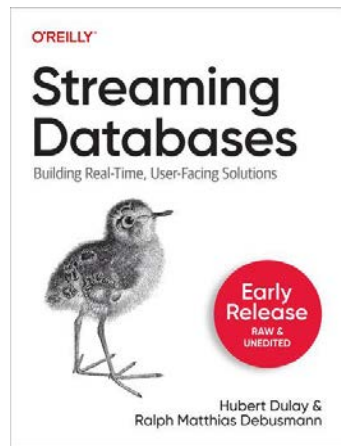
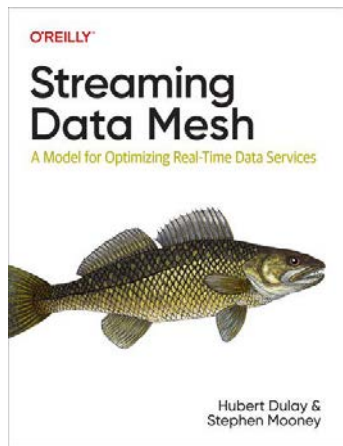


The Streaming Plane

The plane that brings together all data planes

Hubert Dulay

ex-Monolithic Data Engineer
Now Developer Advocate
@StarTree and Author



What You'll Learn

- What is the Data Divide
- What is Data Gravity
- What is the Streaming Plane
- How to consume Real-Time Analytics

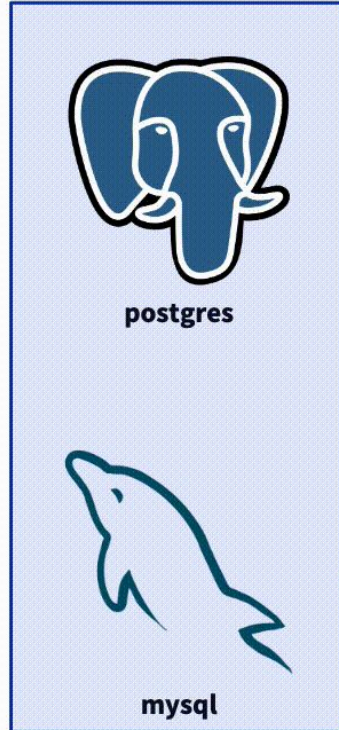
What is the Data Divide?

The Data Divide

Why the Divide?

- Analytical queries are queries like COUNT and AVG or perform JOINS are resource intensive.
- Operational databases support the applications bring revenue. So you don't want to interact with them too much.

Operational Plane



Analytical Plane



-----Downstream----->

-----Upstream-----<

What Happens on the LINE from Left to Right?

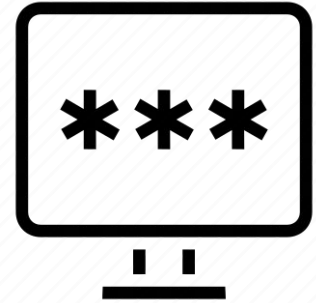
Cleansing



Enrichment



Obfuscation



Row-Based

Most OLTPs store data in row-based format. Optimized for lookups and insert, update, and delete transactions.

