What is Modern Data Architecture?

UNDERSTANDING MODERN DATA ARCHITECTURE

Miller Trujillo
Senior Software Engineer

datacamp

What will we learn?

- Modern data architectures
- Different types of data architectures
- Components
  - Ingestion to serving
  - Data governance
  - Orchestration
- Design a modern data solution

Modern data architectures

Characteristic	Modern data architecture	Traditional data architecture
Volume and variety	Massive volumes and multiple formats	Smaller volumes of structured data
Storage	Data lakes, NO-SQL, data warehouse	Relational databases, data warehouse
Processing approach	Real-time or near real-time processing	Batch only
Scalability and flexibility	Cloud-based solutions	Hardware investments
Democratization	Self-service analytics tools	Heavy dependence on IT departments

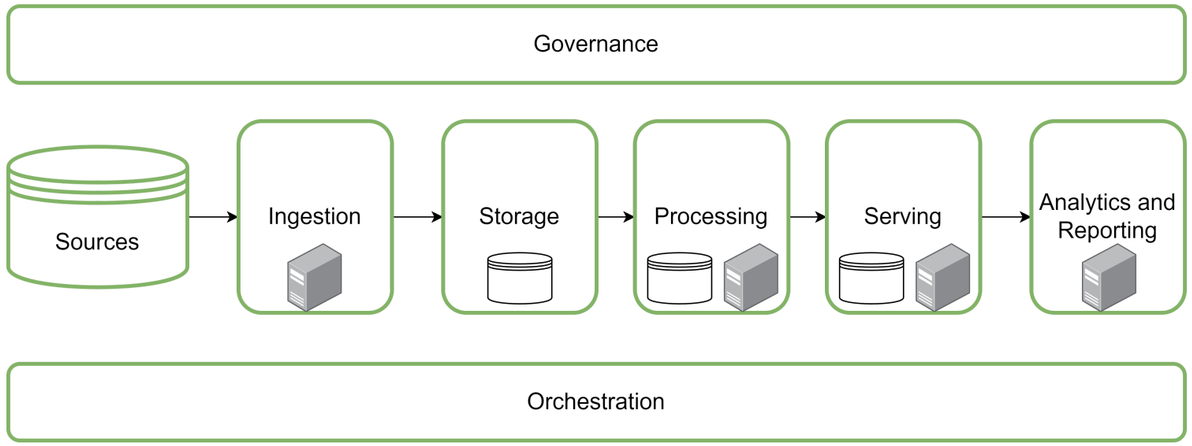
Modern data architectures requirements

- Flexibility and extensibility
- Uses cloud platforms to scale
- Handles rising data volumes
- Critical business paths
  - Billing
- Integrate distributed domains
- Data governance and security

[1] https://www.ibm.com/resources/the-data-differentiator/data-architecture

Components

- Data sources
- Data ingestion
- Data storage
- Data processing
- Data orchestration
- Data governance and security



[1] https://learn.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data

Let's practice!

