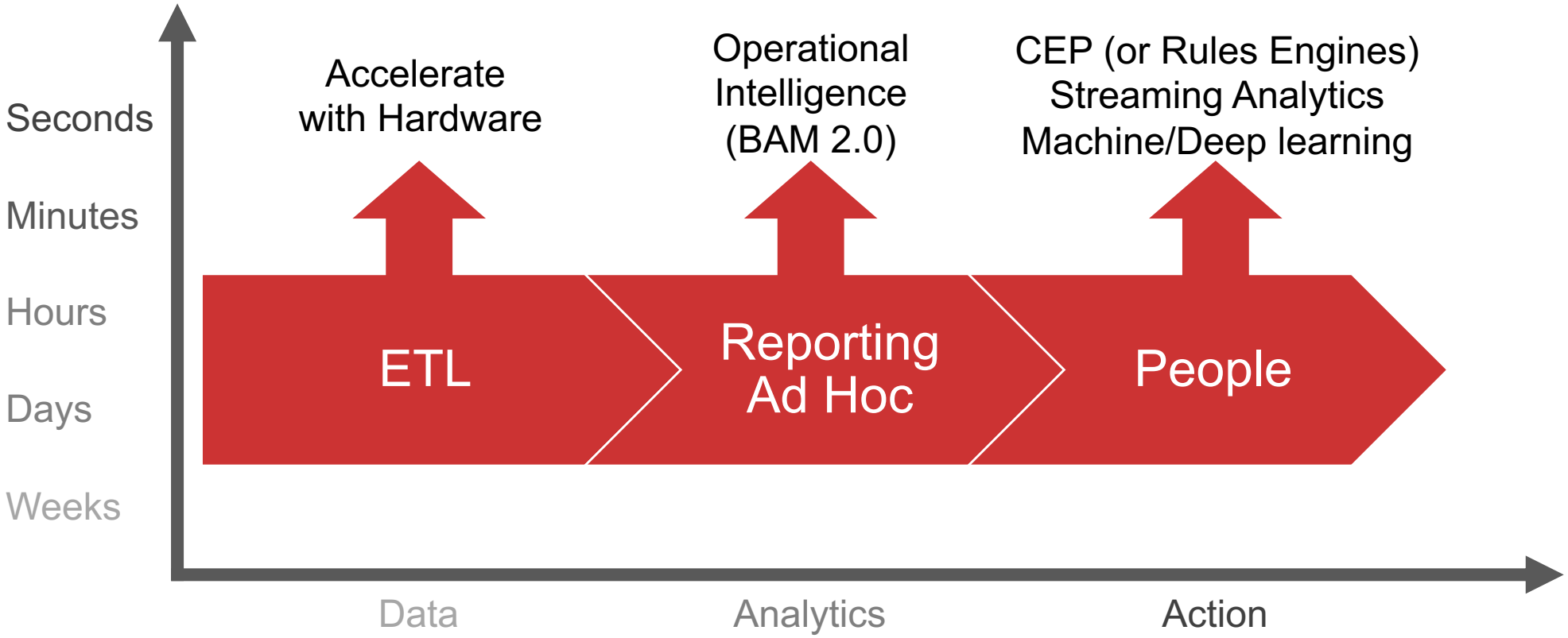
Best Practice 4.1

Architect for Batch and Real-Time Analytics

Traditional Analytics

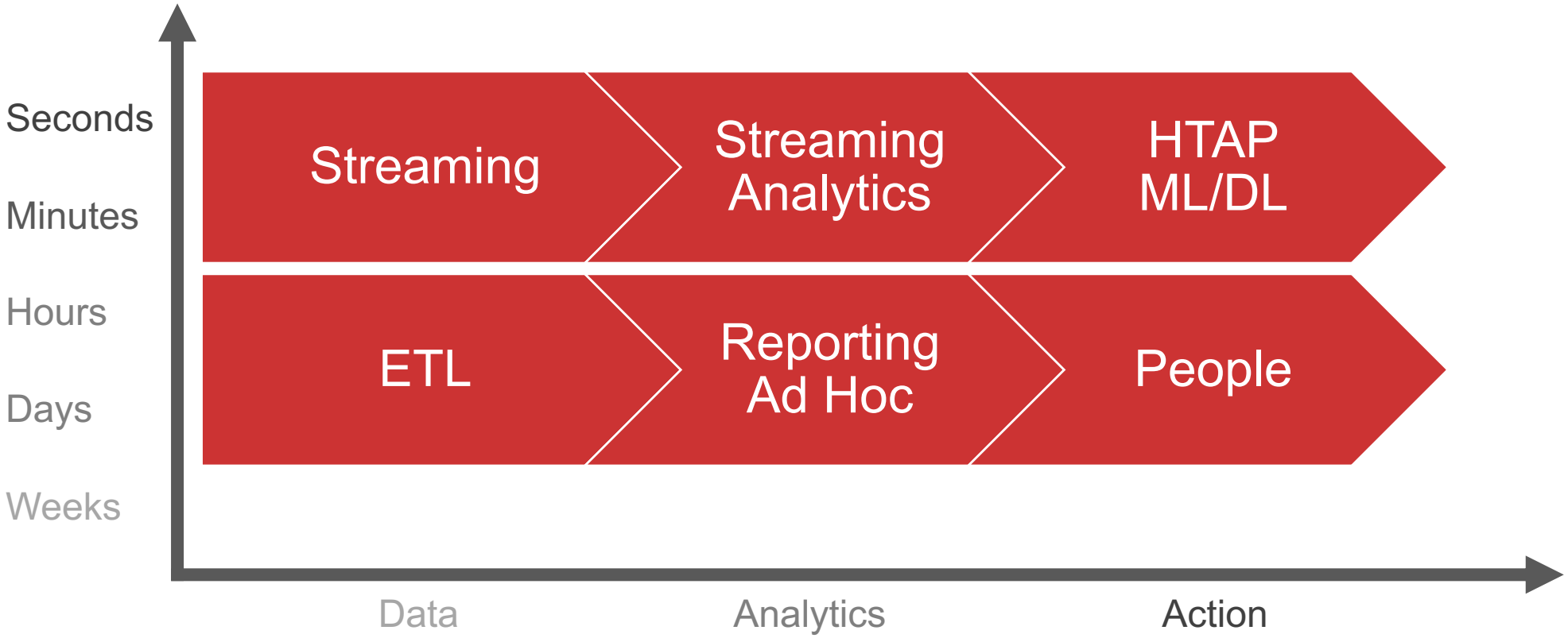


The History of ”One-Off” Real-Time Analytics



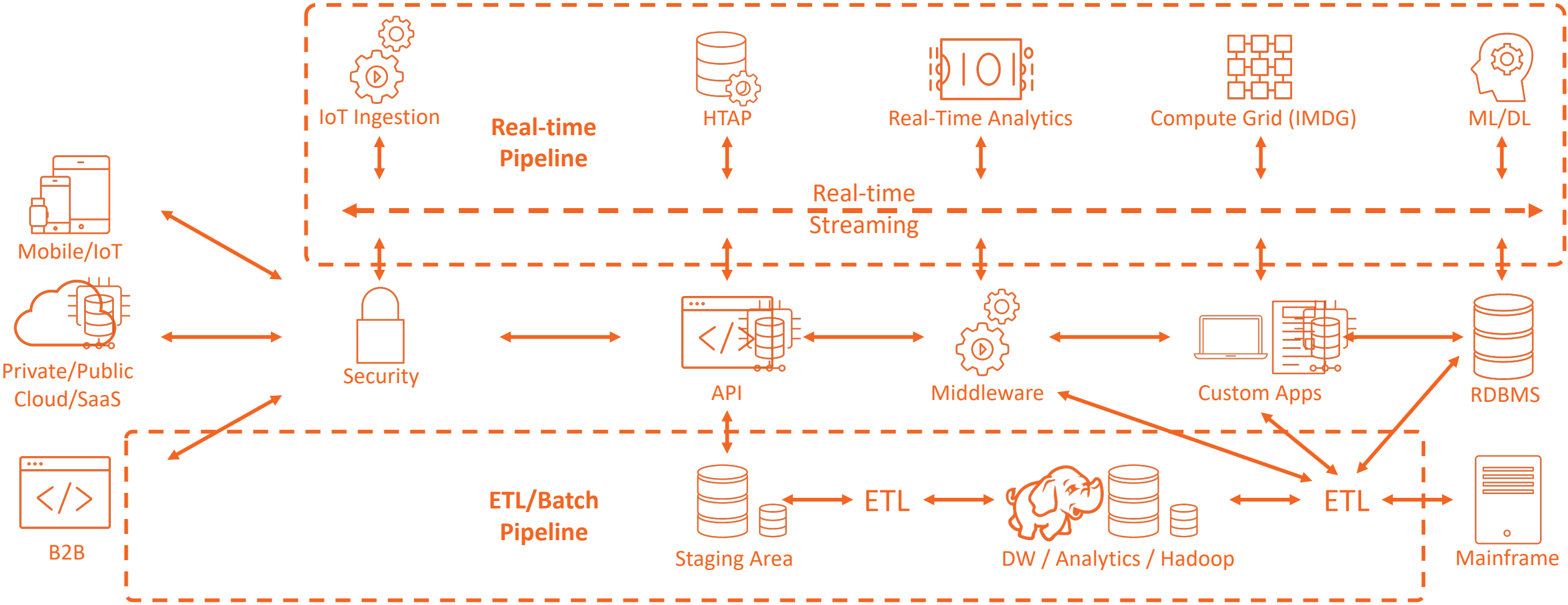
Unified Real-Time Analytics and Automation

