DATA-DRIVEN APPLICATION ARCHITECTURES FOR CONTAINERS

William Benton
Principal Software Engineer
Red Hat, Inc.

FORECAST

Containerized microservices are the future

Data-driven applications are the future

Architectures for data-driven applications

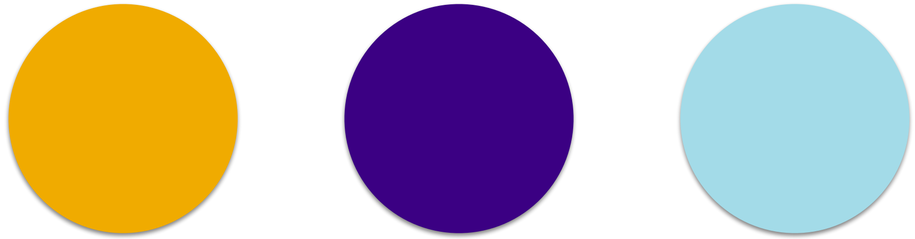
Conclusions and takeaways

MICROSERVICES ARE THE FUTURE

MICROSERVICES, DEFINED

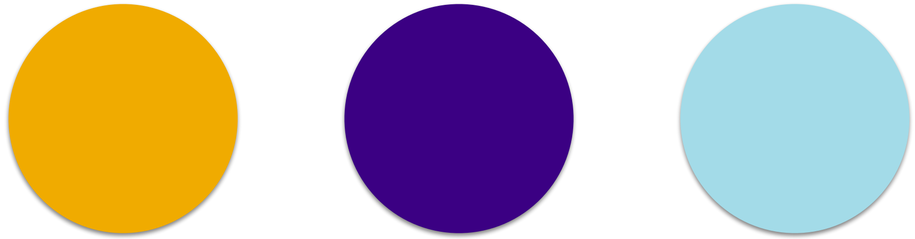
A microservice architecture employs lightweight, modular, and typically stateless components with well-defined interfaces and contracts.

BENEFITS OF MICROSERVICE ARCHITECTURES

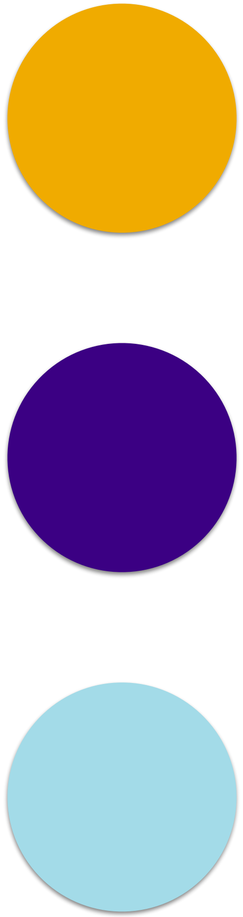


BENEFITS OF MICROSERVICE ARCHITECTURES

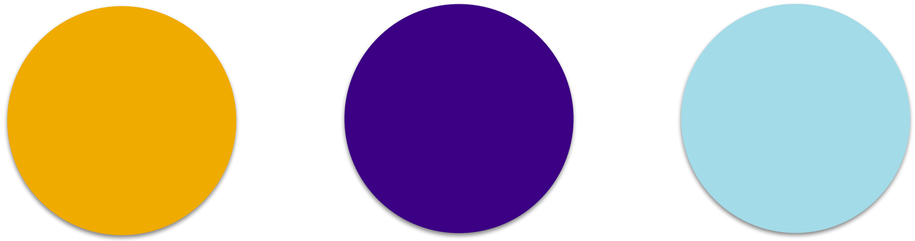
BENEFITS OF MICROSERVICE ARCHITECTURES



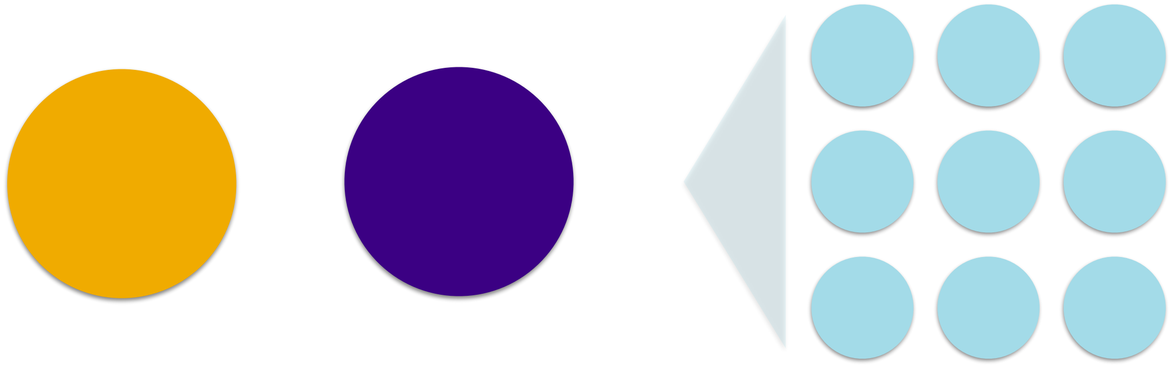
BENEFITS OF MICROSERVICE ARCHITECTURES



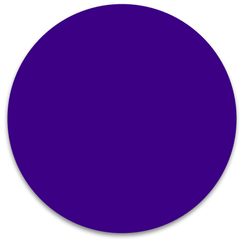
BENEFITS OF MICROSERVICE ARCHITECTURES



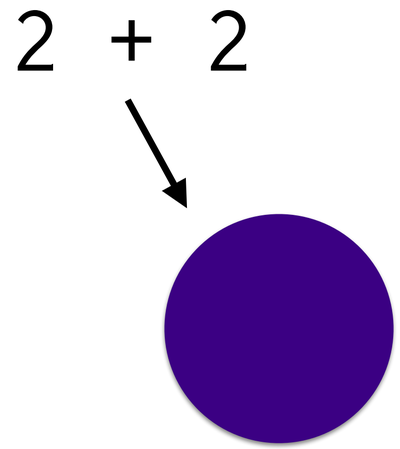
BENEFITS OF MICROSERVICE ARCHITECTURES



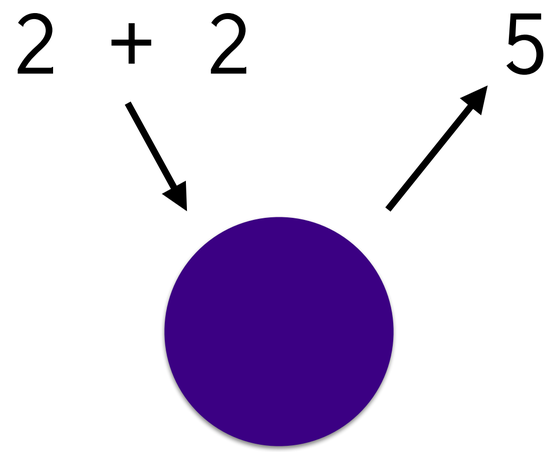
BENEFITS OF MICROSERVICE ARCHITECTURES



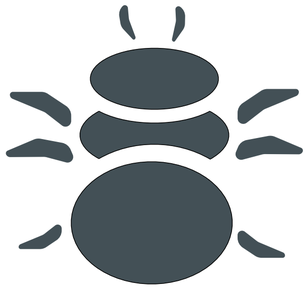
BENEFITS OF MICROSERVICE ARCHITECTURES

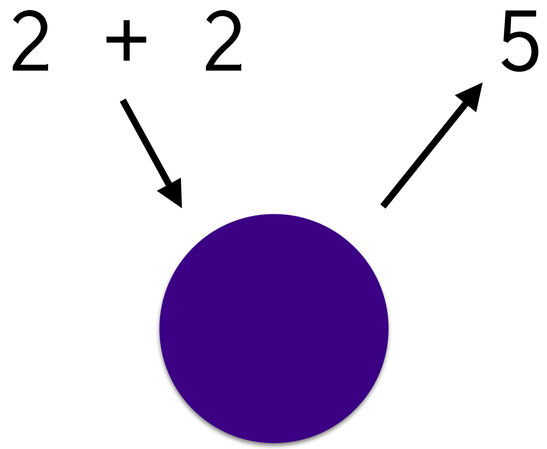


BENEFITS OF MICROSERVICE ARCHITECTURES

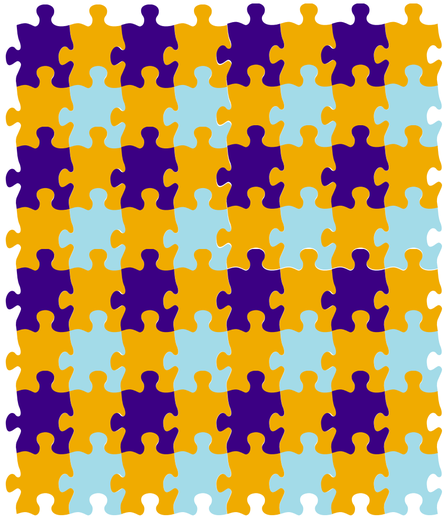


BENEFITS OF MICROSERVICE ARCHITECTURES

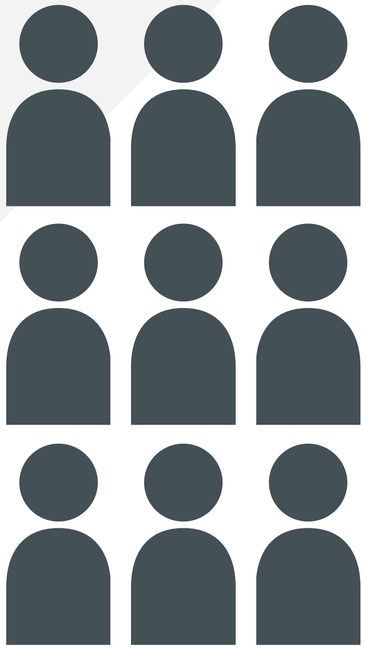


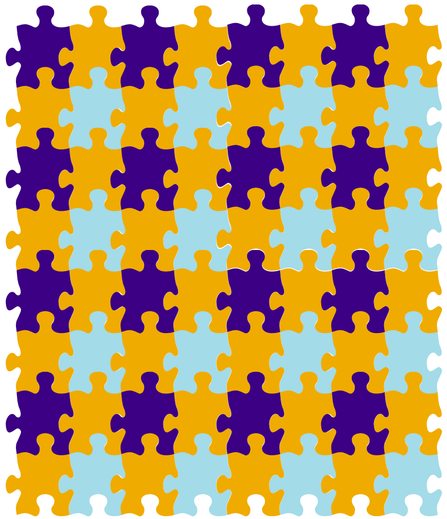


BENEFITS OF MICROSERVICE ARCHITECTURES

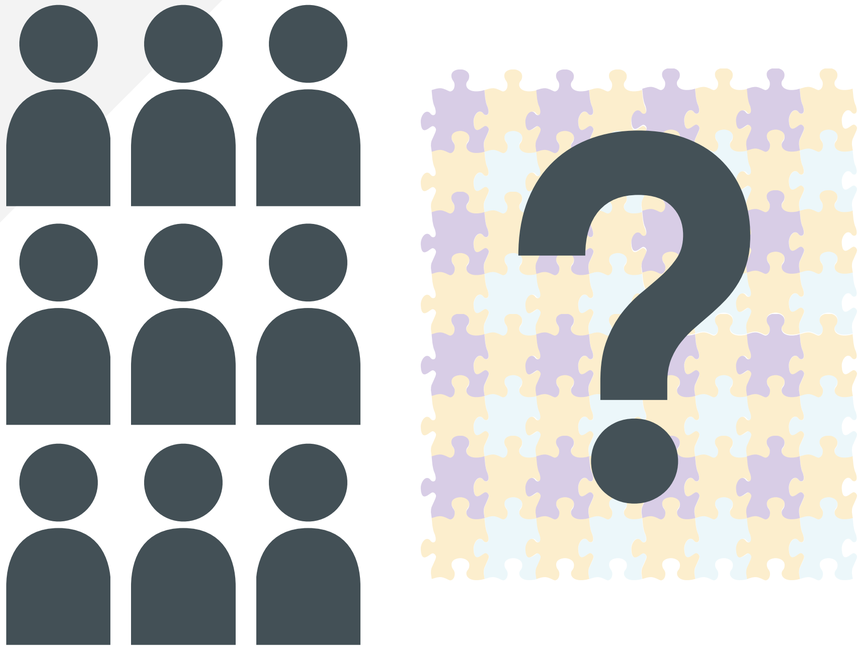


BENEFITS OF MICROSERVICE ARCHITECTURES

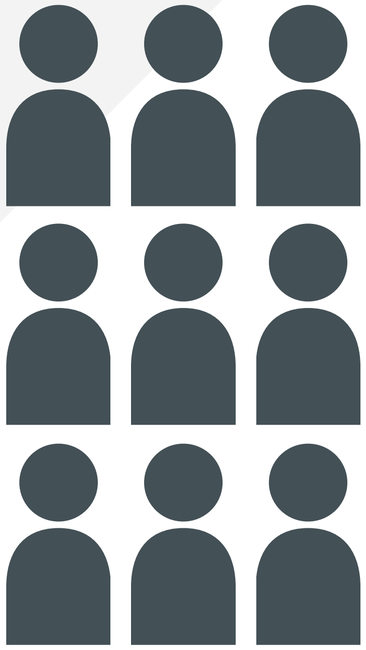


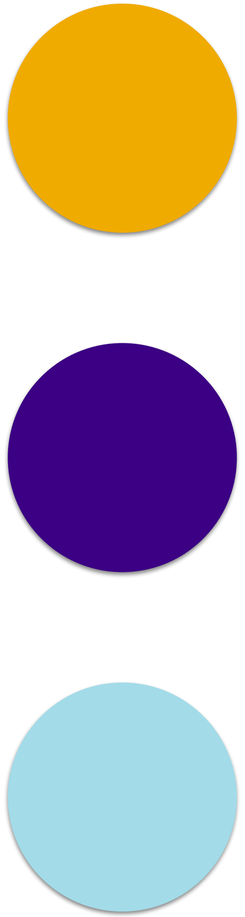


BENEFITS OF MICROSERVICE ARCHITECTURES

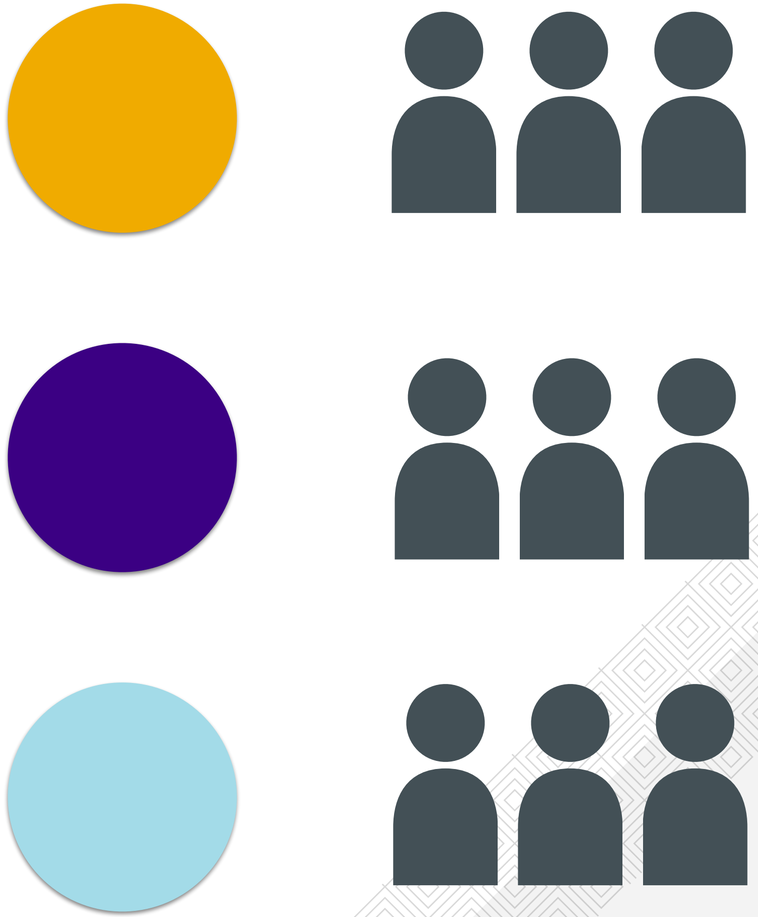


BENEFITS OF MICROSERVICE ARCHITECTURES

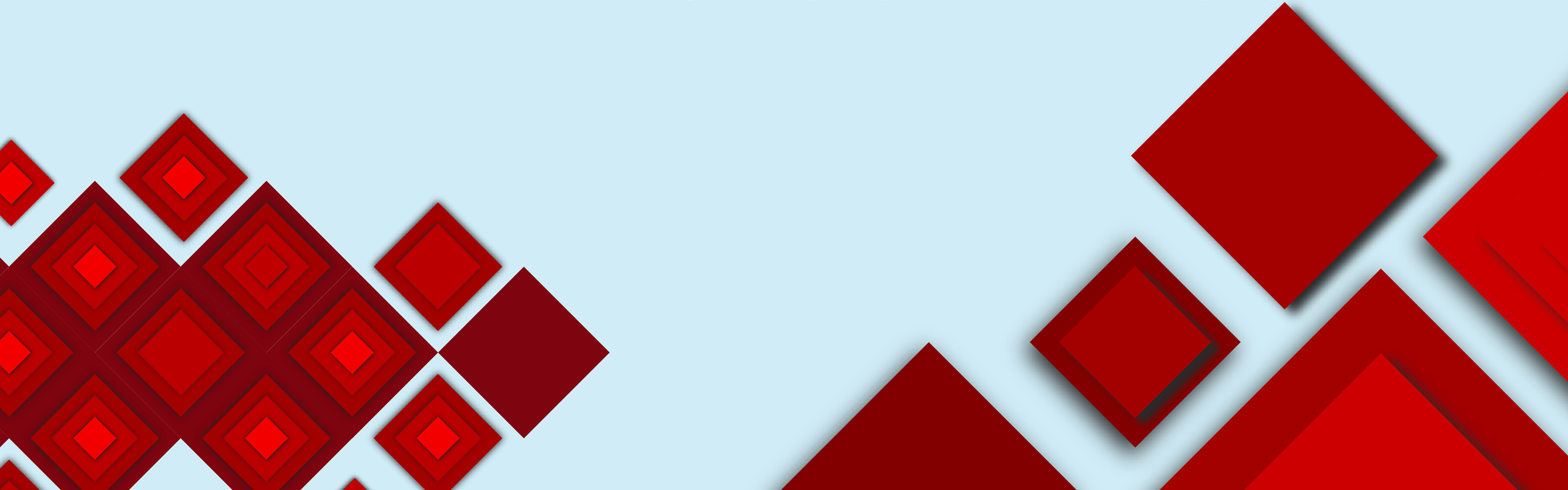




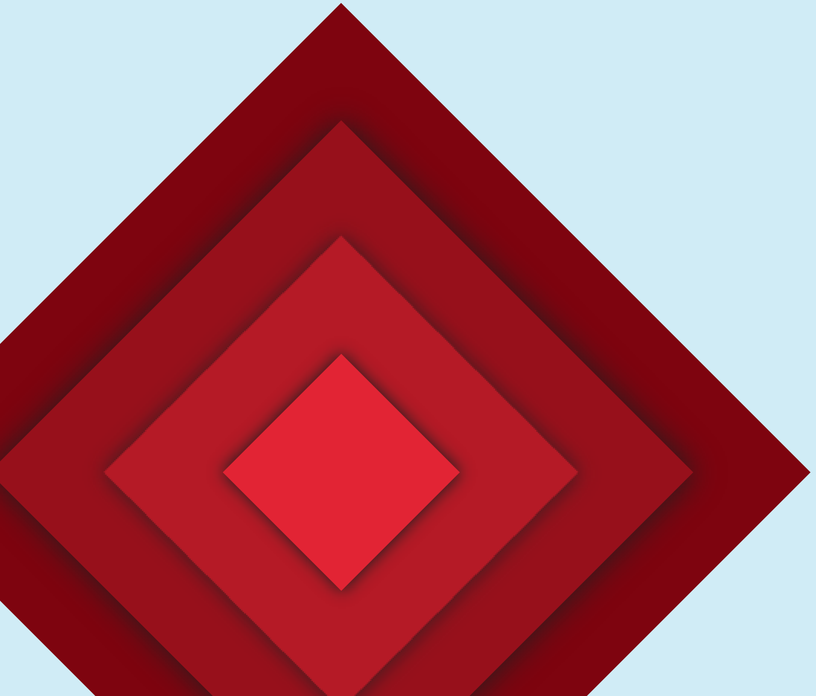
BENEFITS OF MICROSERVICE ARCHITECTURES

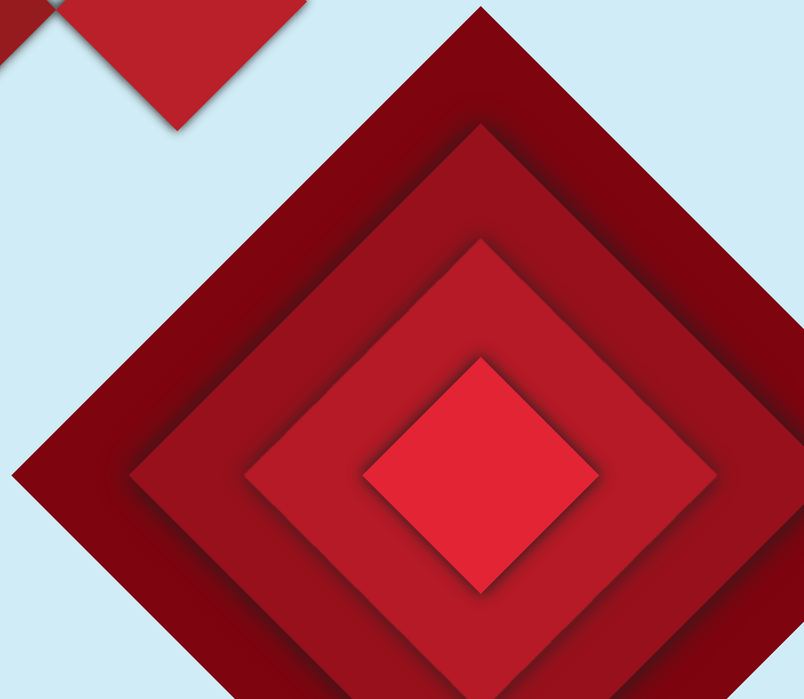


DATA-DRIVEN APPLICATIONS ARE THE FUTURE

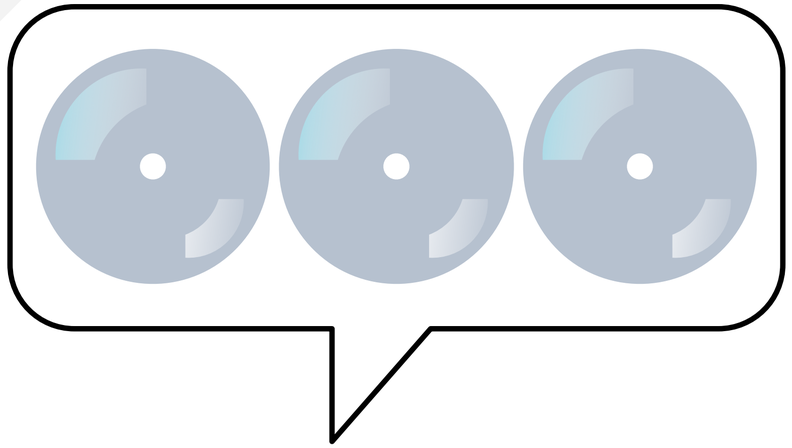


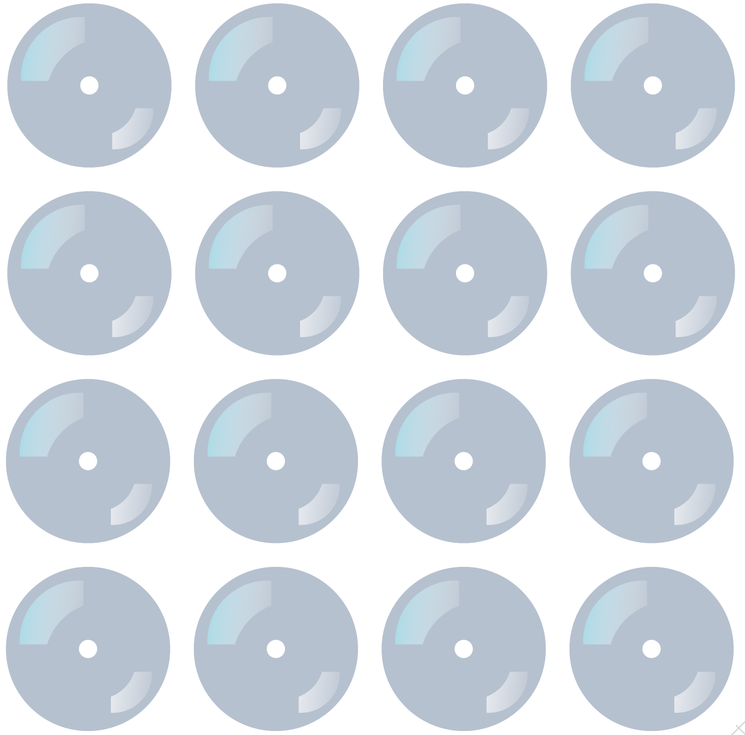
EXAMPLE APP: AMAZON RETAIL



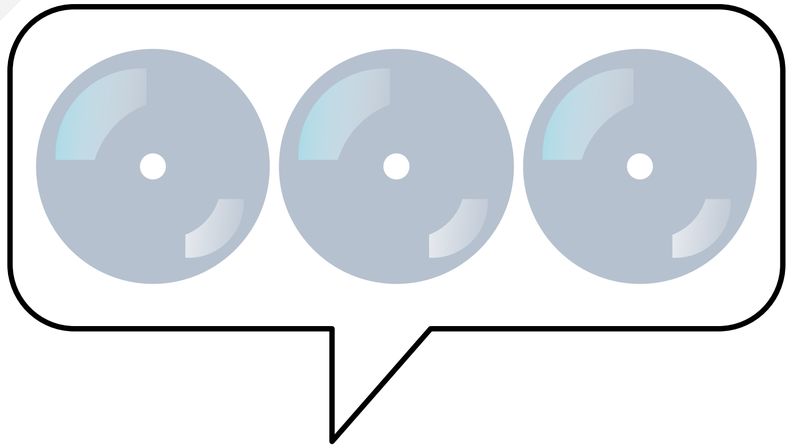


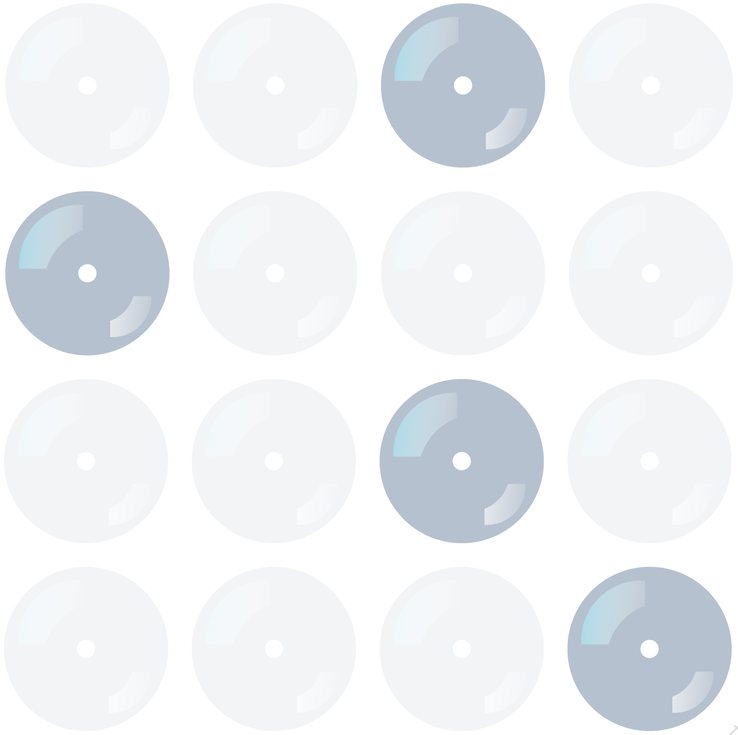
RETAIL IN 1994 (BEFORE AMAZON)



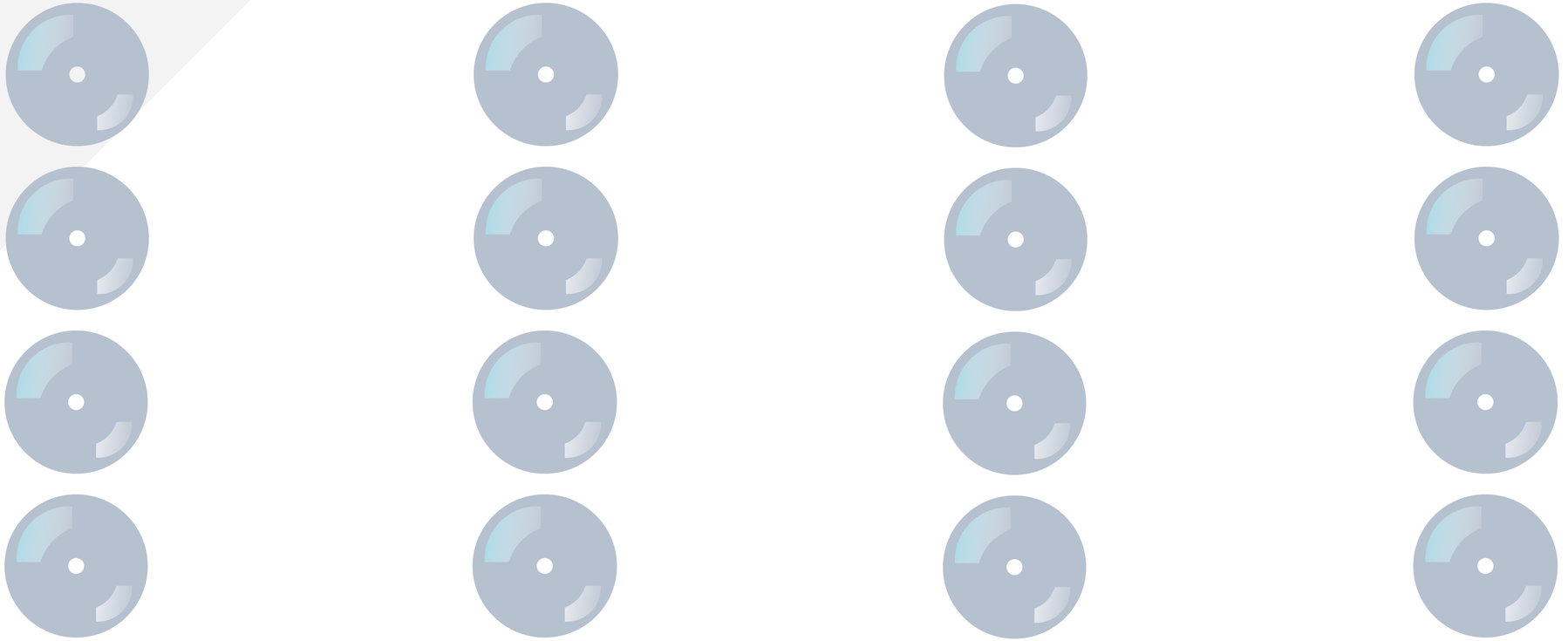


RETAIL IN 1994 (BEFORE AMAZON)





RETAIL IN 1999 (EARLY AMAZON)



RETAIL IN 1999 (EARLY AMAZON)

RETAIL IN 1999 (EARLY AMAZON)

RETAIL IN THE MID-2000s

★★★★★ **1 of 1,024 people found this helpful**

Enim ex perspiciatis provident enim in. Aut ipsa quibusdam itaque deserunt esse ipsum soluta. Aut aliquid tempora cumque dicta perferendis omnis quia. Minima voluptatem qui tenetur dolor. Alias repellendus quis voluptatum. Sint aliquid quibusdam quasi.

★★★ **768 of 960 people found this helpful**

Sed feugiat odio at mauris imperdiet, a suscipit turpis consequat. Donec pretium dignissim purus sit amet tincidunt. In imperdiet, magna et finibus commodo. Aliquam et malesuada dui. Nunc vitae ligula urna.

★★★★ **0 of 53 people found this helpful**

Curabitur purus arcu, tincidunt quis sollicitudin in, molestie in purus. Nam elementum sem vitae purus viverra, non mollis nulla venenatis. Suspendisse at leo et mauris sagittis vehicula. Donec sit amet faucibus enim, quis varius leo.

RETAIL IN THE MID-2000s

★★★★★ 1 of 1,024 people found this helpful

Enim ex perspiciatis provident enim in. Aut ipsa quibusdam itaque deserunt esse ipsum soluta. Aut aliquid tempora cumque dicta perferendis omnis quia. Minima voluptatem qui tenetur dolor. Alias repellendus quis voluptatum. Sint aliquid quibusdam quasi.

★★★ 768 of 960 people found this helpful

Sed feugiat odio at mauris imperdiet, a suscipit turpis consequat. Donec pretium dignissim purus sit amet tincidunt. In imperdiet, magna et finibus commodo. Aliquam et malesuada dui. Nunc vitae ligula urna.

★★★★ 0 of 53 people found this helpful

Curabitur purus arcu, tincidunt quis sollicitudin in, molestie in purus. Nam elementum sem vitae purus viverra, non mollis nulla venenatis. Suspendisse at leo et mauris sagittis vehicula. Donec sit amet faucibus enim, quis varius leo.

RETAIL IN THE MID-2000s

RETAIL IN THE MID-2000s

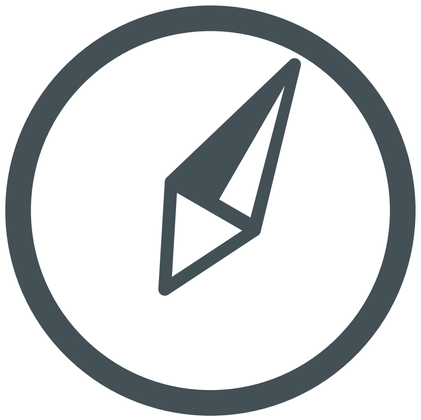
Thanks for your purchase! People who bought this item also bought:

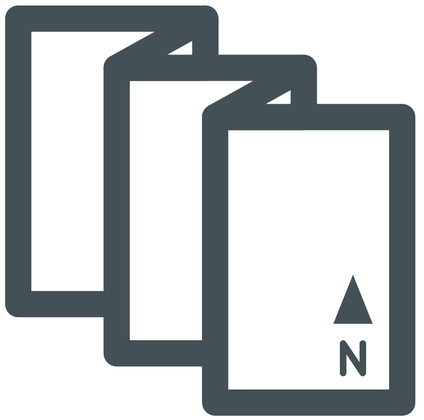
RETAIL IN THE MID-2000s

Thanks for your purchase! People who bought this item also bought:

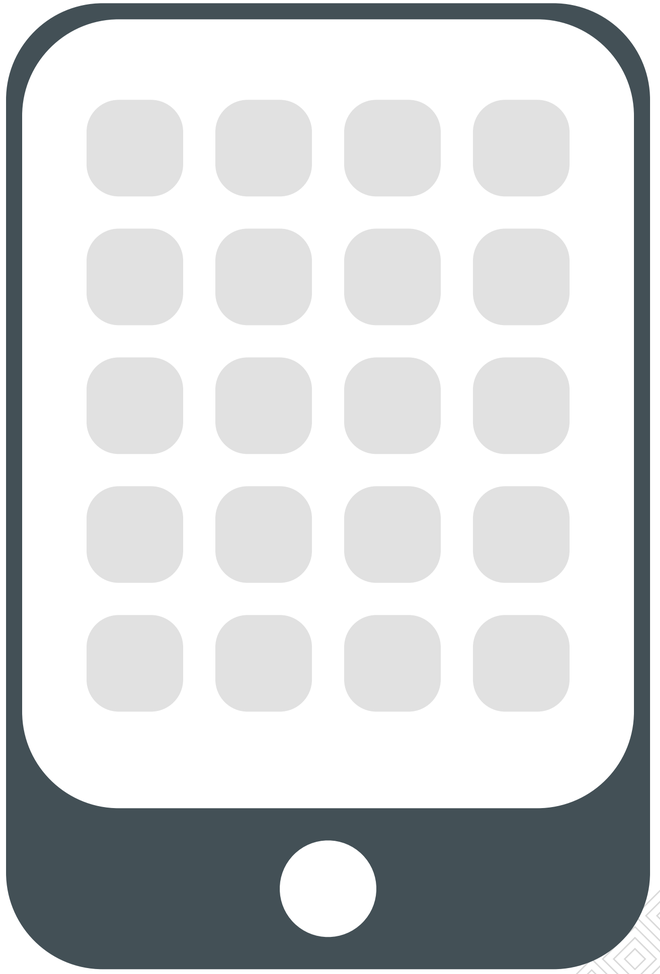
EXAMPLE APP: STRAVA

FITNESS TRACKING TEN YEARS AGO

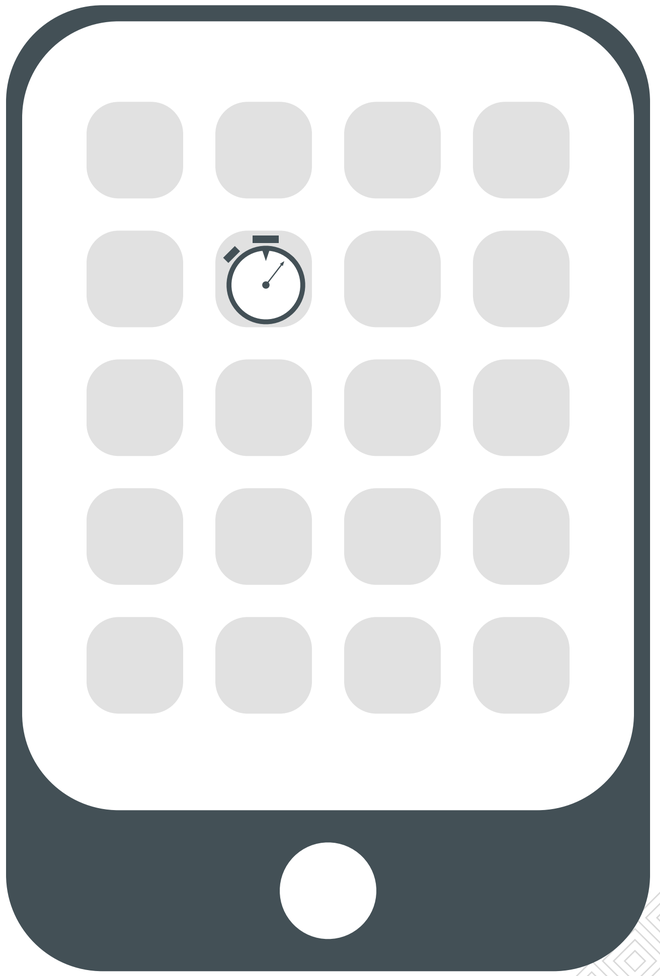




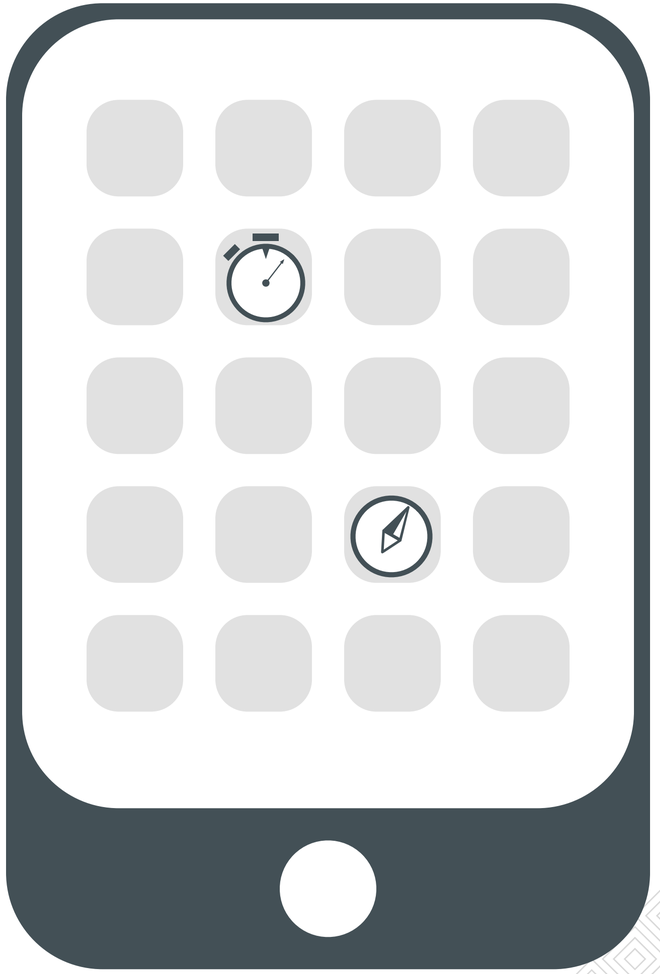
FITNESS TRACKING IN AN AGE OF MOBILE APPS