UNDERSTANDING MODERN DATA ARCHITECTURE

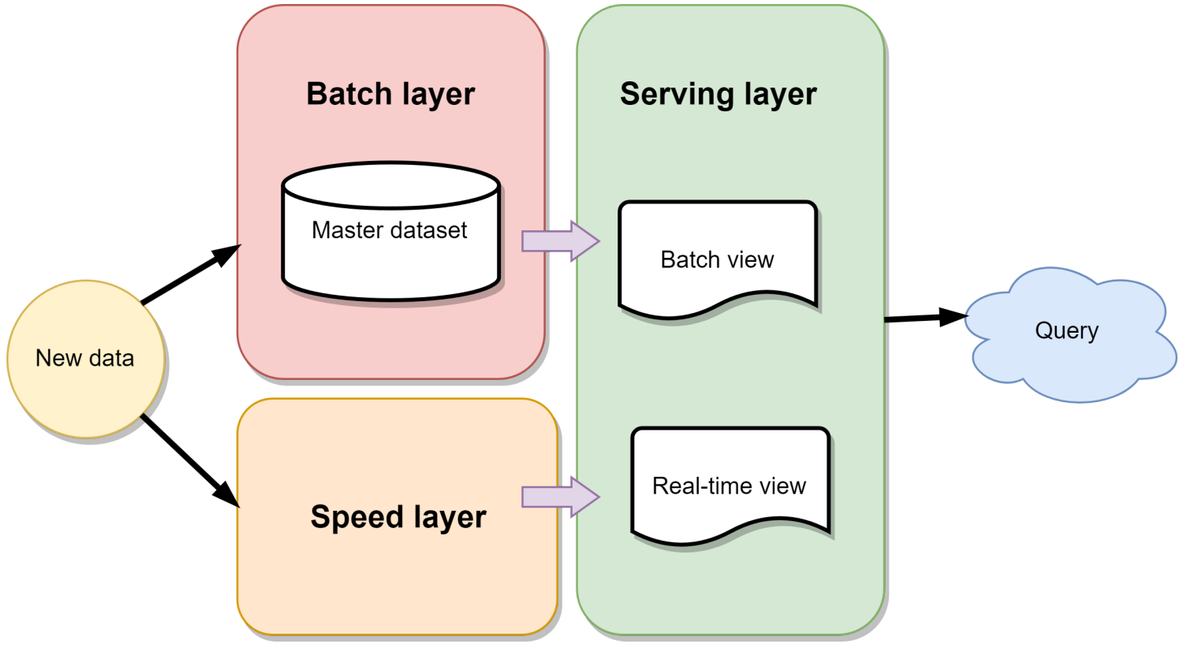
Examples of the modern data architecture

UNDERSTANDING MODERN DATA ARCHITECTURE

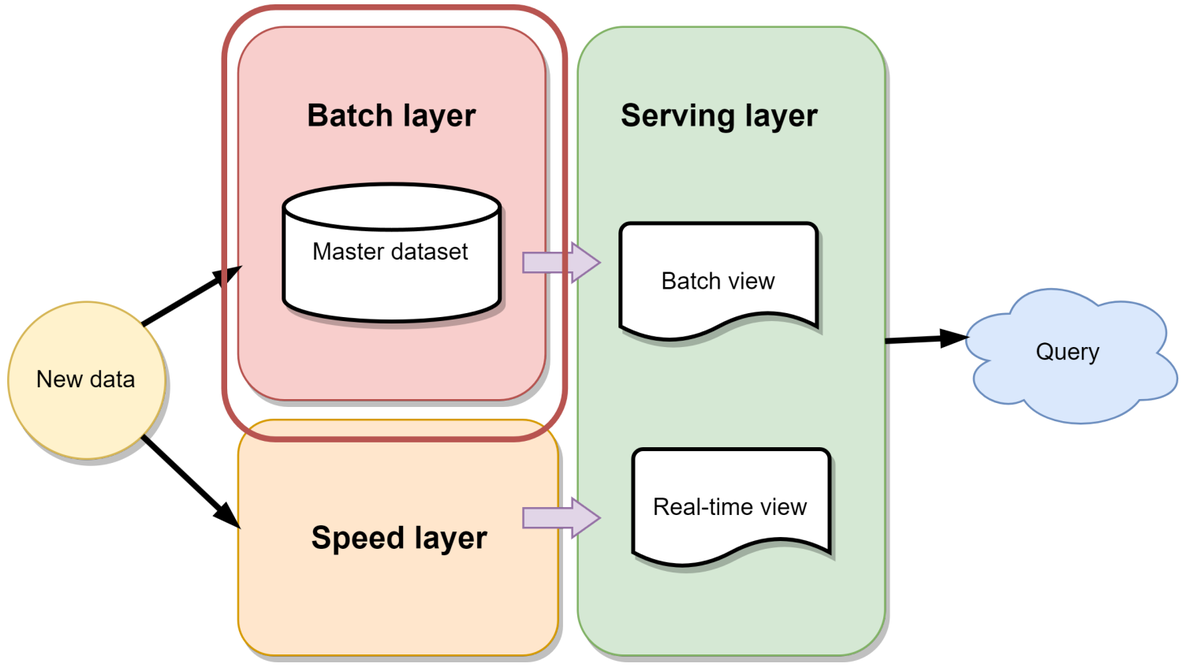
Miller Trujillo
Senior Software Engineer

Lambda architecture

- Data processing architecture
- Enables processing of large-scale datasets
- Enables real-time data processing
- Combines batch and streaming
- 3 main layers:
  - Batch layer
  - Speed layer
  - Serving layer