For a Unified Real-Time Experience and Business



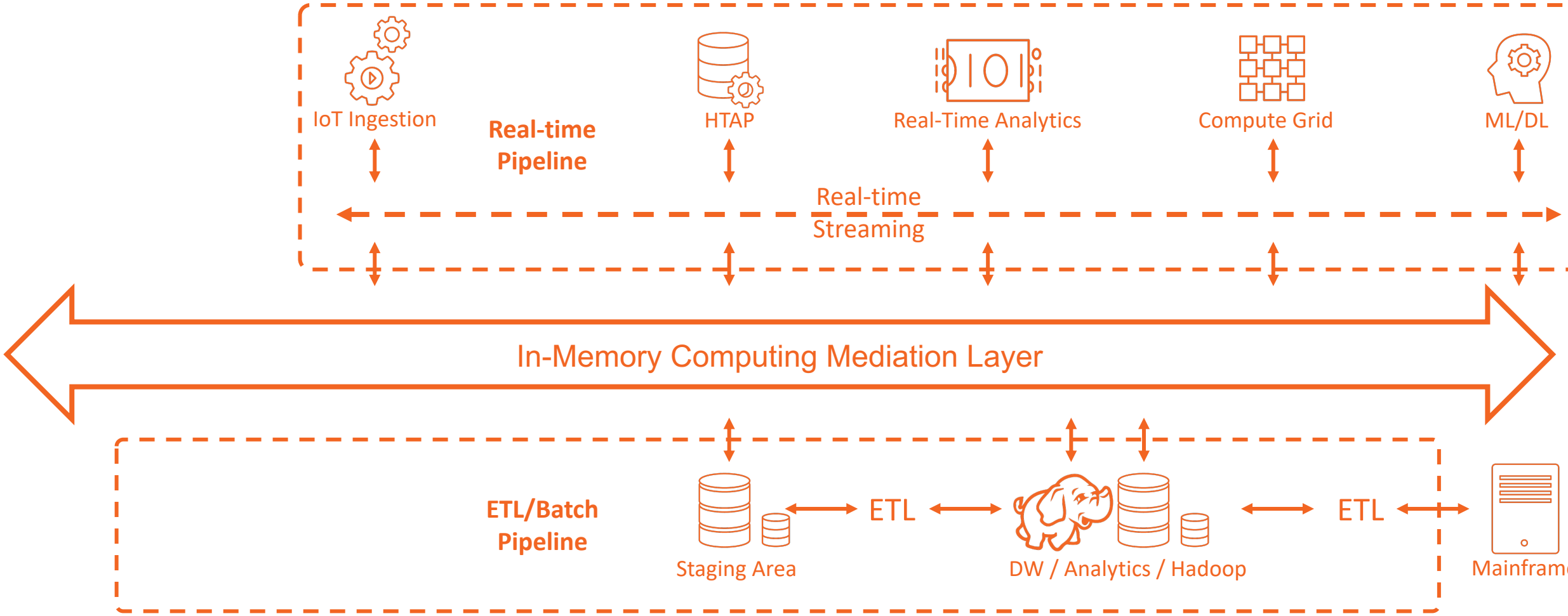
GridGain

Adding a Real-Time Data Stream



GridGain

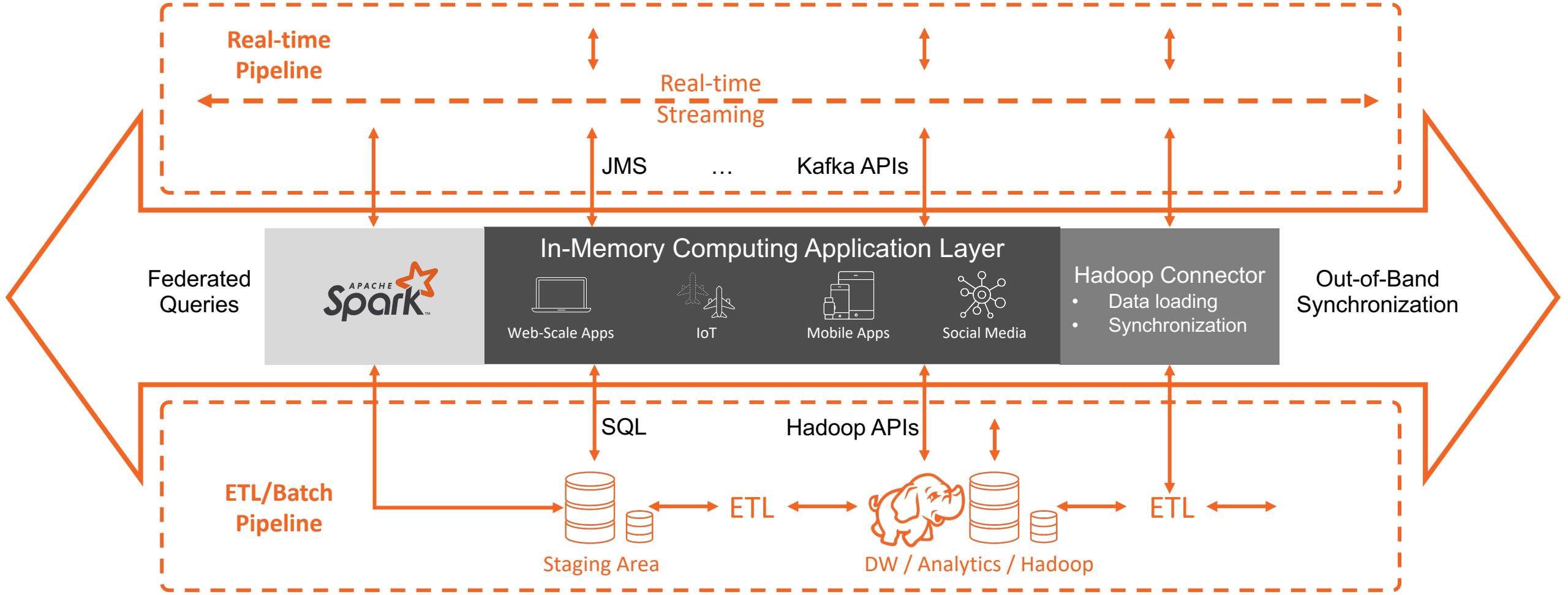
How To Integrate Real-Time and Batch Analytics
With an In-Memory Computing Mediation Layer