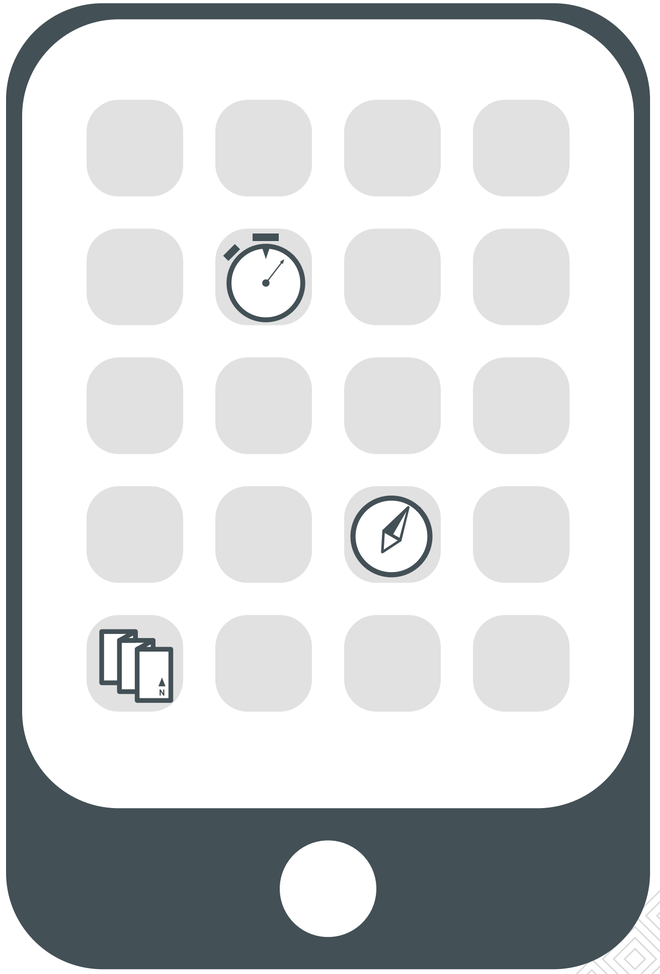
FITNESS TRACKING IN AN AGE OF MOBILE APPS



FITNESS TRACKING IN AN AGE OF MOBILE APPS



FITNESS TRACKING IN AN AGE OF MOBILE APPS

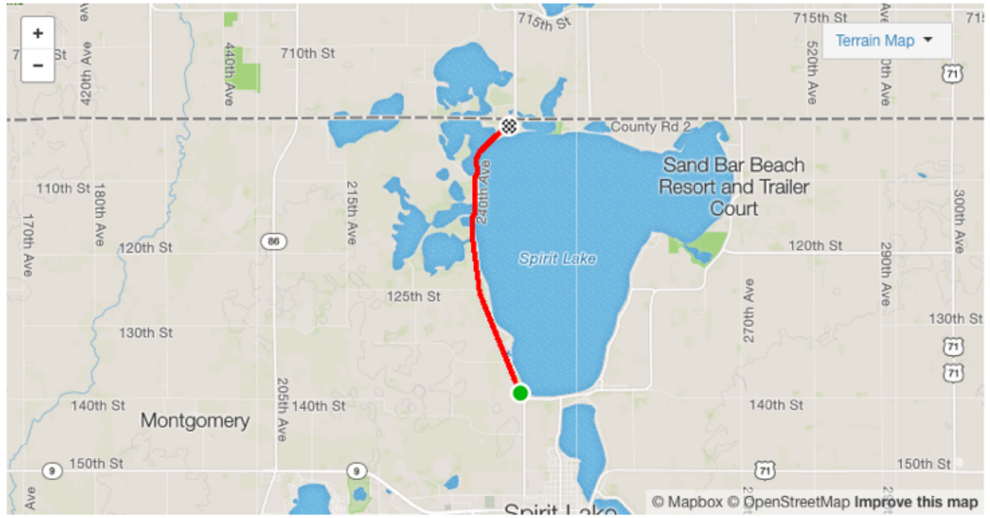


ONE KILLER FEATURE

★ West Shore Whip

Ride Segment Spirit Lake, IA

4.0 mi	0 %	1,401 ft	1,452 ft	51 ft	334 Attempts By 140 People
Distance	Avg Grade	Lowest Elev	Highest Elev	Elev Difference	



Fastest Times

All-Time PR 8:04 Jun 6, 2015

Will Benton | GDVC
KOM 8:04 Jun 6, 2015

QOM 10:55 Jul 9, 2015

Compare Efforts

Set a Goal for this Segment View All ›

Set Goal · Train for this segment

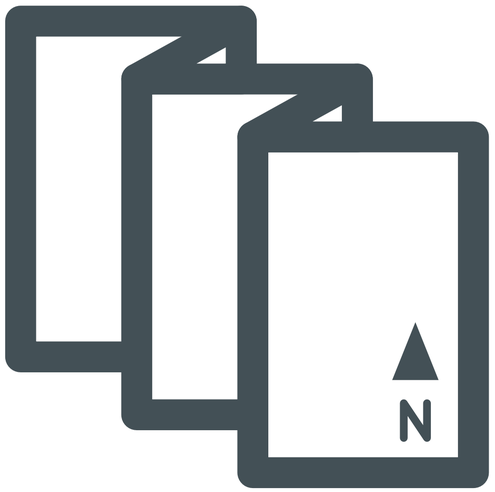
ONE KILLER FEATURE

Overall

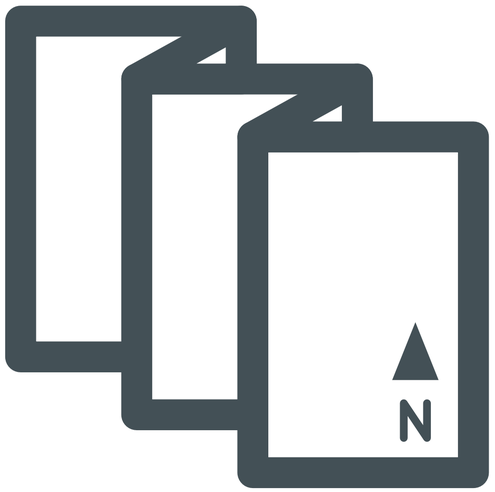
MY CURRENT PLACE	MY BEST TIME	All-Time	Men and Women
1 / 140	8:04		

Rank	Name	Date	Speed	HR	Power	VAM	Time
	Will Benton | GDVC	Jun 6, 2015	29.9mi/h	164bpm	332W	-	8:04
2	[redacted]	Jul 4, 2014	28.1mi/h	-	404W	-	8:34
3	[redacted]	Aug 4, 2015	27.2mi/h	-	351W	-	8:51
4	[redacted]	Jul 5, 2015	25.4mi/h	-	313W	-	9:29
5	[redacted]	Jul 19, 2014	24.7mi/h	161bpm	418W	-	9:45
6	[redacted]	Jul 5, 2013	24.4mi/h	-	296W	-	9:51
7	[redacted]	Aug 13, 2015	24.4mi/h	-	272W	-	9:52