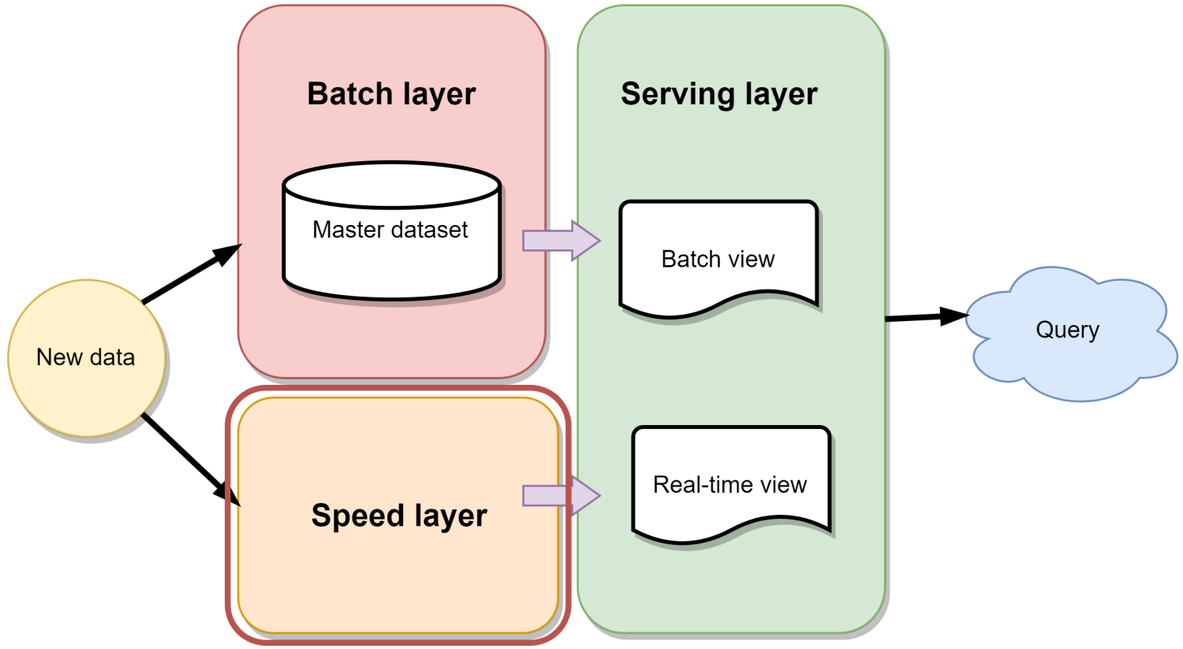


Lambda architecture layers: Batch



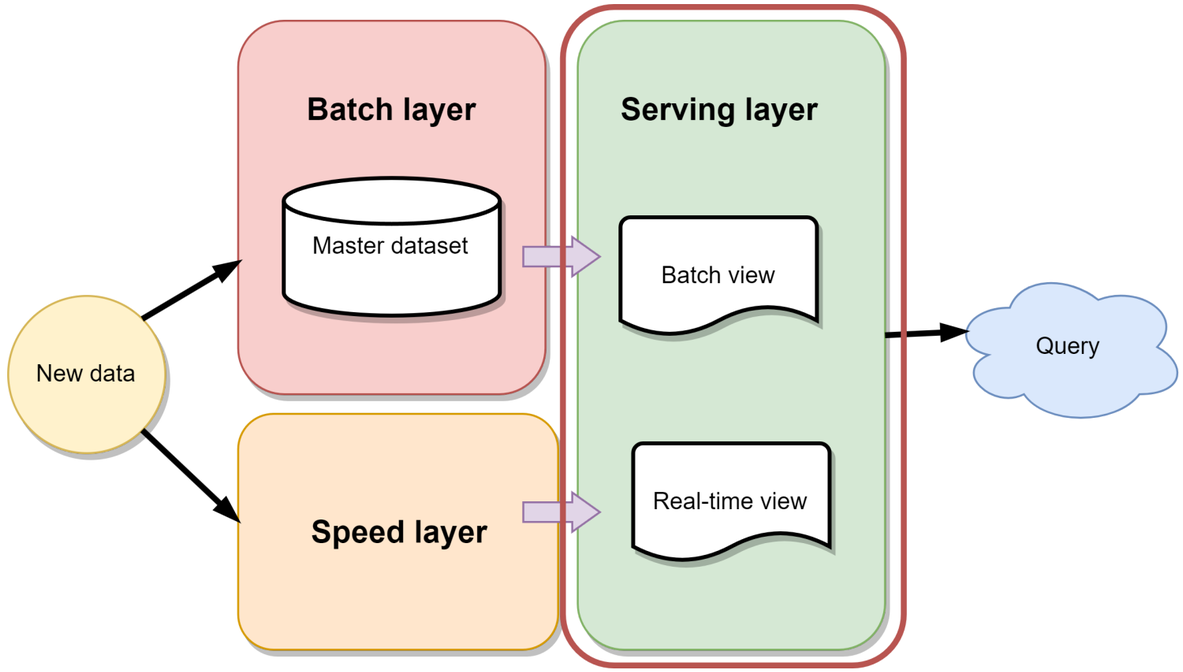
- Master dataset
  - Immutable data
  - Append-only
- Pre-computed views
- Bank accounts
  - Master dataset: List of transactions
  - Views: Transactions applied to get the balance
- Perfect accuracy