GridGain

In-Memory Computing Mediation Layer

The Different Types of Integration



GridGain

Best Practice 4.1 – Architect for Batch and Real-Time Analytics

- Build a New Real-Time Analytics Layer on Top of ETL
  - Batch-based ETL is here to stay (for a long time)
  - Avoid departmental real-time analytics projects that aren't integrated
- Integrate Batch and Real-Time Analytics Together with In-Memory Computing
  - Real-time integration using streaming (messaging) or APIs
  - SQL for data warehouses, Hadoop APIs to access Hadoop-based data lakes
  - Federated queries (using SparkSQL or SQL) across batch and real-time

Best Practice 4.2
Move Towards Integrated Lambda

Case Study: Wellington
Next Generation, Real-Time IBOR

WELLINGTON MANAGEMENT®

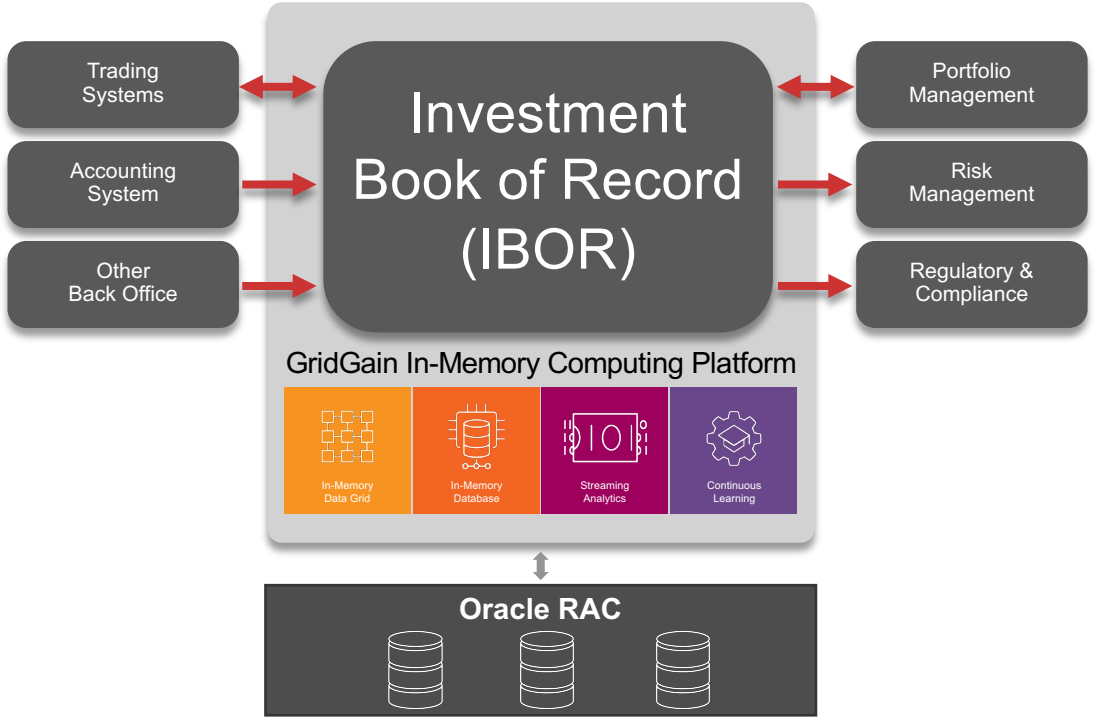
A top 20 worldwide asset management firm with over $1 trillion under management

Challenges

- Current systems no longer scaled to handle volumes
- Didn't comply with regulations following financial crisis
- Needed to introduce new asset classes faster

GridGain Enables

- Investment Book of Record (IBOR), a single real-time version of the truth for positions, exposure, valuations and performance for all customers, teams and trades.
- 10x performance gains, linear horizontal scalability
- Enabled transactions and analytics on a single platform
- Collocated computing for complex calculations, analytics

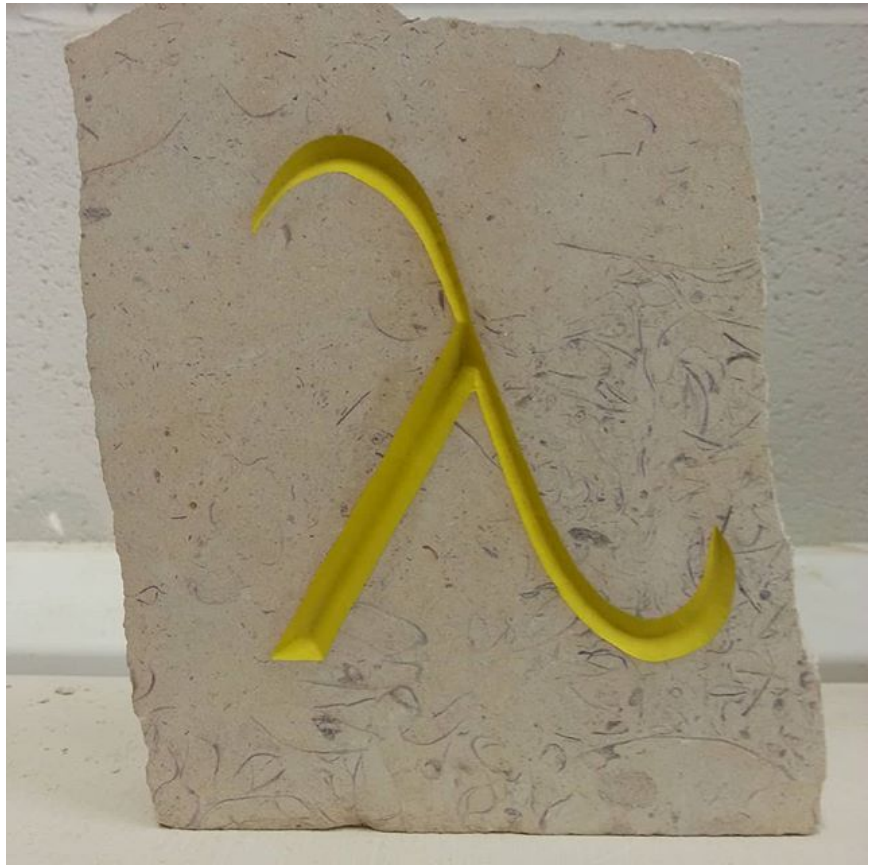


https://www.imcsummit.org/2018/us/session/optimized-memory-ibor-architecture-cloud-environment-using-apache-ignite