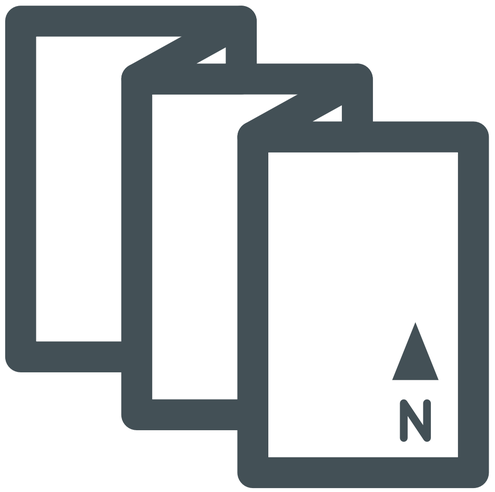
DATA-DRIVEN FEATURE: ROUTING



ROUTING FOR TRANSPORTATION

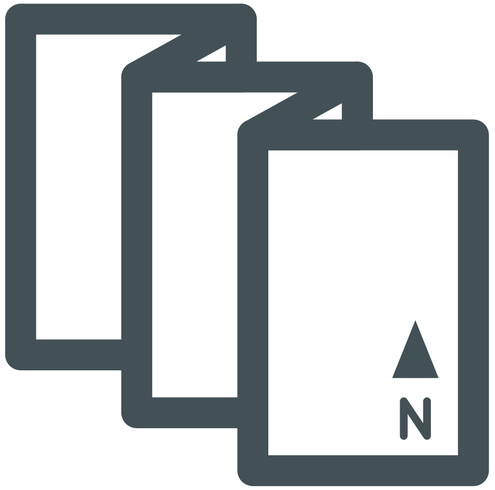


ROUTING FOR TRANSPORTATION



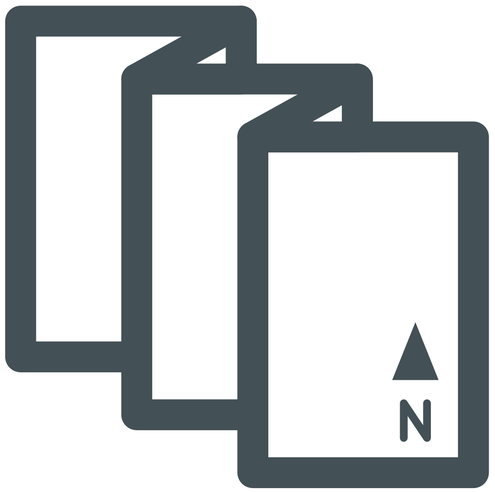
- distance

ROUTING FOR TRANSPORTATION



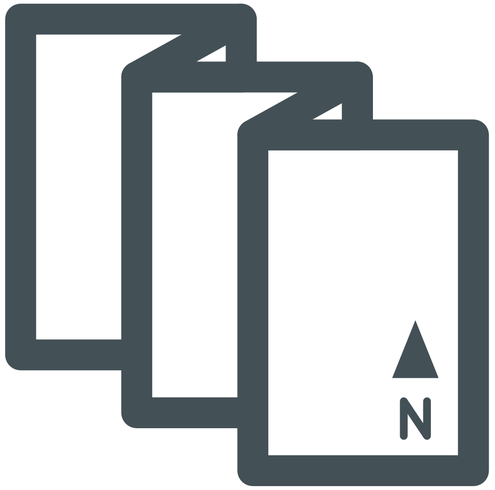
- distance
- traffic

ROUTING FOR TRANSPORTATION



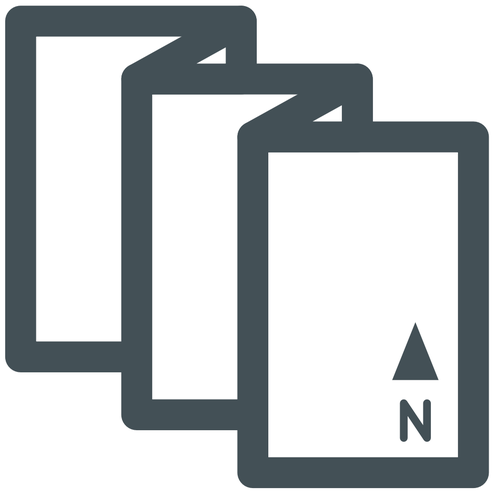
- distance
- traffic
- construction

ROUTING FOR TRANSPORTATION

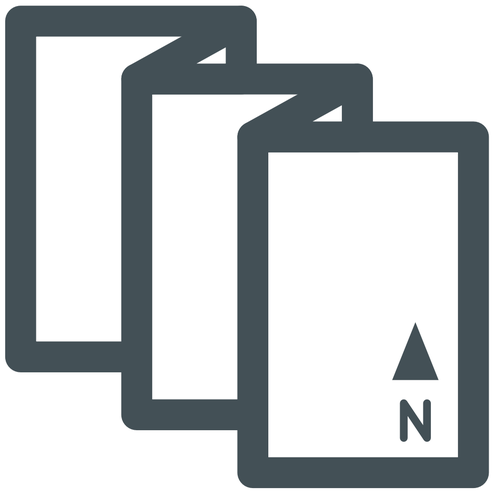


- distance
- traffic
- construction
- speed limits

ROUTING FOR FITNESS ACTIVITIES

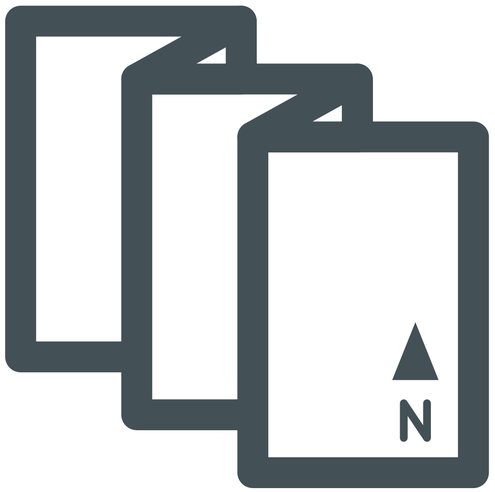


ROUTING FOR FITNESS ACTIVITIES



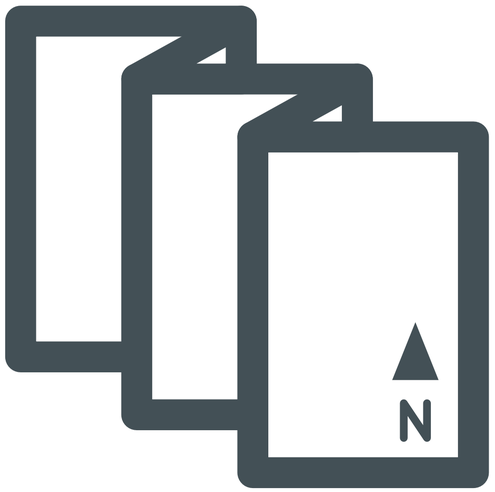
- distance

ROUTING FOR FITNESS ACTIVITIES



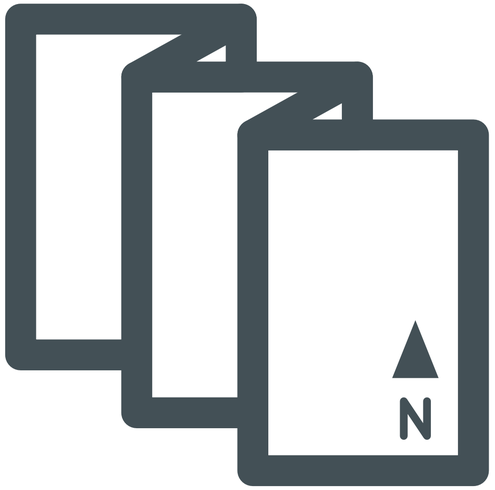
- distance
- elevation

ROUTING FOR FITNESS ACTIVITIES



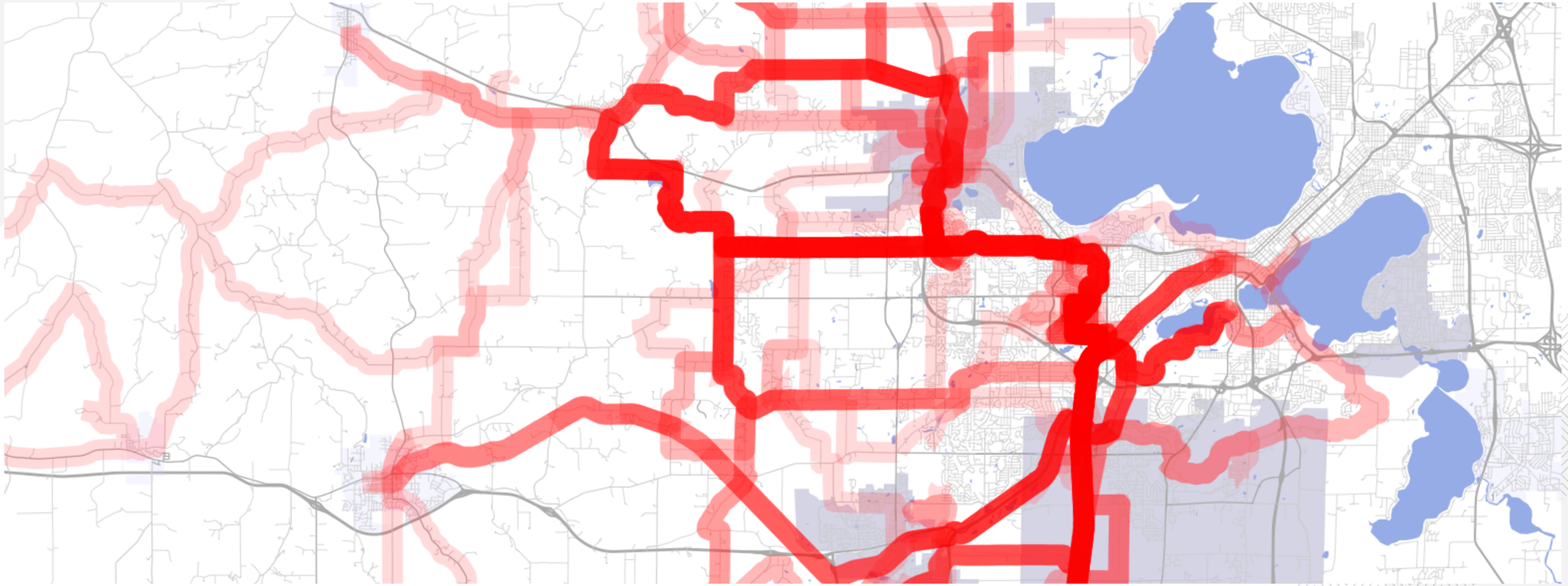
- distance
- elevation
- safety

ROUTING FOR FITNESS ACTIVITIES

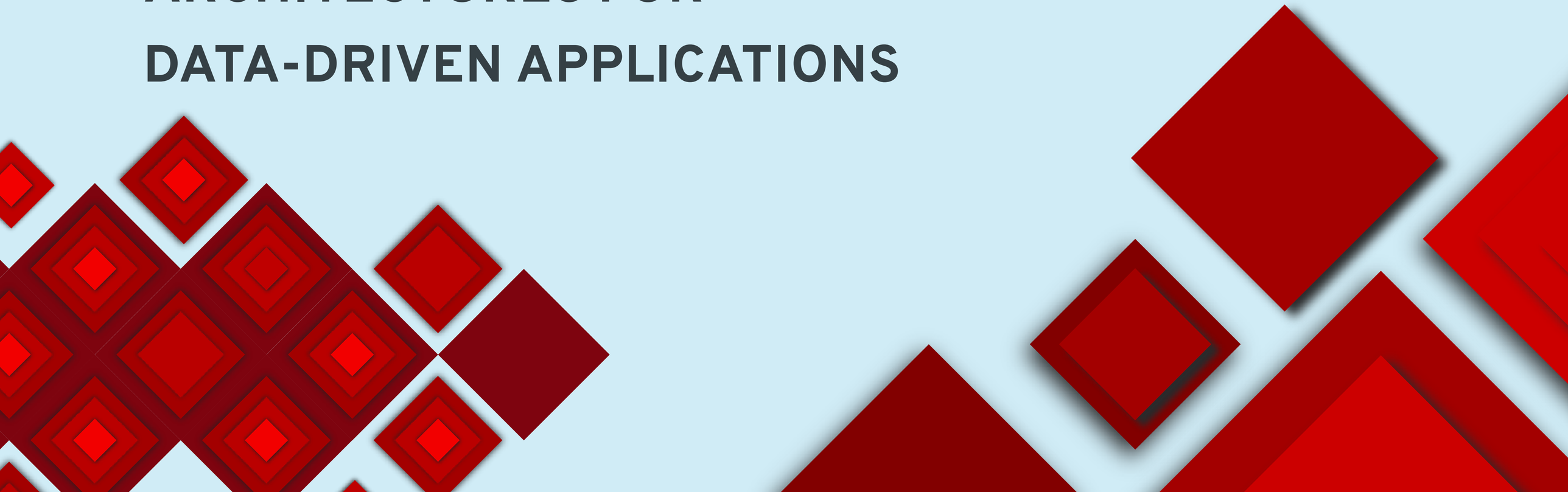


- distance
- elevation
- safety
- interruptions

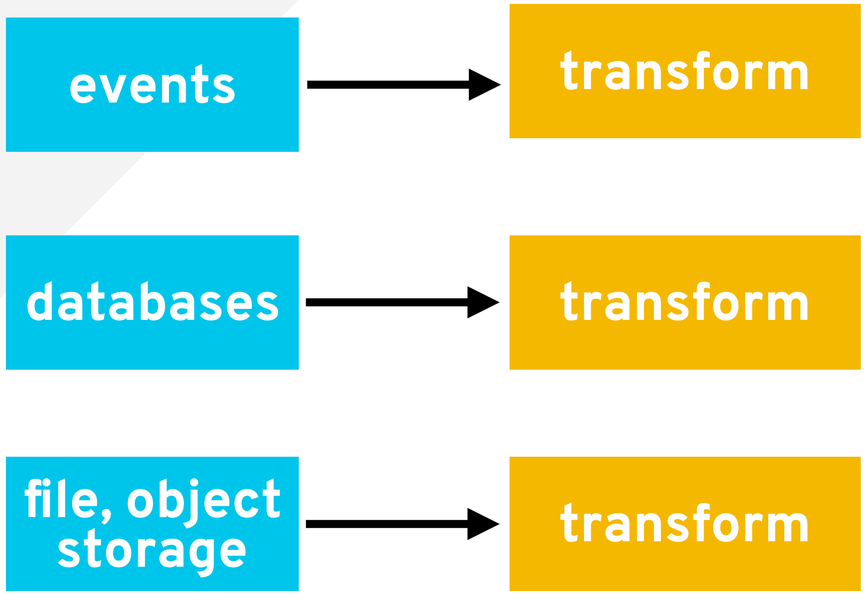
POPULARITY-BASED ROUTING



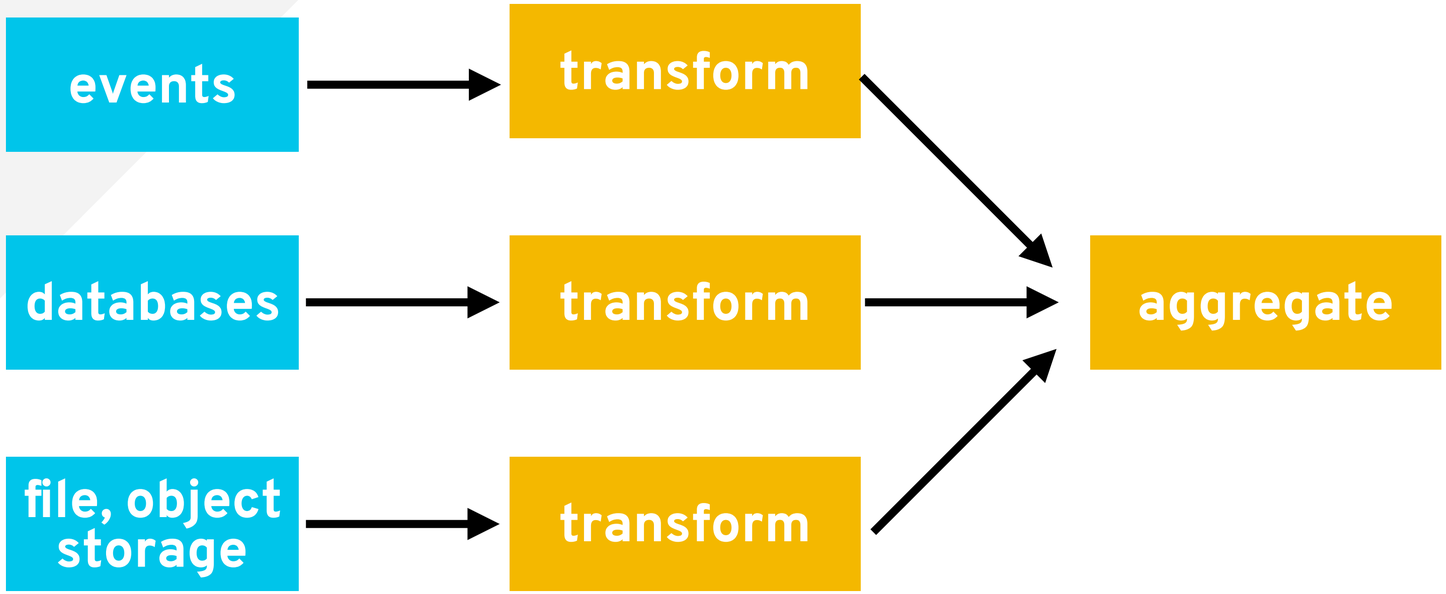
ARCHITECTURES FOR DATA-DRIVEN APPLICATIONS



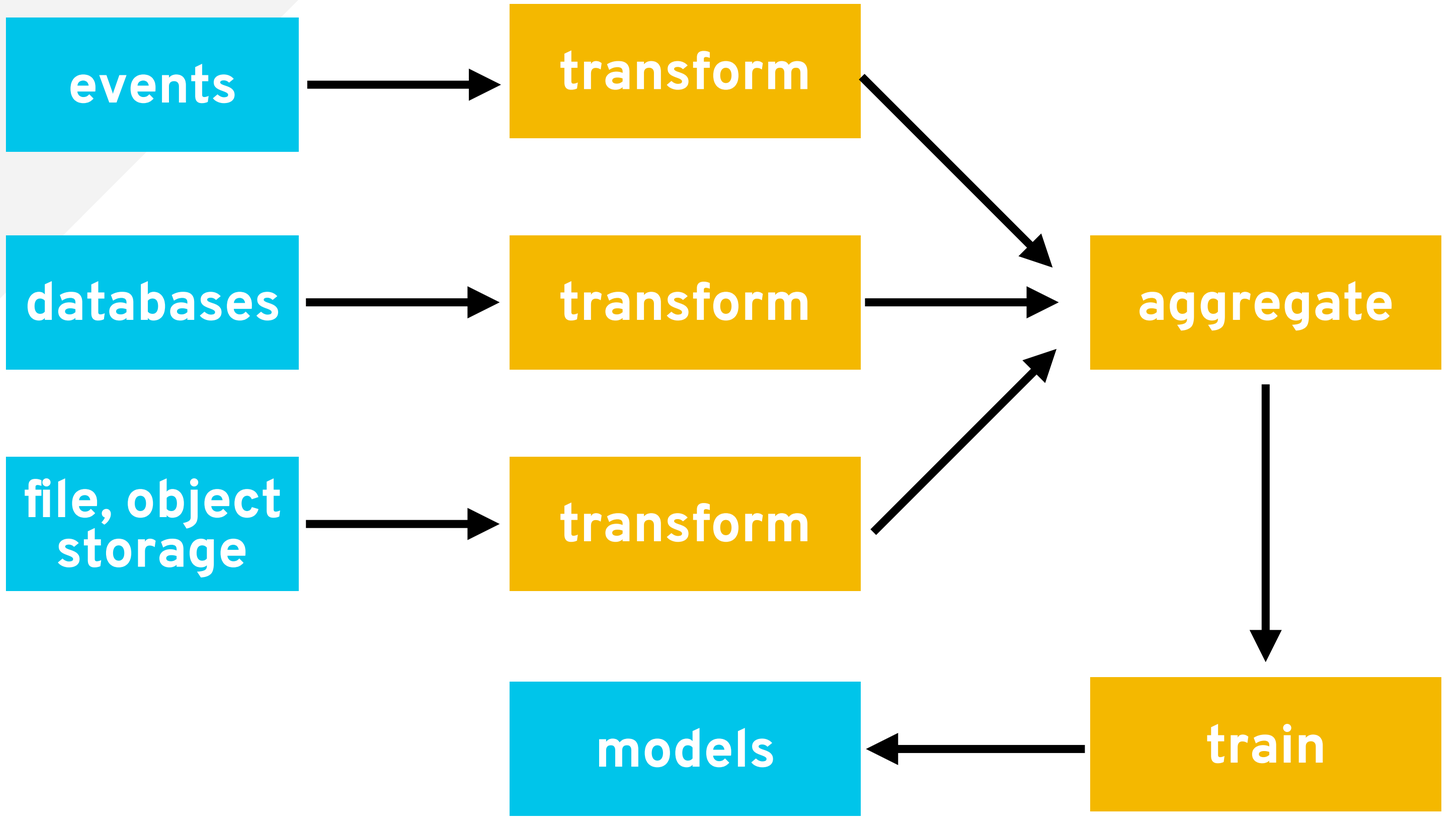
APPLICATION RESPONSIBILITIES



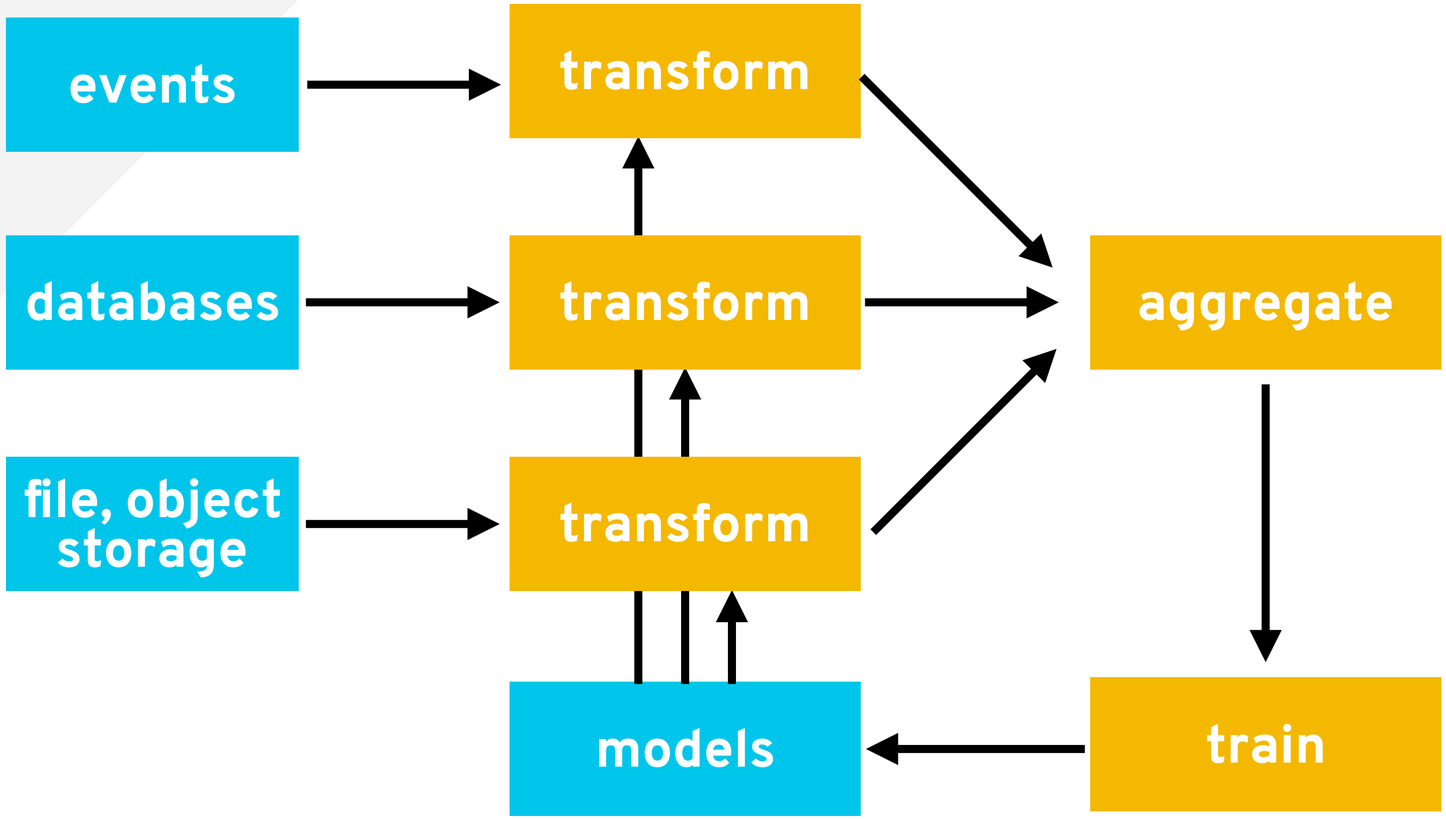
APPLICATION RESPONSIBILITIES



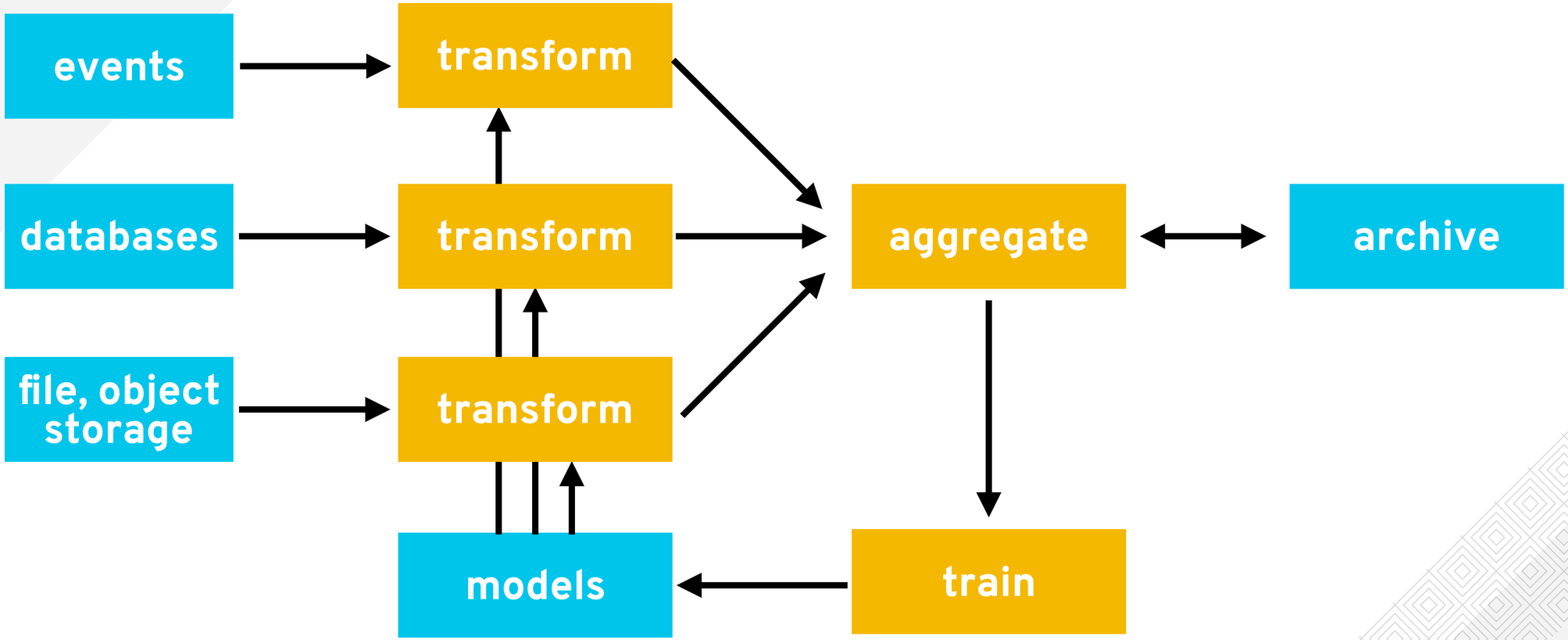
APPLICATION RESPONSIBILITIES



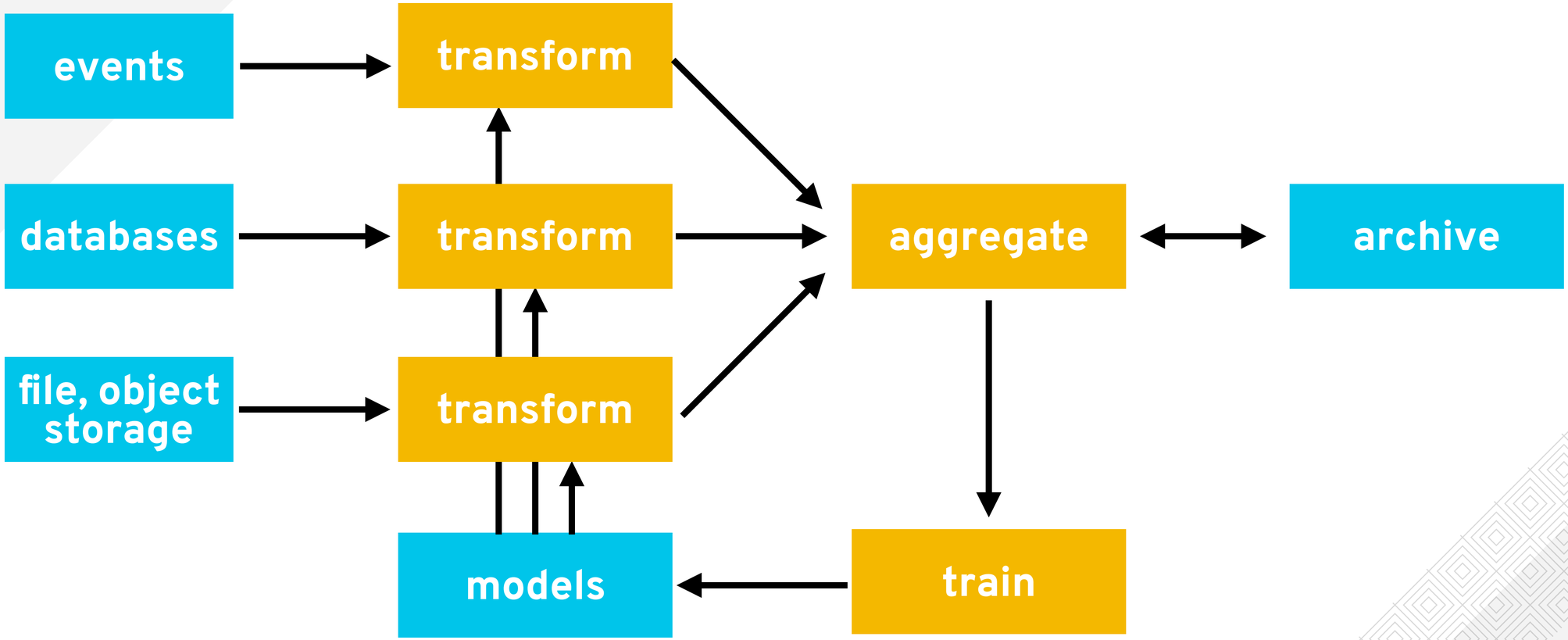
APPLICATION RESPONSIBILITIES



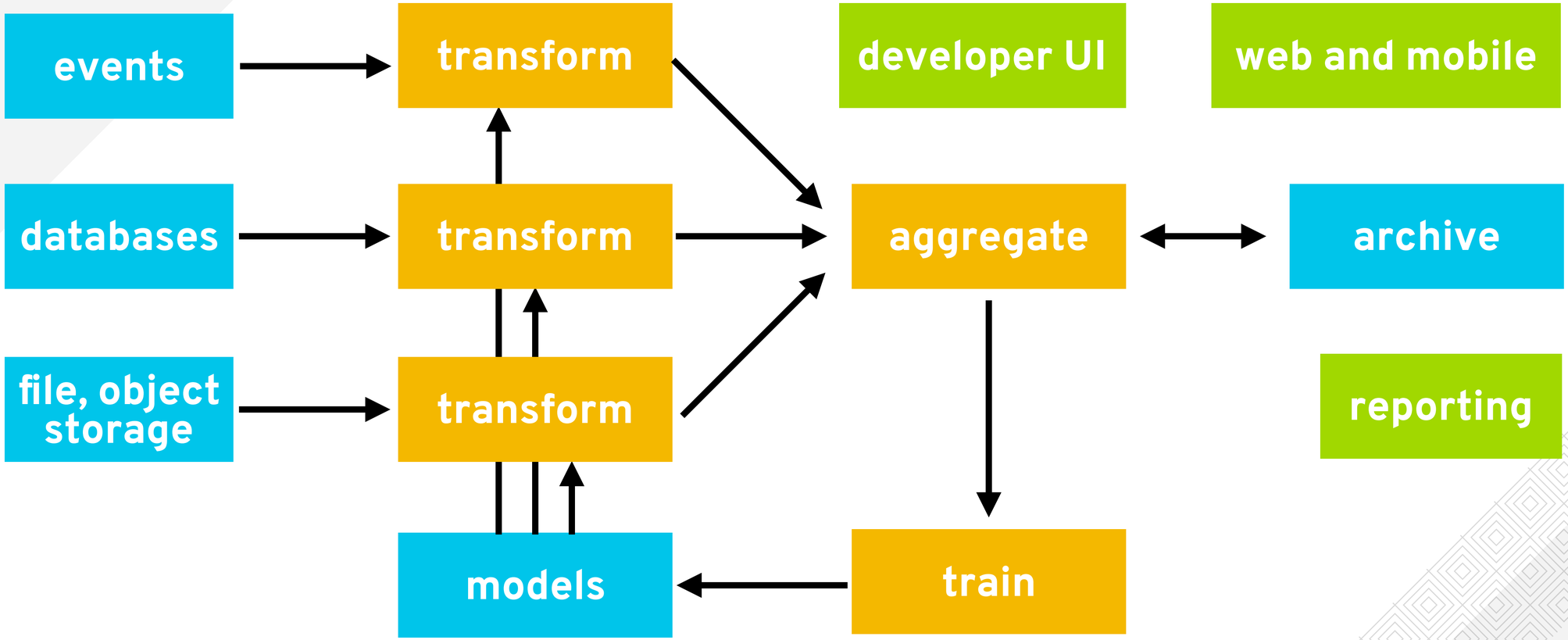
APPLICATION RESPONSIBILITIES



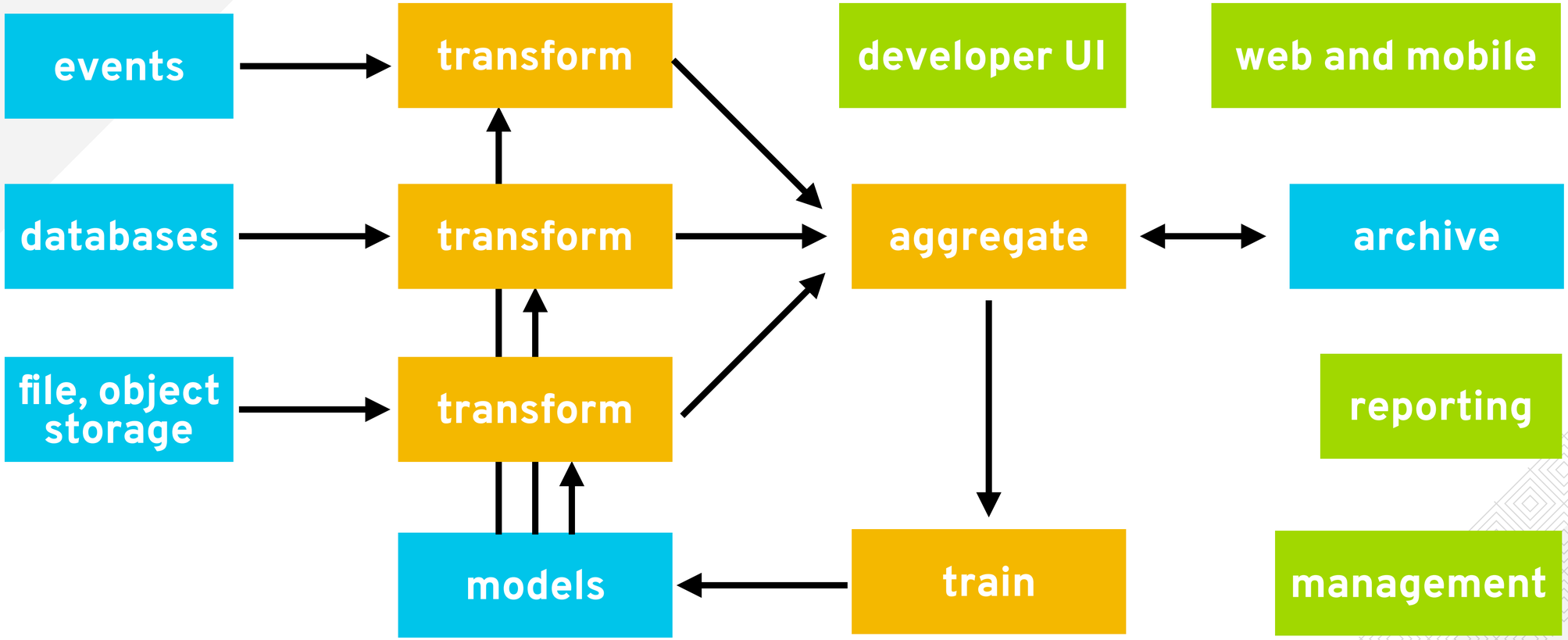
APPLICATION RESPONSIBILITIES



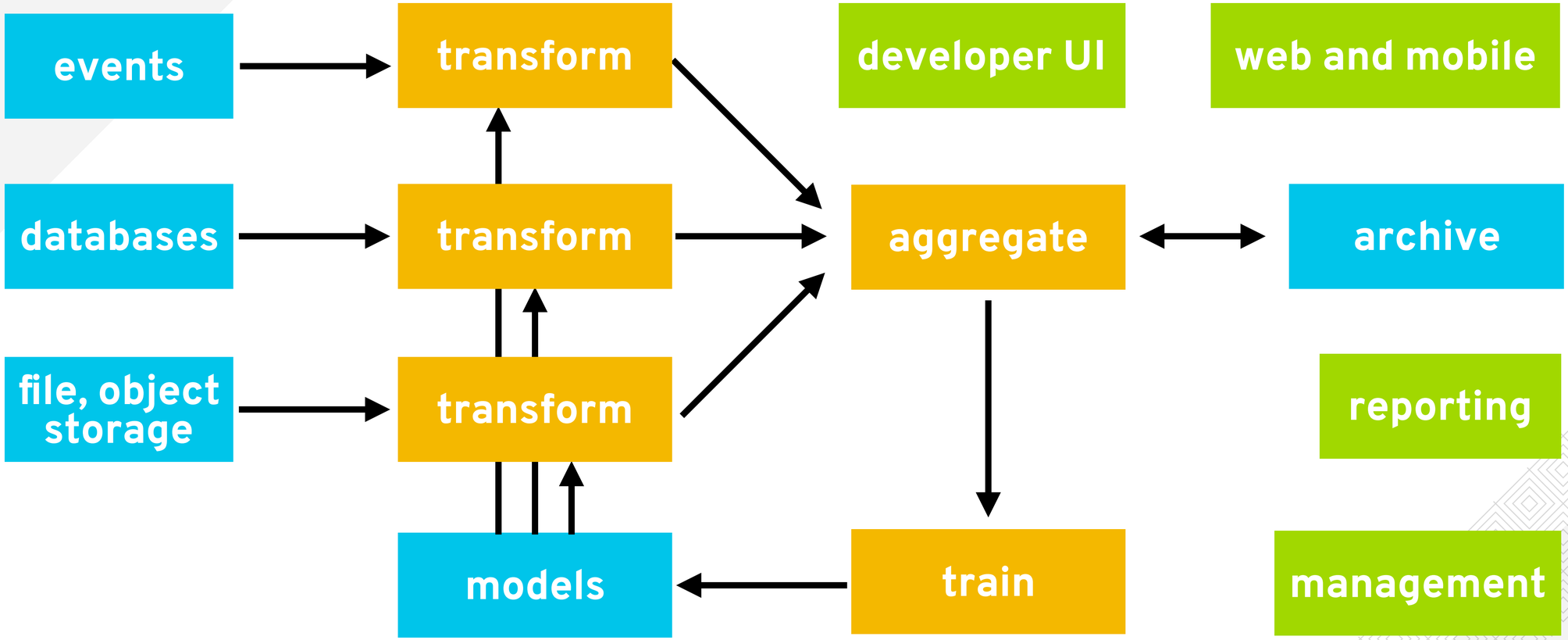
APPLICATION RESPONSIBILITIES



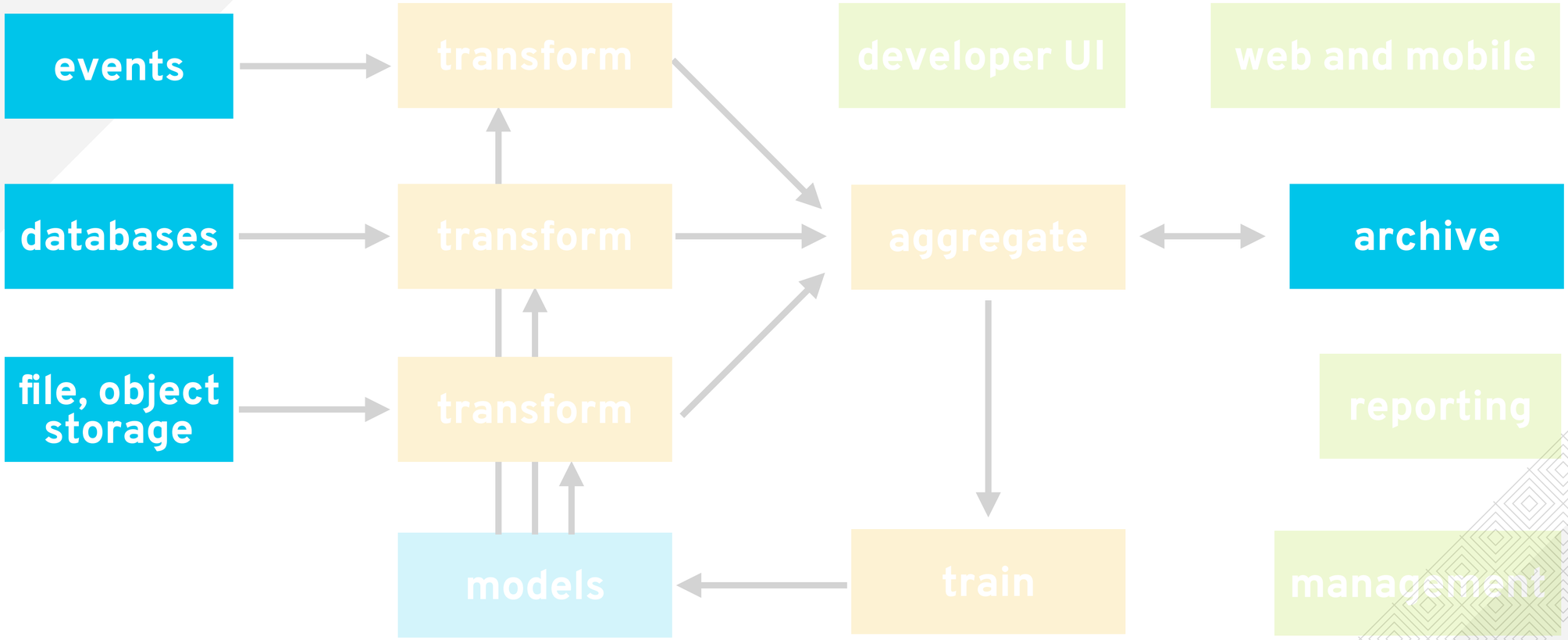
APPLICATION RESPONSIBILITIES



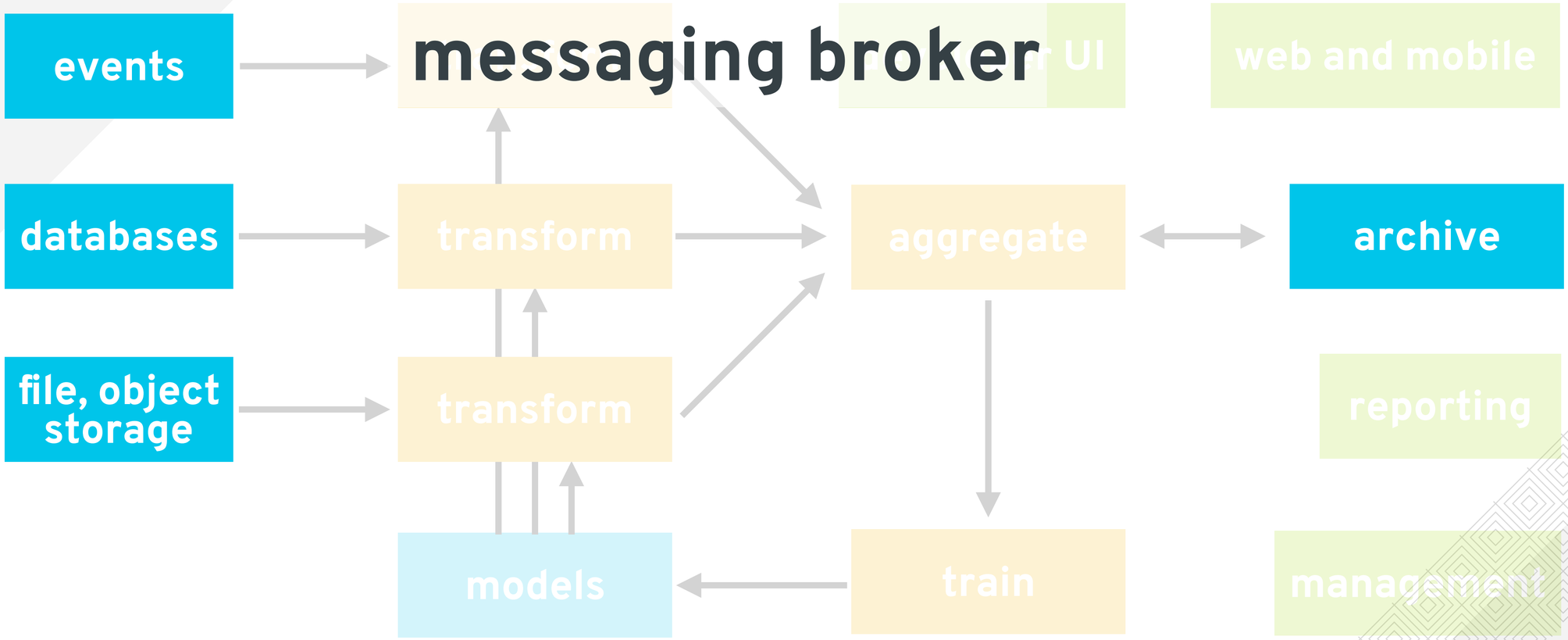
APPLICATION RESPONSIBILITIES



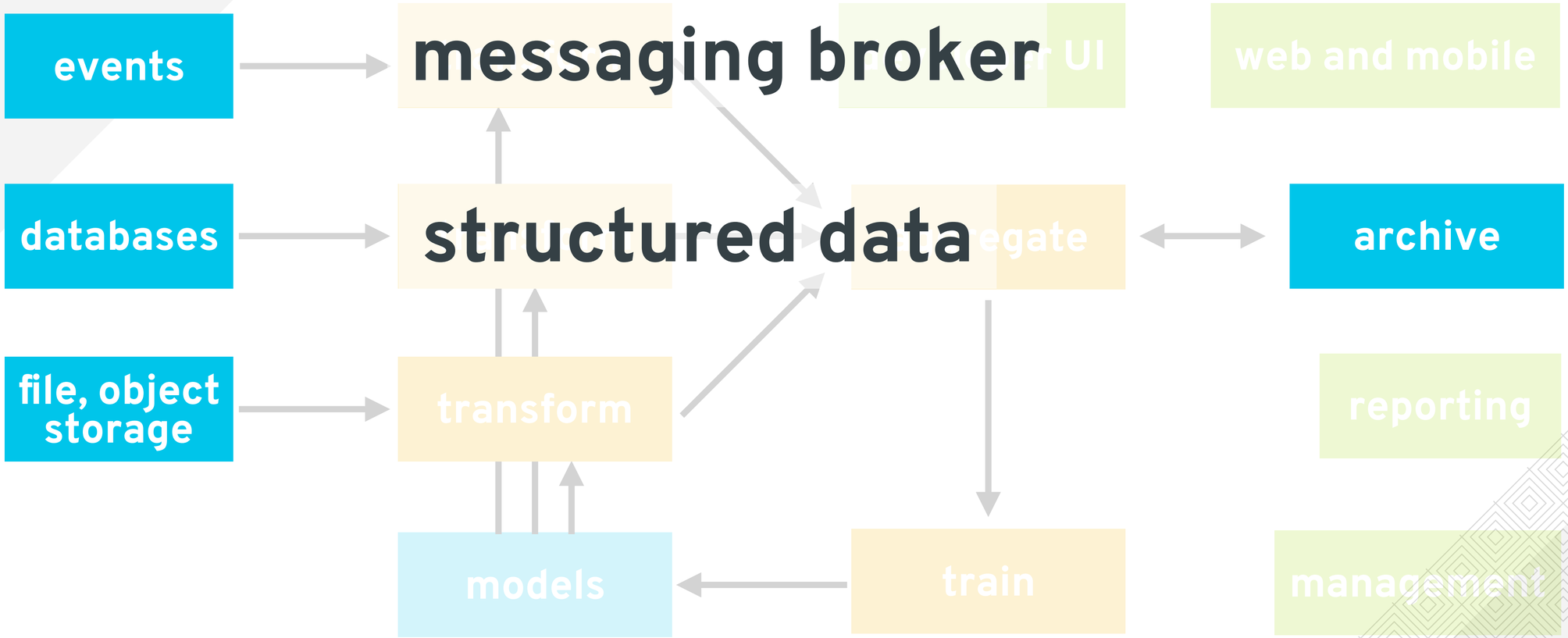
APPLICATION RESPONSIBILITIES



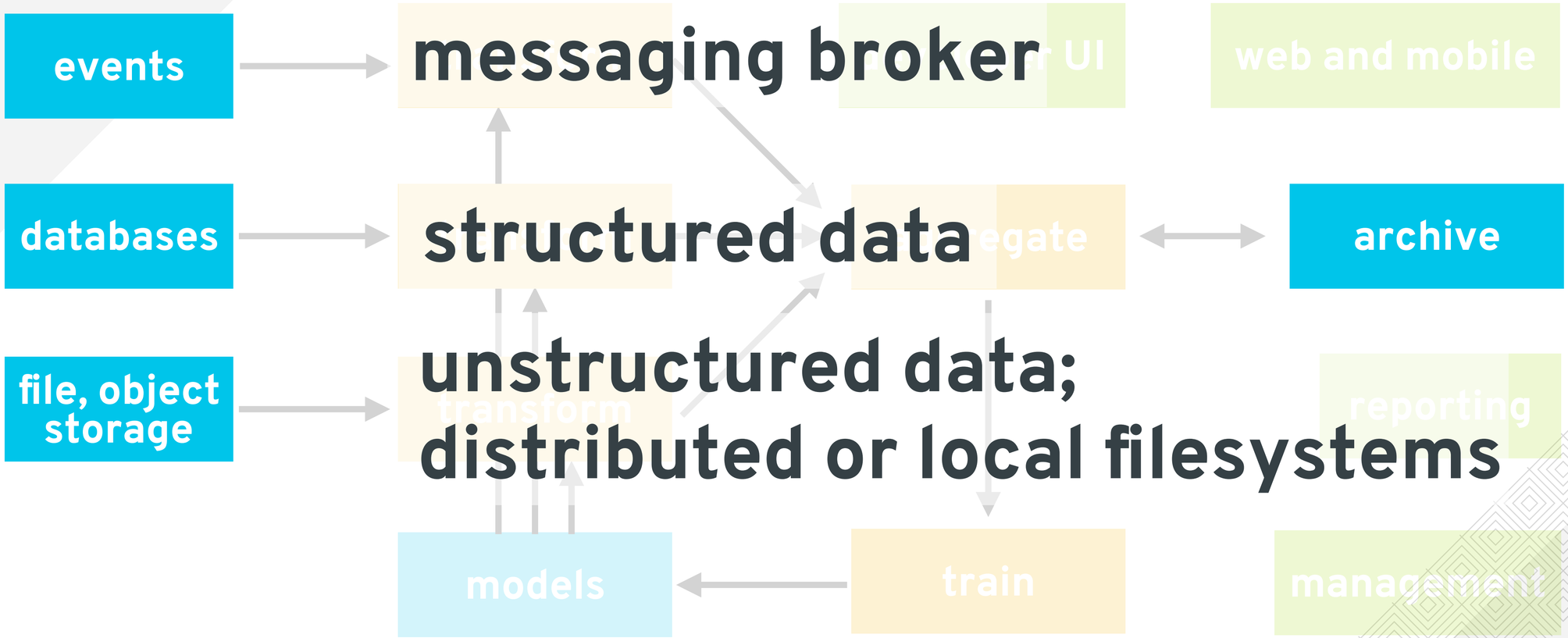
APPLICATION RESPONSIBILITIES



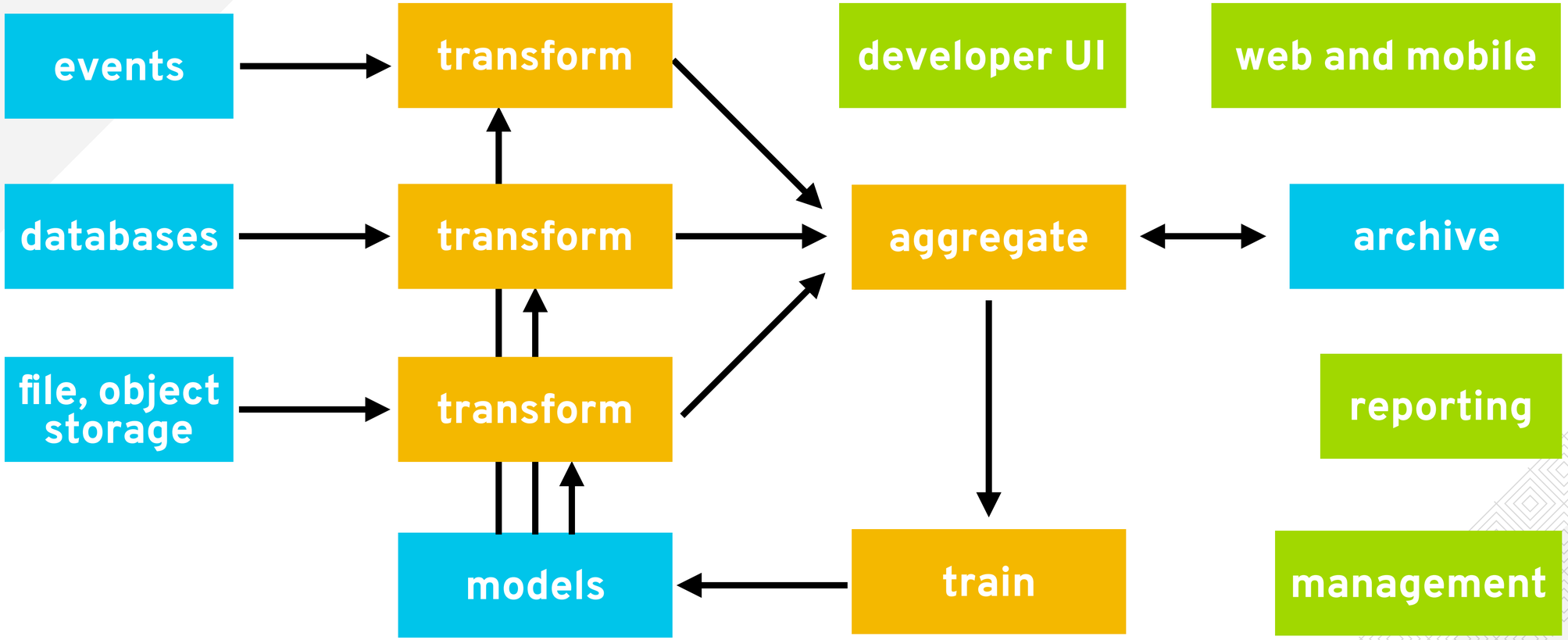
APPLICATION RESPONSIBILITIES



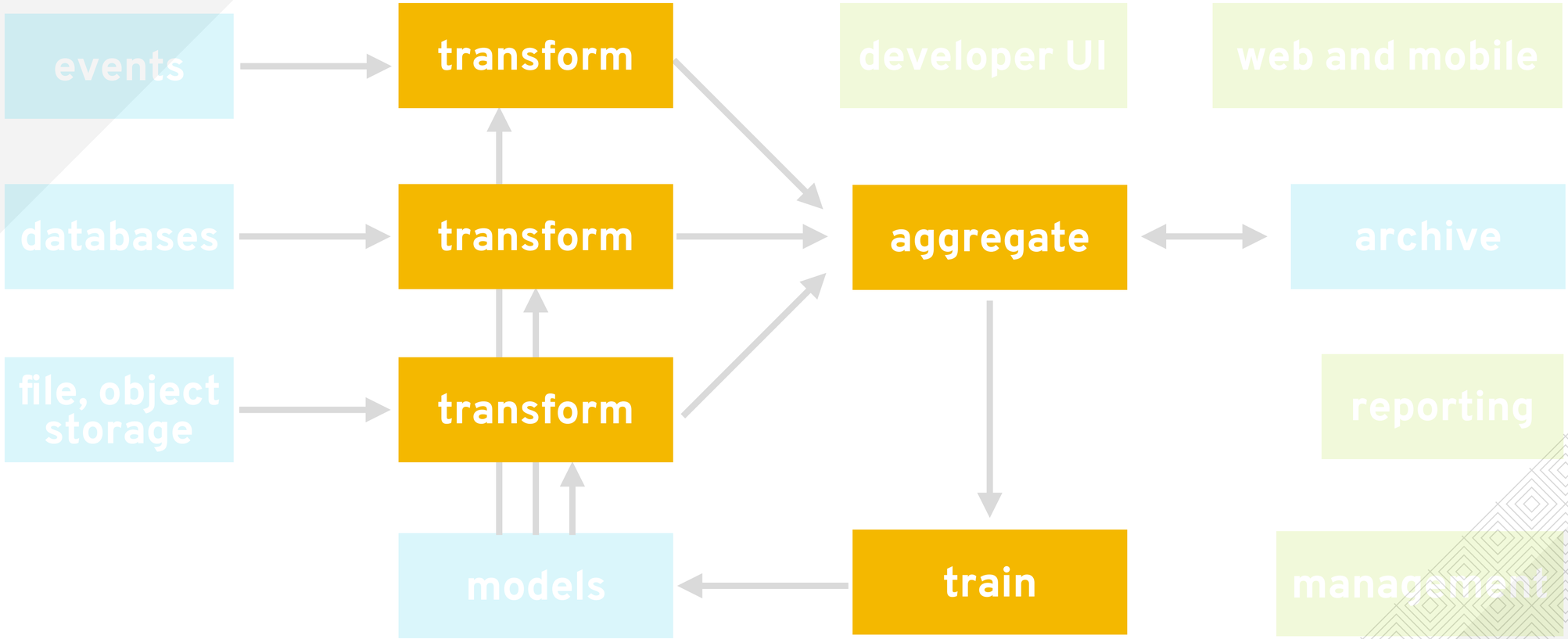
APPLICATION RESPONSIBILITIES



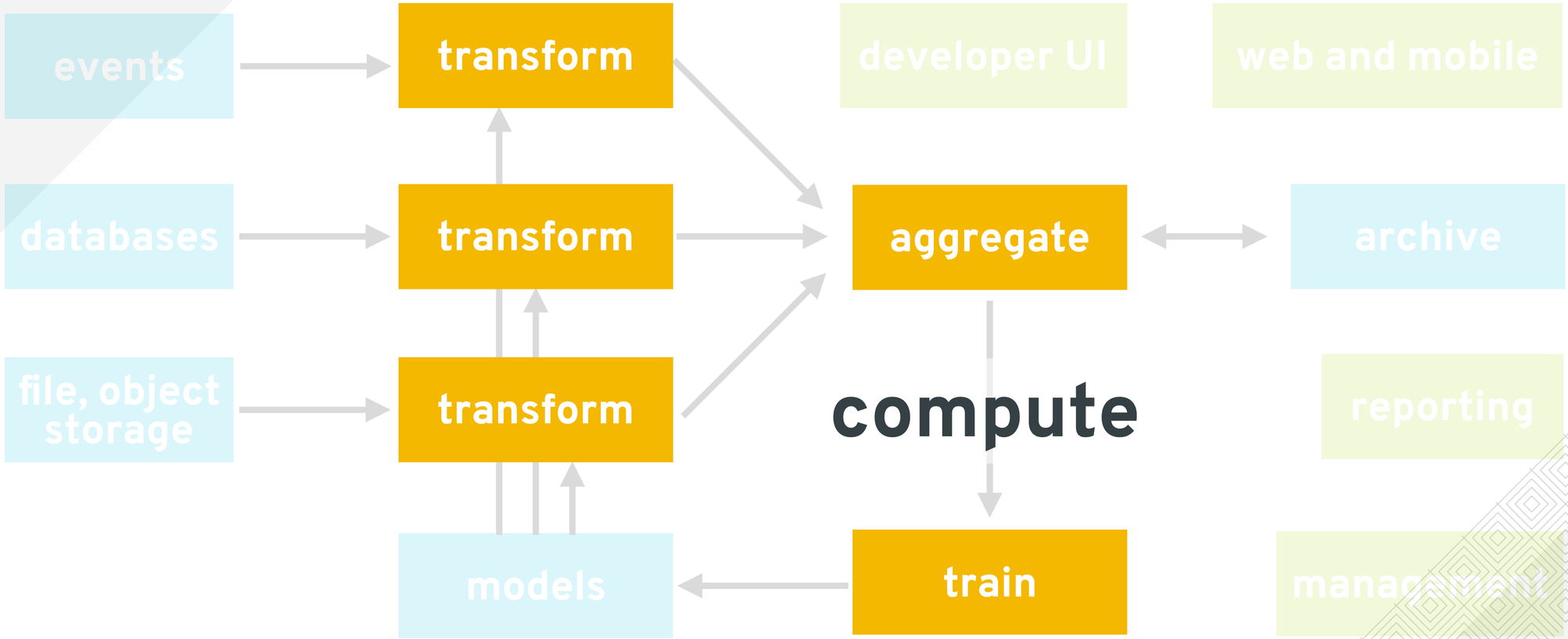
APPLICATION RESPONSIBILITIES



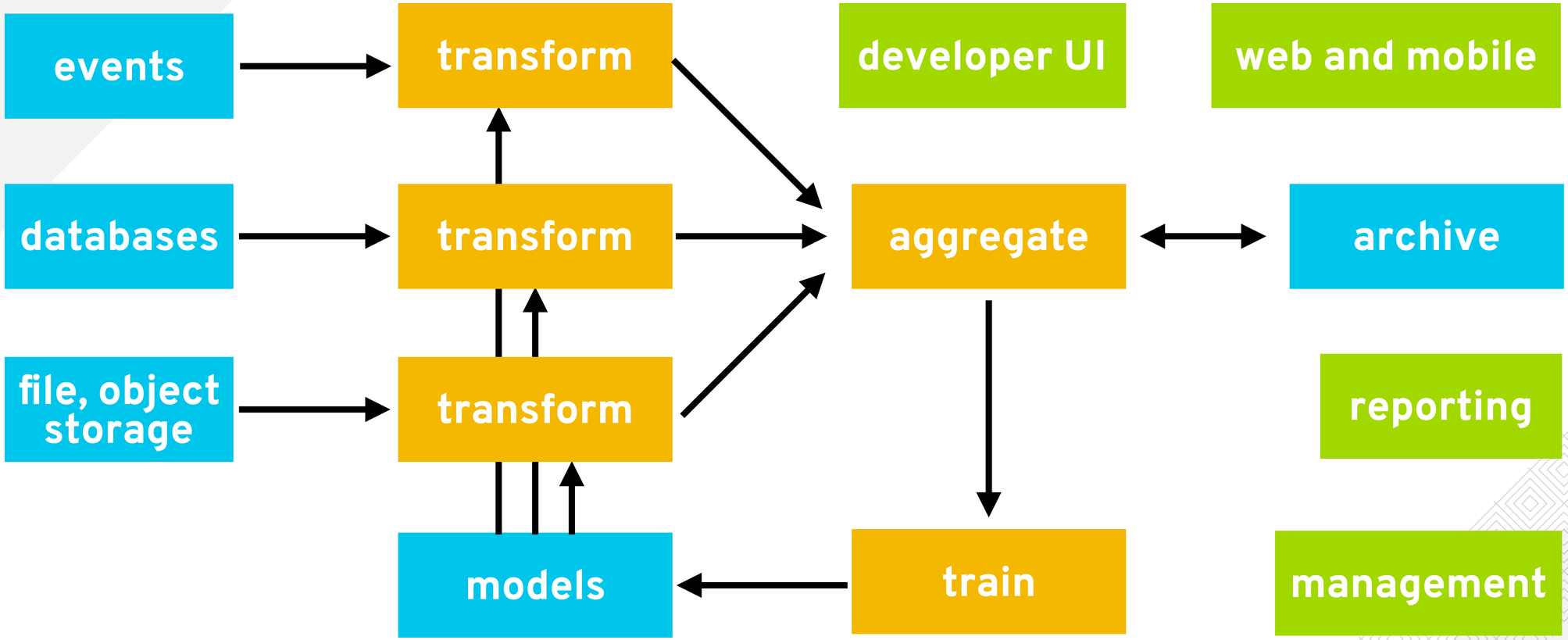
APPLICATION RESPONSIBILITIES



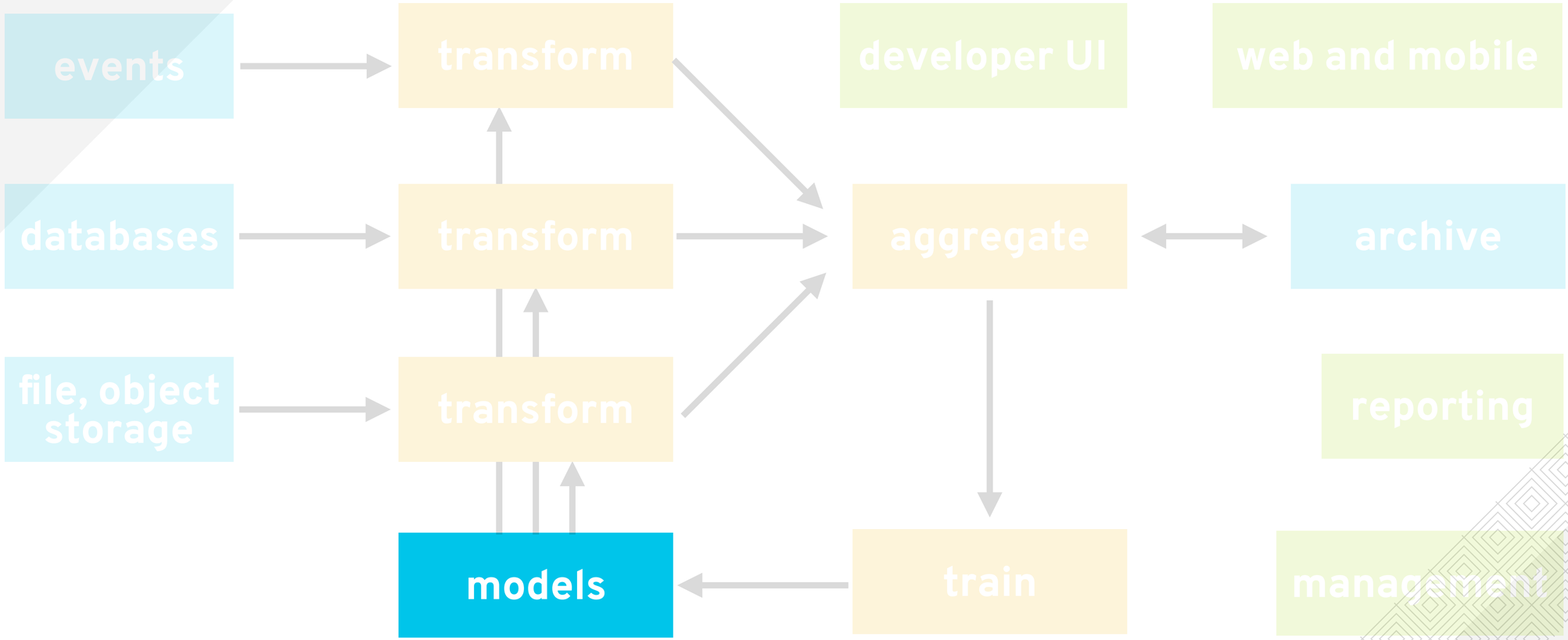
APPLICATION RESPONSIBILITIES



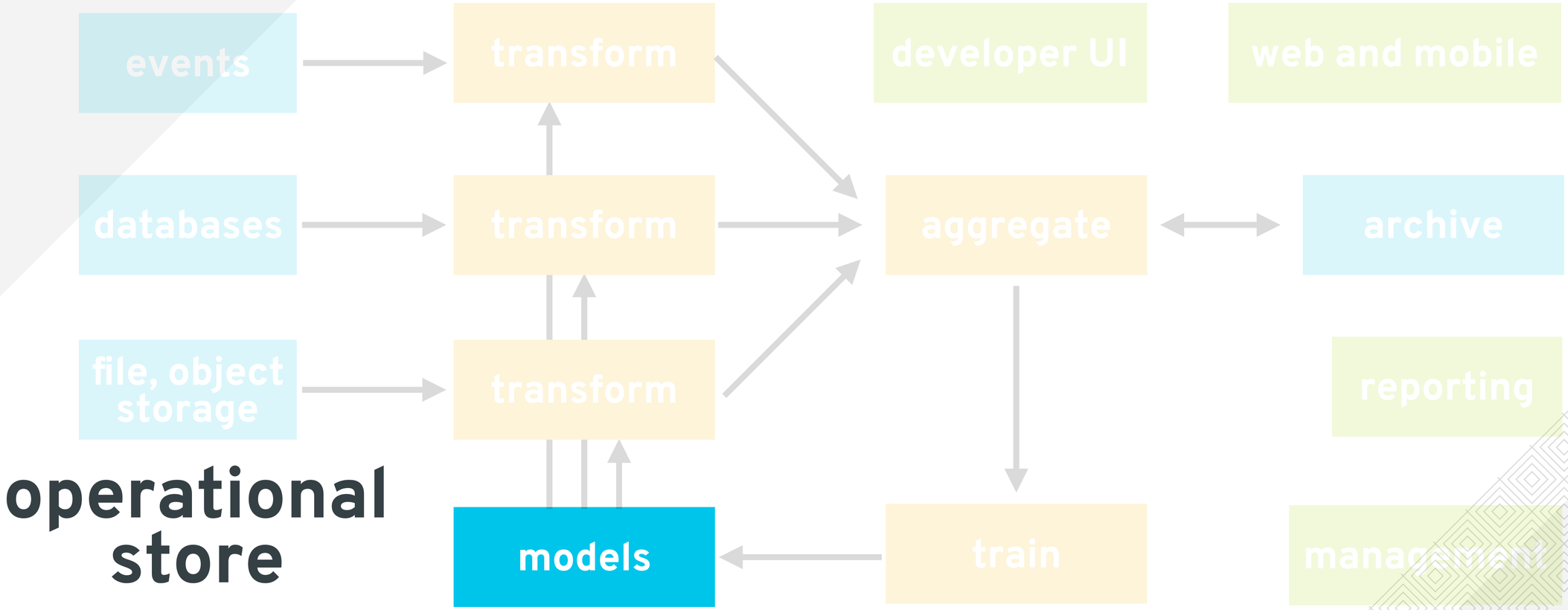
APPLICATION RESPONSIBILITIES



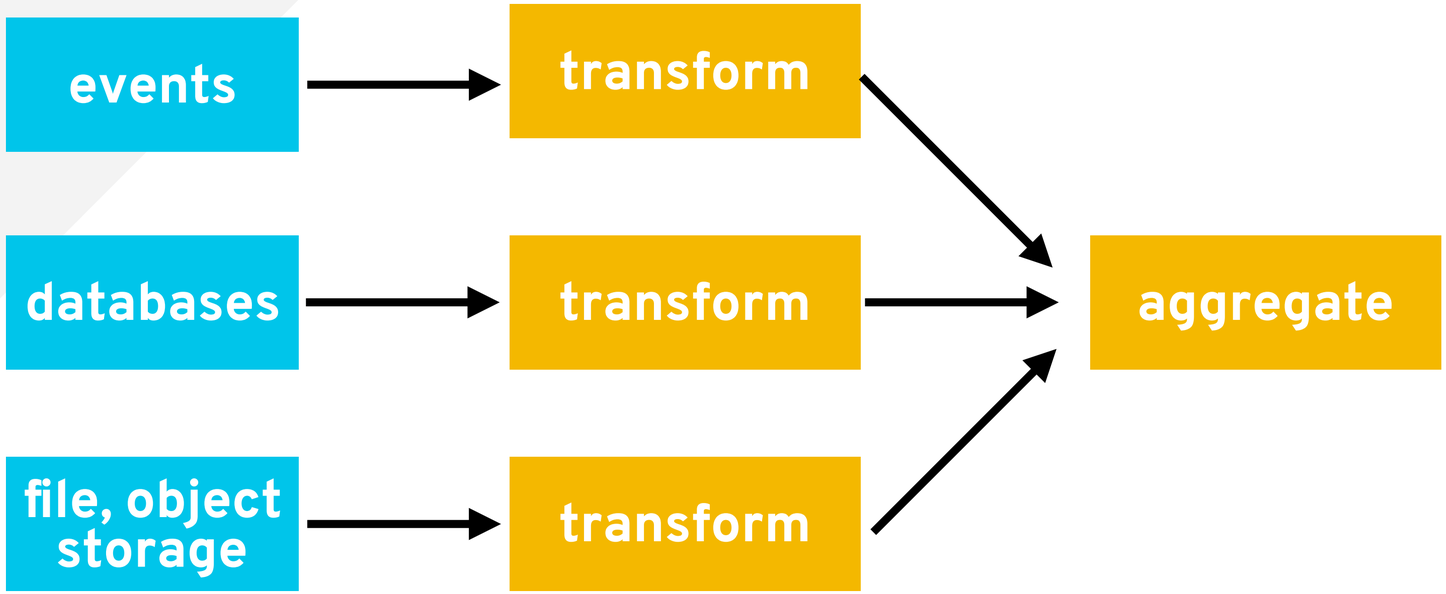
APPLICATION RESPONSIBILITIES



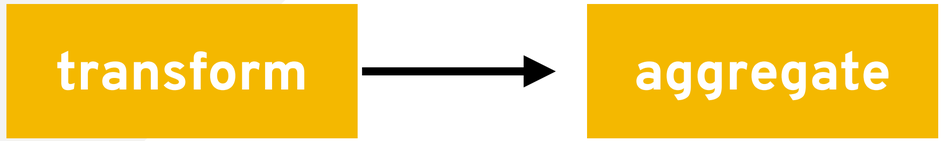
APPLICATION RESPONSIBILITIES



DATA FEDERATION IS FUNDAMENTAL



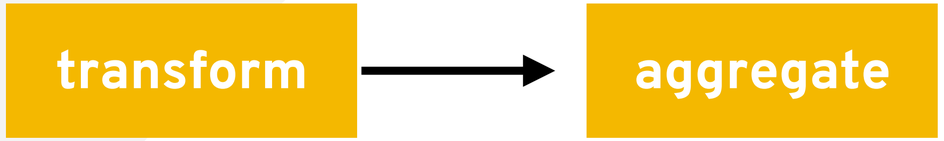
DATA FEDERATION IS FUNDAMENTAL



For many data science projects and data-driven applications, **80% or more of engineering effort** is devoted to **data wrangling tasks.**

A **data federation solution** allows us to **normalize and manipulate** various **data sources** and then access them from a **unified interface.**

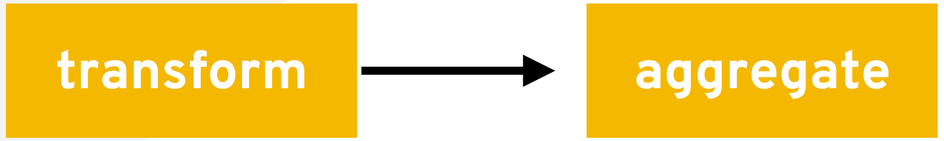
DATA FEDERATION IS FUNDAMENTAL



We can evaluate data federation solutions by asking several questions:

- What is the **source of truth**?