Operational Plane



Analytical Plane



-Easy-

-Hard-

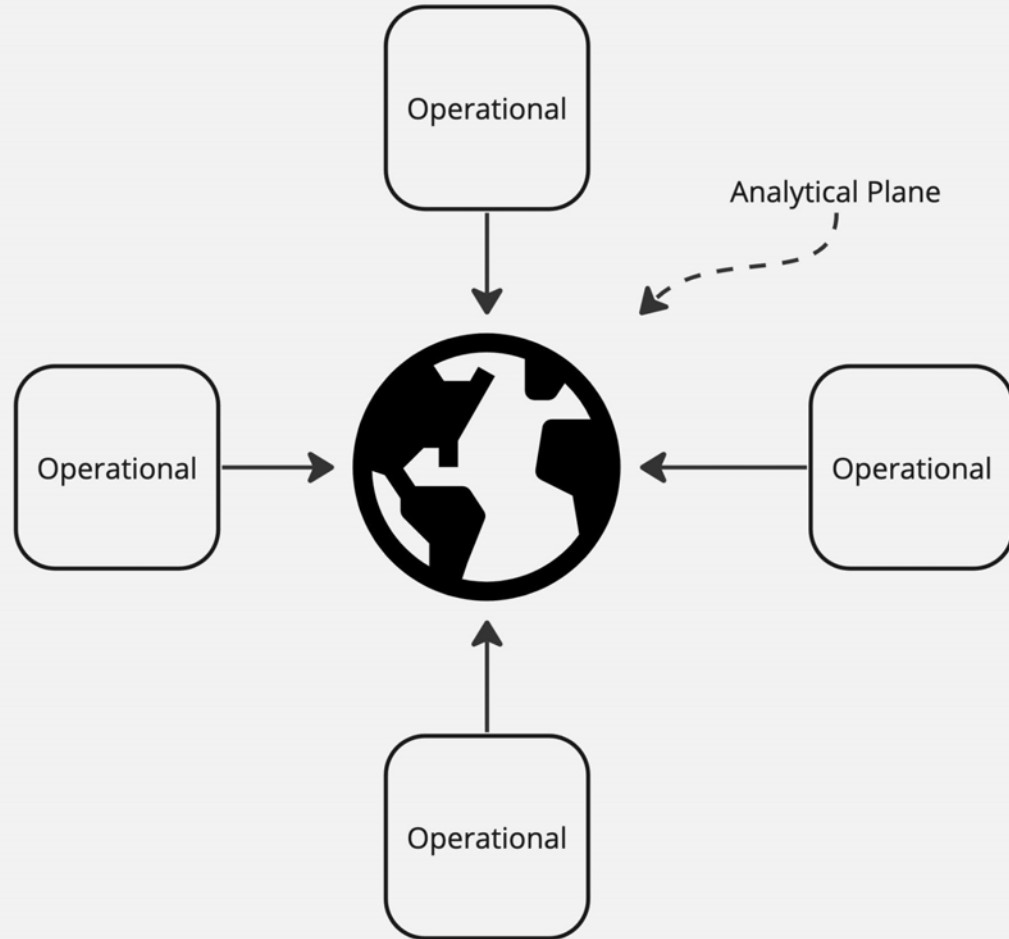
Columnar-Based

OLAP databases store their data in columnar-based formats like Parquet, which are optimized for analytical queries.