GridGain

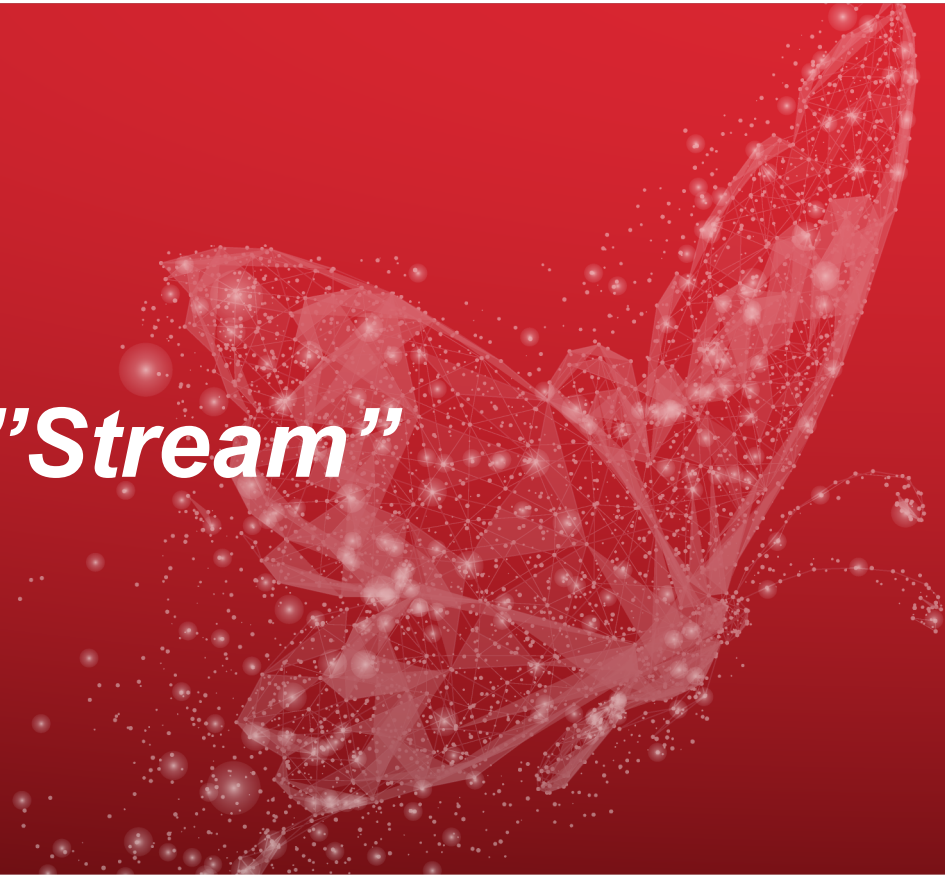
Best Practice 4.2 – Move Towards Integrated Lambda

- There are two types of Lambda architectures
  - Integrated: Real-time aggregates and feeds batch
  - Independent: different streams and logic for real-time and batch
- Most will start with Independent
  - Existing ETL is often done
- But you should implement integrate for new data sources and move to integrated over time
  - Eliminates duplicate logic
  - Helps make reporting more consistent



Best Practice 4.3

Build a New Real-Time Data ”Stream” One Engine at a Time



Case Study: Bandwidth
Call Detail Record (CDR) Processing

Telecommunications company that specializes in telephony (VoIP) and wireless solutions including 911 and SMS, IoT, and CCTV solutions that are integrated and accessed via an easy-to-use web-based portal

Challenges

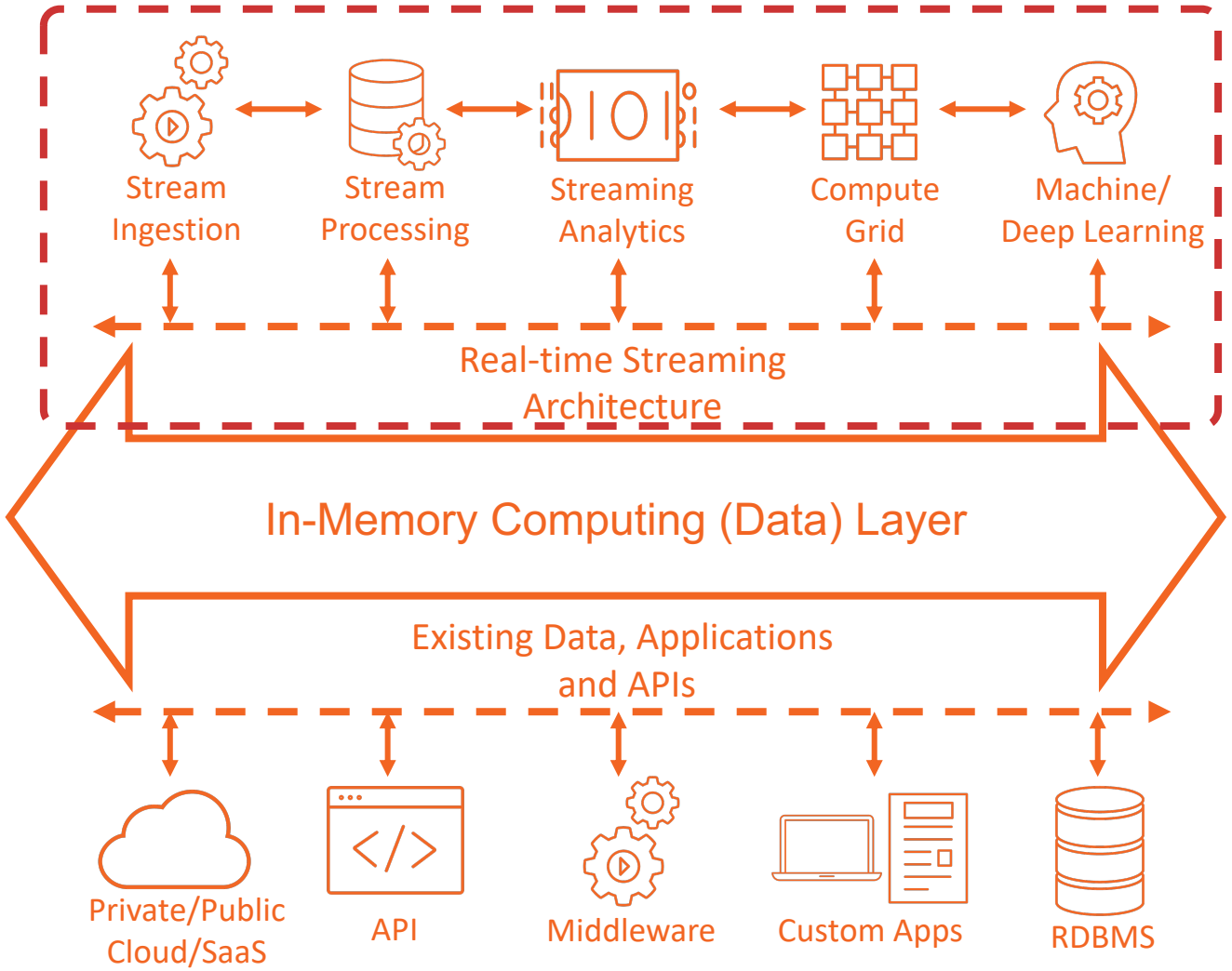
- Process all call detail records (CDR) in real-time during calls to create a unified CDR that can be used by all downstream applications including fraud detection, contacting the Police (911) and billing systems

GridGain Enables

- Real-time CDR processing, with GridGain processing Kafka streams, storing and correlating different records together, and then publishing unified CDRs to Kafka for downstream systems

Best Practice 4.3 – Build a New Stream One Engine at a Time

- Build an IMC layer
  - For real-time/historical data at rest
  - To add fast access and compute
- Design Flows Incrementally
  - Events: messages
  - Data: messages
  - Tasks: processing engines
- Group Projects Strategically
  - Reuse work in the next project
- Design as Microservices
  - Event-driven (real-time “asynch”)
  - Reusable (pub-sub)