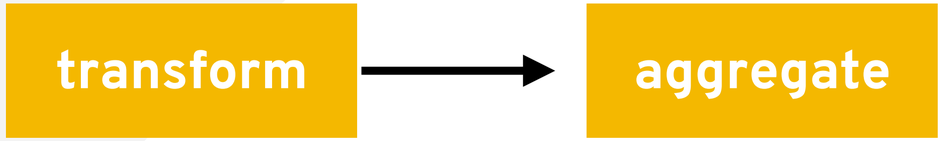
DATA FEDERATION IS FUNDAMENTAL



We can evaluate data federation solutions by asking several questions:

- What is the **source of truth**?
- How does the solution **scale**?

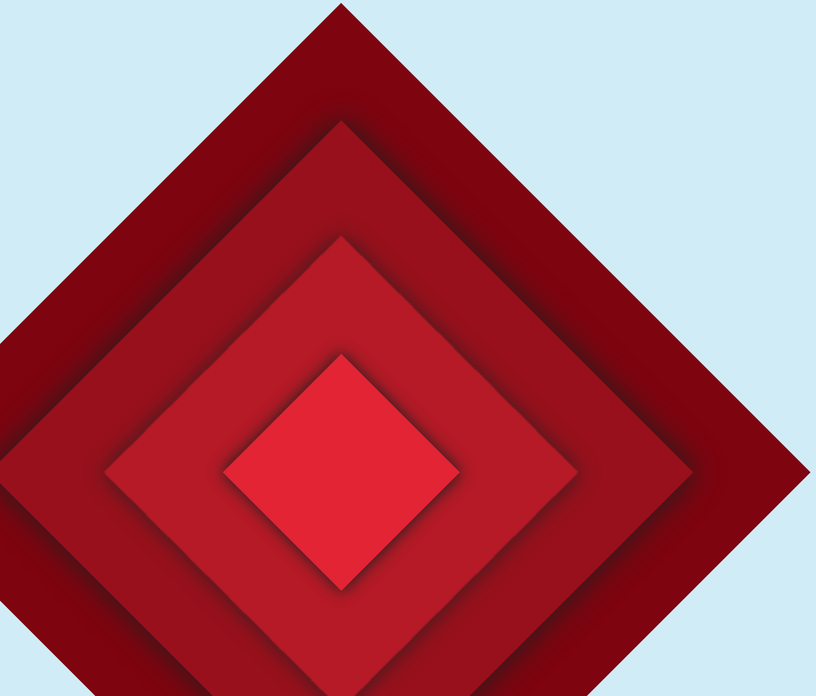
DATA FEDERATION IS FUNDAMENTAL

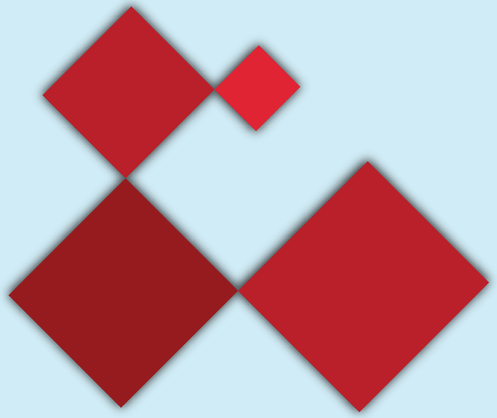


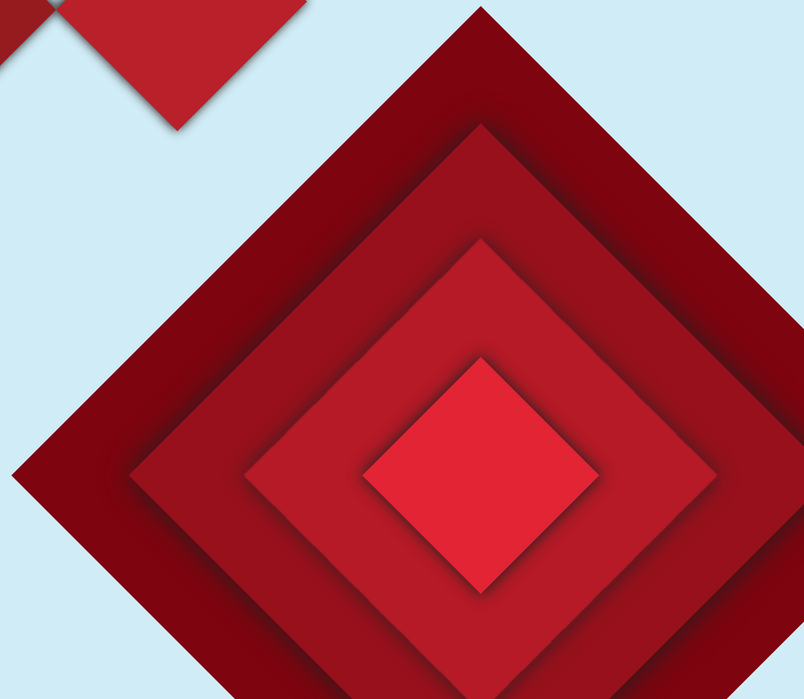
We can evaluate data federation solutions by asking several questions:

- What is the **source of truth**?
- How does the solution **scale**?
- What **restrictions** does it impose on **how we interact with our data**?