Reverse ETL

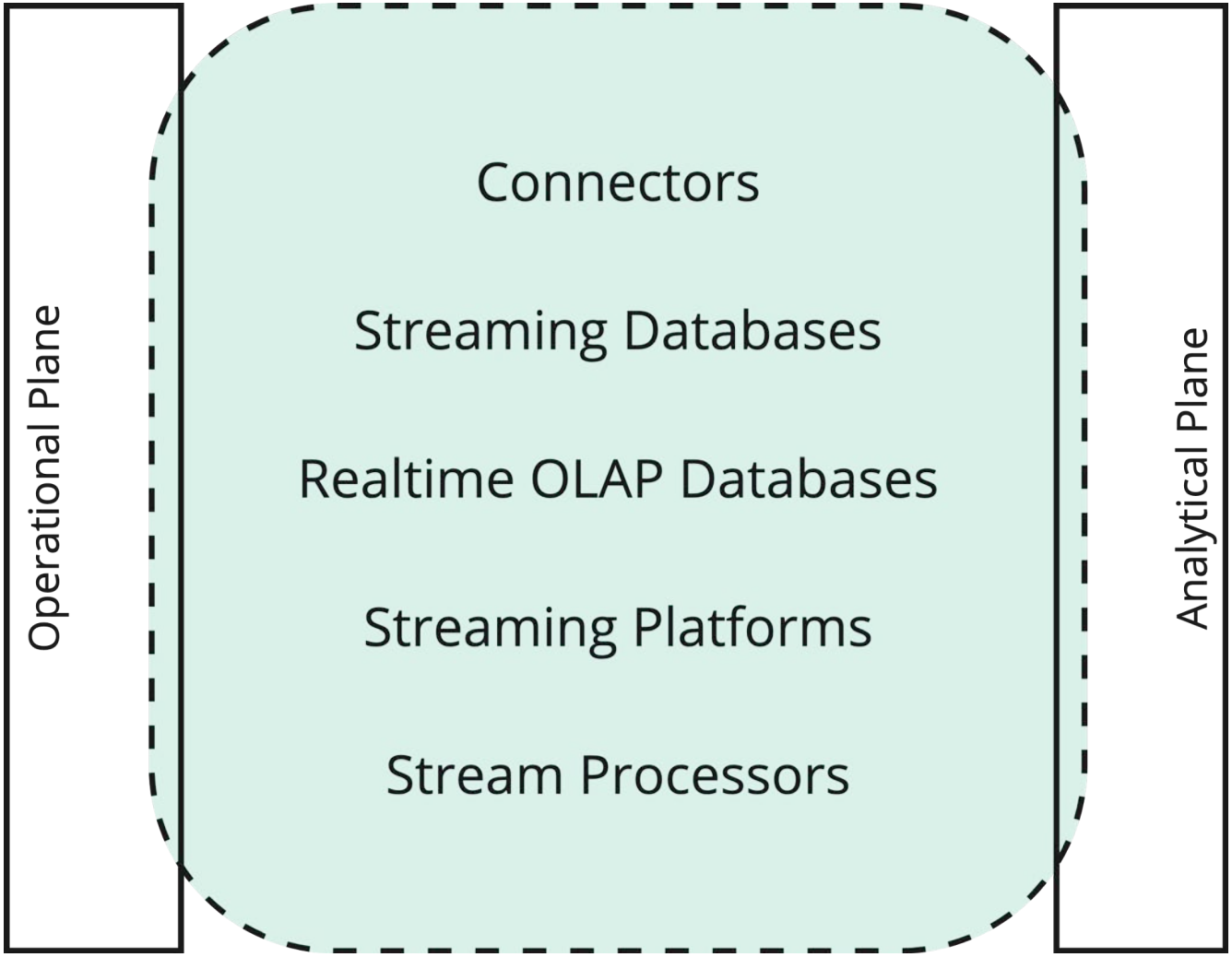
OLAPs don't like to return all rows and columns. They prefer to serve analytical queries which return aggregations of the data.