GridGain

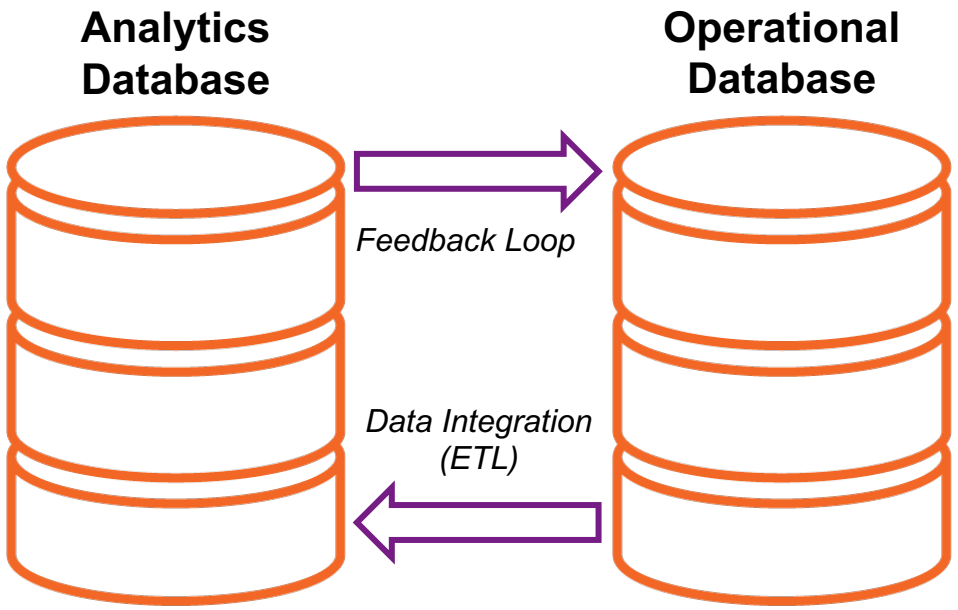
Best Practice 4.4
Architect for HTAP

Traditional IT Architectures Will Struggle to Meet the Needs of the Digital Enterprise

Traditional Architecture

Business Analysis

Analytics
Machine Learning
& Deep Learning/AI



Business Transactions

Operations

Manual Decision Making

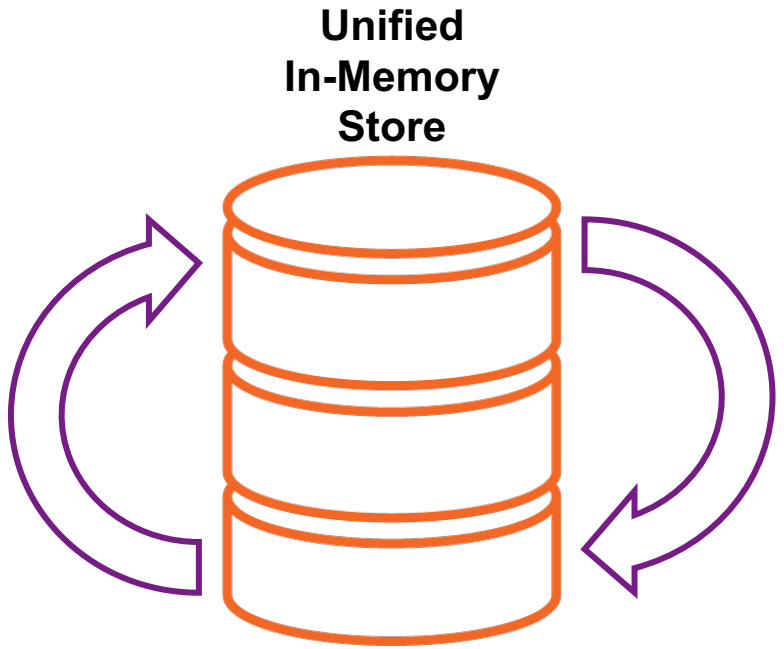
GridGain

IMC Platforms Will Power the Digital Enterprise

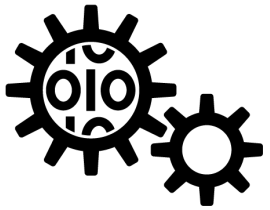
Unified In-Memory Computing HTAP/HOAP Architecture

Business Analysis & Transactions

Analytics
Machine Learning
Deep Learning/AI
& Operations



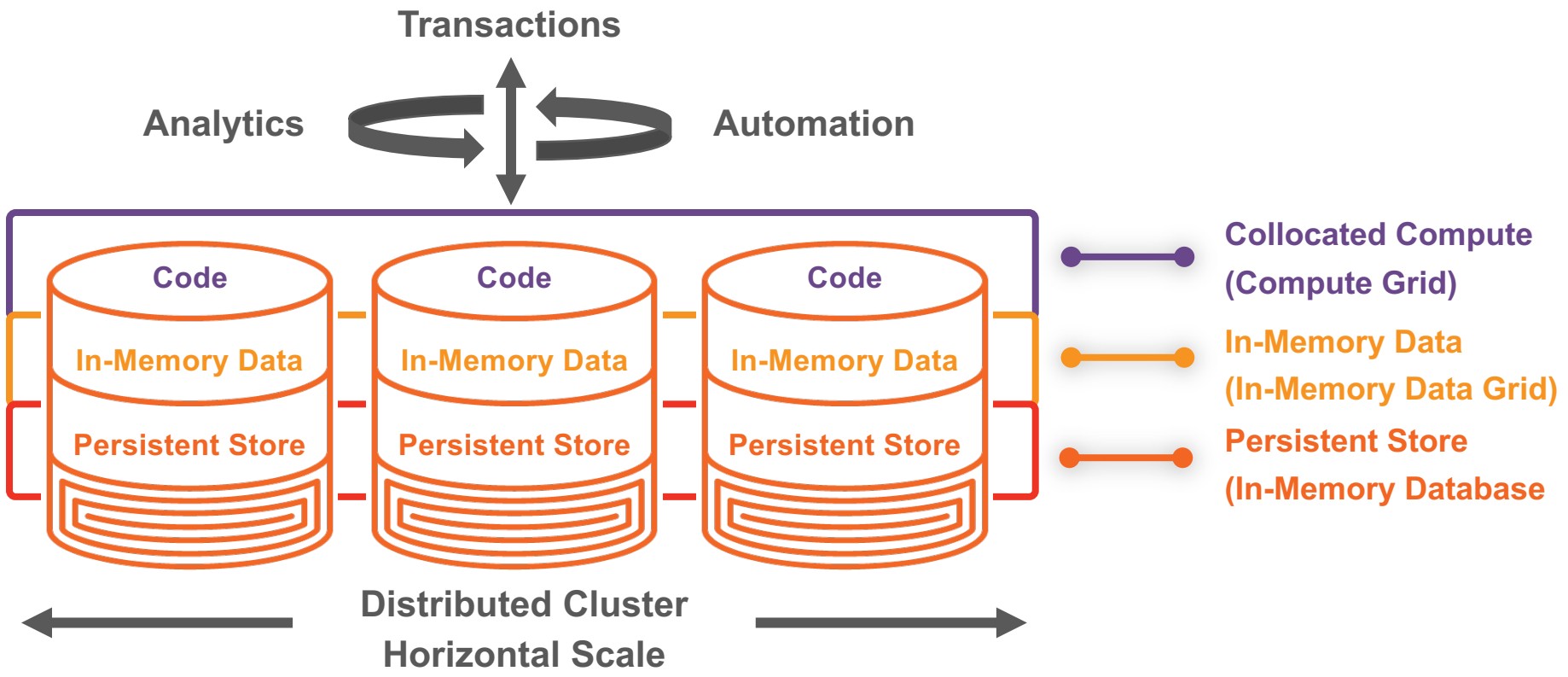
Automated Decision Making



GridGain

How HTAP Combines Transactions, Analytics and Automation

- In-Memory Speed
  - Data in RAM
  - Saved on disk or database
- Horizontal Scale
  - Shared-nothing architecture
  - Partitioning and replication
  - Collocated compute to run transactions, analytics and automation together on same data