LEGACY ARCHITECTURES



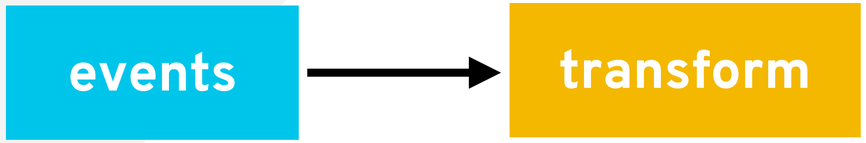




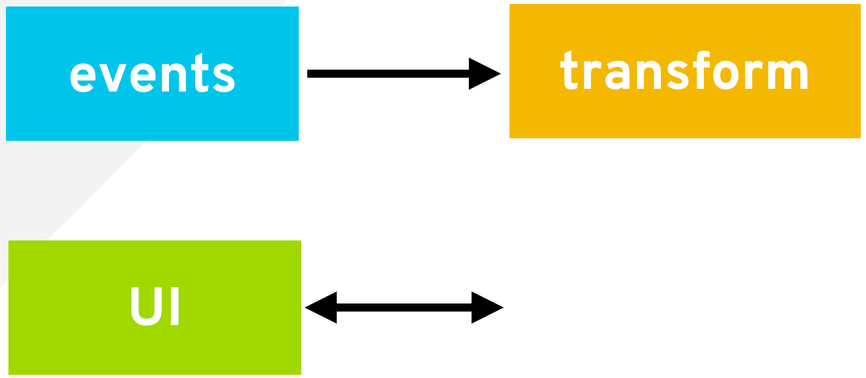
CONVENTIONAL DATA WAREHOUSE

events

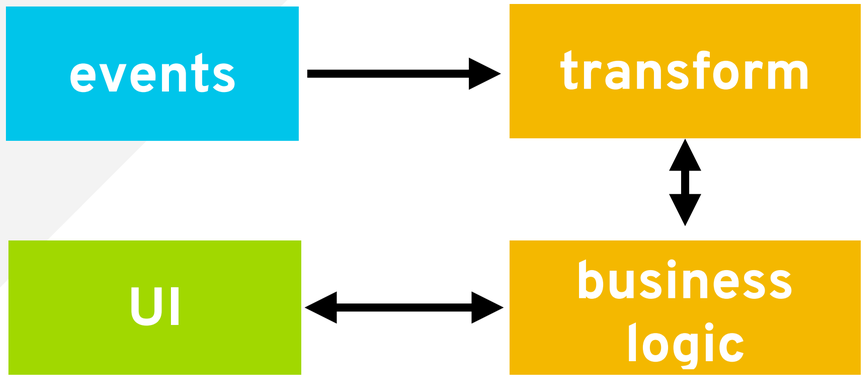
CONVENTIONAL DATA WAREHOUSE



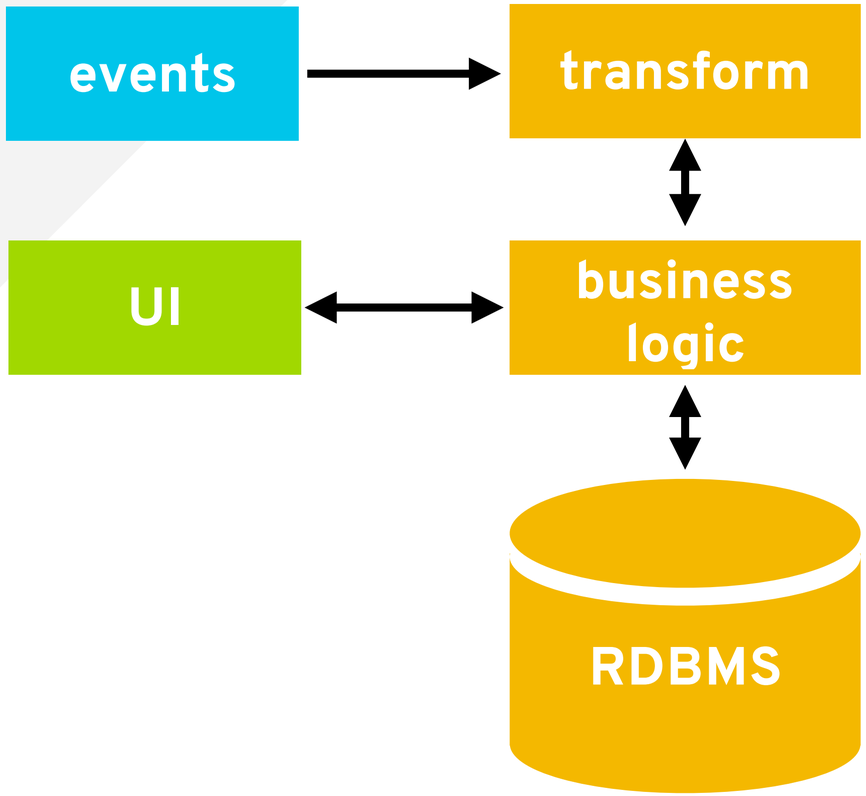
CONVENTIONAL DATA WAREHOUSE



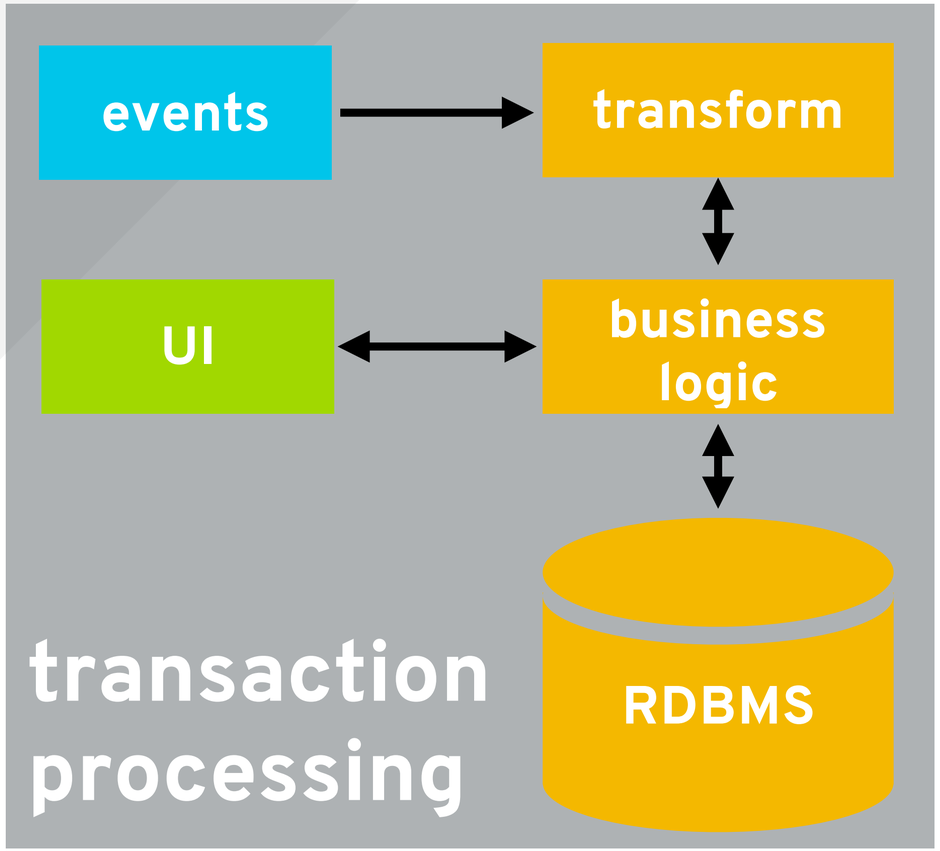
CONVENTIONAL DATA WAREHOUSE



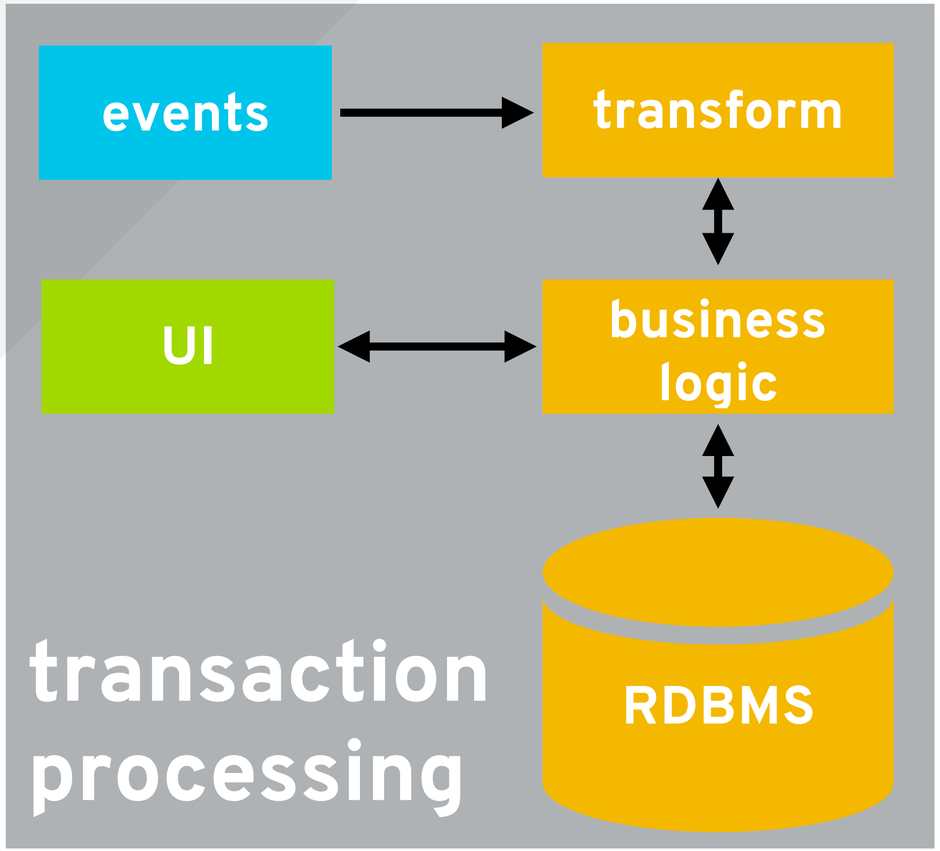
CONVENTIONAL DATA WAREHOUSE



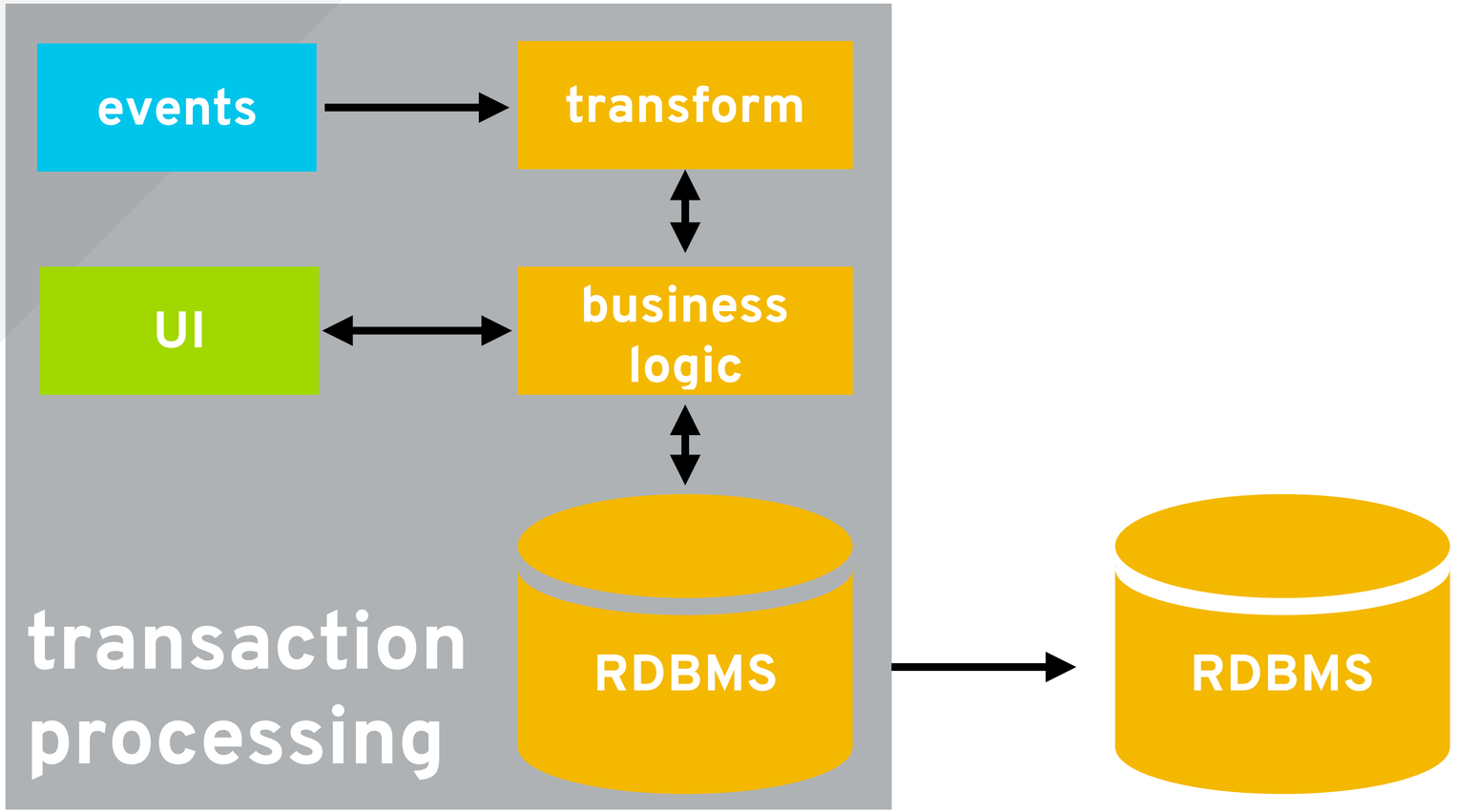
CONVENTIONAL DATA WAREHOUSE



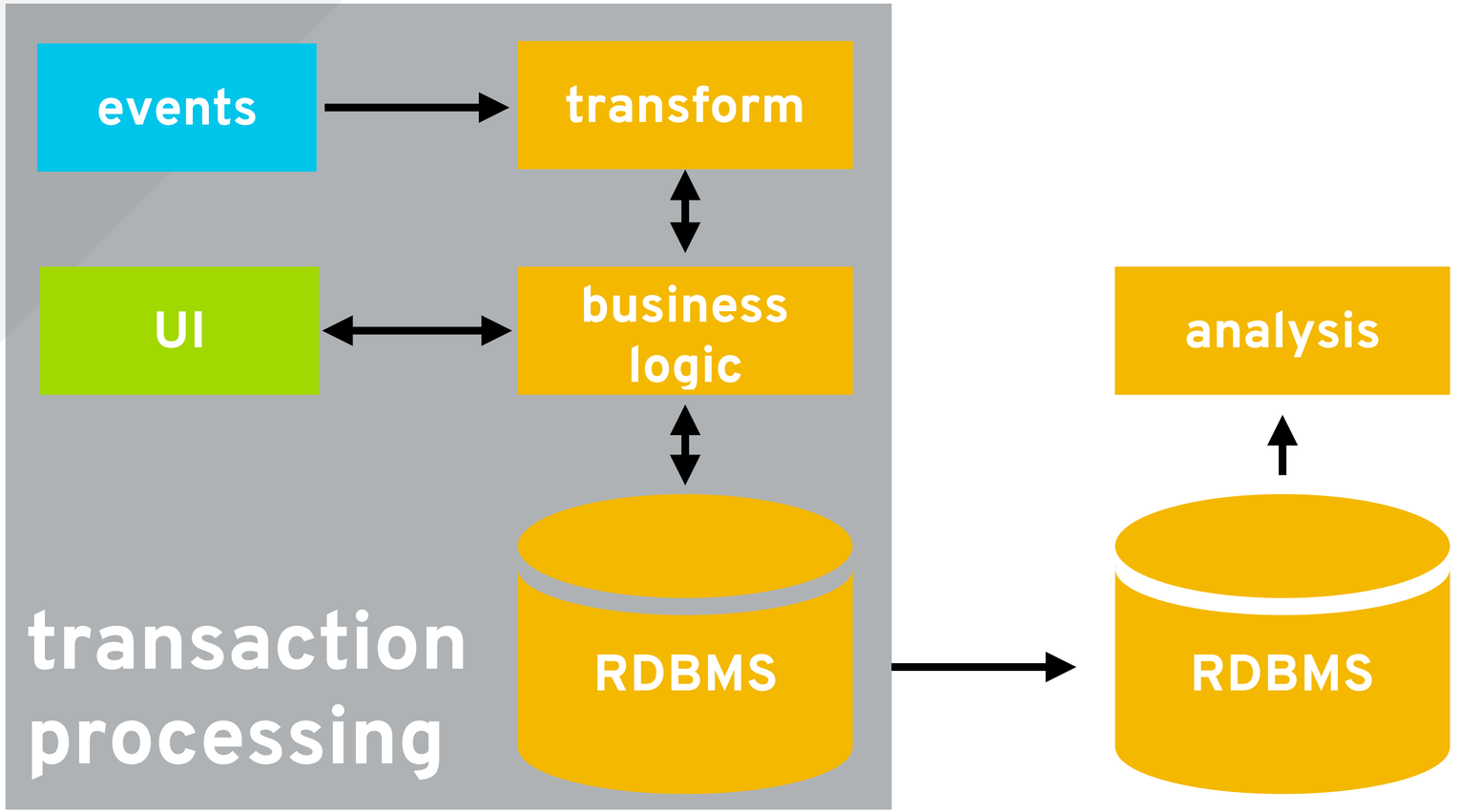
CONVENTIONAL DATA WAREHOUSE



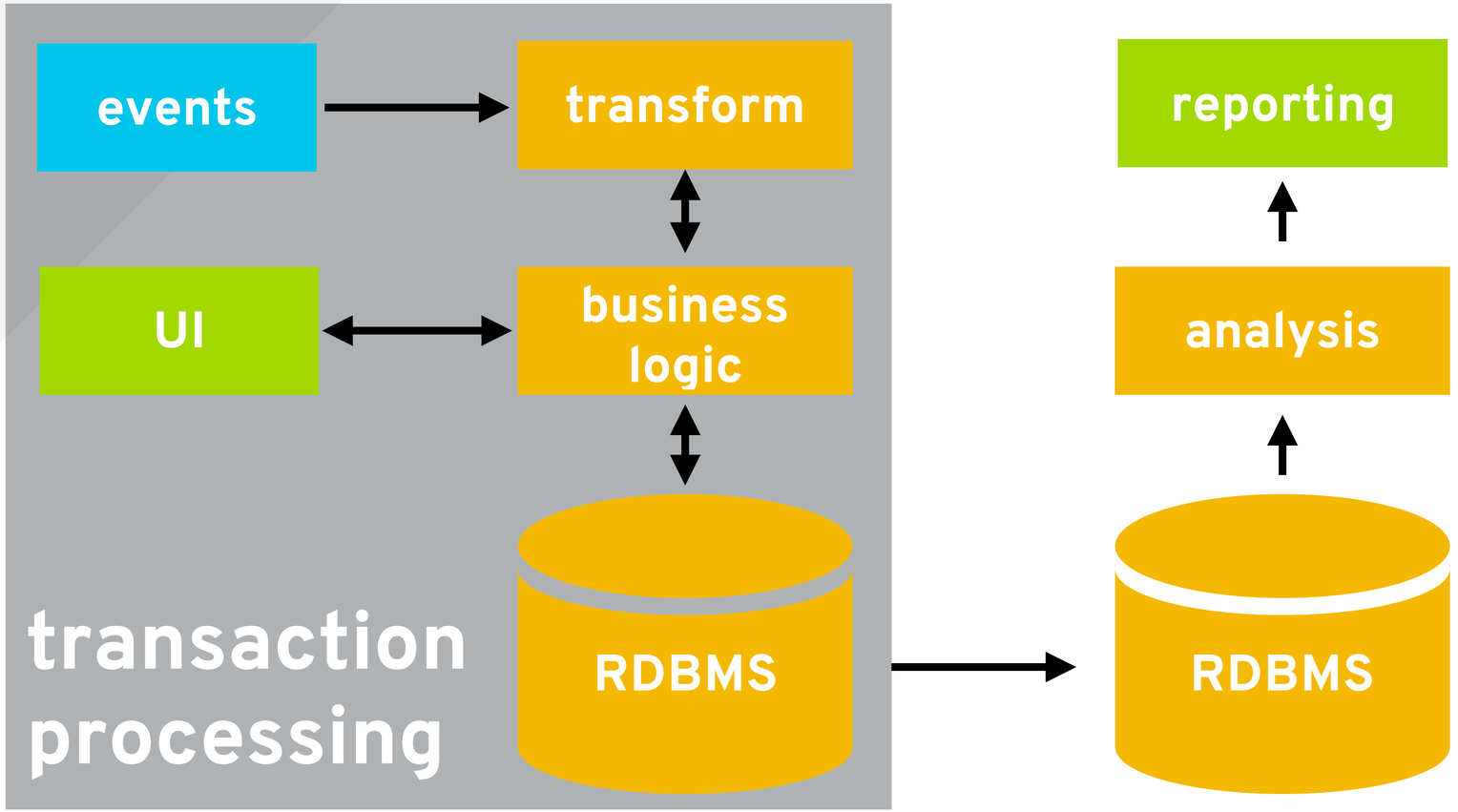
CONVENTIONAL DATA WAREHOUSE



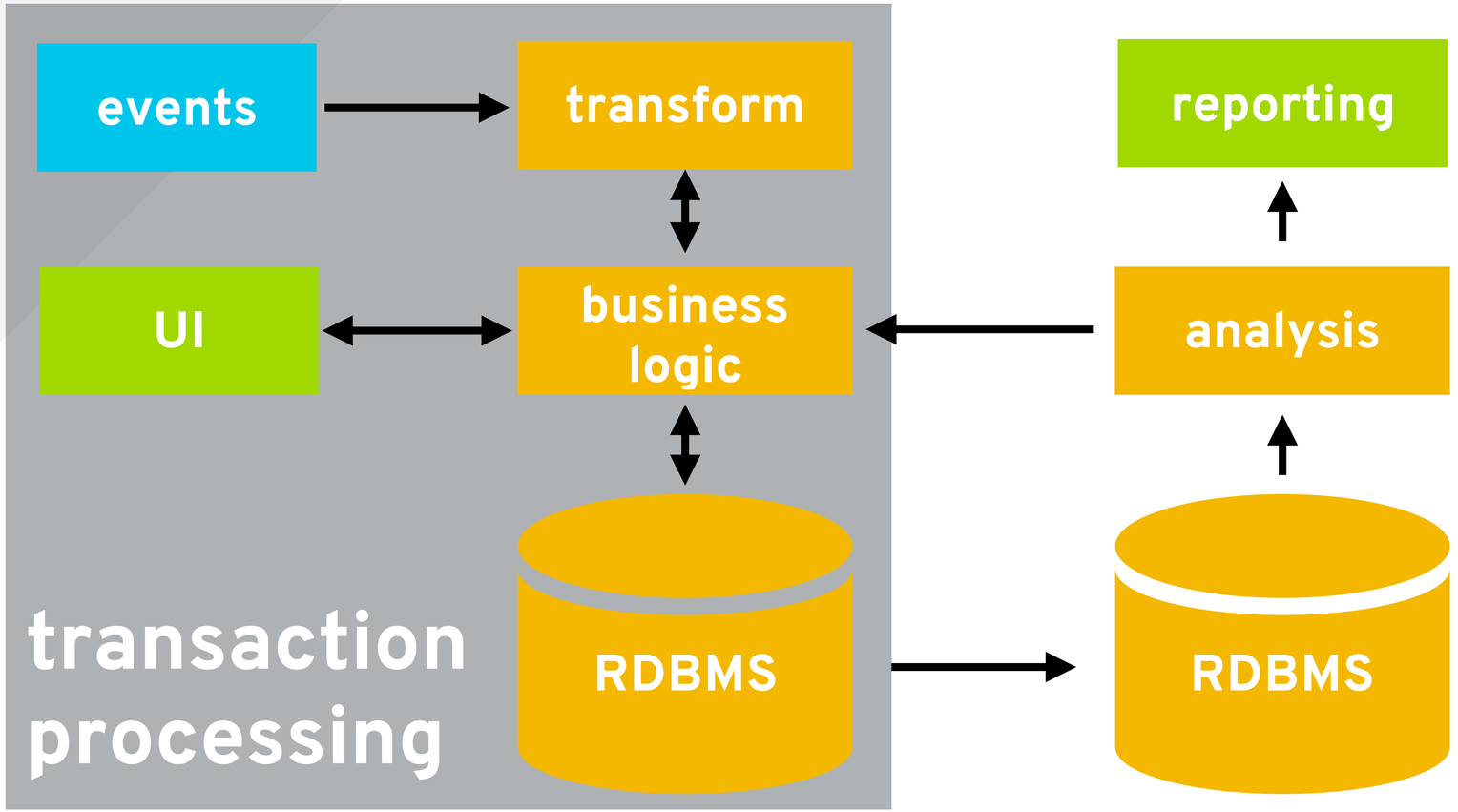
CONVENTIONAL DATA WAREHOUSE



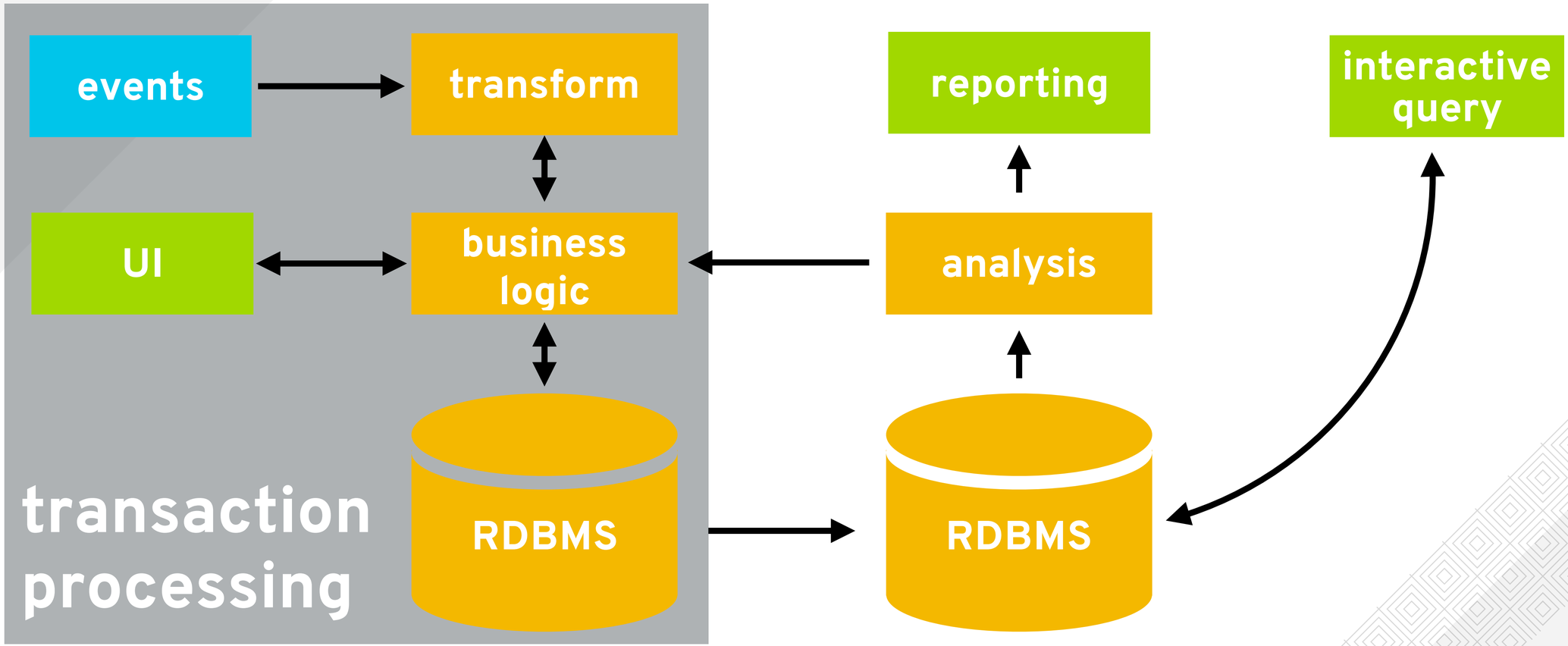
CONVENTIONAL DATA WAREHOUSE



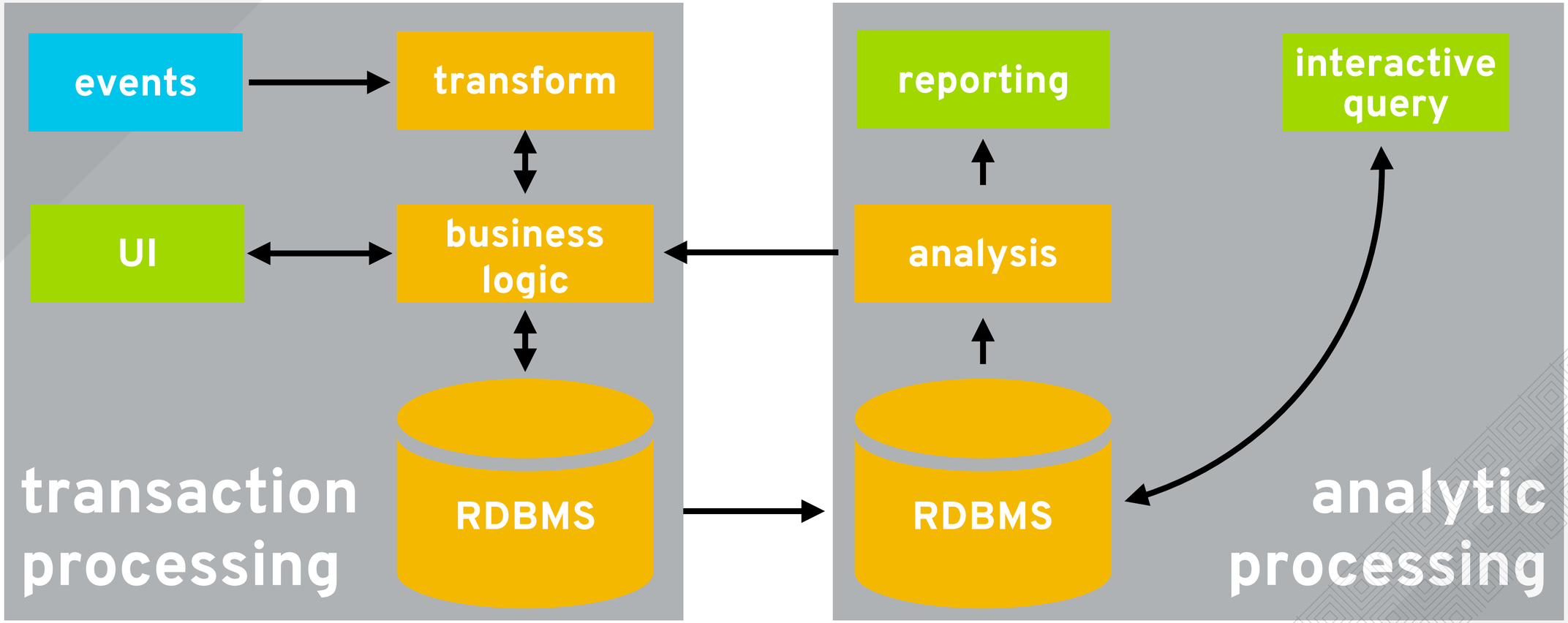
CONVENTIONAL DATA WAREHOUSE



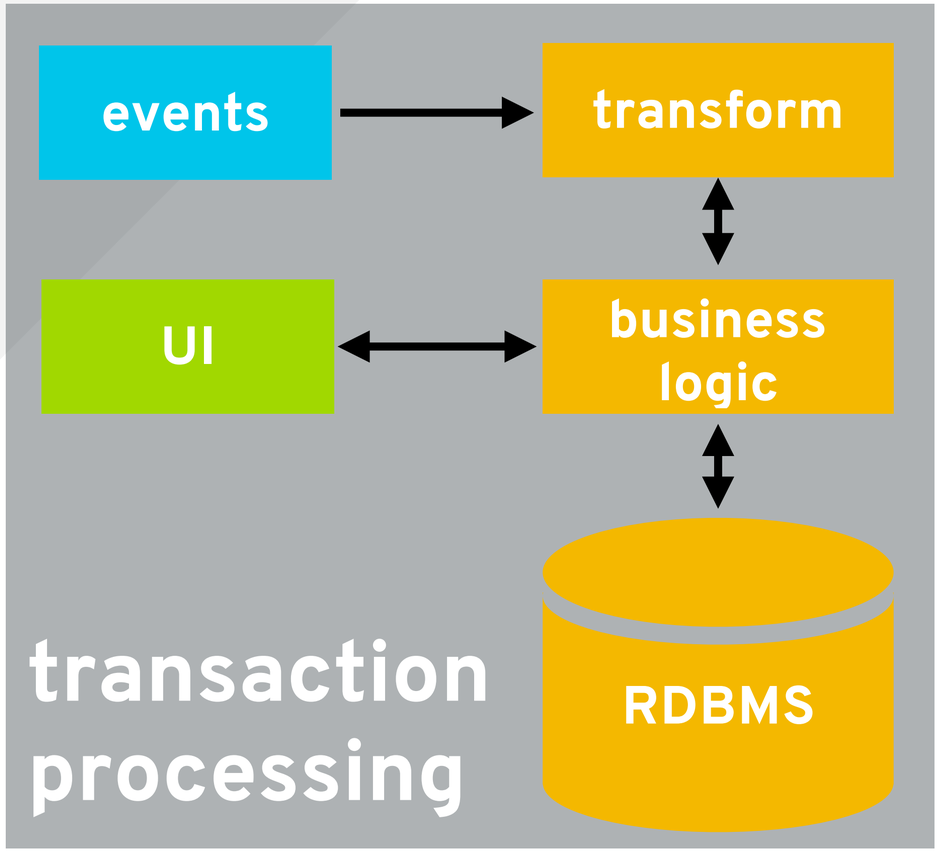
CONVENTIONAL DATA WAREHOUSE



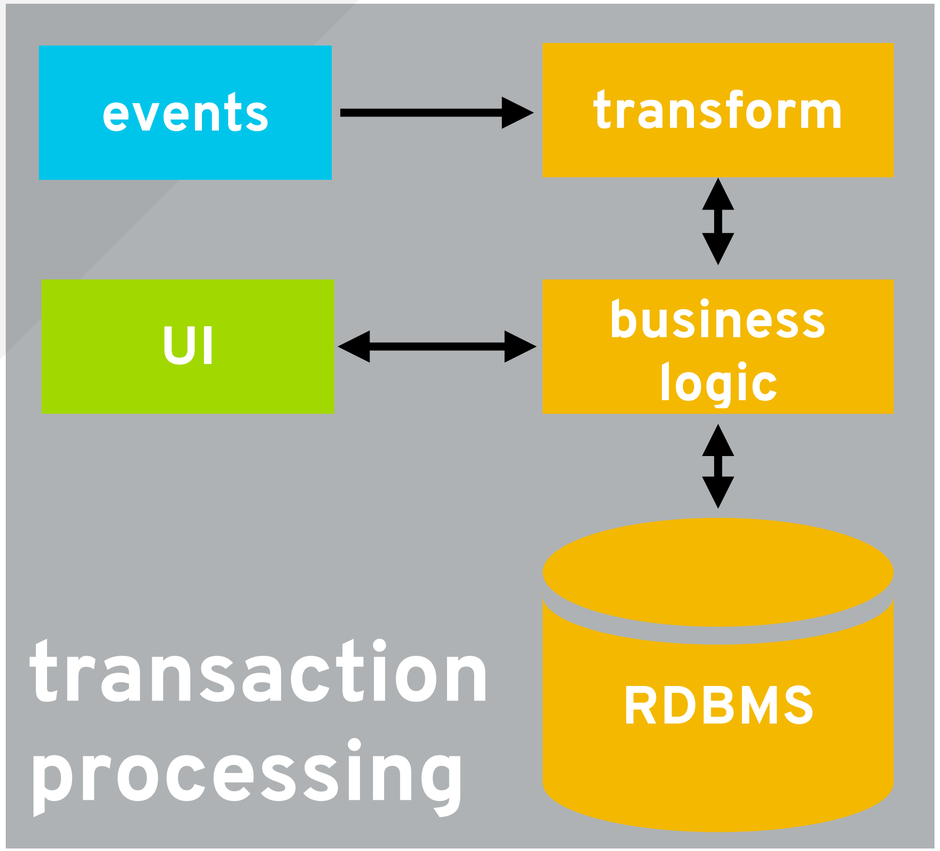
CONVENTIONAL DATA WAREHOUSE



CONVENTIONAL DATA WAREHOUSE

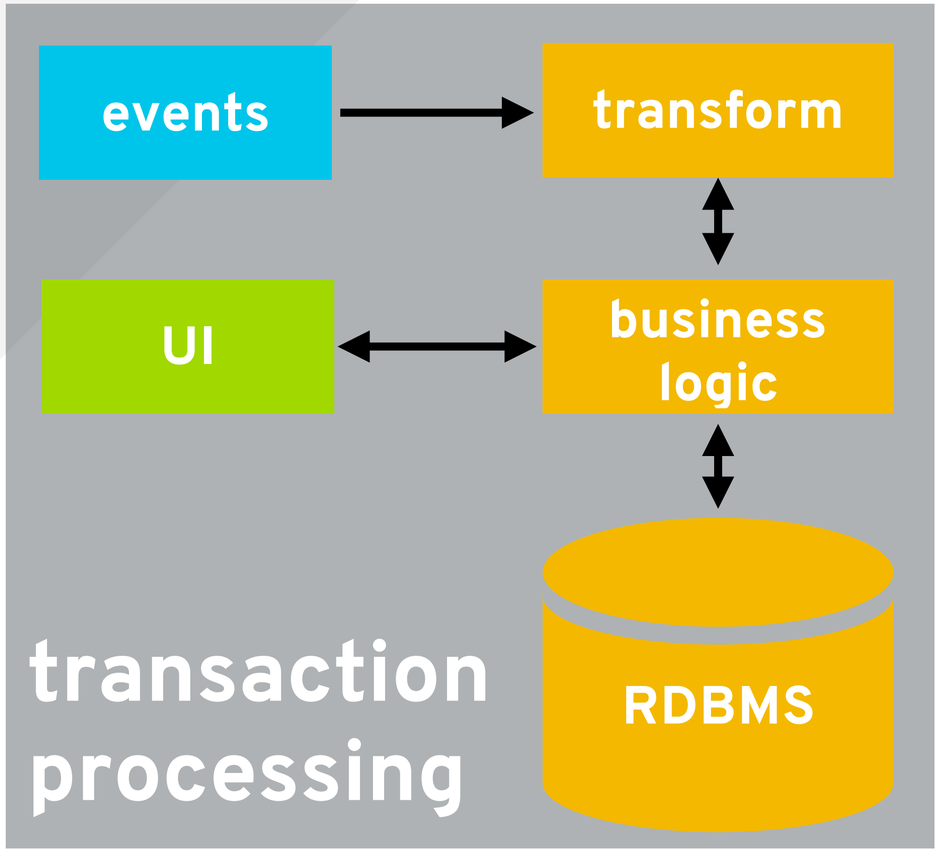


CONVENTIONAL DATA WAREHOUSE



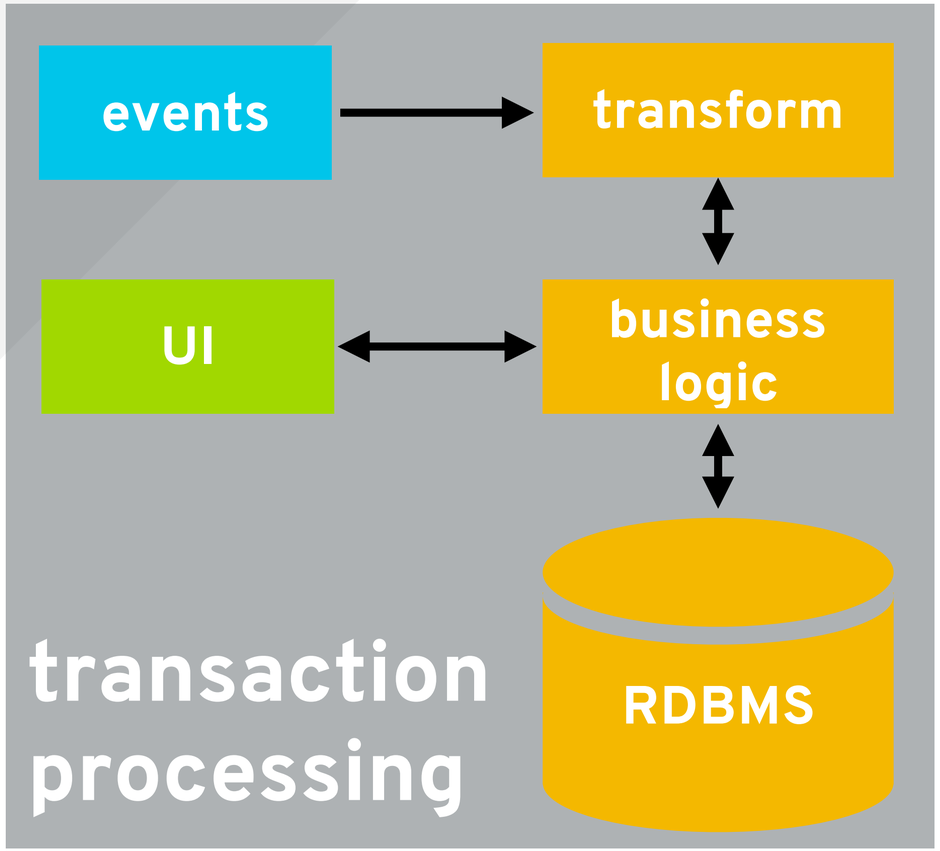
- stateful

CONVENTIONAL DATA WAREHOUSE



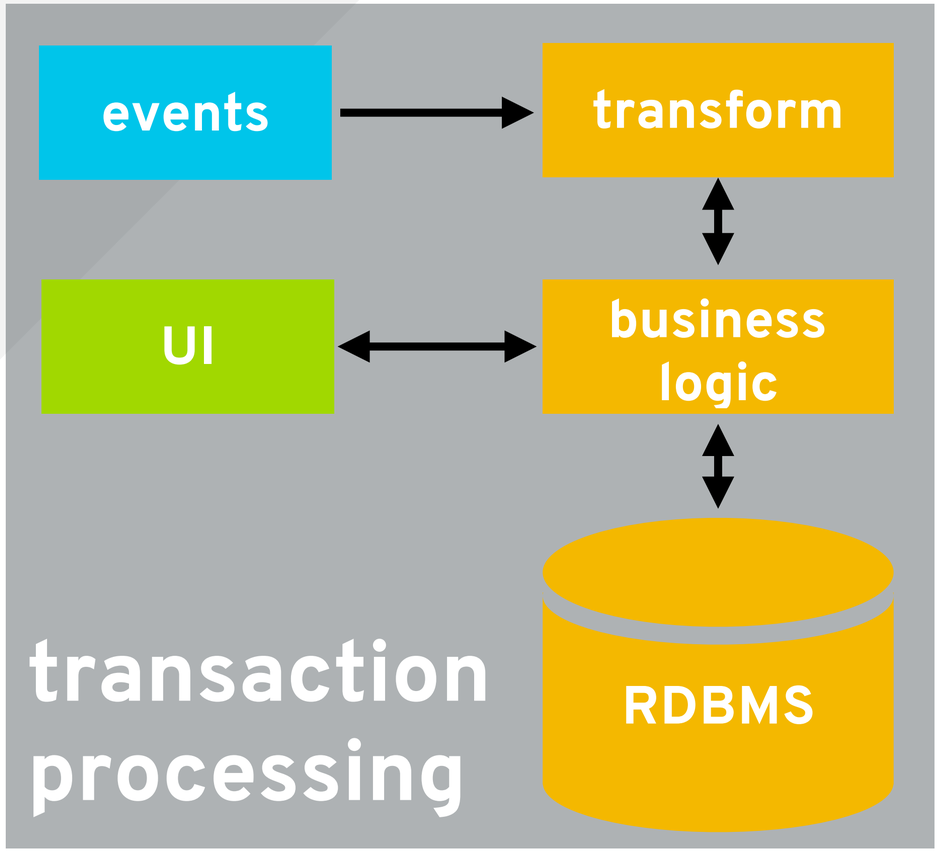
- stateful
- difficult to scale out

CONVENTIONAL DATA WAREHOUSE



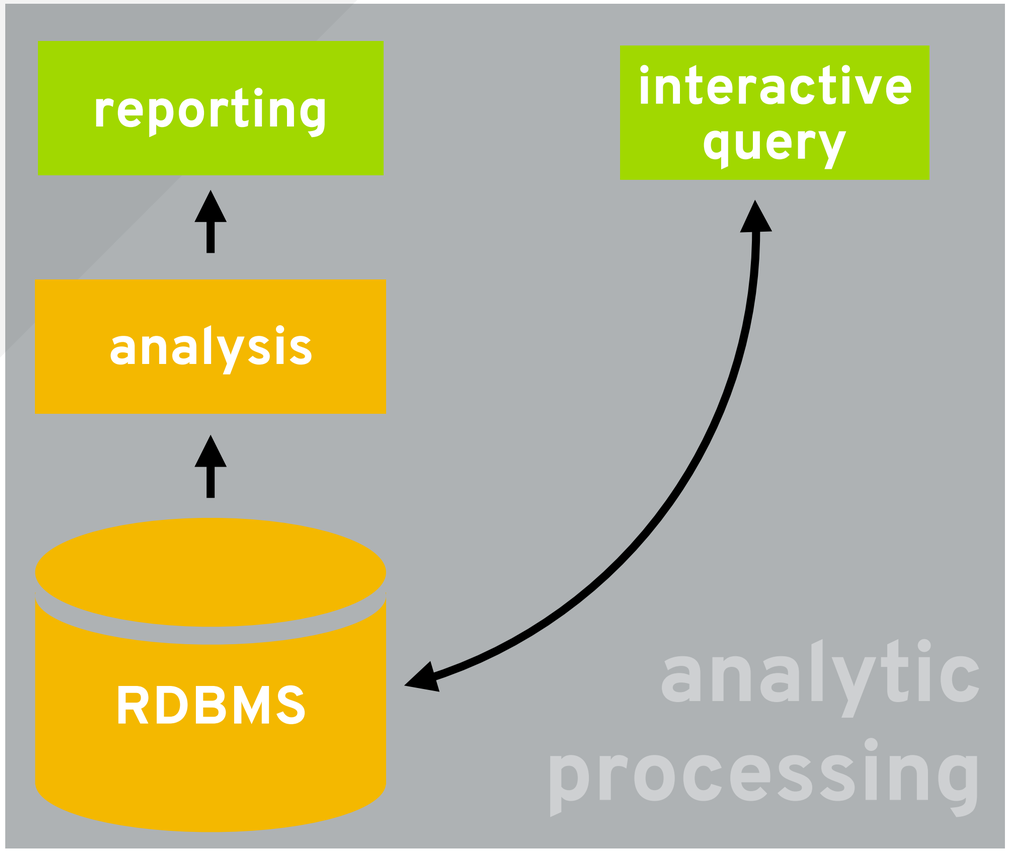
- stateful
- difficult to scale out
- difficult to manage

CONVENTIONAL DATA WAREHOUSE

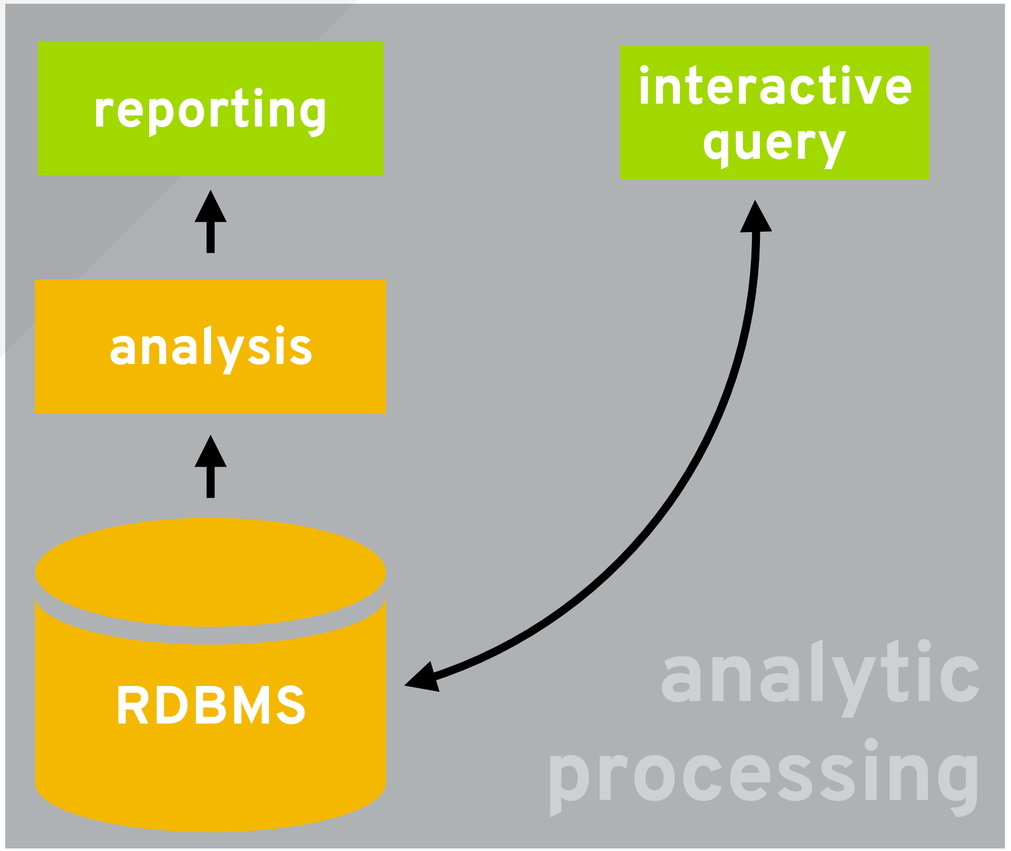


- stateful
- difficult to scale out
- difficult to manage
- no raw data in "source of truth"

CONVENTIONAL DATA WAREHOUSE

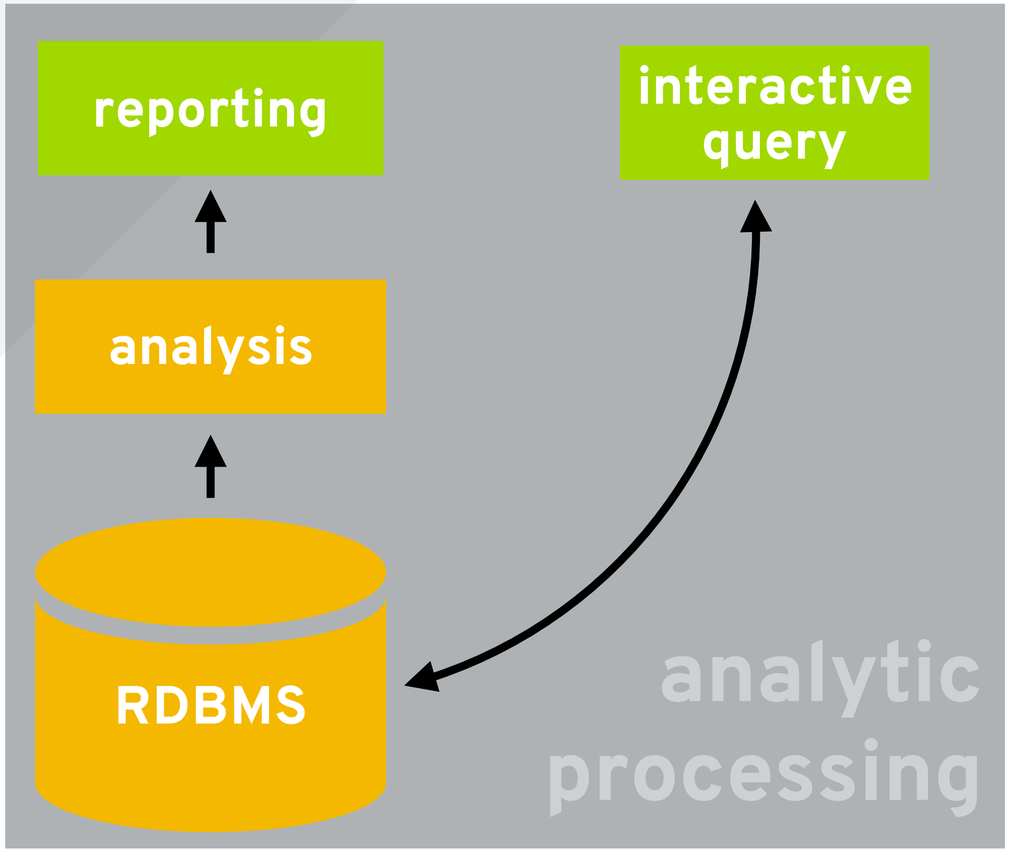


CONVENTIONAL DATA WAREHOUSE



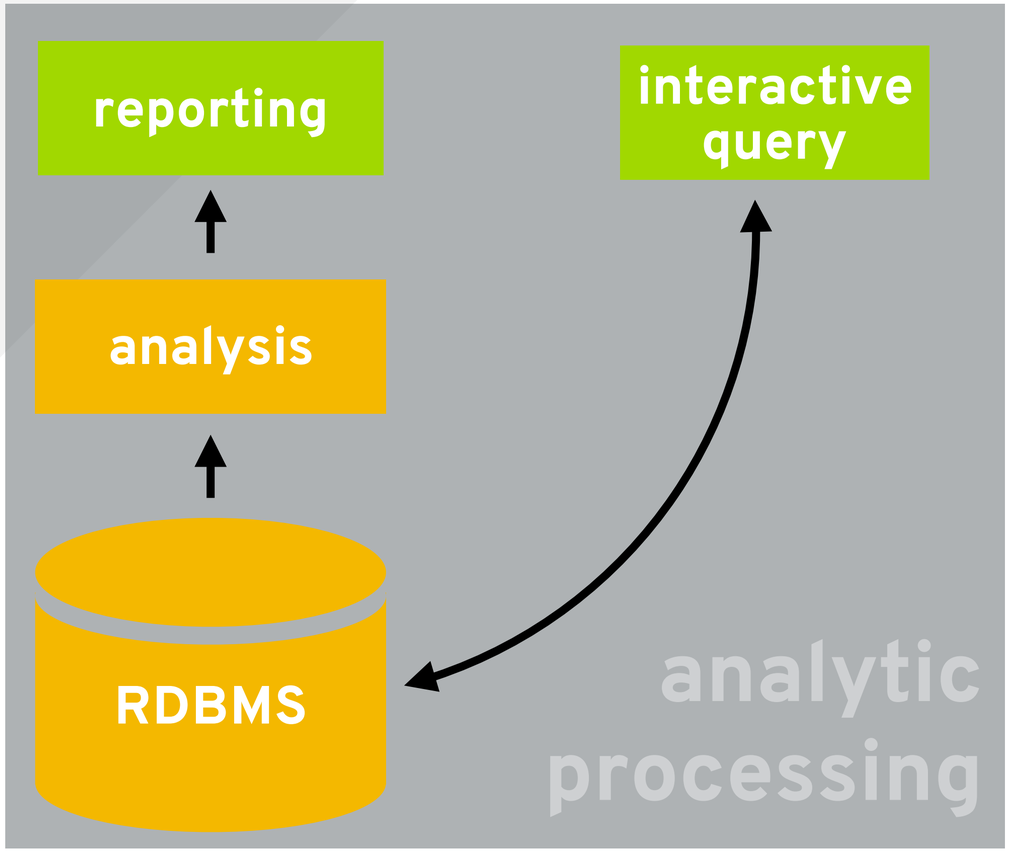
- results lag behind

CONVENTIONAL DATA WAREHOUSE



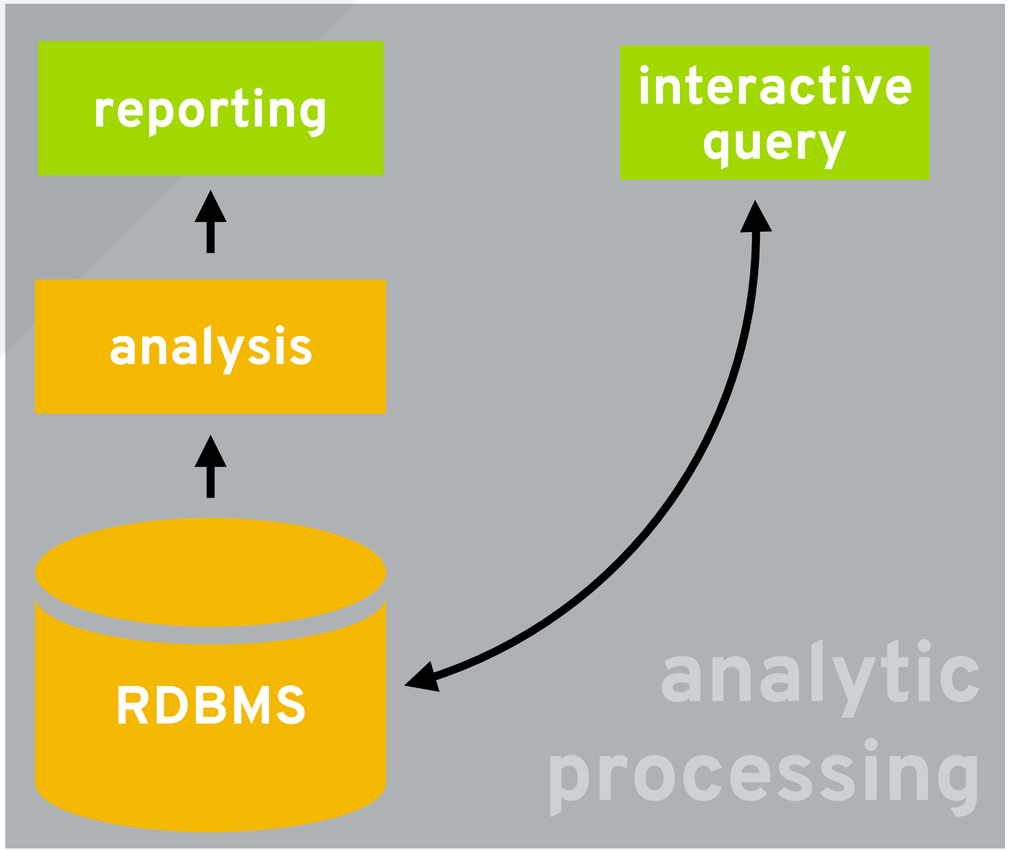
- results lag behind
- difficult to scale out

CONVENTIONAL DATA WAREHOUSE



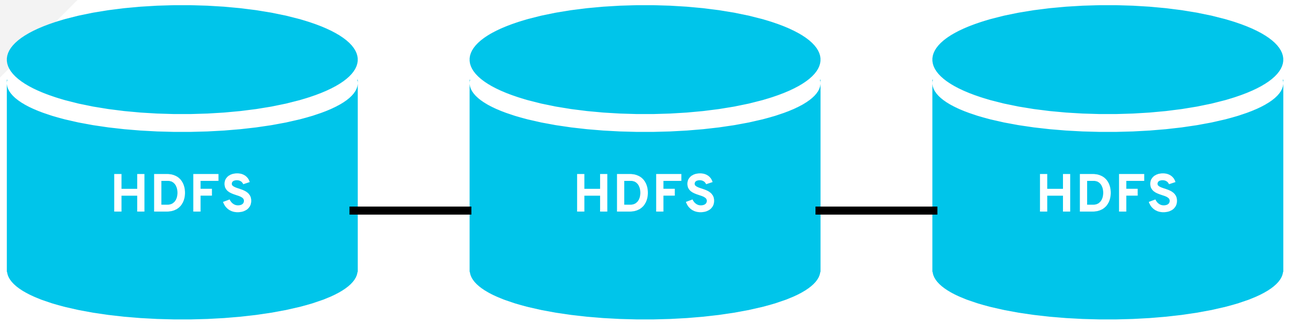
- results lag behind
- difficult to scale out
- limited to RDBMS capabilities/interface

CONVENTIONAL DATA WAREHOUSE

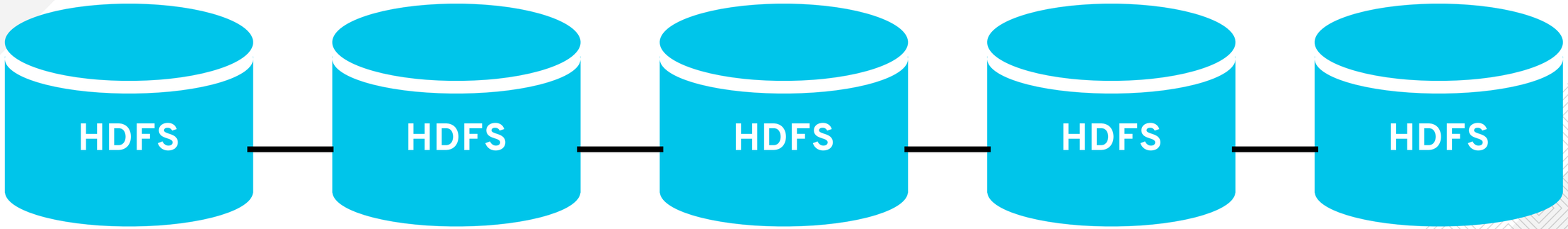


- results lag behind
- difficult to scale out
- limited to RDBMS capabilities/interface
- difficult to implement model-training code

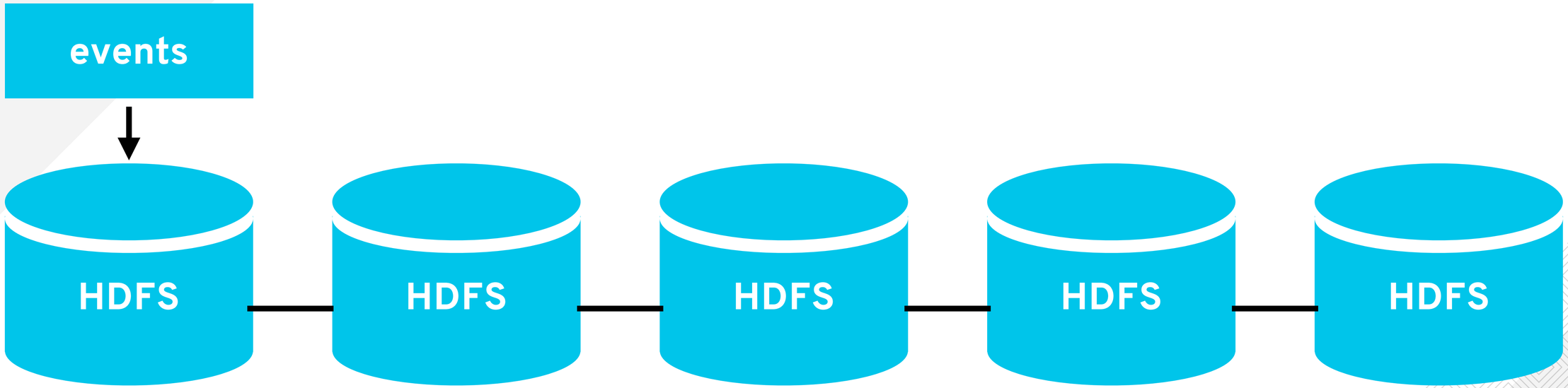
HADOOP-STYLE “DATA LAKE”



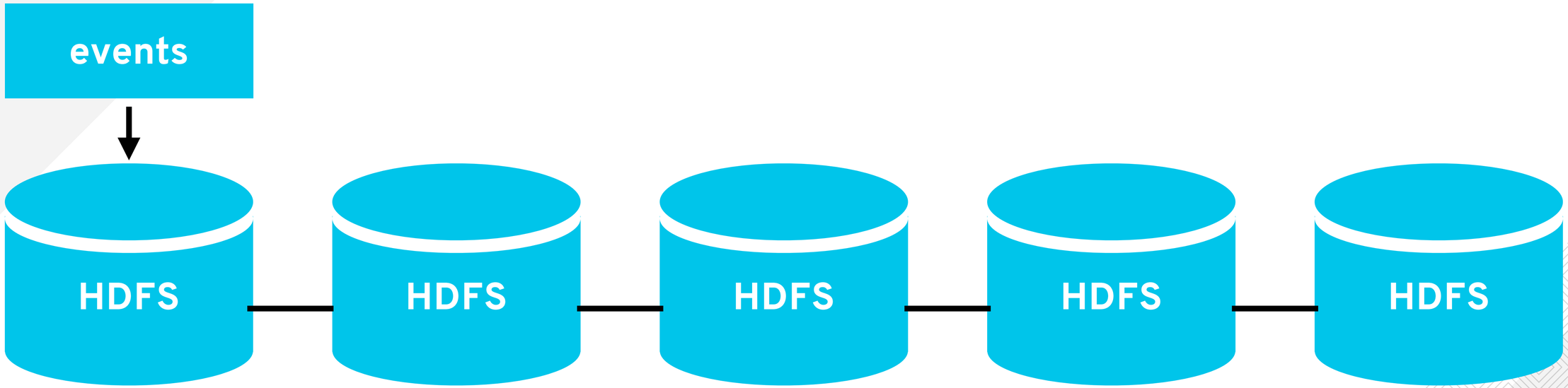
HADOOP-STYLE “DATA LAKE”



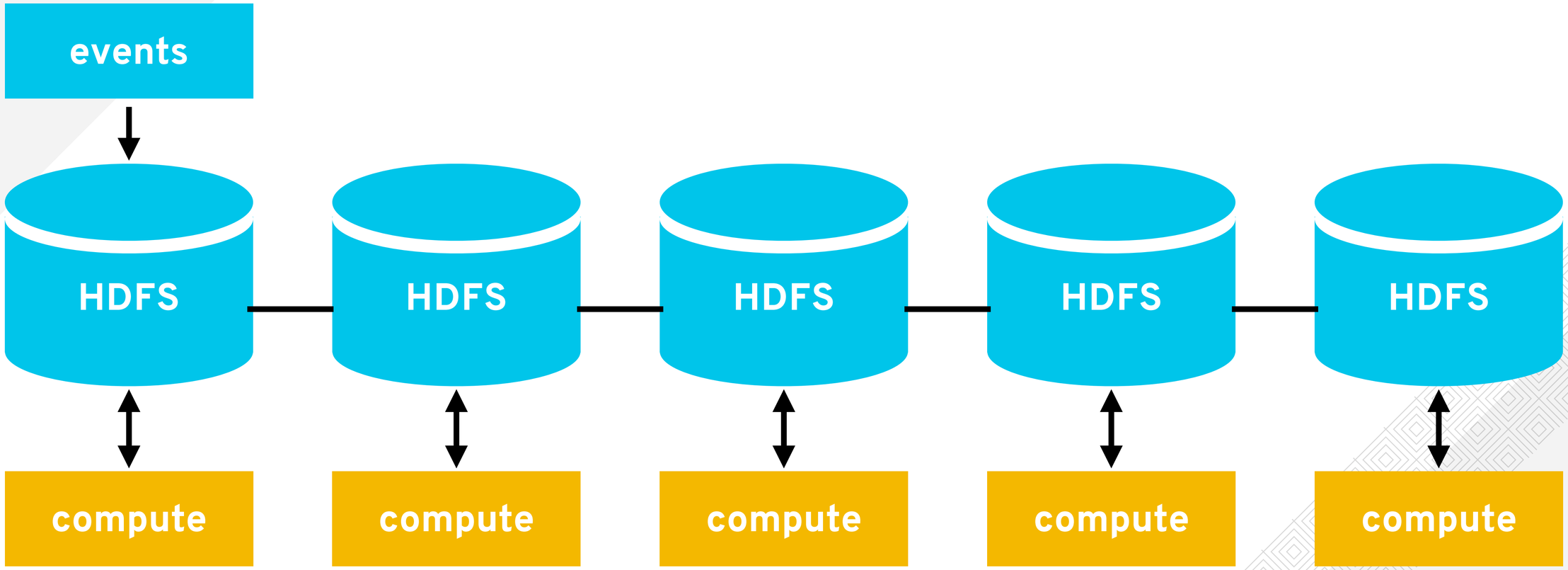
HADOOP-STYLE “DATA LAKE”



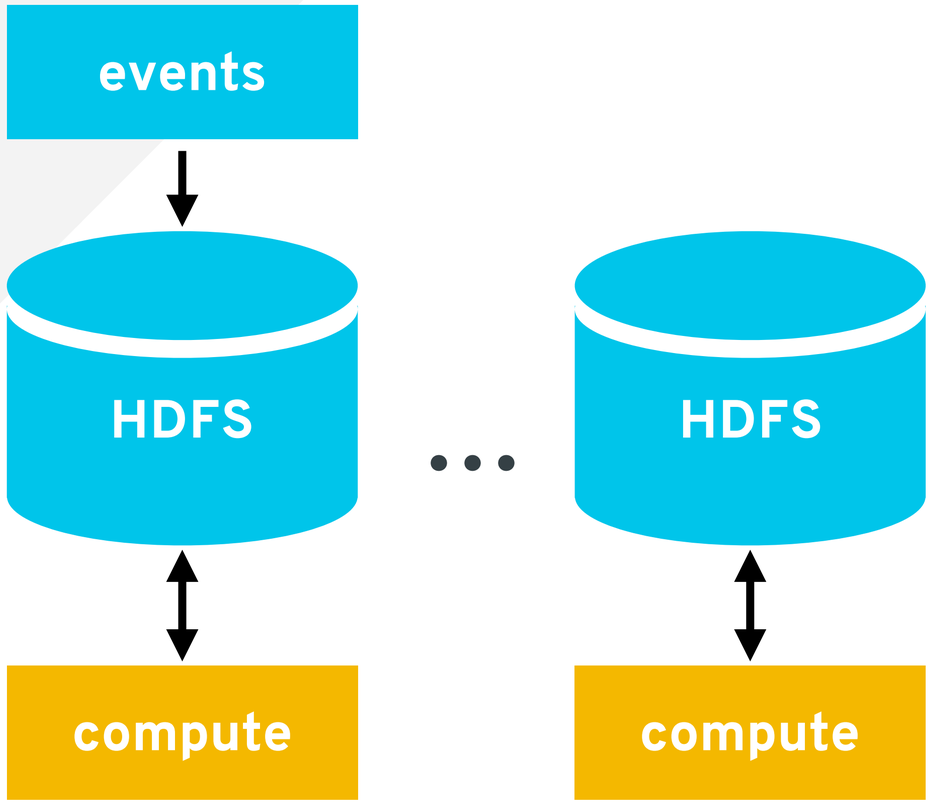
HADOOP-STYLE “DATA LAKE”



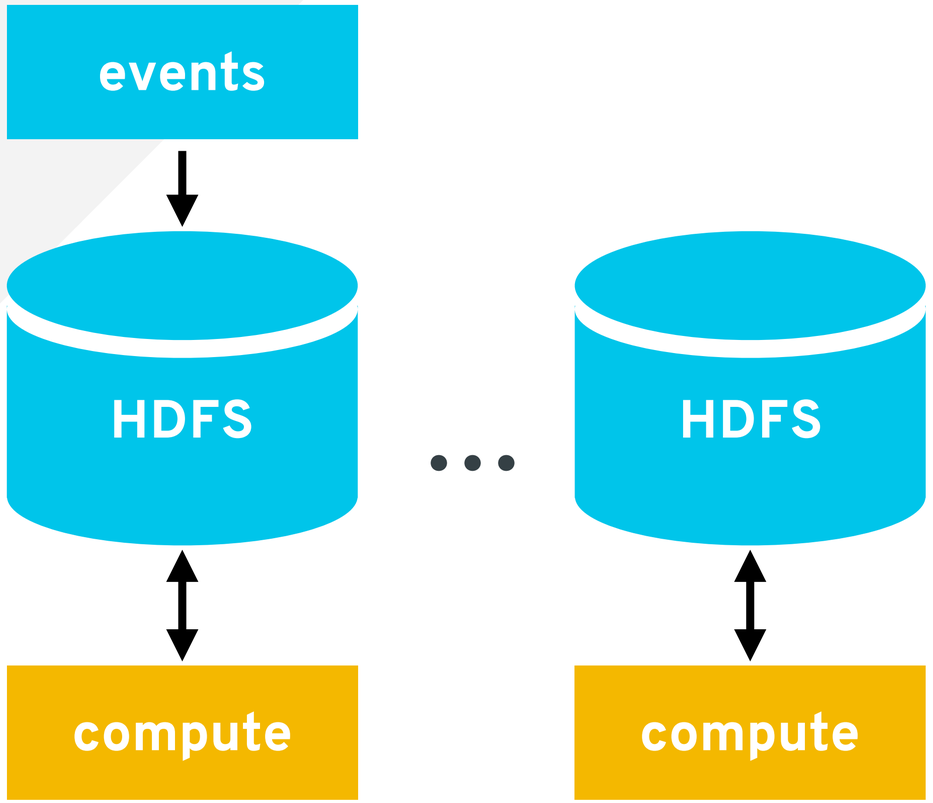
HADOOP-STYLE “DATA LAKE”



HADOOP-STYLE “DATA LAKE”

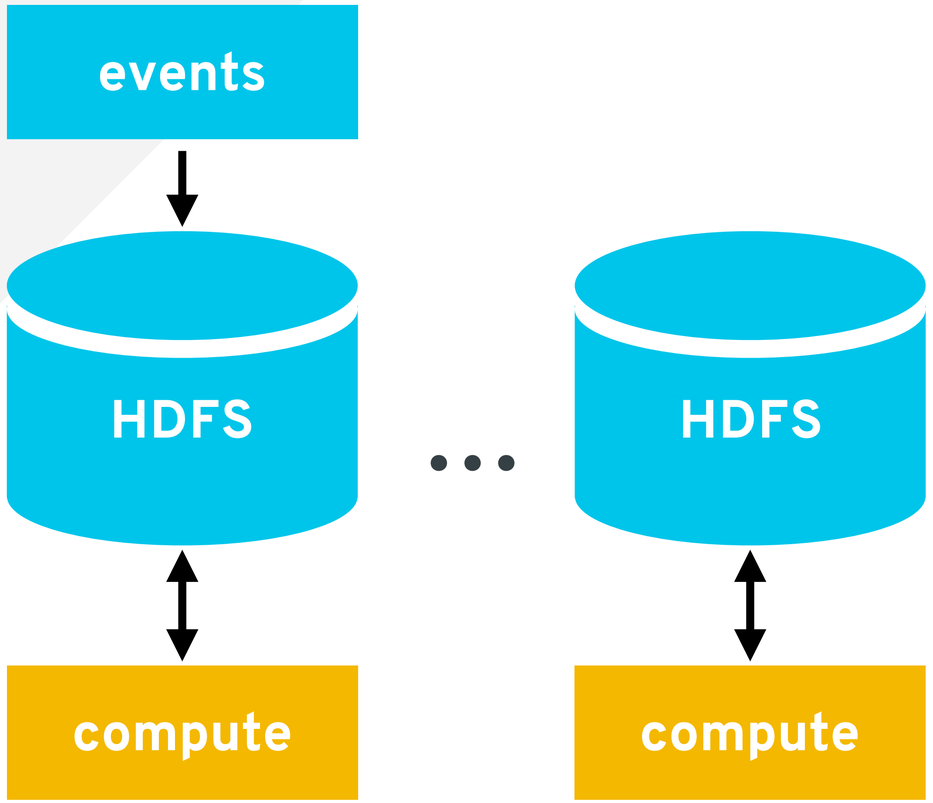


HADOOP-STYLE “DATA LAKE”