Data Gravity

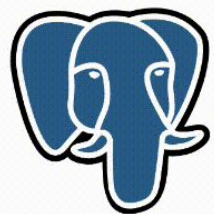


What is the Streaming Plane?

The Stuff Inside
the
Streaming Plane

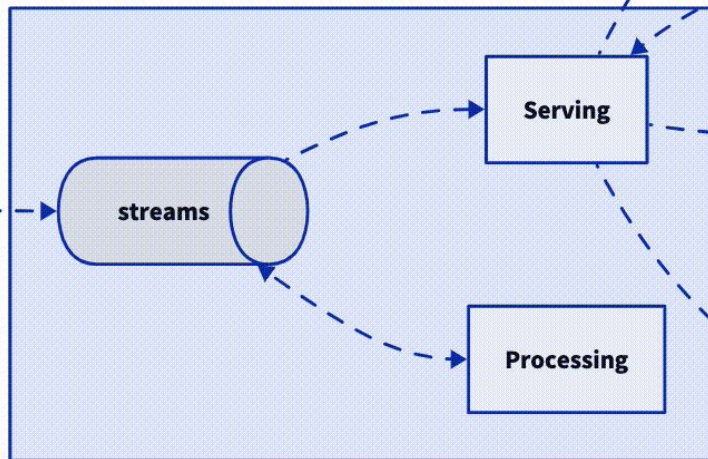


Streaming Plane



Operational Plane

CDC



Streaming Plane

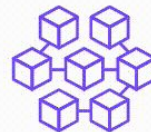
Serving

Processing

Loading



Analytical Plane



Non-Human Facing



Human Facing

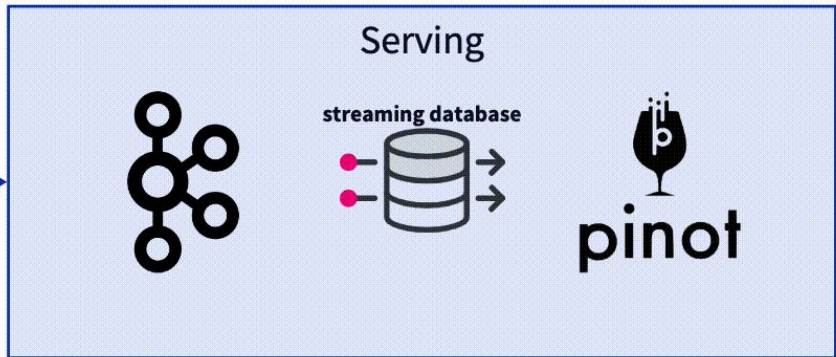
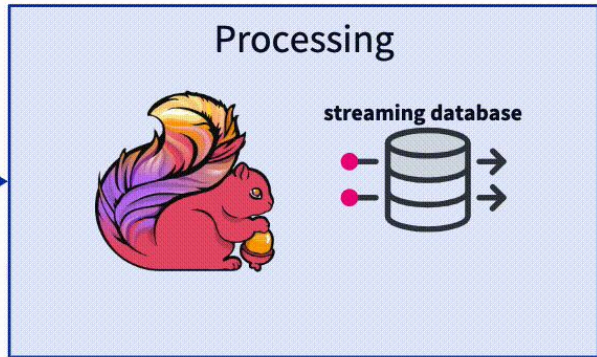
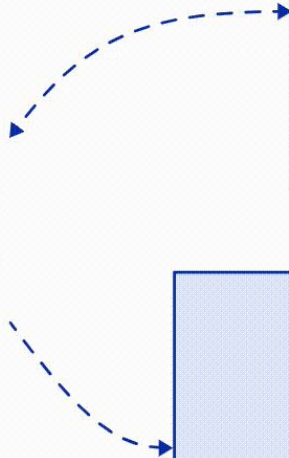
Inside the Streaming Plane

Processing

Transforms the streams places the results back into the streaming platform.

Serving

Serving provides synchronous and asynchronous endpoints for applications to consume.