Lambda architecture layers: Speed



- Batch jobs at a given schedule
- Gap between known data and reality
- Speed layer close that gap
- Not perfect accuracy

Lambda architecture layers: Serving

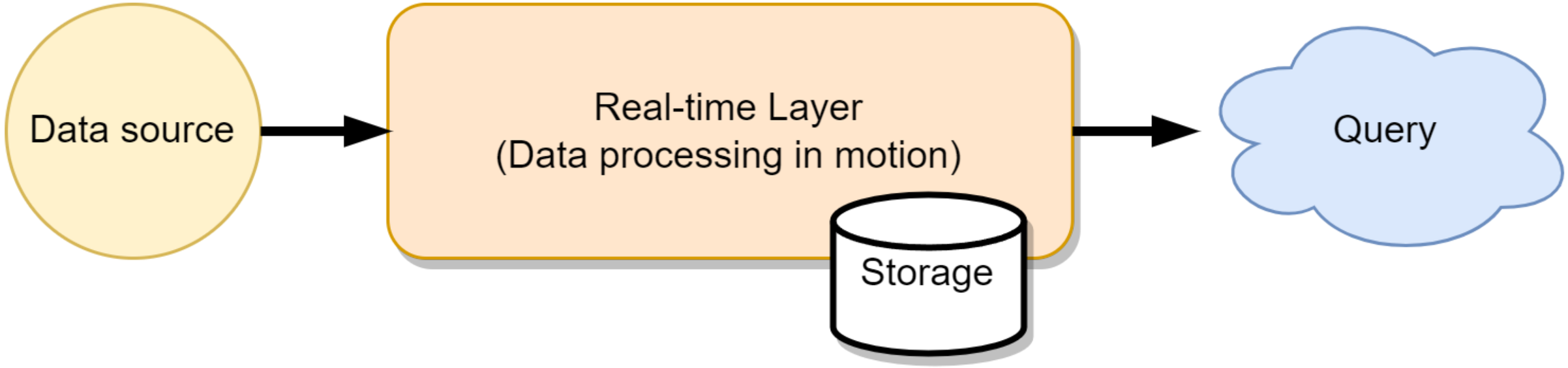


- Merge batch and real-time views
- Unified view of data

Complexity of Lambda architecture

- Two layers increase complexity
- Duplicated logic, code, algorithms
- Multiple stacks: Batch and streaming

Kappa architecture



Lambda vs. Kappa

Aspect	Lambda Architecture	Kappa Architecture
Batch Processing	Present	By processing the source of events again
Stream Processing	Present	Present
Complexity	Higher	Lower
Real-time Processing	Delayed by batch layer	Direct and streamlined
Use Cases	Complex analysis, historical data	Real-time analytics, real-time monitoring

Let's practice!