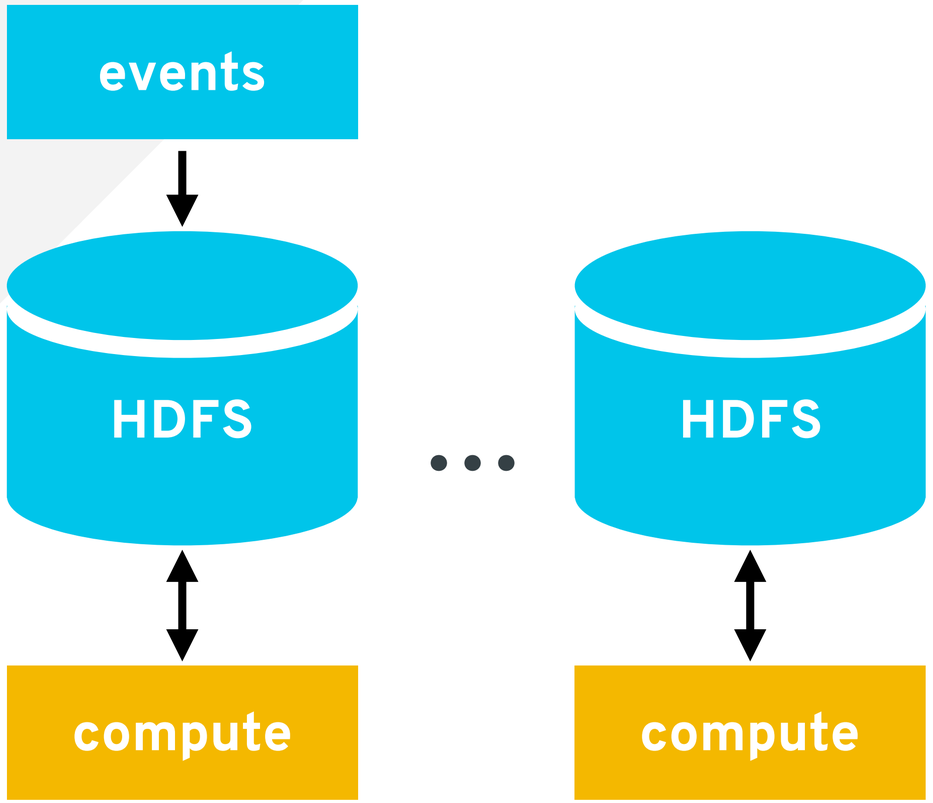
- tightly-coupled compute and storage

HADOOP-STYLE “DATA LAKE”



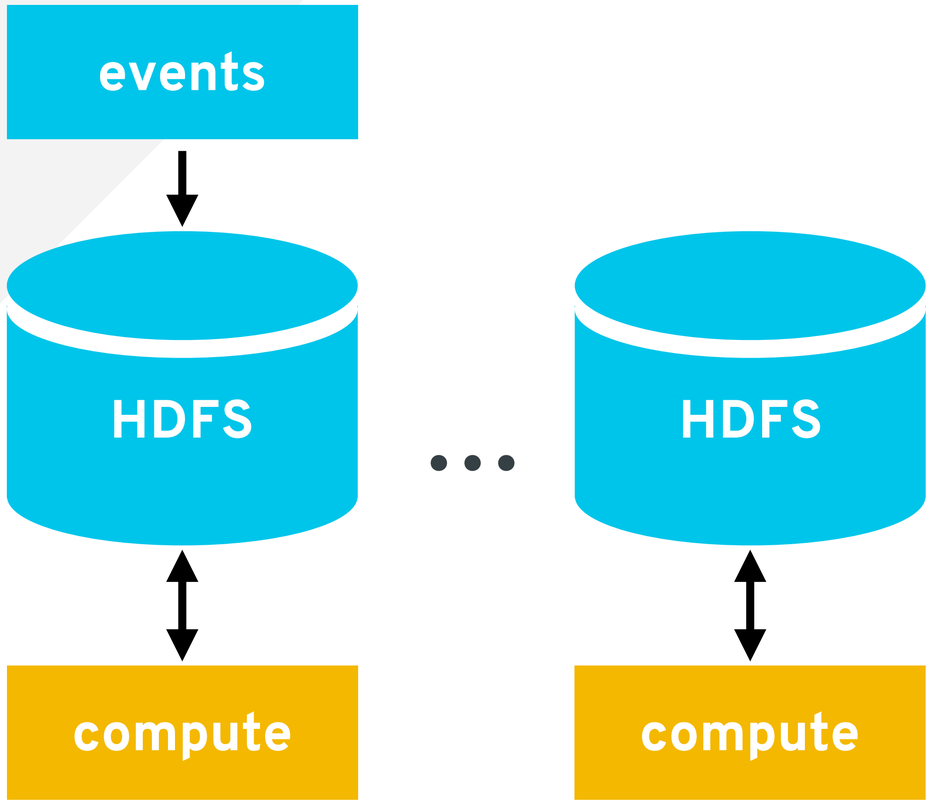
- tightly-coupled compute and storage
- dedicated resources

HADOOP-STYLE “DATA LAKE”



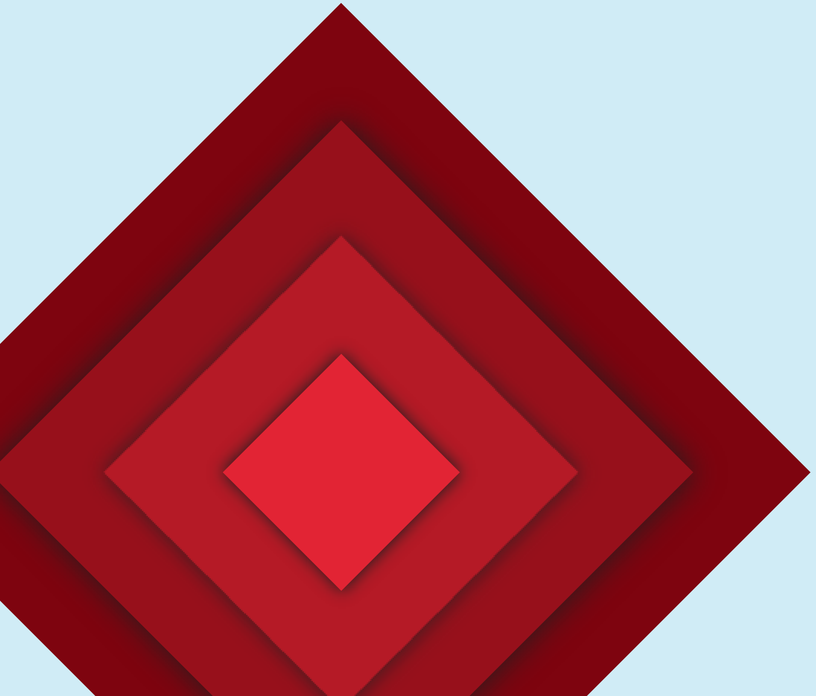
- tightly-coupled compute and storage
- dedicated resources
- low-level API

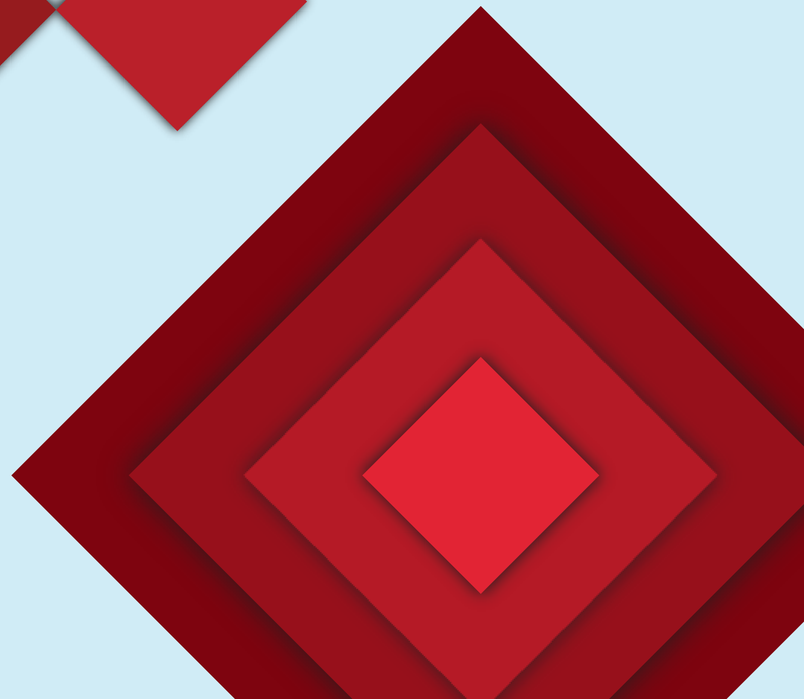
HADOOP-STYLE “DATA LAKE”



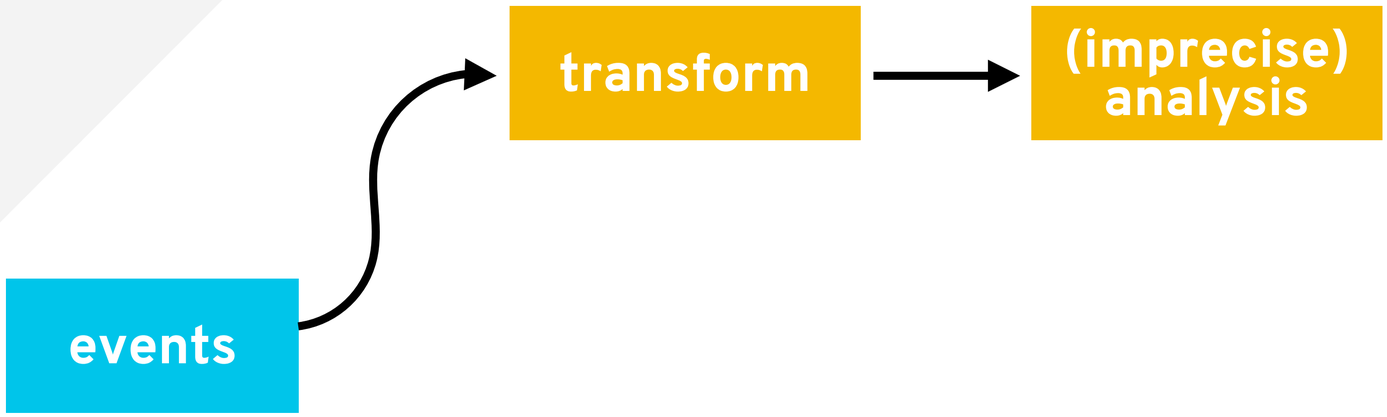
- tightly-coupled compute and storage
- dedicated resources
- low-level API
- high-latency results

MICROSERVICE-FRIENDLY ARCHITECTURES

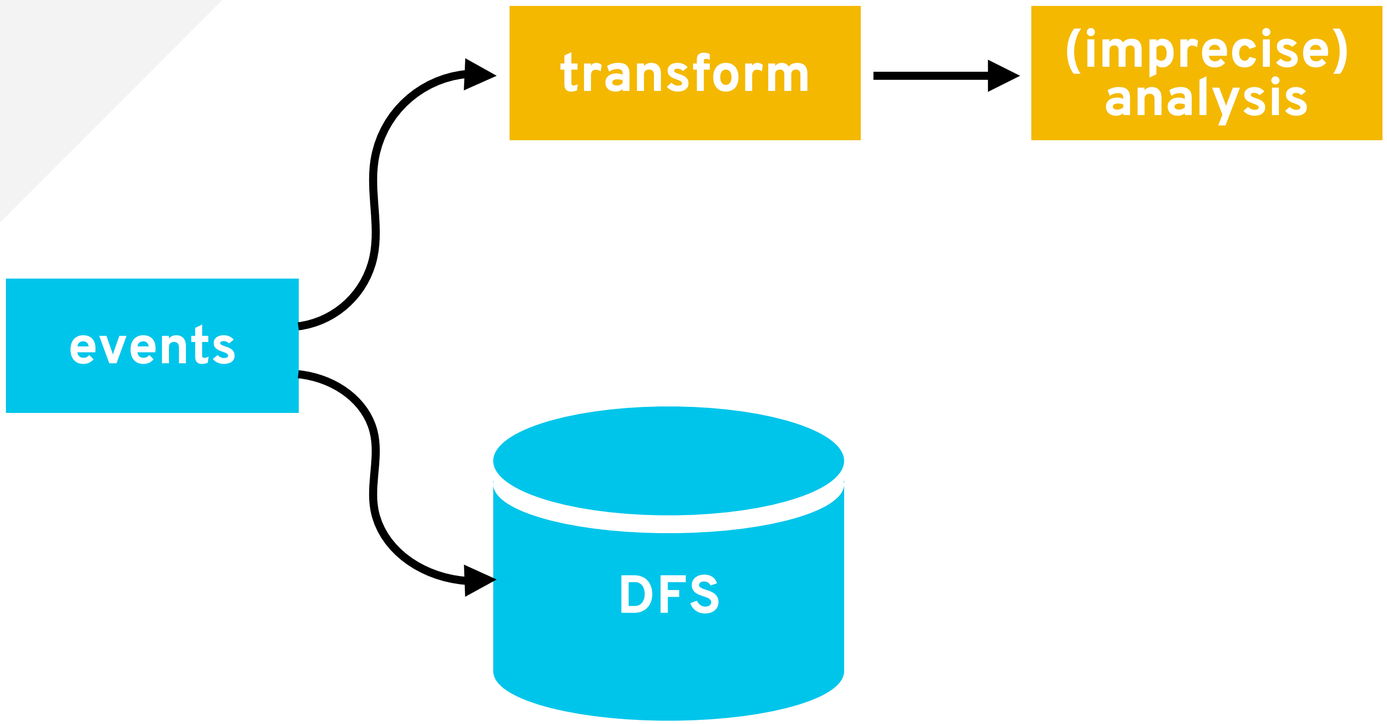




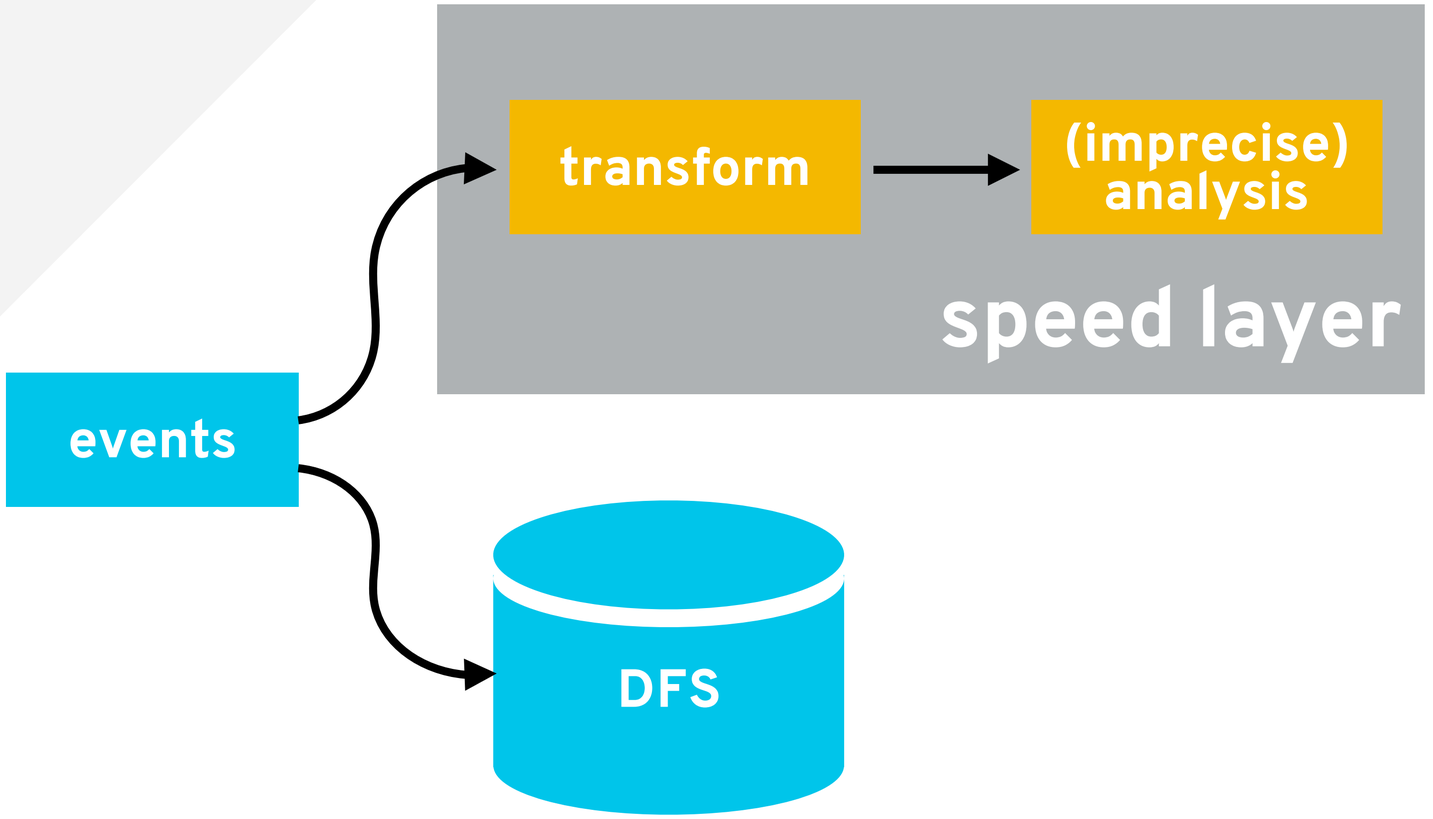
THE LAMBDA ARCHITECTURE



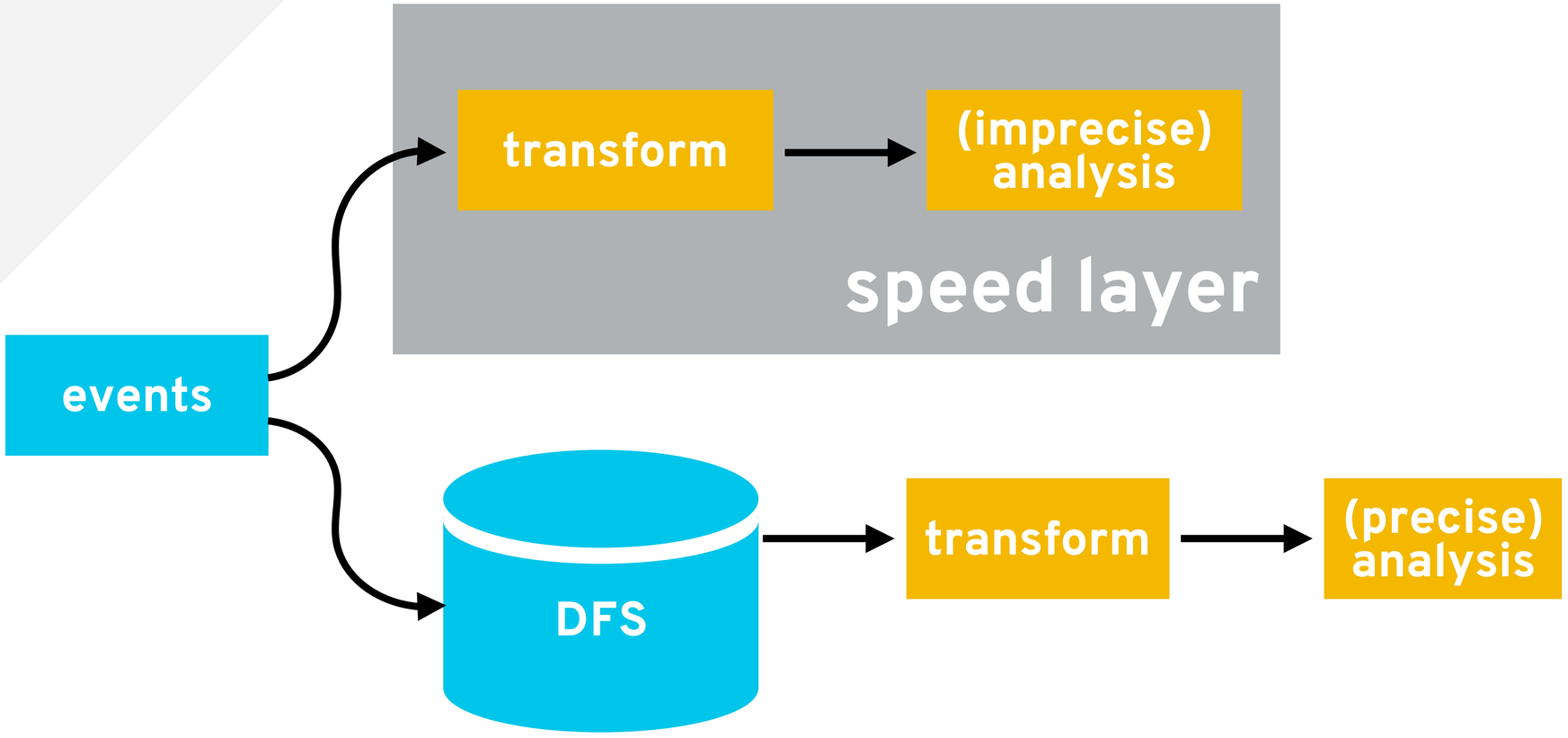
THE LAMBDA ARCHITECTURE



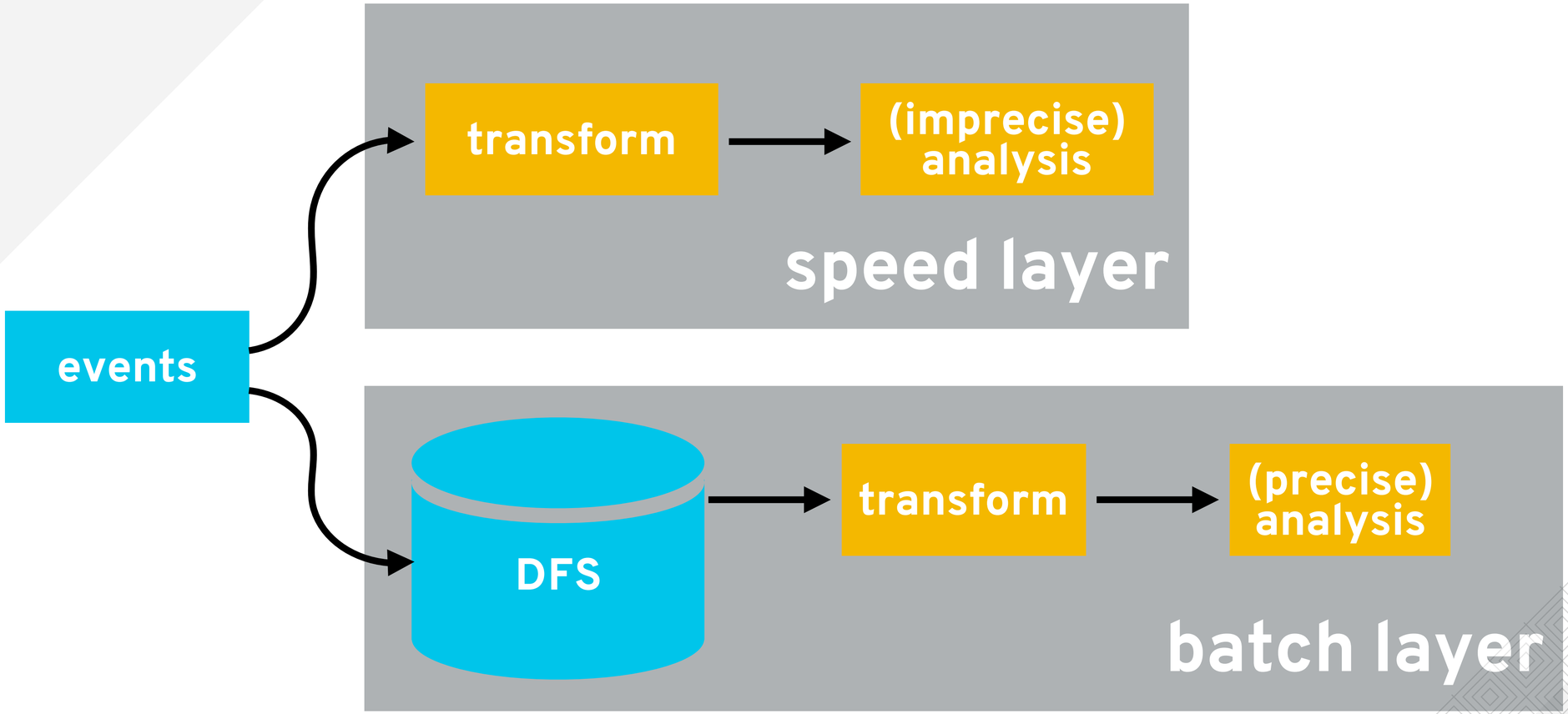
THE LAMBDA ARCHITECTURE



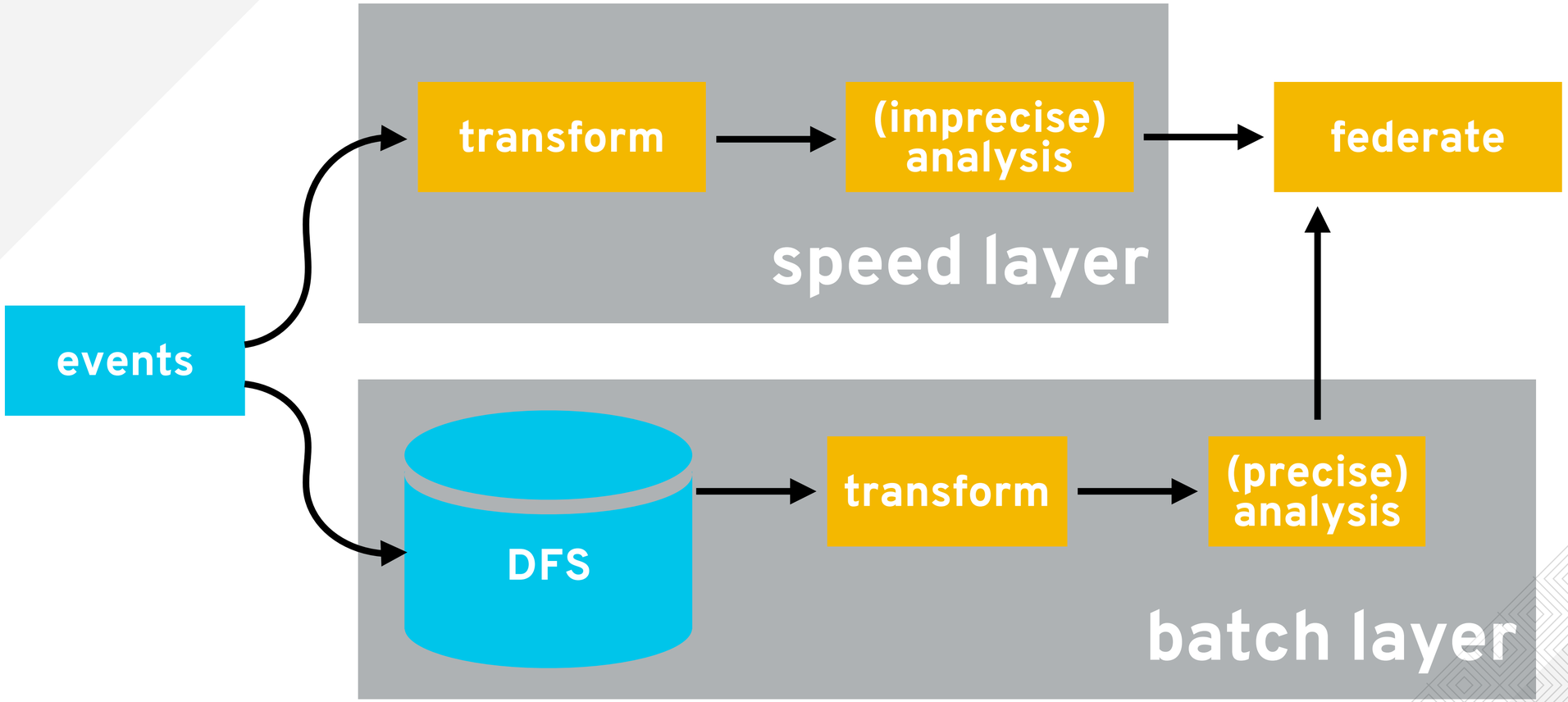
THE LAMBDA ARCHITECTURE



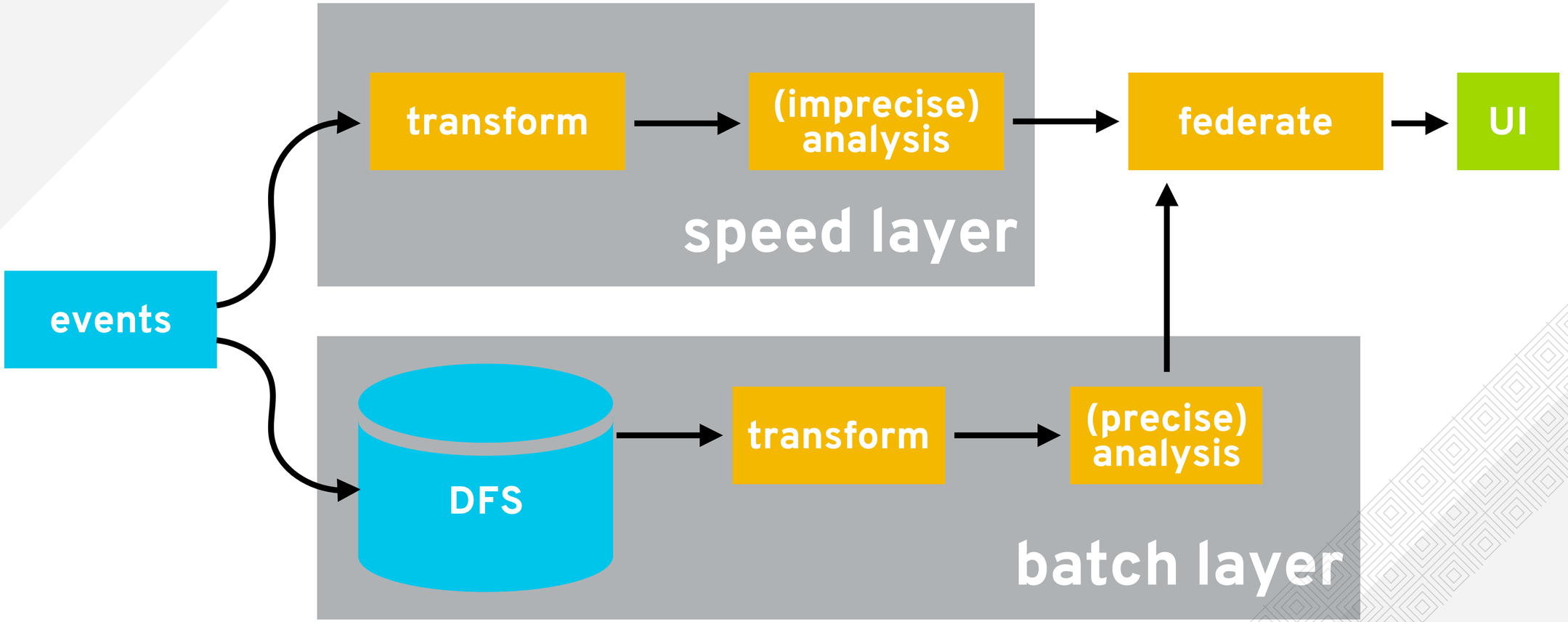
THE LAMBDA ARCHITECTURE



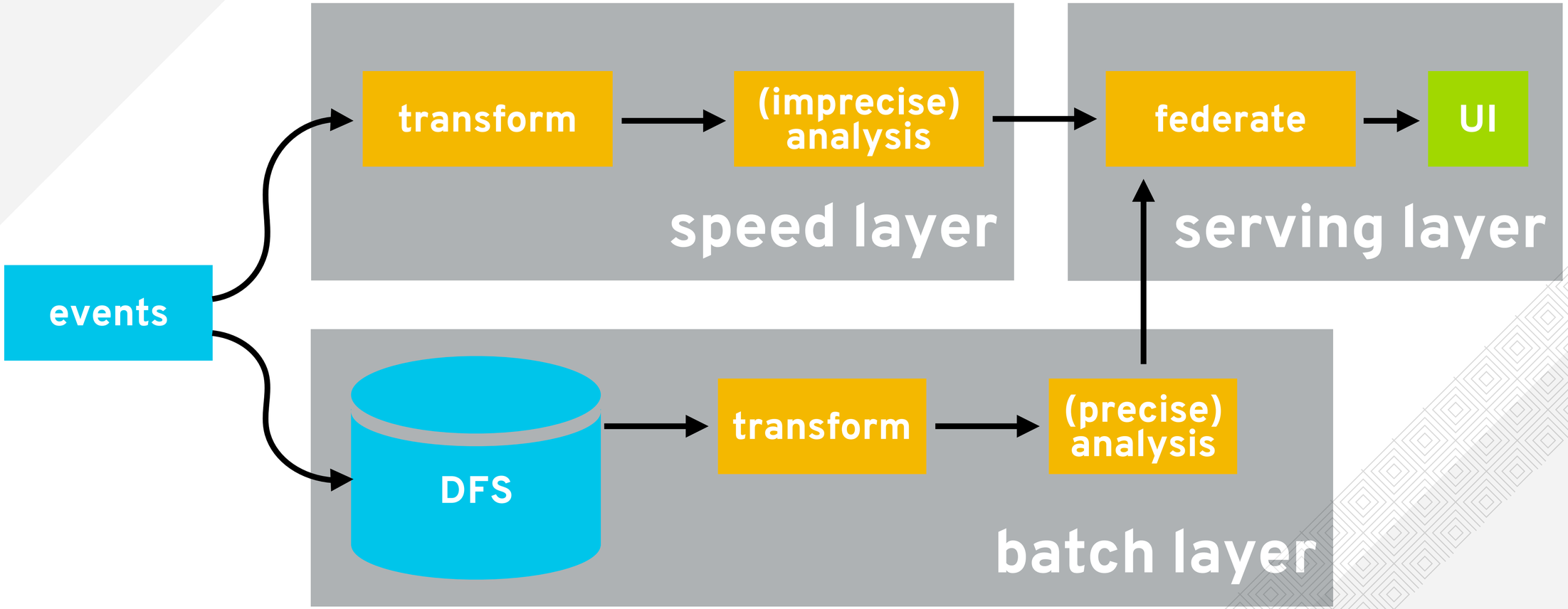
THE LAMBDA ARCHITECTURE



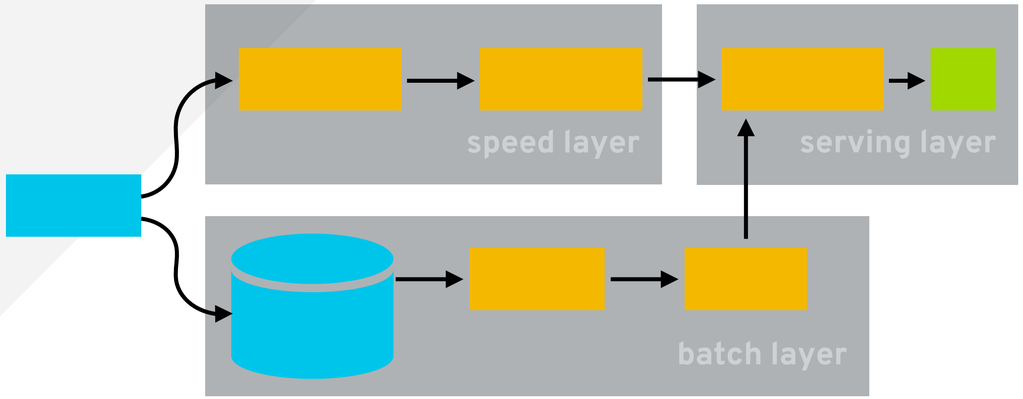
THE LAMBDA ARCHITECTURE



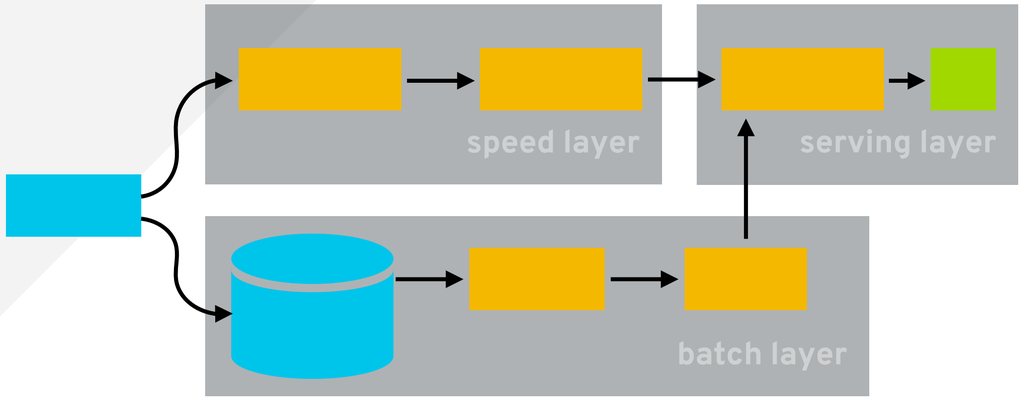
THE LAMBDA ARCHITECTURE



THE LAMBDA ARCHITECTURE

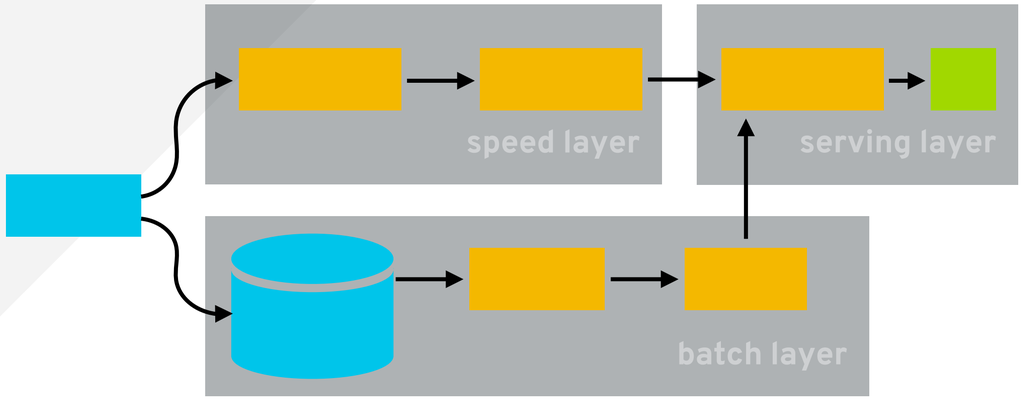


THE LAMBDA ARCHITECTURE



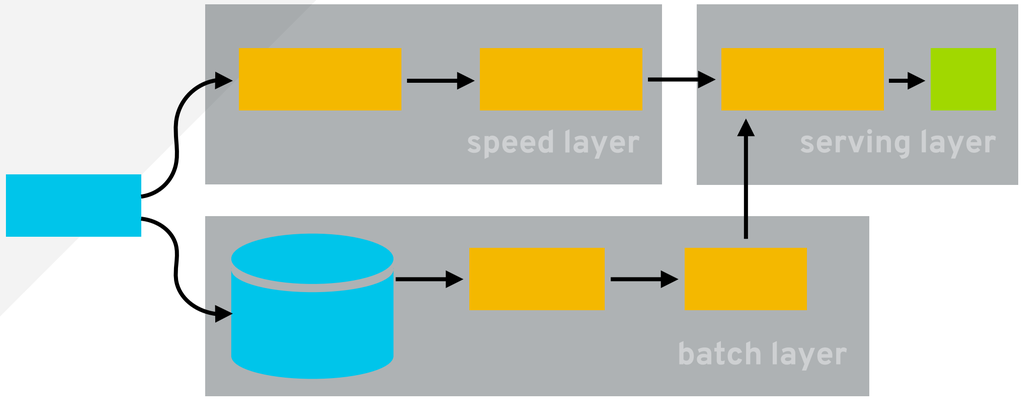
- suitable for micro-service architectures