GridGain

Case Study: HomeAway
Real-Time Pricing and Personalization

HomeAway®

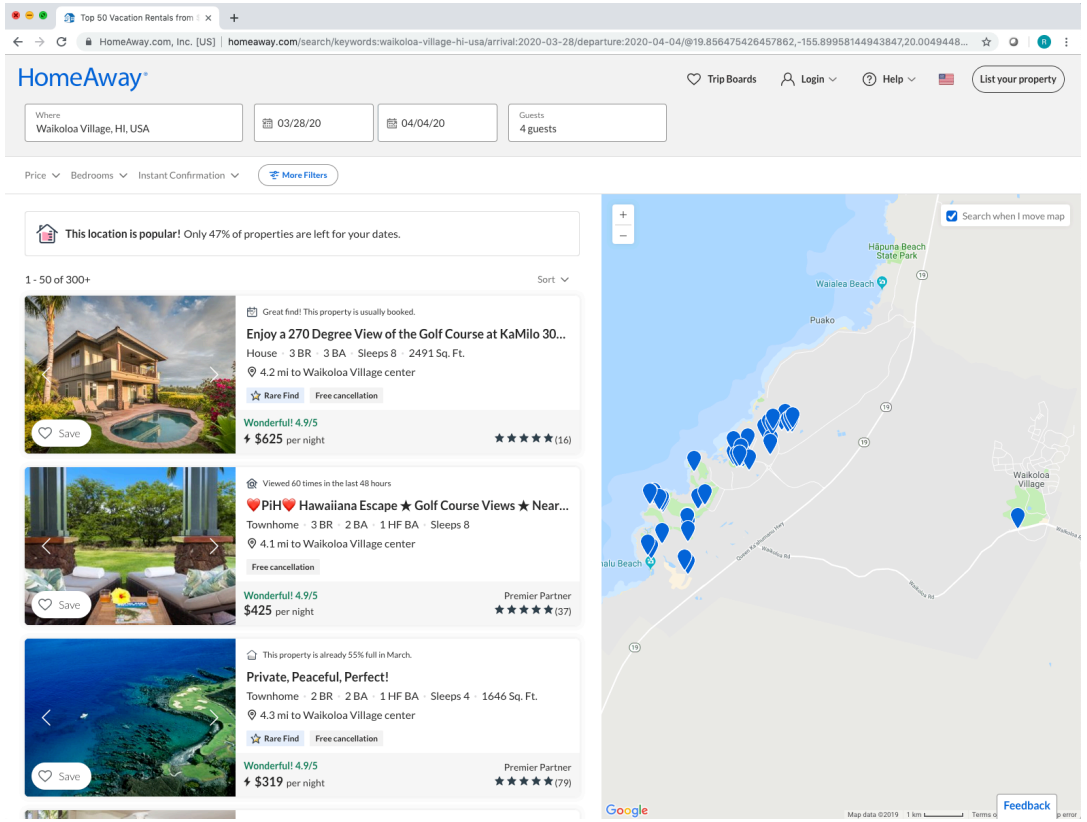
A world leader in vacation rentals with more than two million places to stay across 190 countries. Part of Expedia, HomeAway includes VRBO.

Challenges

- Perform over 450,000 pricing calculations per second at peak times
- Access over 110GB of pricing data per second
- Deliver sub-second response times every time

GridGain Enables

- Collocated processing for pricing calculations without the need for extra networking capacity
- Sub-second response times, even at peak loads

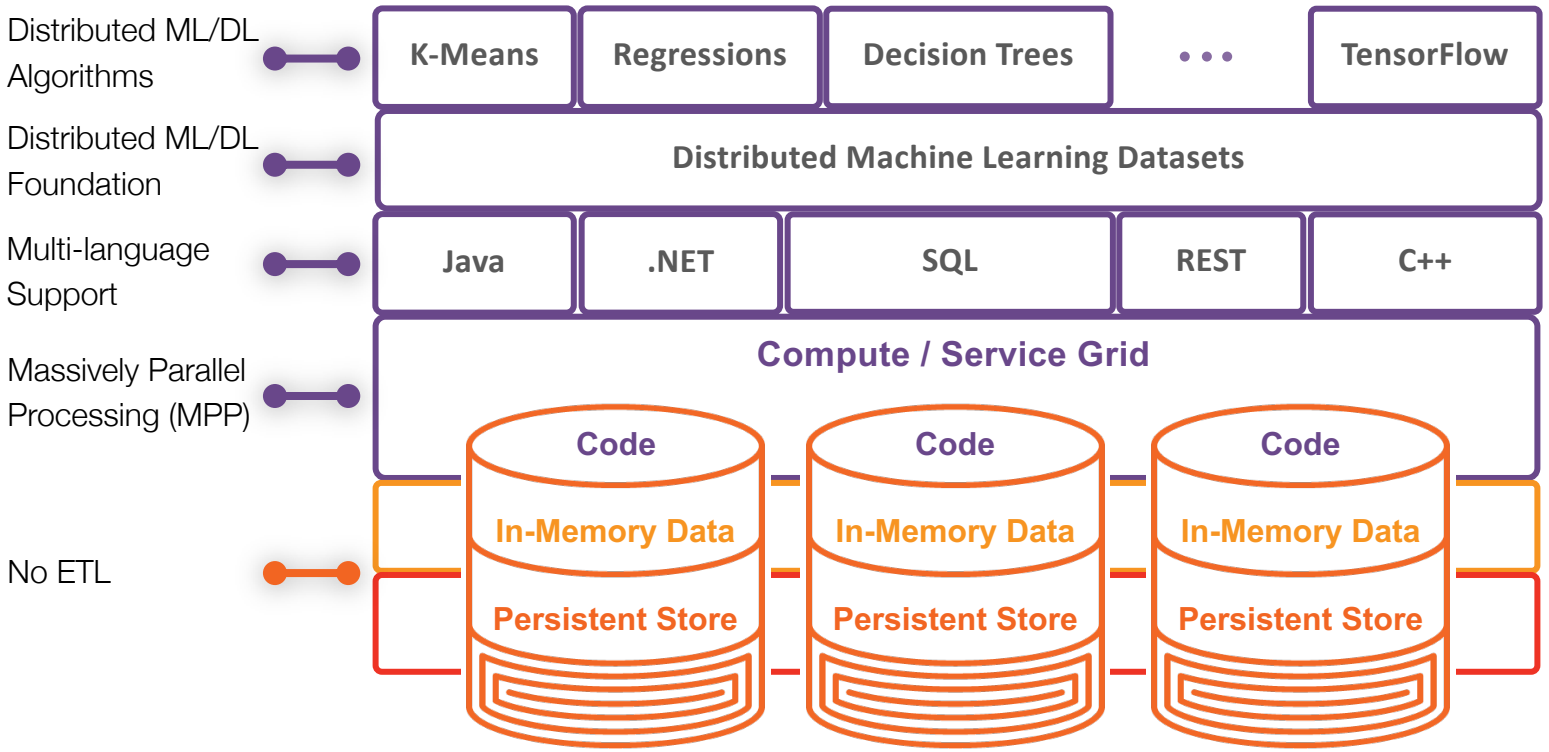


https://www.imcsummit.org/2017/us/sessions/ignite-compute-grid-in-cloud

GridGain

In-Memory Computing As A Foundation
For General Purpose Analytics and Automation

- General-Purpose Compute Grid
  - Transactions + Analytics
  - SQL (Analytics Tools)
  - Multi-language compute
  - Data-intensive
  - Massively Parallel Processing
- Platform for Continuous Machine and Deep Learning
  - No ETL (Partitioning)
  - Distributed algorithms
  - Stream processing (incremental)



GridGain

Best Practice 4.4 – Architect for HTAP

- The Only Way to Perform Real-Time Analytics and Automation is HTAP
  - ETL adds too much delay – takes too long and uses stale data
  - Perform transactions and analytics against the same data
- In-Memory Computing and Streaming is the Foundation for In-Process HTAP
  - Keep data in RAM
  - Scale horizontally to ensure cost-effective scaling
  - Collocate processing to lower latency and avoid the network as a bottleneck
- Choose an IMC platform that supports all your different workloads
  - Transactions
  - Streaming analytics
  - Machine and deep learning