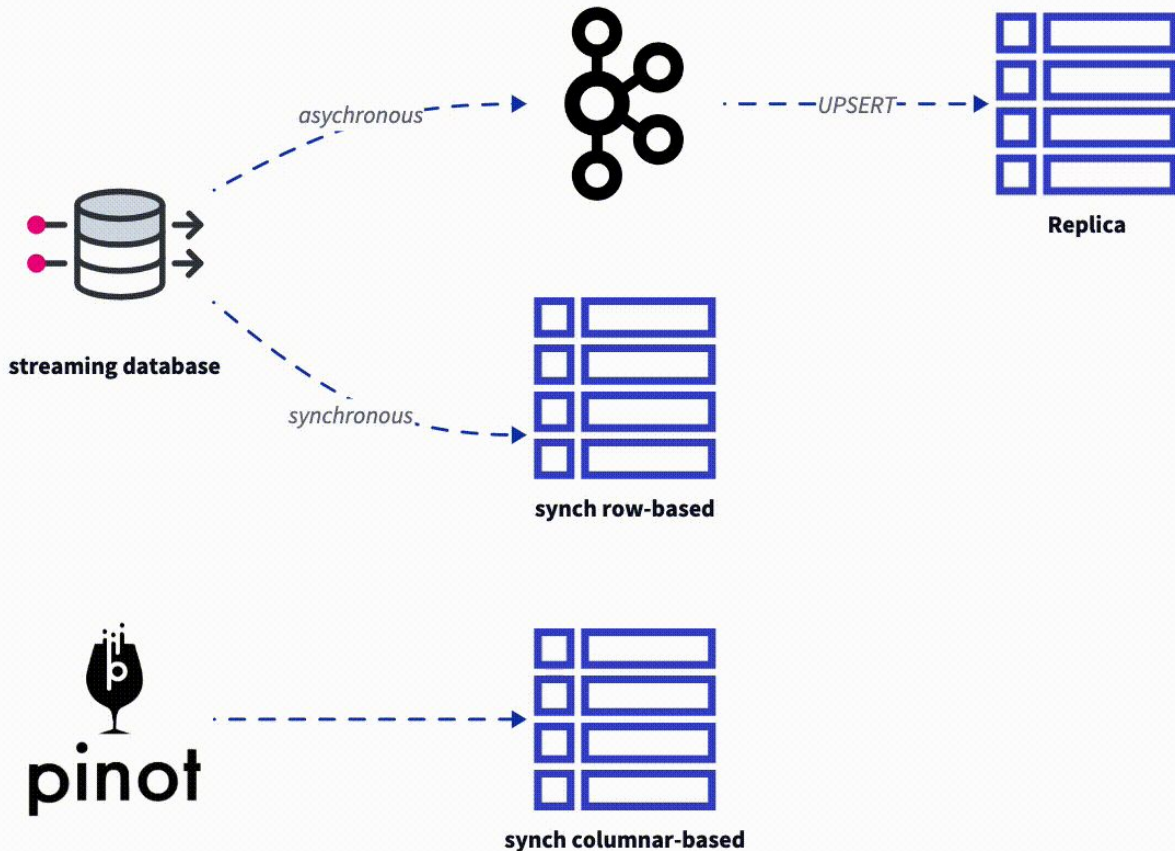
Inside the Serving Layer

Serving

The serving layer serves data synchronously and asynchronously through tabular constructs called MATERIALIZED VIEWS.

Asynchronous

Consumers of these data APIs need to build a replica of the original materialized view.



How does the Streaming Plane Work: Pushing and Pulling Materialized Views

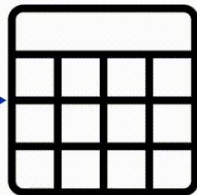
Serving Materialized Views

Push Queries

Try to off load some of the work so that the pull query does not have to perform heavy-lifting workloads. These queries run asynchronously. Examples are Joins and Aggregations.



Push Query



Materialized View

Pull Query



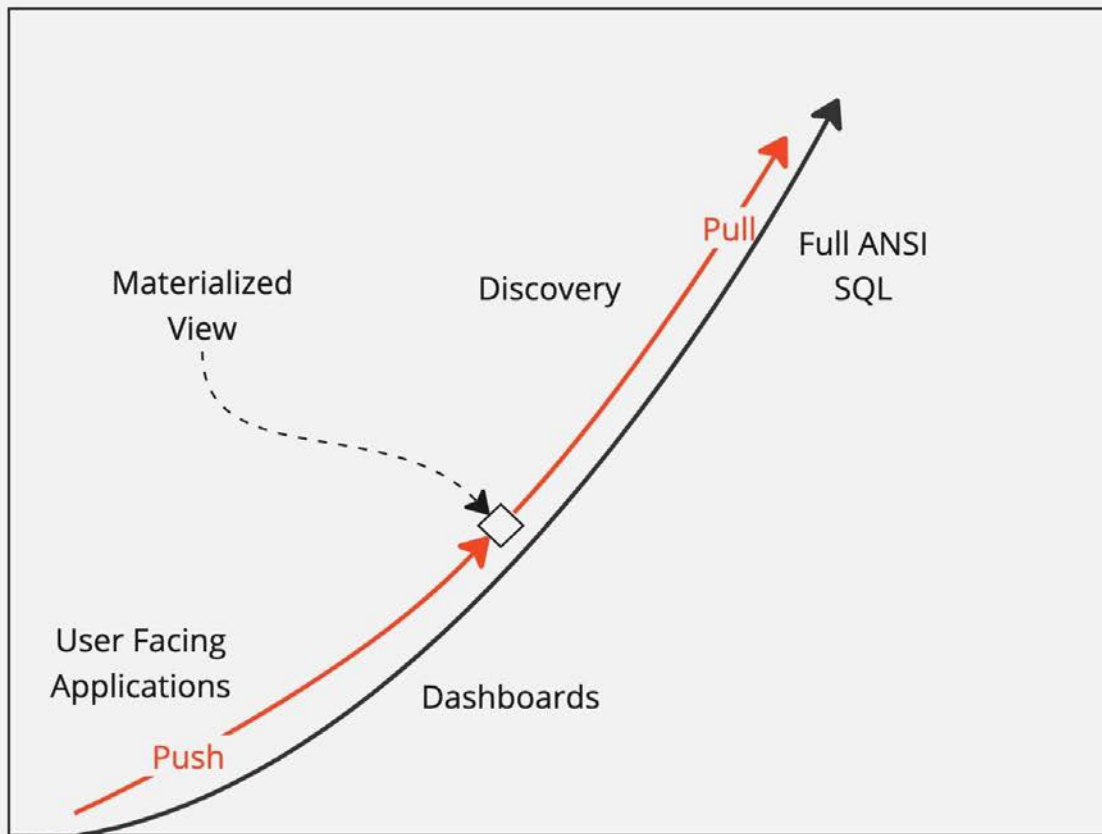
Pull Queries

Are analytical queries that are authored by a Human. Using defined in SQL, these queries need to be executed with low latency. Aggregations are done here and sometimes called slicing-and-dicing.