THE LAMBDA ARCHITECTURE



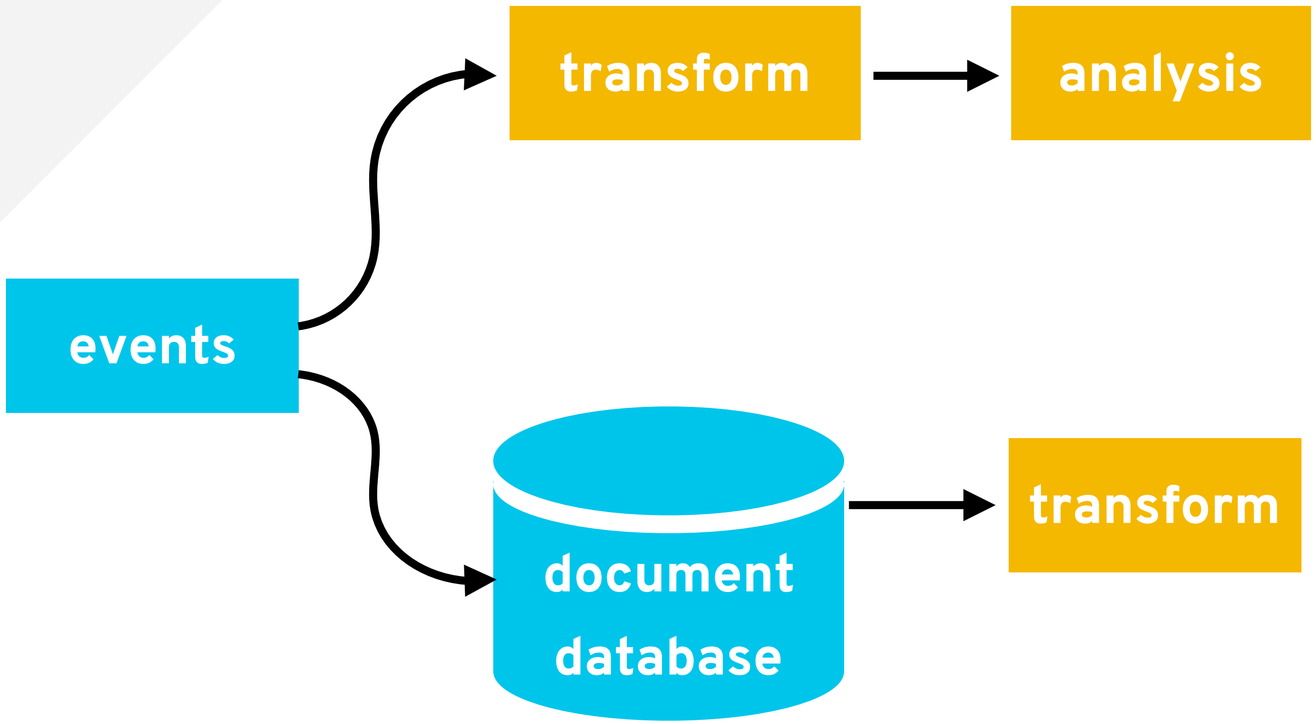
- suitable for micro-service architectures
- raw data remain available

THE LAMBDA ARCHITECTURE

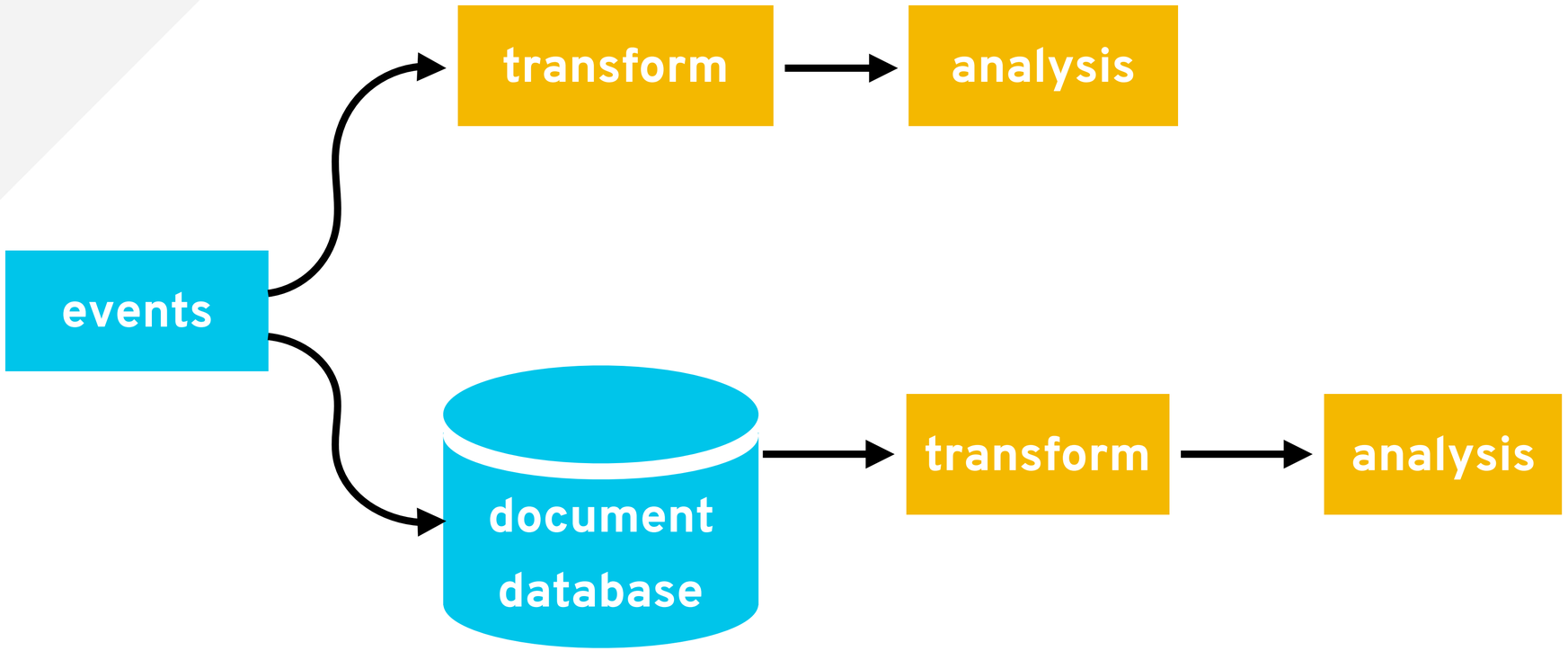


- suitable for micro-service architectures
- raw data remain available
- requires analyses to be implemented twice

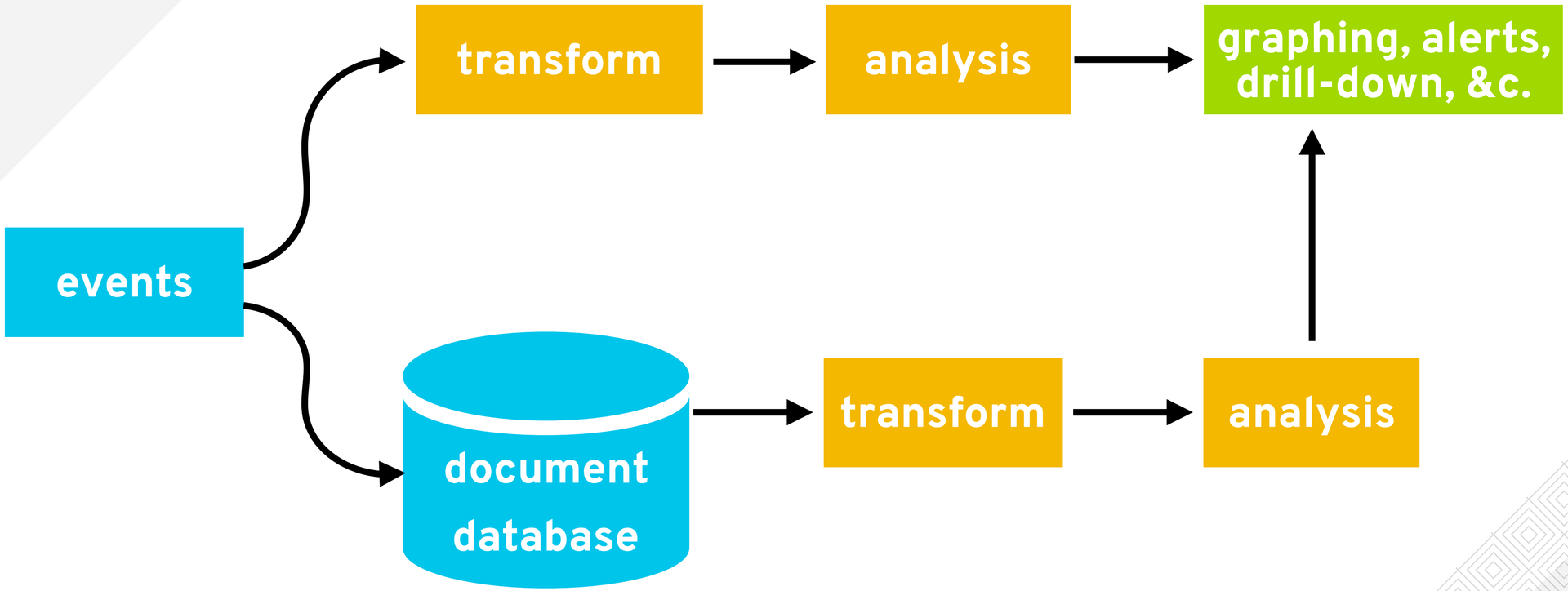
EXAMPLE: INFRASTRUCTURE LOG PROCESSING



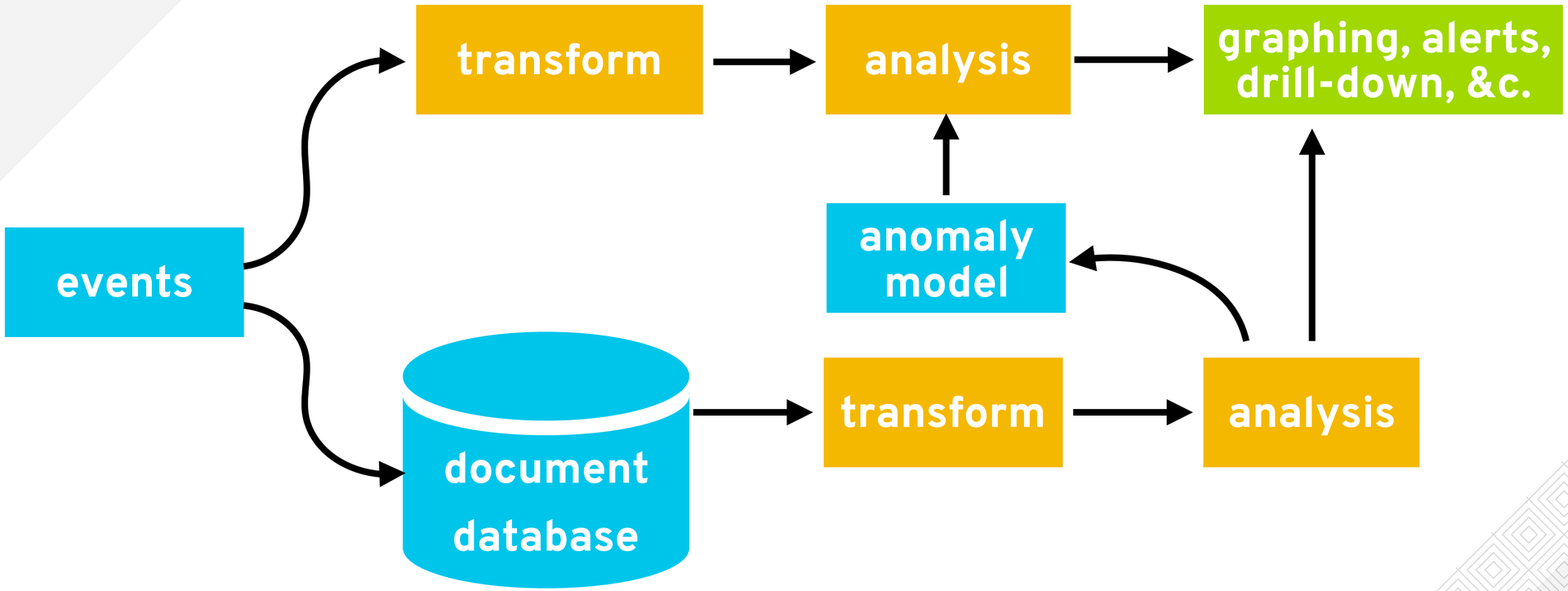
EXAMPLE: INFRASTRUCTURE LOG PROCESSING



EXAMPLE: INFRASTRUCTURE LOG PROCESSING



EXAMPLE: INFRASTRUCTURE LOG PROCESSING



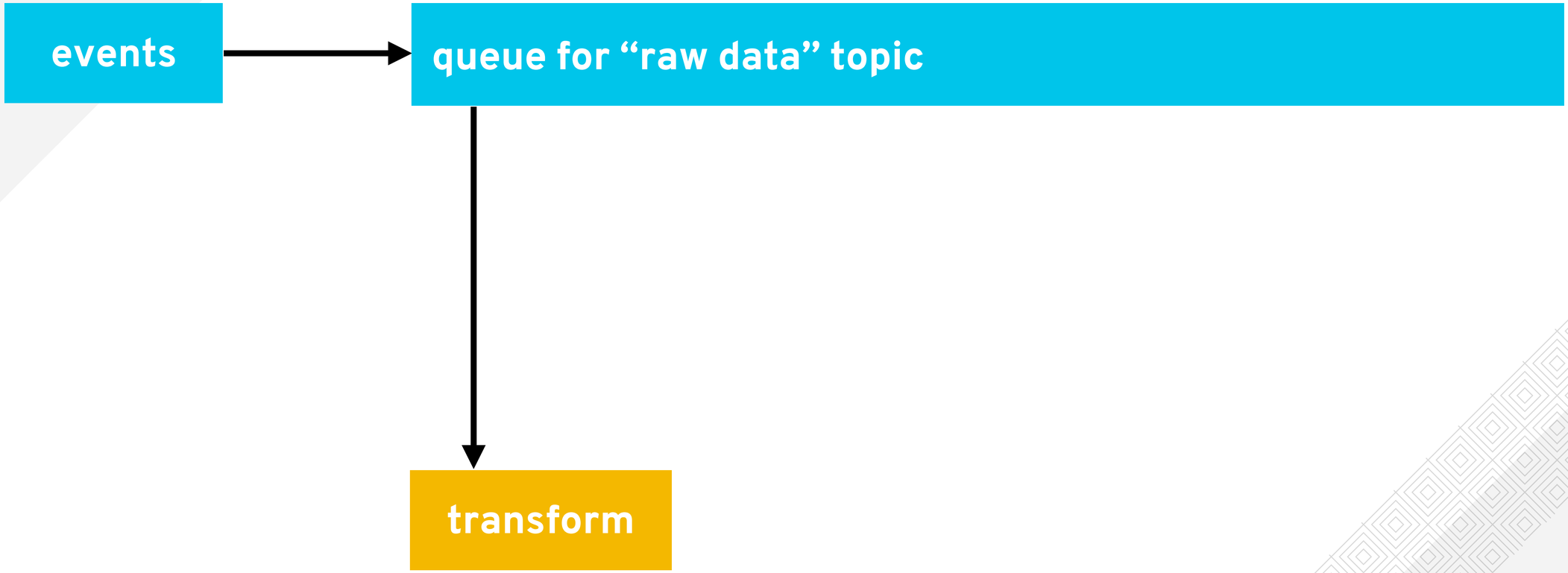
THE KAPPA ARCHITECTURE

events

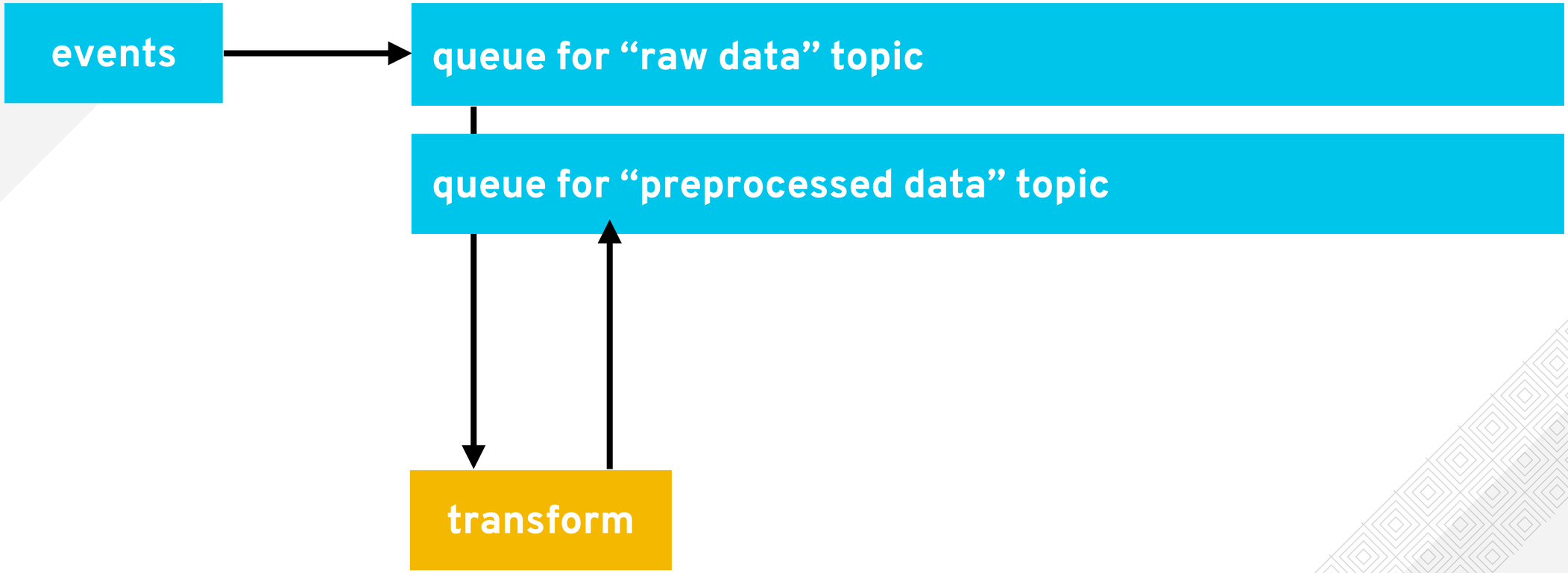
THE KAPPA ARCHITECTURE

THE KAPPA ARCHITECTURE

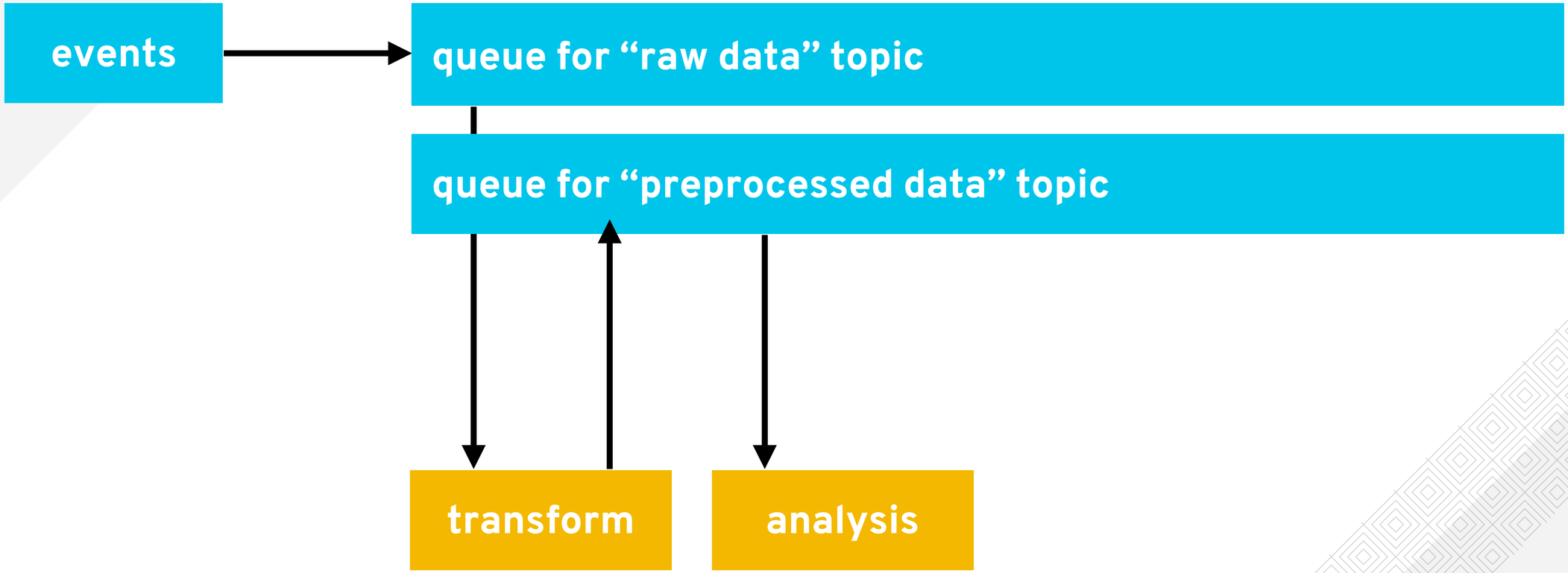
THE KAPPA ARCHITECTURE



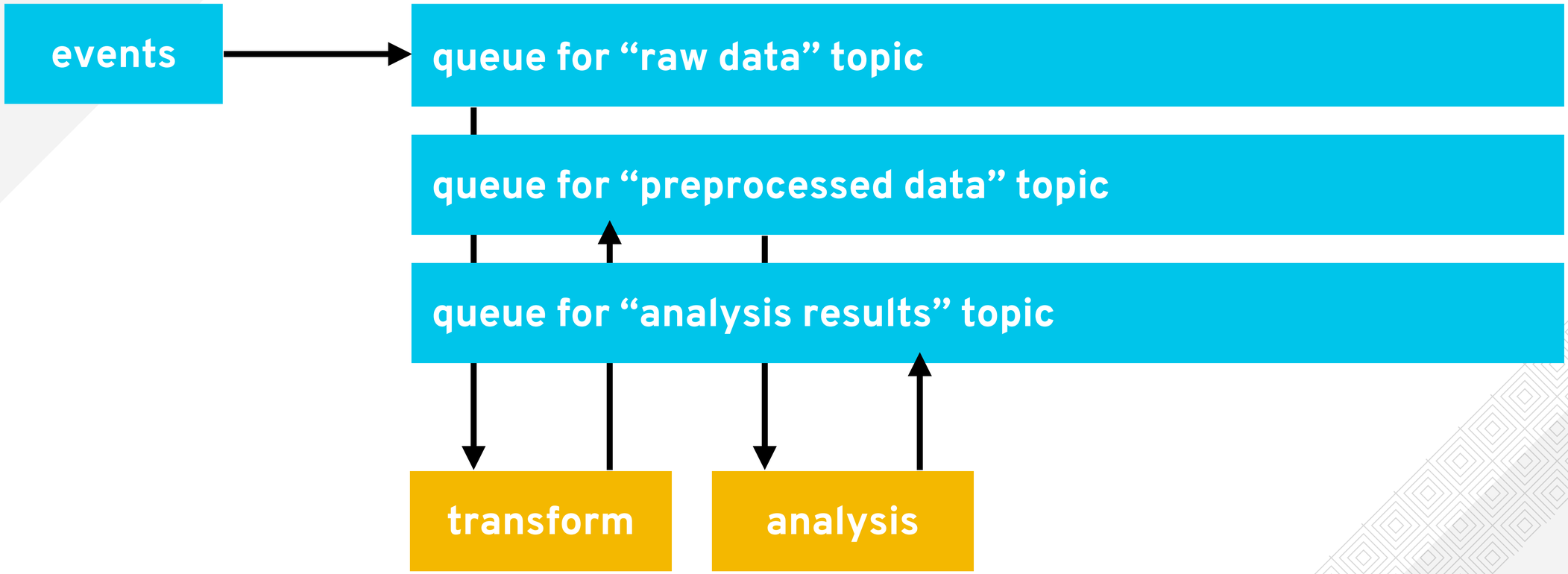
THE KAPPA ARCHITECTURE



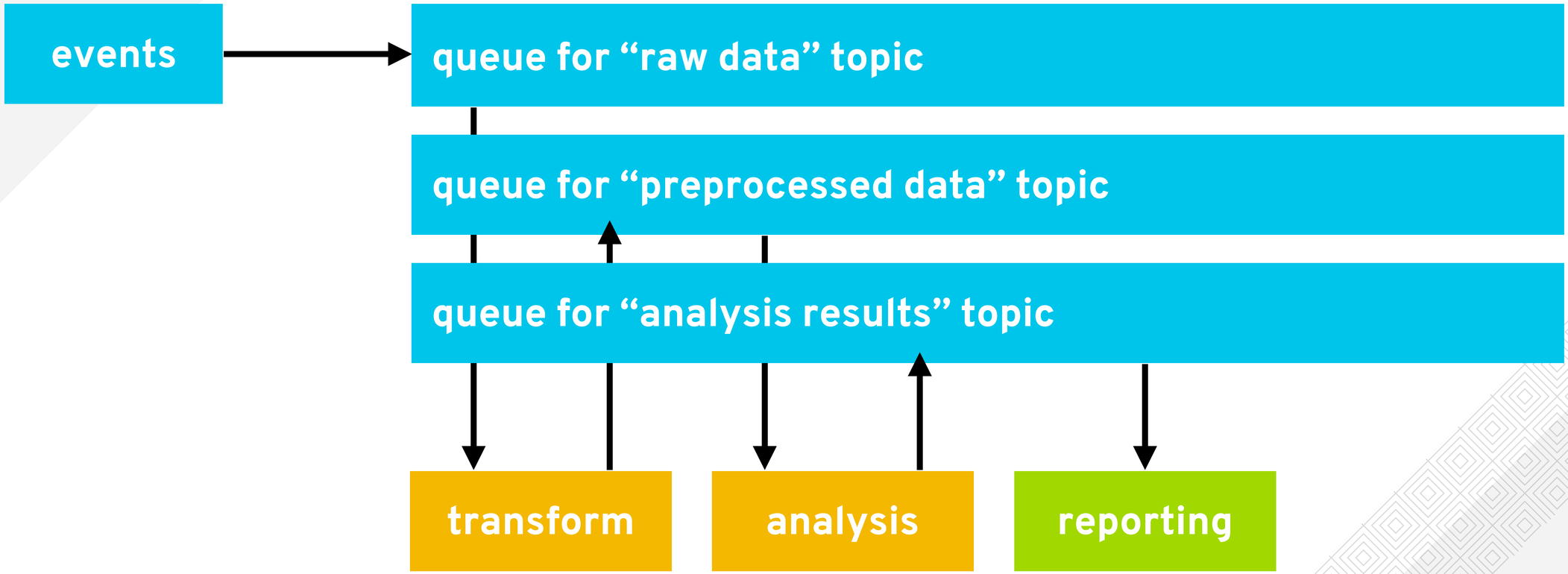
THE KAPPA ARCHITECTURE



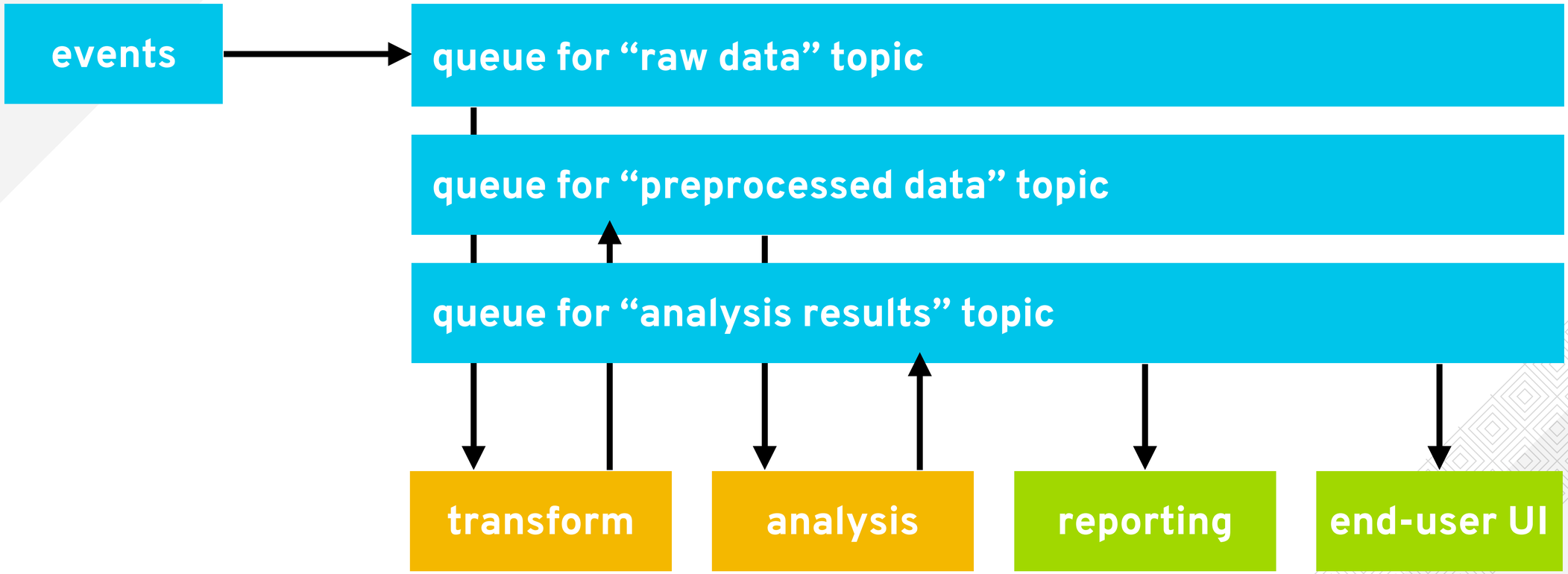
THE KAPPA ARCHITECTURE



THE KAPPA ARCHITECTURE



THE KAPPA ARCHITECTURE



THE KAPPA ARCHITECTURE

THE KAPPA ARCHITECTURE

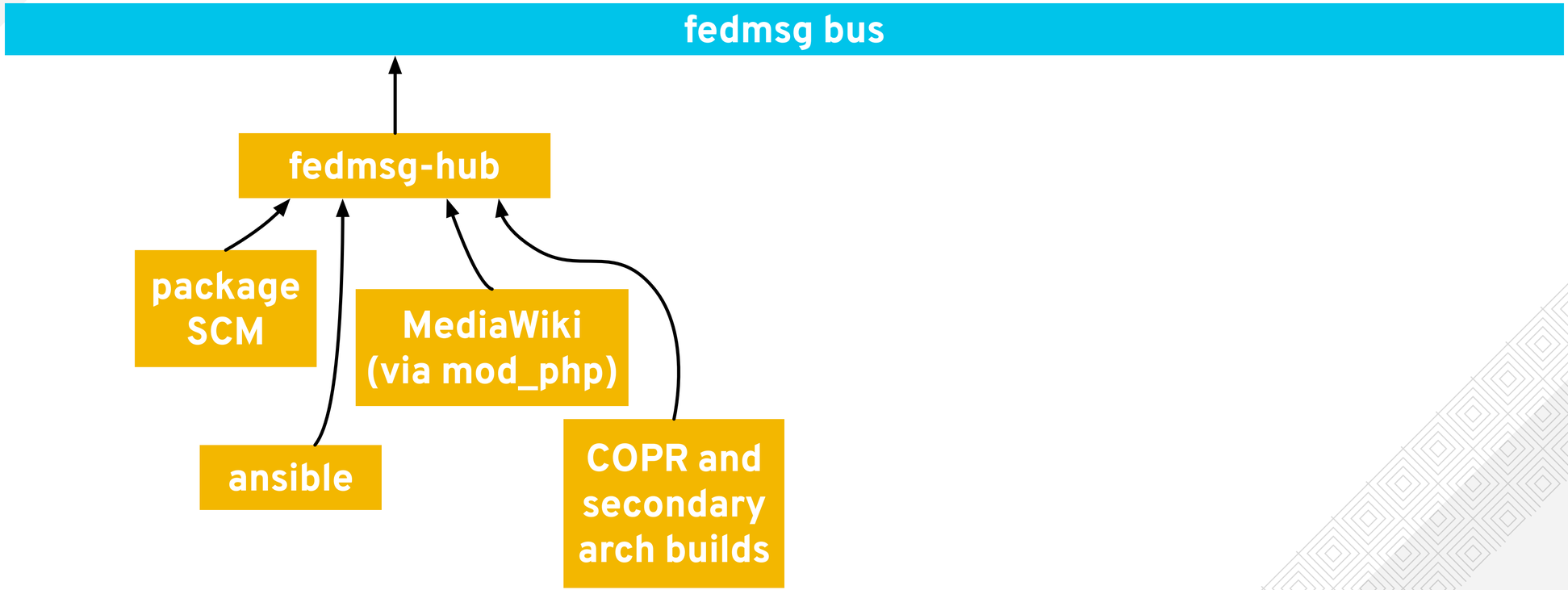
- suitable for micro-service architectures

THE KAPPA ARCHITECTURE

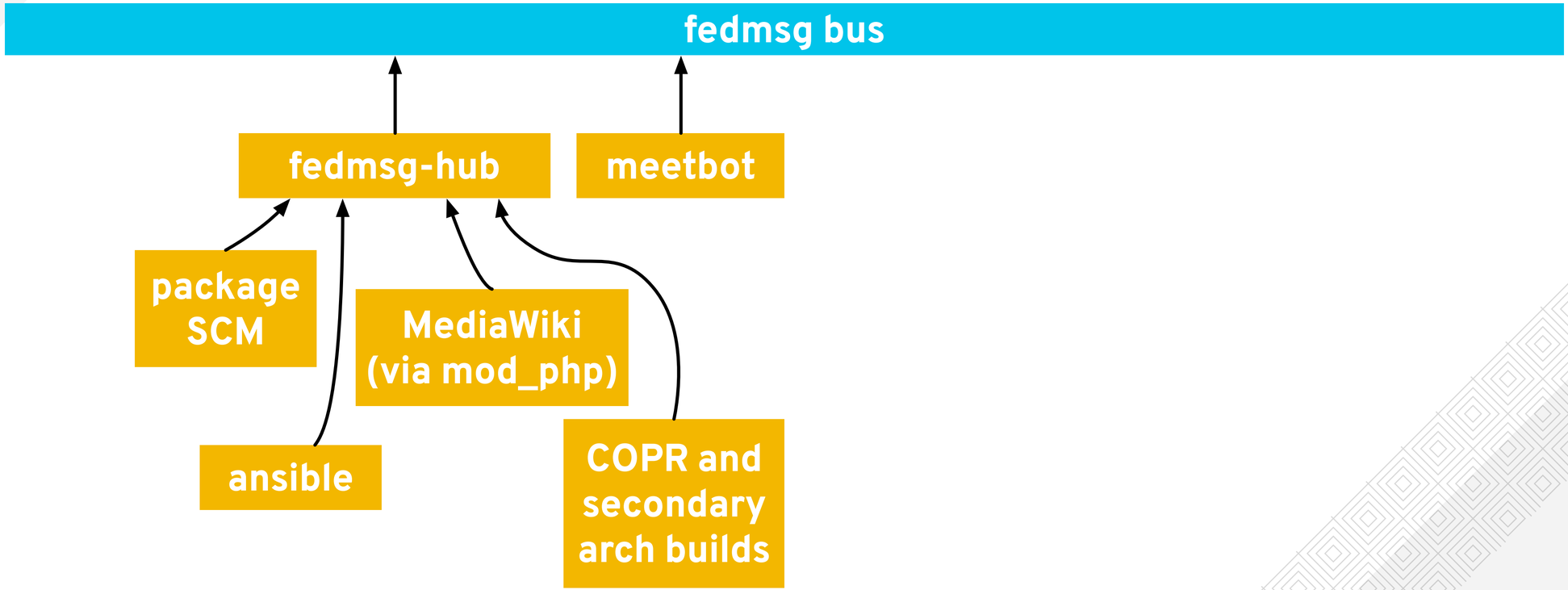
- suitable for micro-service architectures
- requires sophisticated streaming framework, streaming algorithms