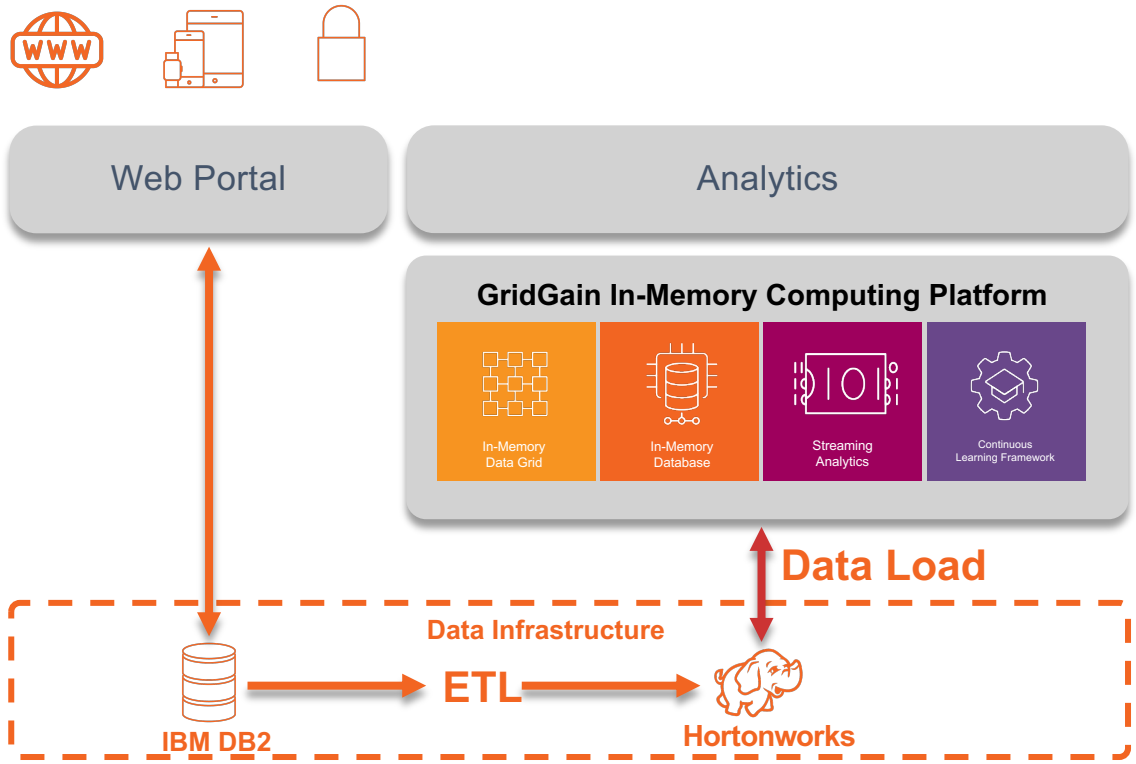
Best Practice 4.5

Embrace Data Lakes, and Plan to Move to Real-Time

Government Agency – Fraud Analytics
Started with Data Lake Acceleration

Leading health agency that administers health programs

- Problems
  - Query and reporting times for fraud analytics too slow due to slow Hadoop (HIVE) performance
  - Desire to modernize database (DB2)
  - New need for Machine Learning
- GridGain Solution
  - In-memory computing for fraud analytics that eliminated performance bottlenecks
  - Supports future machine learning needs
  - Supports moving to more real-time analytics and automation



Best Practice 4.5 – Embrace Data Lakes, and Plan to Move to Real-Time

- Embrace Data Lakes, The Current Version of The Truth
  - Real-time analytics and SQL acceleration is a common use case
  - Data quality and consistency (master data management) is a big investment
- Do not try to Hive Accelerate or Federation
  - A “Data Mart” architecture or separate synchronized store works better
- Plan Ahead to Move Upstream to Real-Time (Live) Data
  - Understand whether real-time data is important
  - The only way to get real-time data is to move upstream to the live streams
  - You will already need to stream some new data in directly

Best Practice 4.6

Invest in Real-Time Data Governance Now

One Competency is Never Enough for Transformation

ML / DL

API Management

Business-to-Business (B2B)

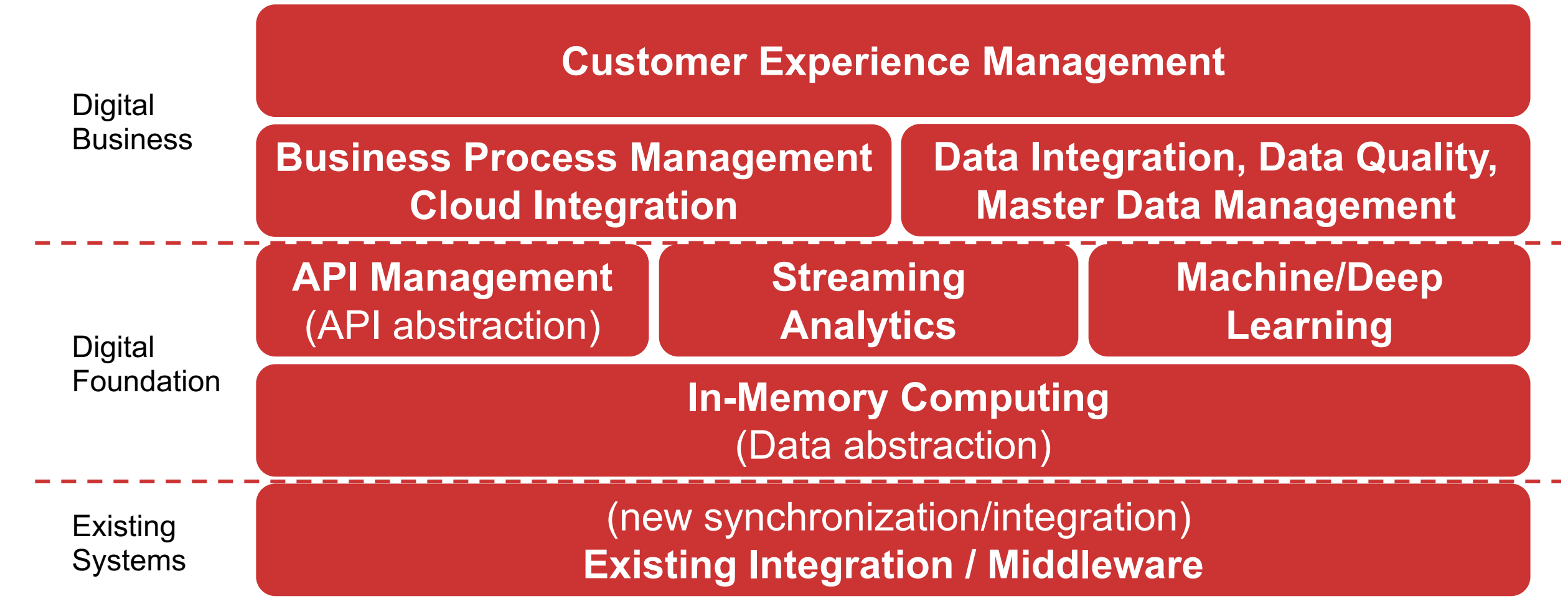
Streaming Analytics (BAM 2.0)