UNDERSTANDING MODERN DATA ARCHITECTURE

Data Mesh and Data Fabric

UNDERSTANDING MODERN DATA ARCHITECTURE

Miller Trujillo
Senior Software Engineer

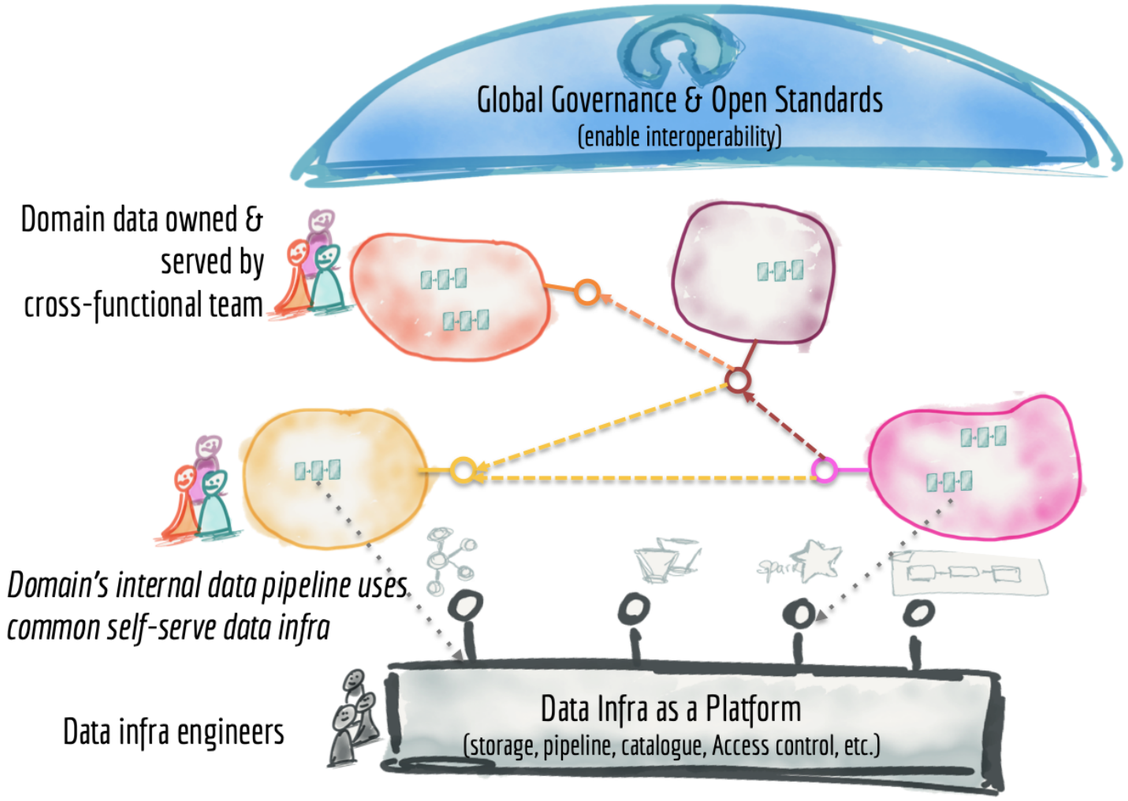
datacamp

Data Mesh

- Architectural approach to decentralize data ownership and management
- Ownership and management:
  - Full responsibility for the data
  - Ensure quality, security and privacy

Data Mesh architecture

- Owners for each dataset
- Expose data as products
- Sharing a common infrastructure
- Microservices
- Mindset shift!



[1] https://martinfowler.com/articles/data-monolith-to-mesh.html

How to process data then?

- Organizational and cultural aspects of data management
- Domain data products are first-class citizens
- Tools are second-class citizens
- Data lakes and warehouses are just nodes, details within a domain

Data Mesh benefits and challenges

- **Benefits:**
  - *Data democratization*
  - Decentralization
  - Clear ownership
  - Scalability and agility
  - Faster Time to Market
  - Flexibility in Technology Stack
  - ...
- **Challenges:**
  - Cultural Shift
  - Data Governance
  - Collaboration and Coordination
  - Technical Complexity

Data Fabric

- Unified and integrated view
- Democratization of data
- **Metadata**
  - Knowledge graph
  - AI/ML algorithms
  - Active metadata
- Intelligent integration
- Better data management and protection

[1] https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration [2] https://www.ibm.com/topics/data-fabric

A Data Fabric should support...

- ETL
- Streaming
- Replication
- Messaging
- Data virtualization
- Data microservices
- ...

Let's practice!

UNDERSTANDING MODERN DATA ARCHITECTURE