EXAMPLE: FEDORA INFRASTRUCTURE

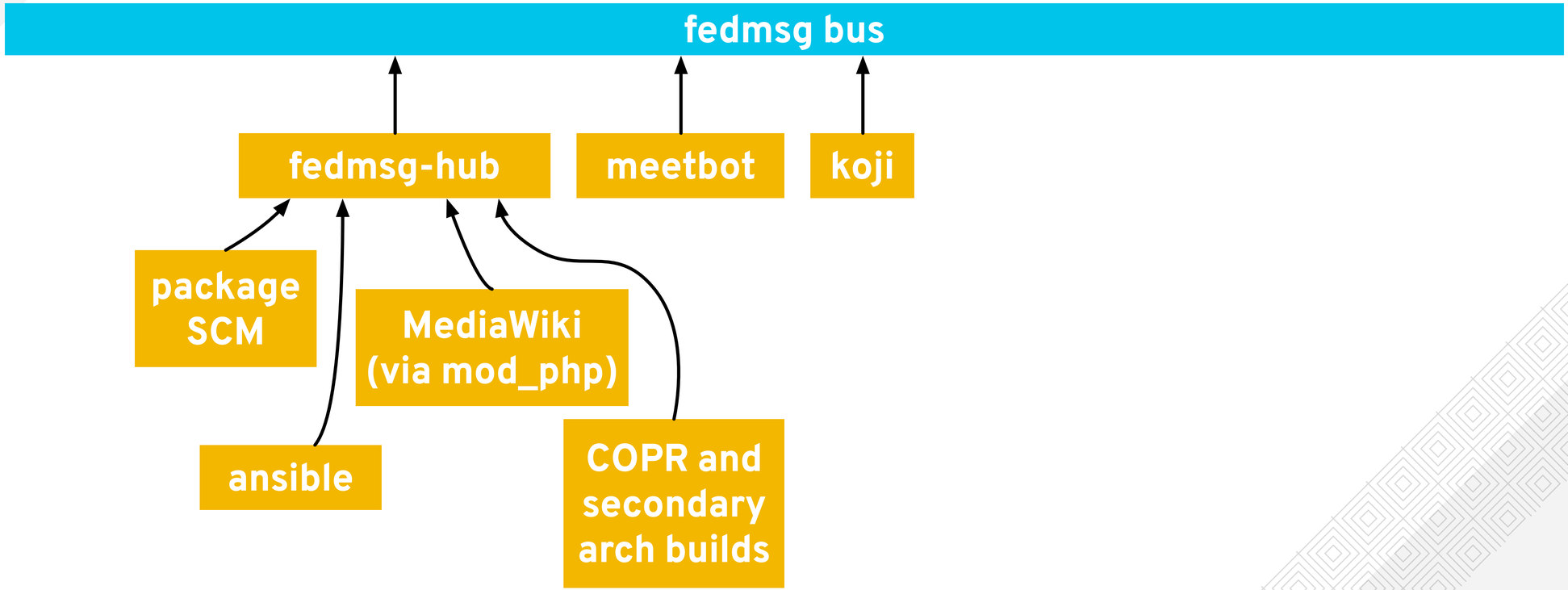
EXAMPLE: FEDORA INFRASTRUCTURE



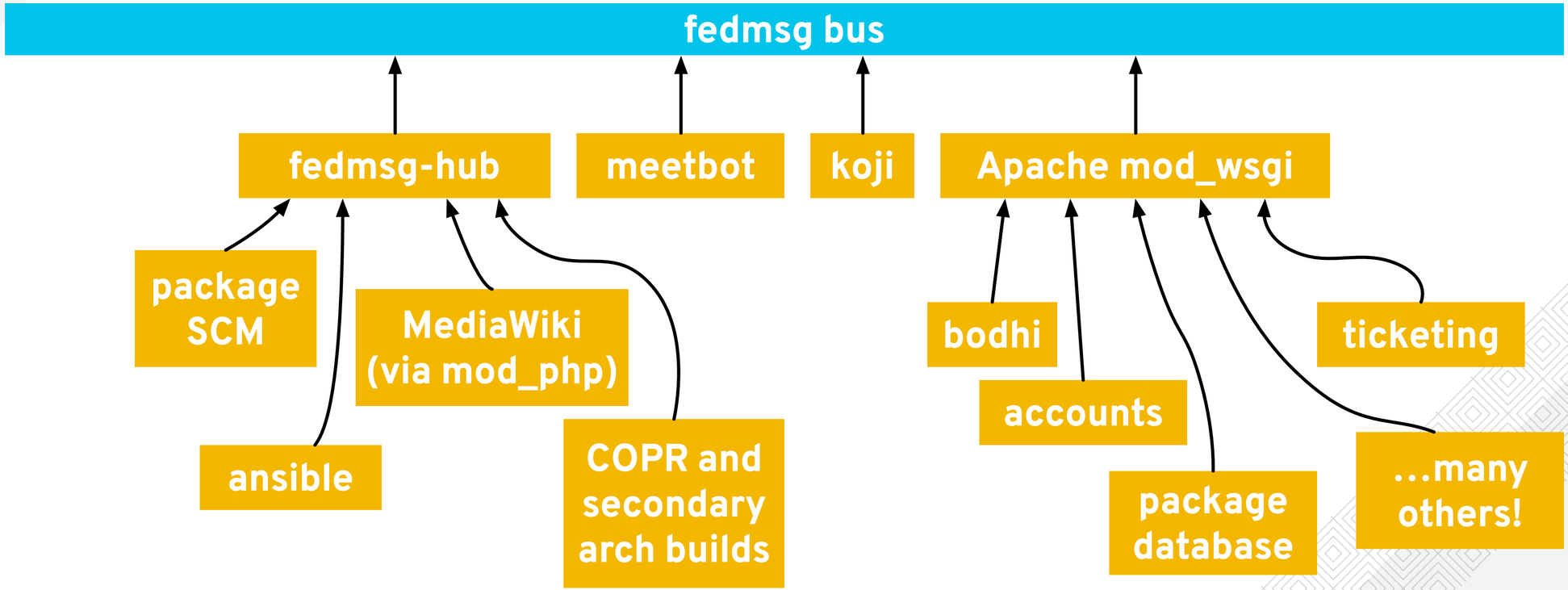
EXAMPLE: FEDORA INFRASTRUCTURE



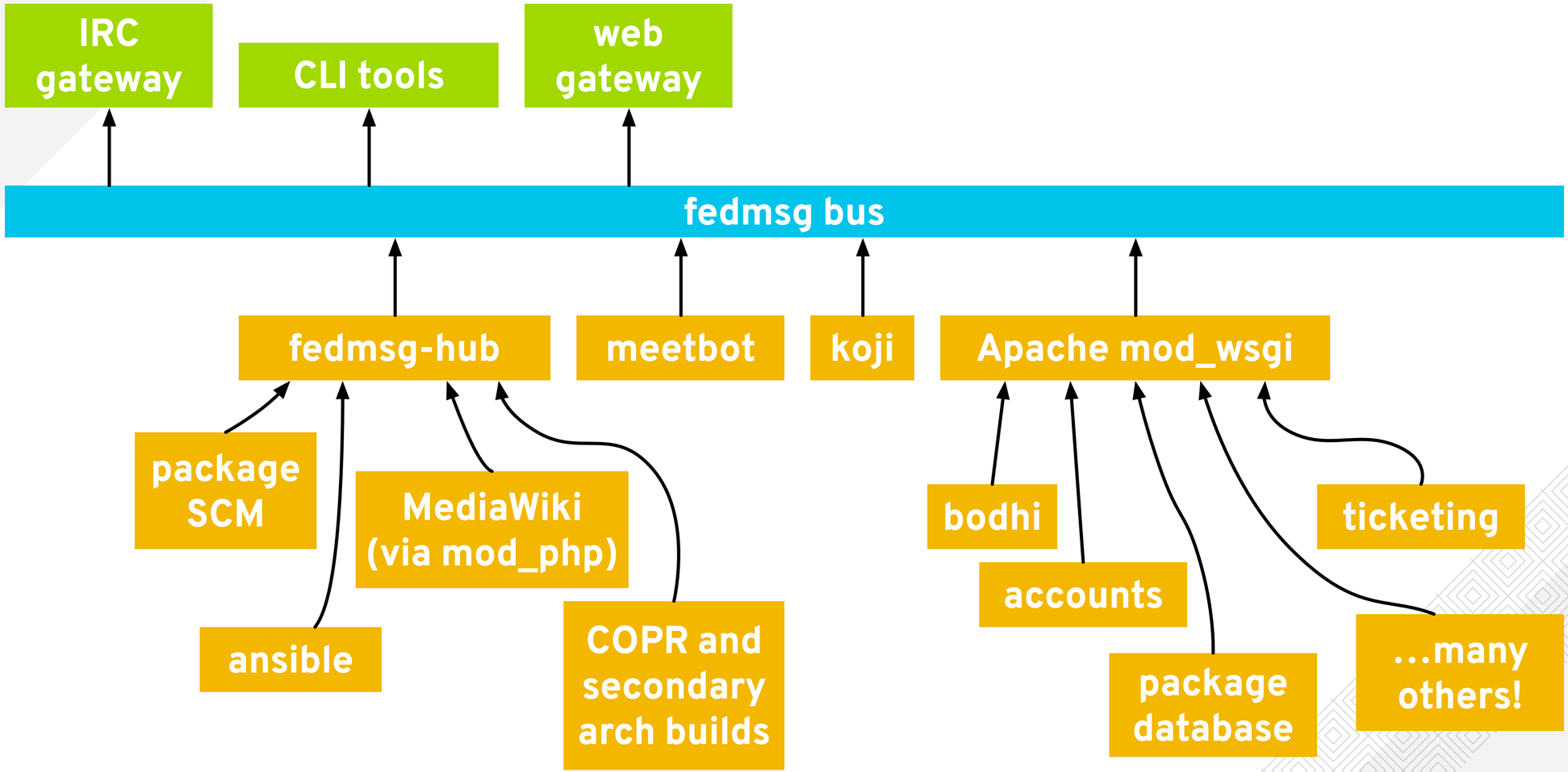
EXAMPLE: FEDORA INFRASTRUCTURE



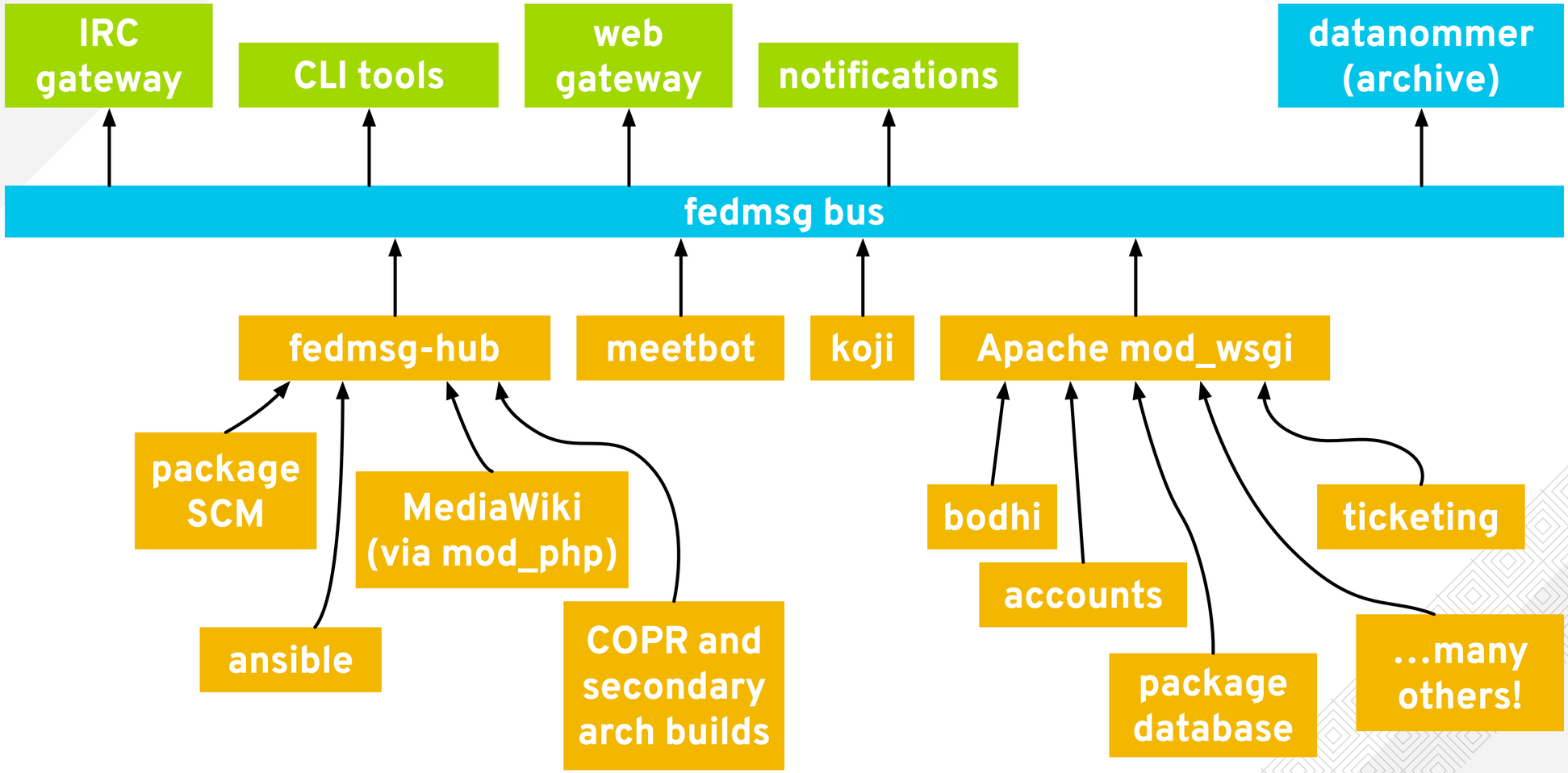
EXAMPLE: FEDORA INFRASTRUCTURE



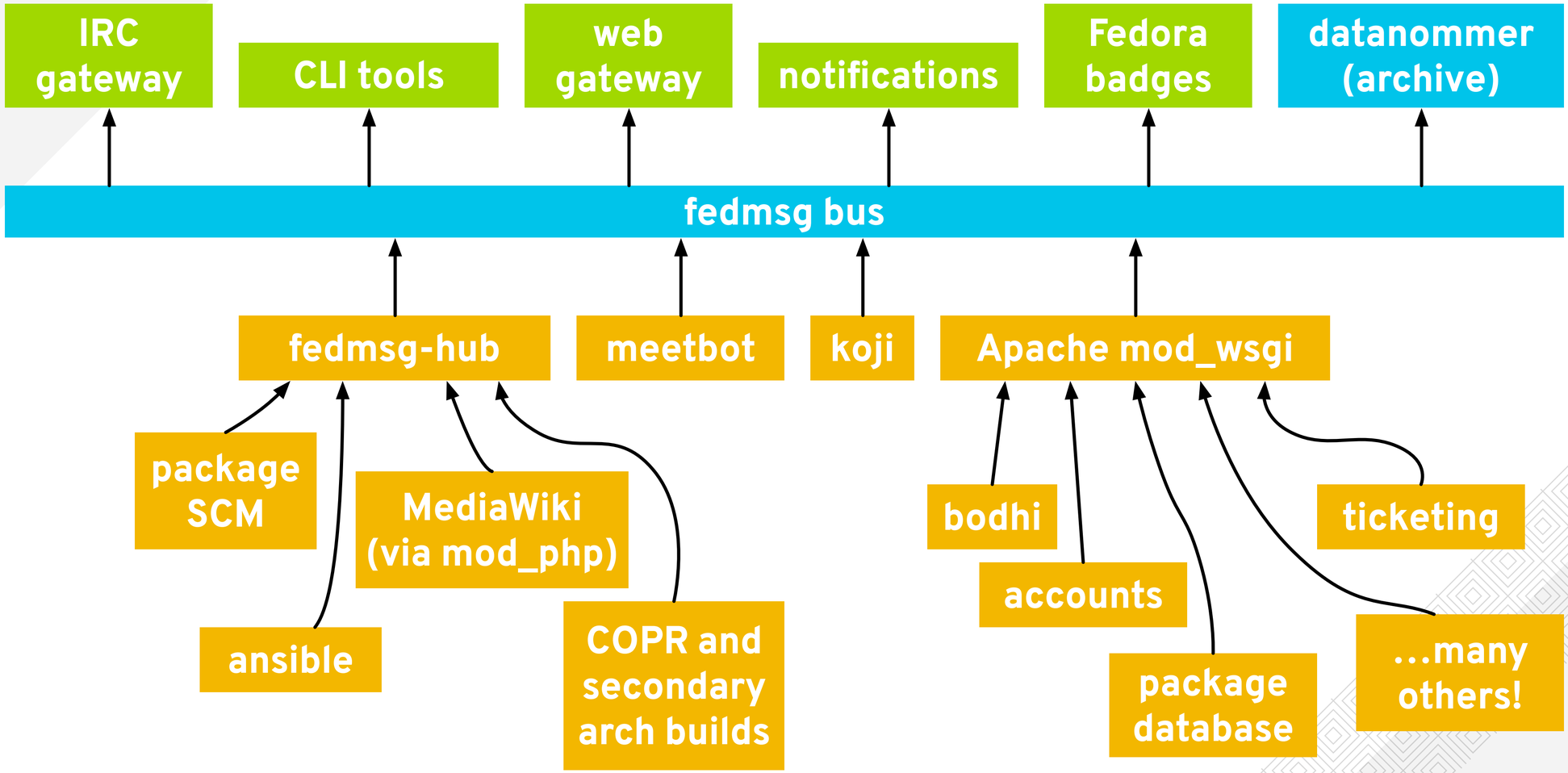
EXAMPLE: FEDORA INFRASTRUCTURE



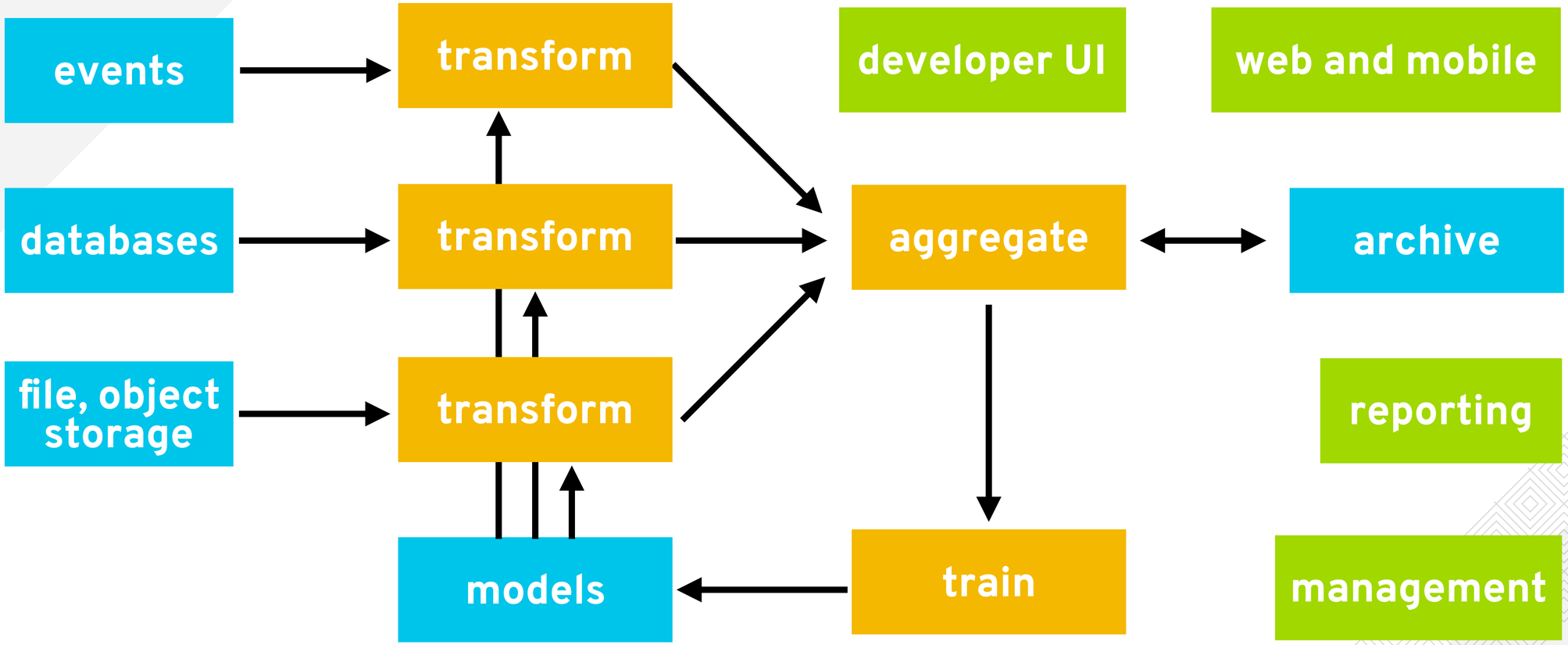
EXAMPLE: FEDORA INFRASTRUCTURE



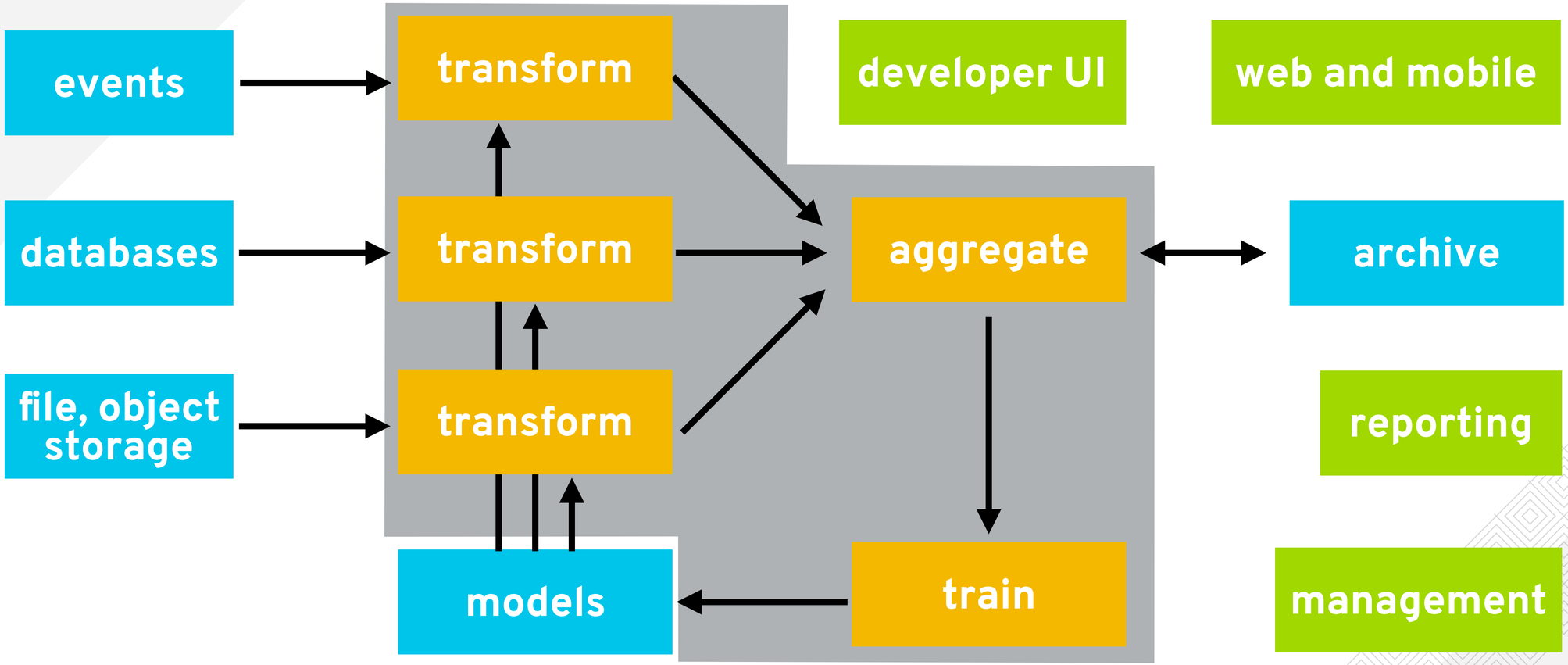
EXAMPLE: FEDORA INFRASTRUCTURE



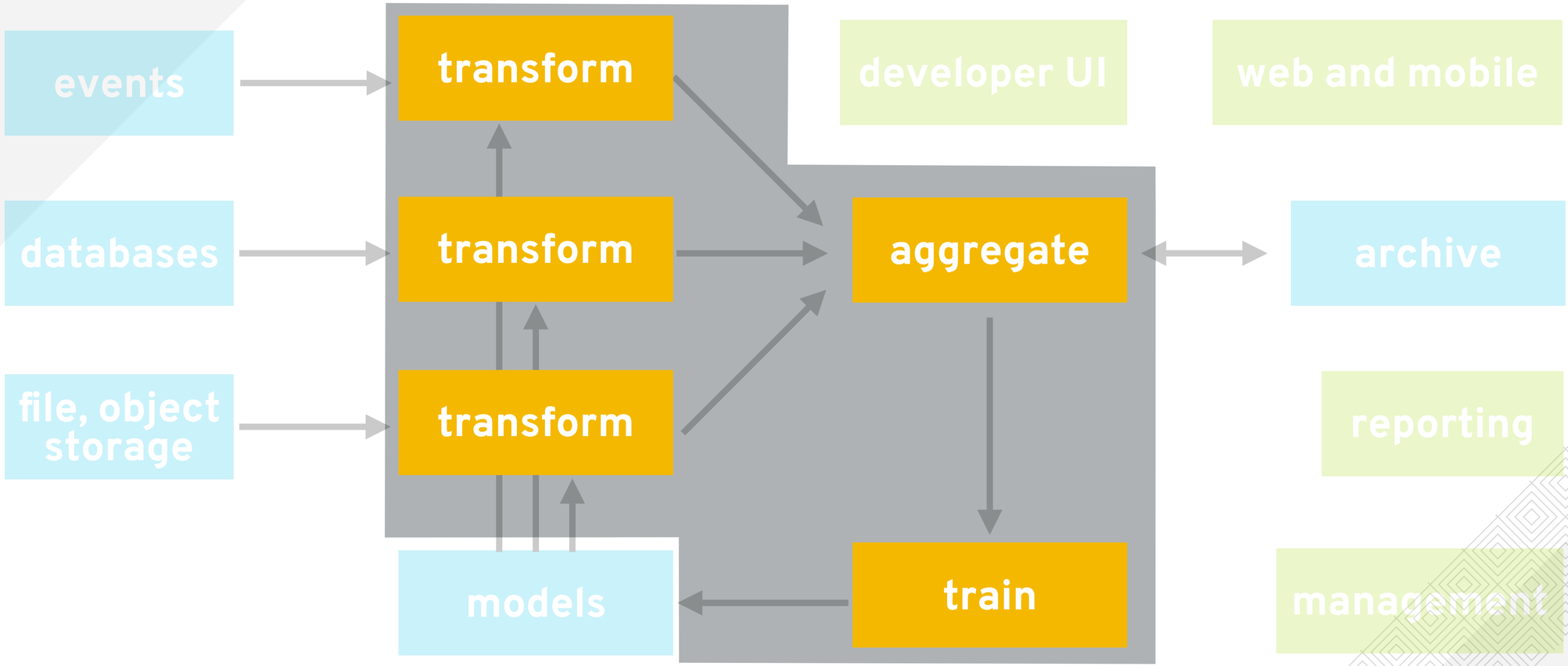
DATA FEDERATION IN THE COMPUTE LAYER



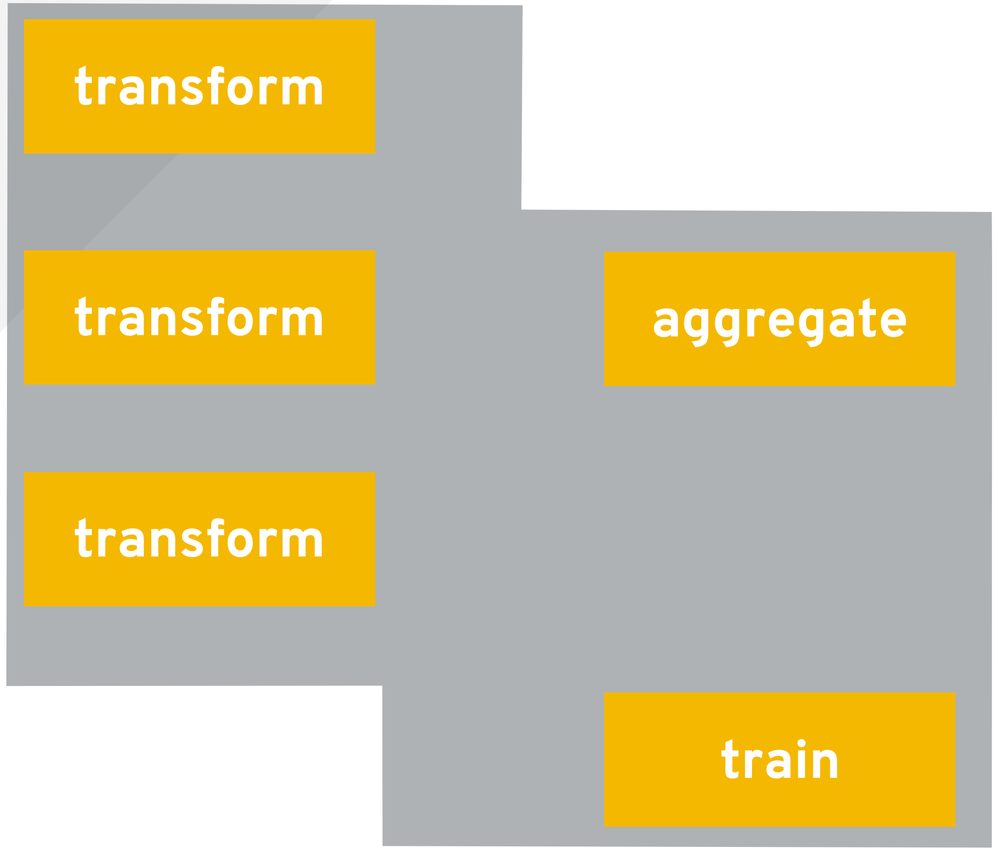
DATA FEDERATION IN THE COMPUTE LAYER



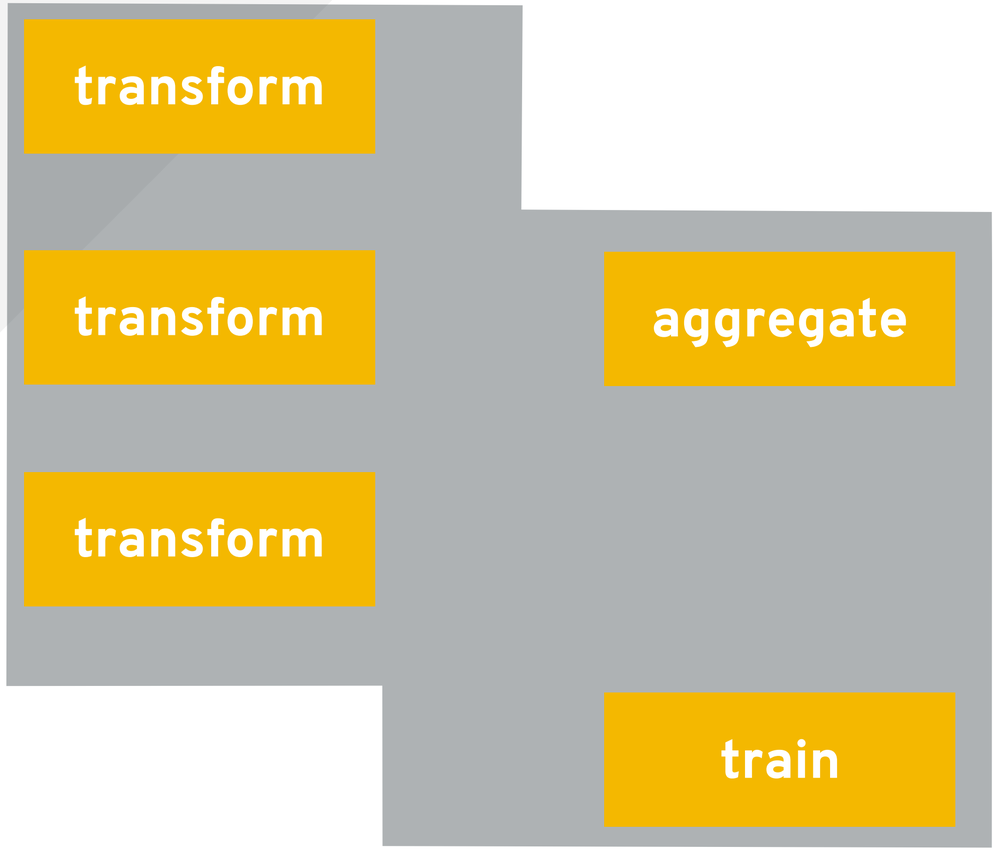
DATA FEDERATION IN THE COMPUTE LAYER



DATA FEDERATION IN THE COMPUTE LAYER

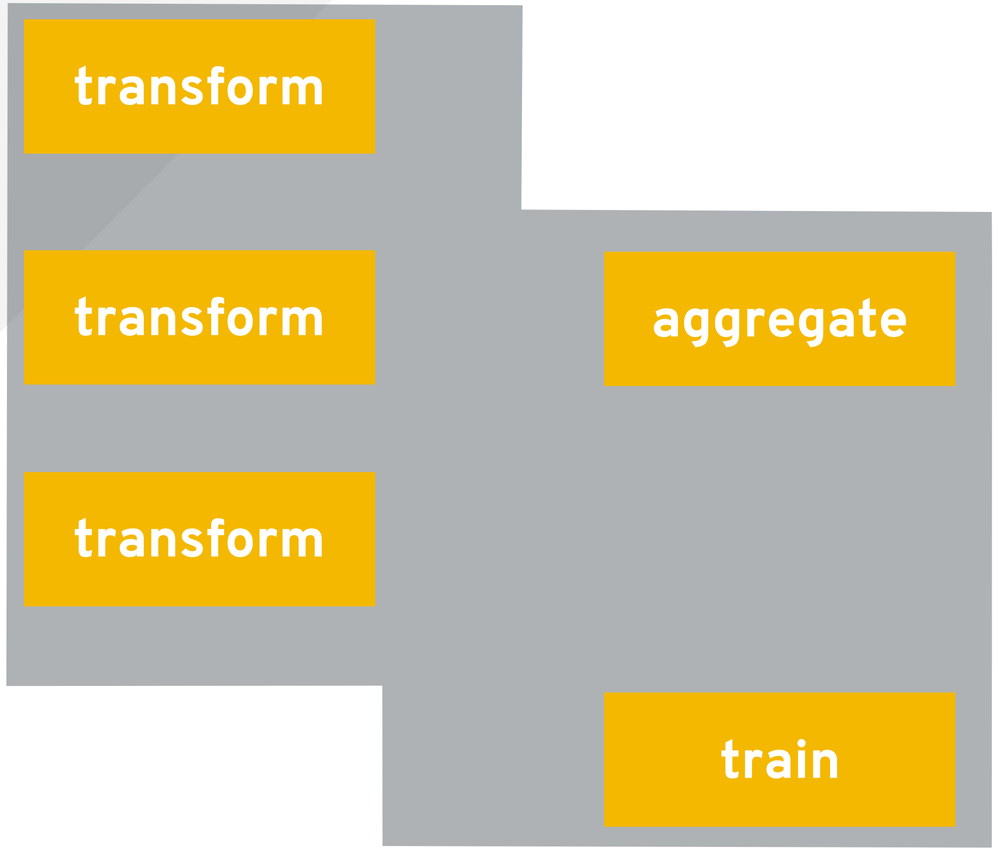


DATA FEDERATION IN THE COMPUTE LAYER



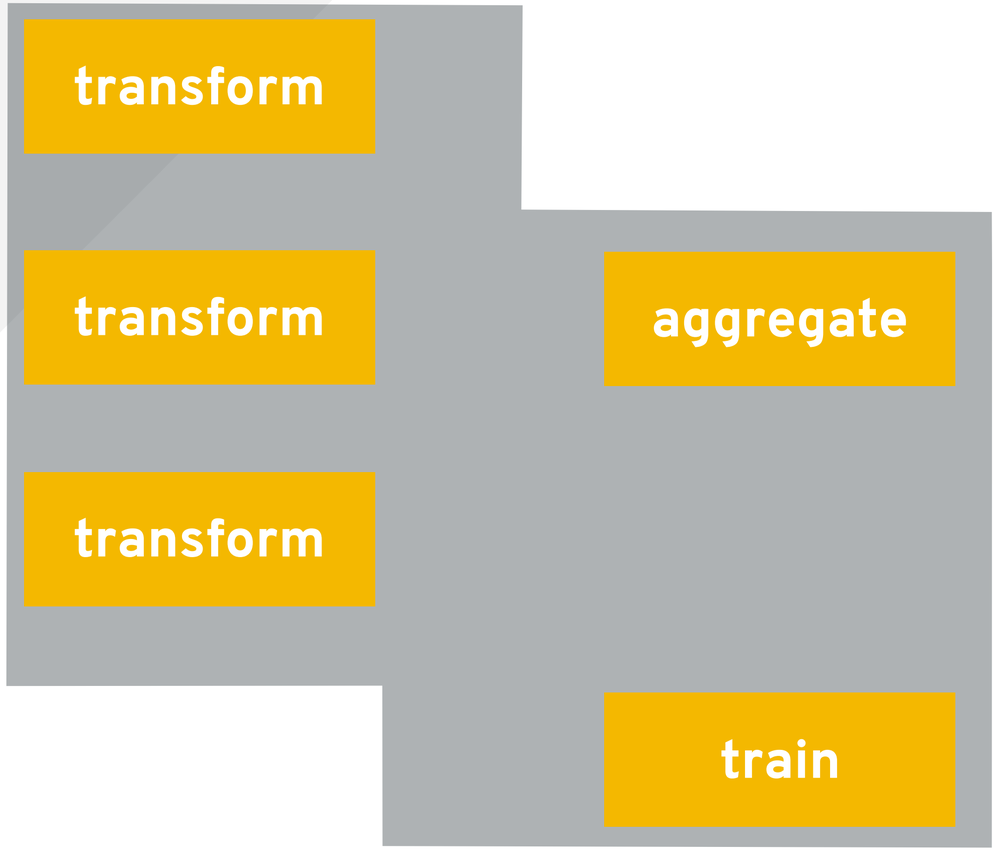
- suitable for micro-service architectures

DATA FEDERATION IN THE COMPUTE LAYER



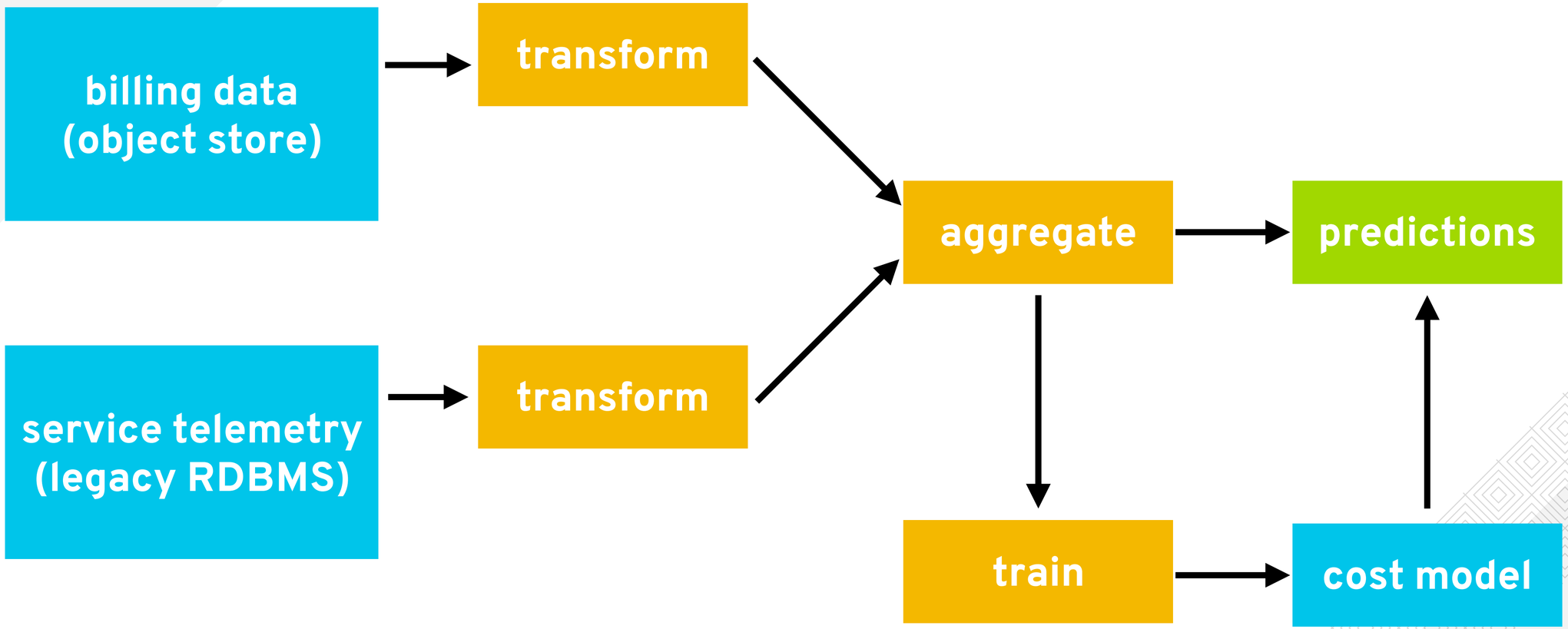
- suitable for micro-service architectures
- flexible and interoperable with other systems

DATA FEDERATION IN THE COMPUTE LAYER

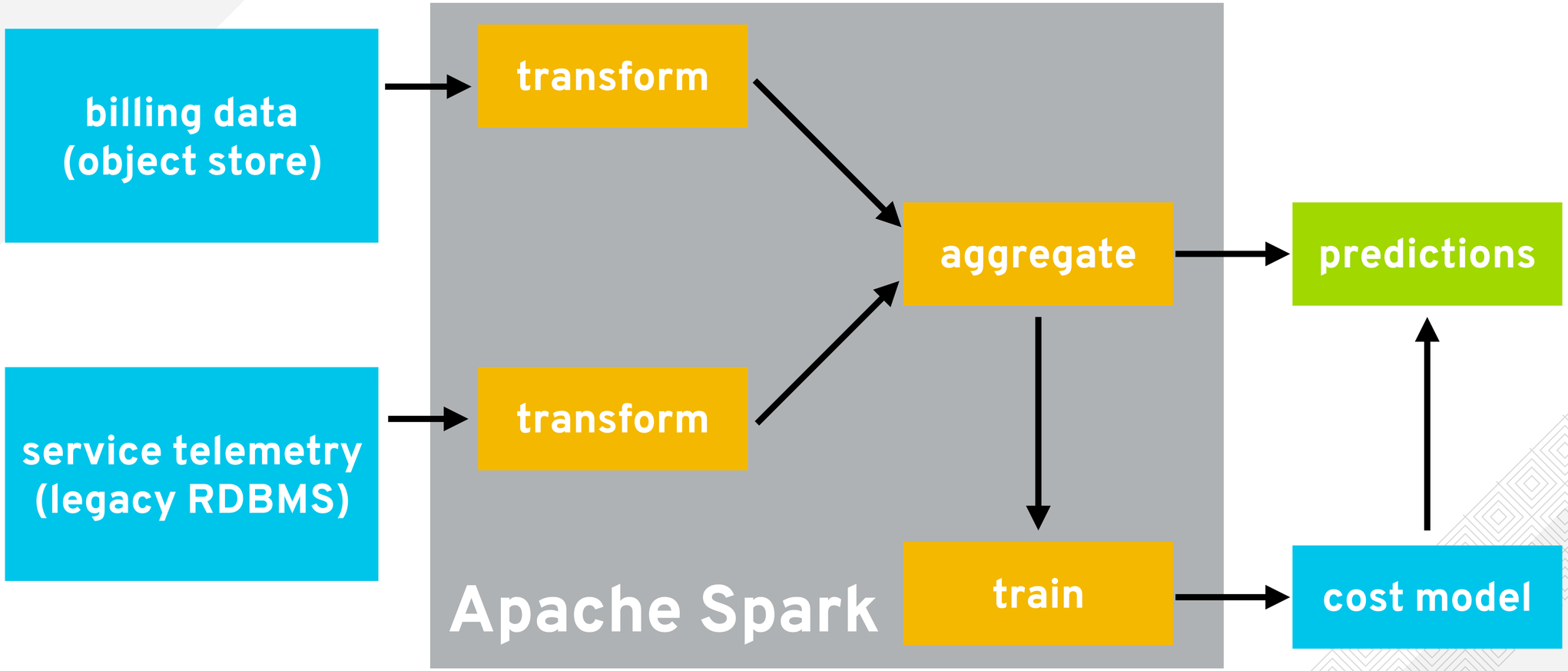


- suitable for micro-service architectures
- flexible and interoperable with other systems
- supports both batch and streaming workloads

EXAMPLE: INFRASTRUCTURE COST MODELING



EXAMPLE: INFRASTRUCTURE COST MODELING



CONCLUSIONS AND TAKEAWAYS

KEY TAKEAWAYS

Analytics is no longer a separate workload; instead, it is a key component of contemporary applications.

You don’t need a heavyweight scheduler and dedicated resources to run data-driven applications in production.

Choosing the right architecture will let you develop and deploy data-driven applications in containers in the public cloud or in your datacenter.

THANKS!

@willb • willb@redhat.com
https://chapeau.freevariable.com