Data Integration
Data Quality
Master Data Management

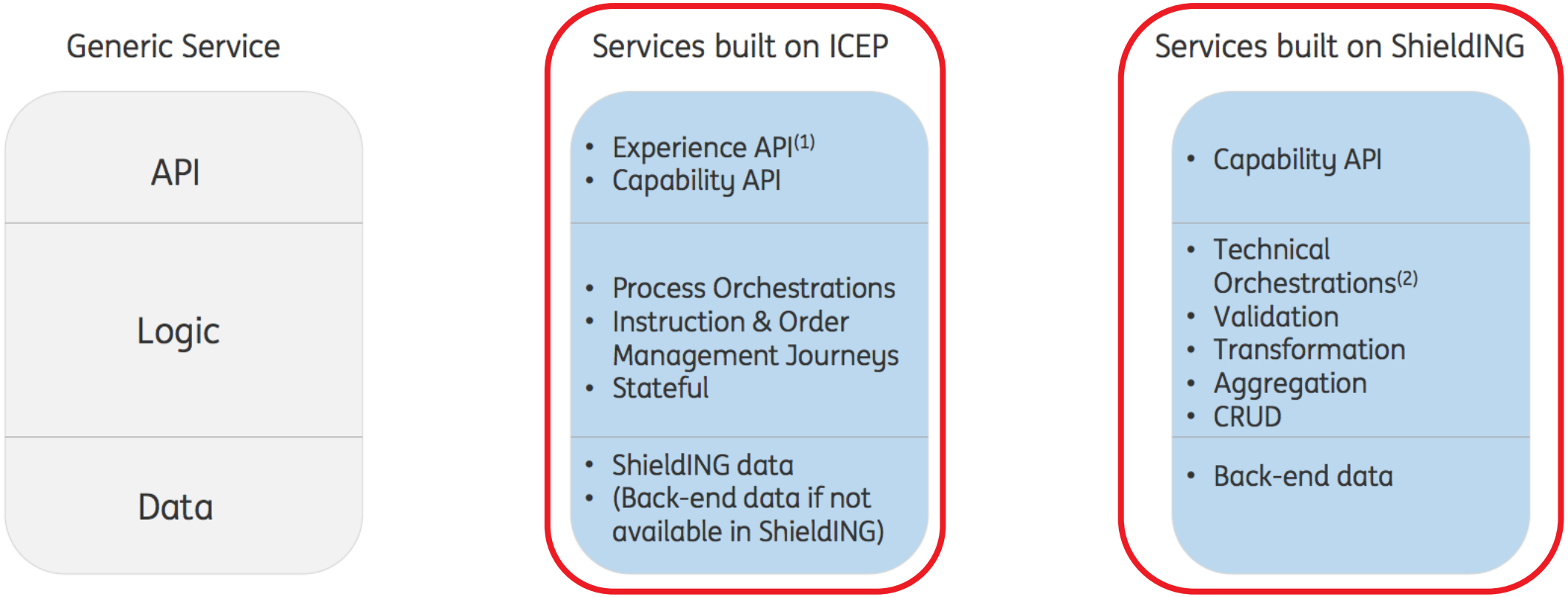
Business Process Management (BPM)

SOA/EAI Messaging

Digital Business Requires Many Components



Key concepts & definitions – Difference between ICEP and ShieldING Services

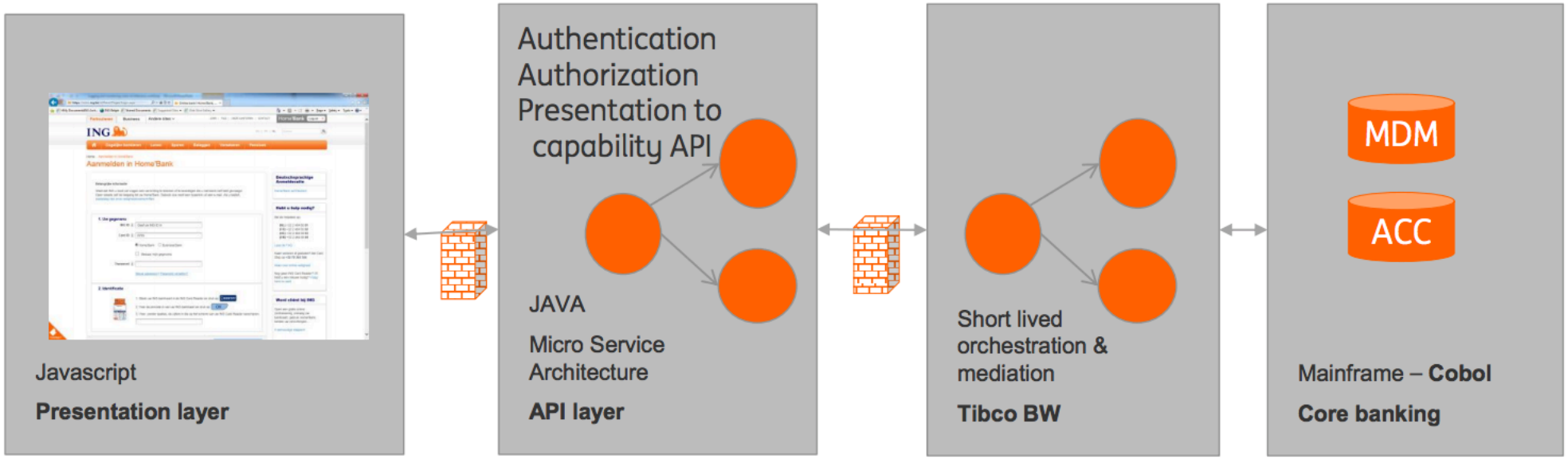


(1) Experience API provide the operations directly supporting the Distribution Application Dialogs and typically correspond to stateless Business Journeys. Currently these API are developed as NSP SL components. These API will gradually be migrated to API on ICEP and ShieldING. See Appendix for description of TPA API.

(2) Technical orchestrations typically involve multiple service calls to retrieve the requested information (where the output of one call is used as input for the other)

14

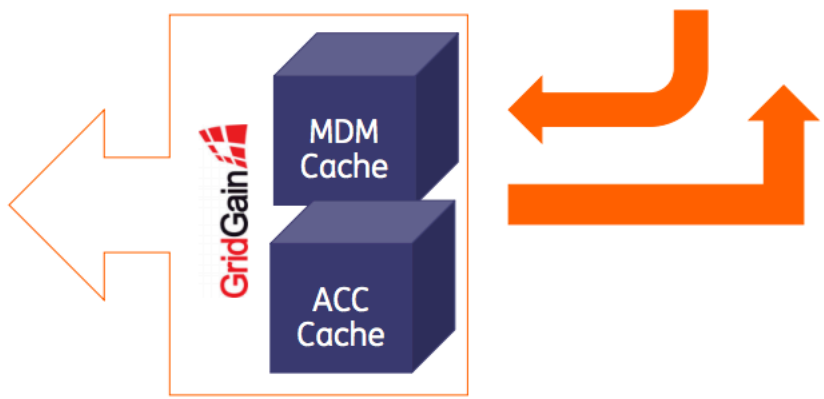
Shield-ING : Data services approach for mainframe offload



AS-IS

Benefits

- Increase availability of the essential data
- Decouple Mainframe availability and performance
- Increase performance
- Data transformation
- Linear horizontal scalable



TO-BE

ING

Best Practice 4.6 (Related to 1.3, 3.3) – Build Real-time Data Governance Now

- Real-Time Data Governance is Critical to Real-Time Analytics and Operations NOW
  - Data governance for ETL could evolve over time (from DI to DQ then MDM)
  - Real-time data governance cannot because people cannot be used to fix data
- Implement self-service data validation at the point of entry, where people are involved
  - Requires APIs and IMC (collocated computing) for real-time validation
  - Requires Data Quality (DQ) and Master Data Management (MDM)
  - Need to perform Data Integration (DI) to include all relevant data
- Add Real-Time DQ and MDM to support data validation at the point of entry
- Be Sure to Integrate with Traditional Data Governance Over Time
  - You can treat real-time data governance as just for new real-time use cases

GridGain

In-Memory Computing Best Practices

Implement Real-Time Analytics, HTAP and Automation

Rob Meyer

Outbound Product Management

August 21, 2019

GridGain Resources

- Webinars
  - Visit https://www.gridgain.com/resources/webinars
  - Visit https://www.imcsummit.org/
- White Papers
  - Visit https://www.gridgain.com/resources/papers
- Videos
  - Visit https://www.gridgain.com/resources/videos
- Sign Up for The In-Memory Computing Summit (Nov 13-14)
  - Visit https://www.imcsummit.org/
- Free 30-Day Ultimate, Enterprise or Professional Edition Trial
  - Visit https://www.gridgain.com/resources/download

GridGain