Push Pull Balance

Data Engineers need to balance between pull and push. The more work don on the push, the less flexibility the pull query (and vica versa).

Query
Flexibility



Query
Latency

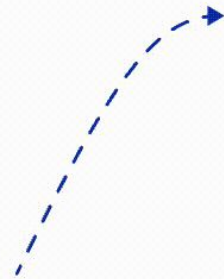
Globally Replicated Serving Layer

Processing Once

Transform near the source and replicate the stream to other regions.

Serving

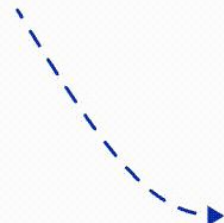
All data can be served locally to the region.



APAC

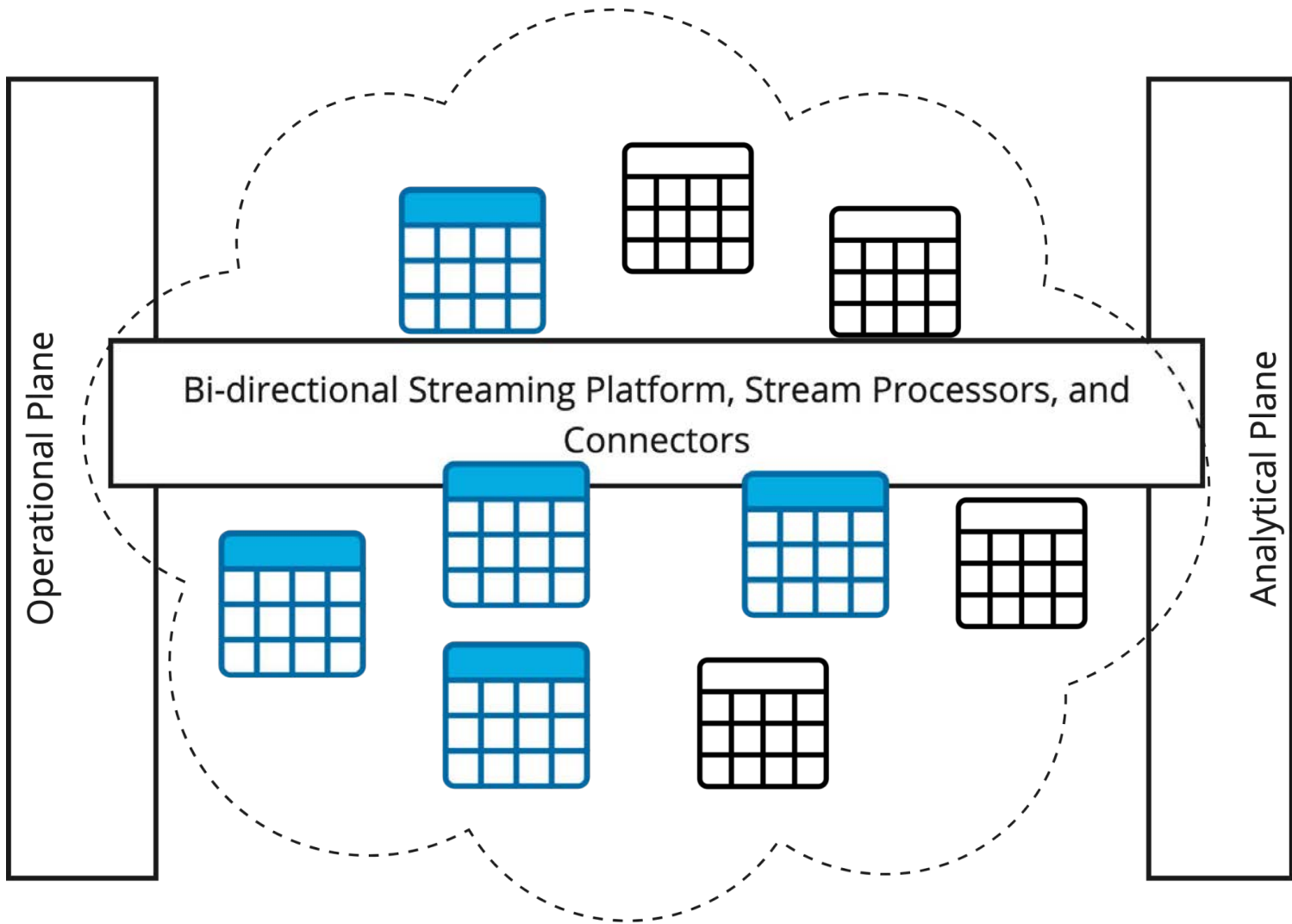


Americas

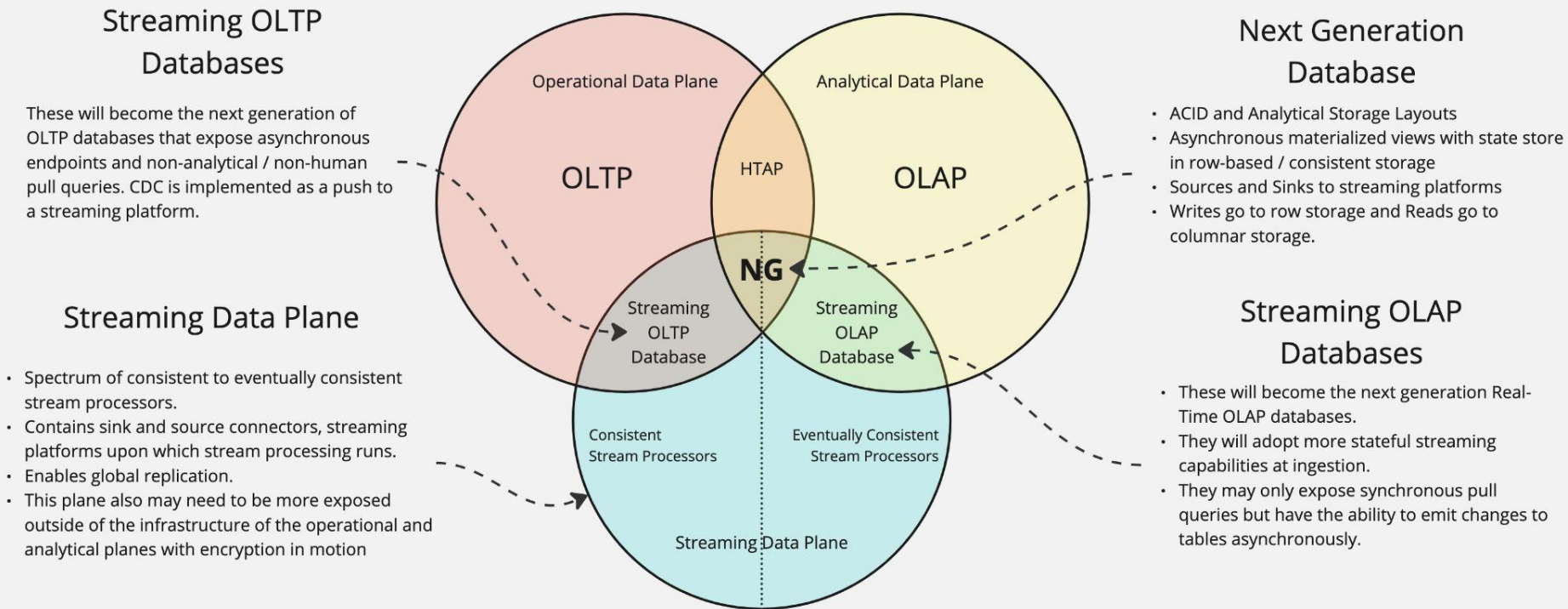


EMEA

Local Consumption



Venn Model for Real Time Analytical Systems



Streaming Plane Summary:

- Streaming Platforms and Connectors provide access to both data in motion and at rest. They also replicate data globally.
- Stream processors transform data and build Materialized Views.
- Realtime OLAP databases provide analytical views to data products.
- SQL is the universal language for data. Especially in the Streaming Plane.
- Data Products exist as both Asynchronous Streams AND Synchronous SQL queries against Materialized Views.

star⁺tree

Thank you!
StarTree Cloud
Free Tier!

