



Demystifying the Data Lakehouse in Fabric

Data Saturday, Copenhagen, 8/2-2025

twoday

Sponsors



GOLD  redgate



SILVER CatMan Solution



BRONZE  BUSINESS INTELLIGENCE



Just Blindbæk

- **Principal Architect at twoday**
 - Pre-sales, workshops, events, marketing
 - Internal practice development
 - Academy: External training
- **Microsoft Data Platform MVP**
- **Found and organizer of**
 - Danish Microsoft BI Community (MsBIP.dk)
 - Power BI UG Denmark (PowerBI.dk)
 - Data Platform Next Step & Power BI Next Step



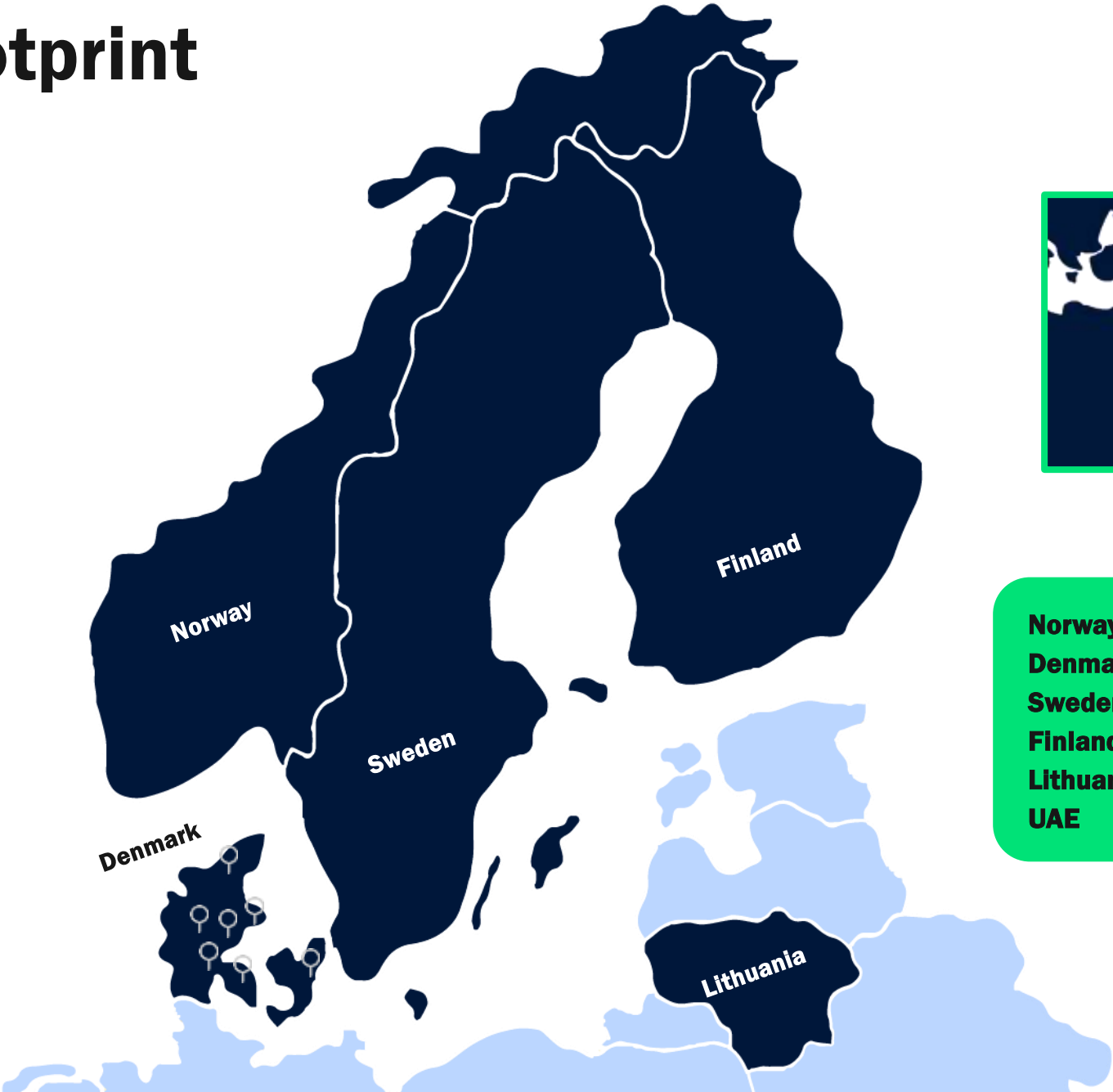
twodays footprint

3,000
Colleagues

6
Countries

35
Offices

40 %
Co-owners

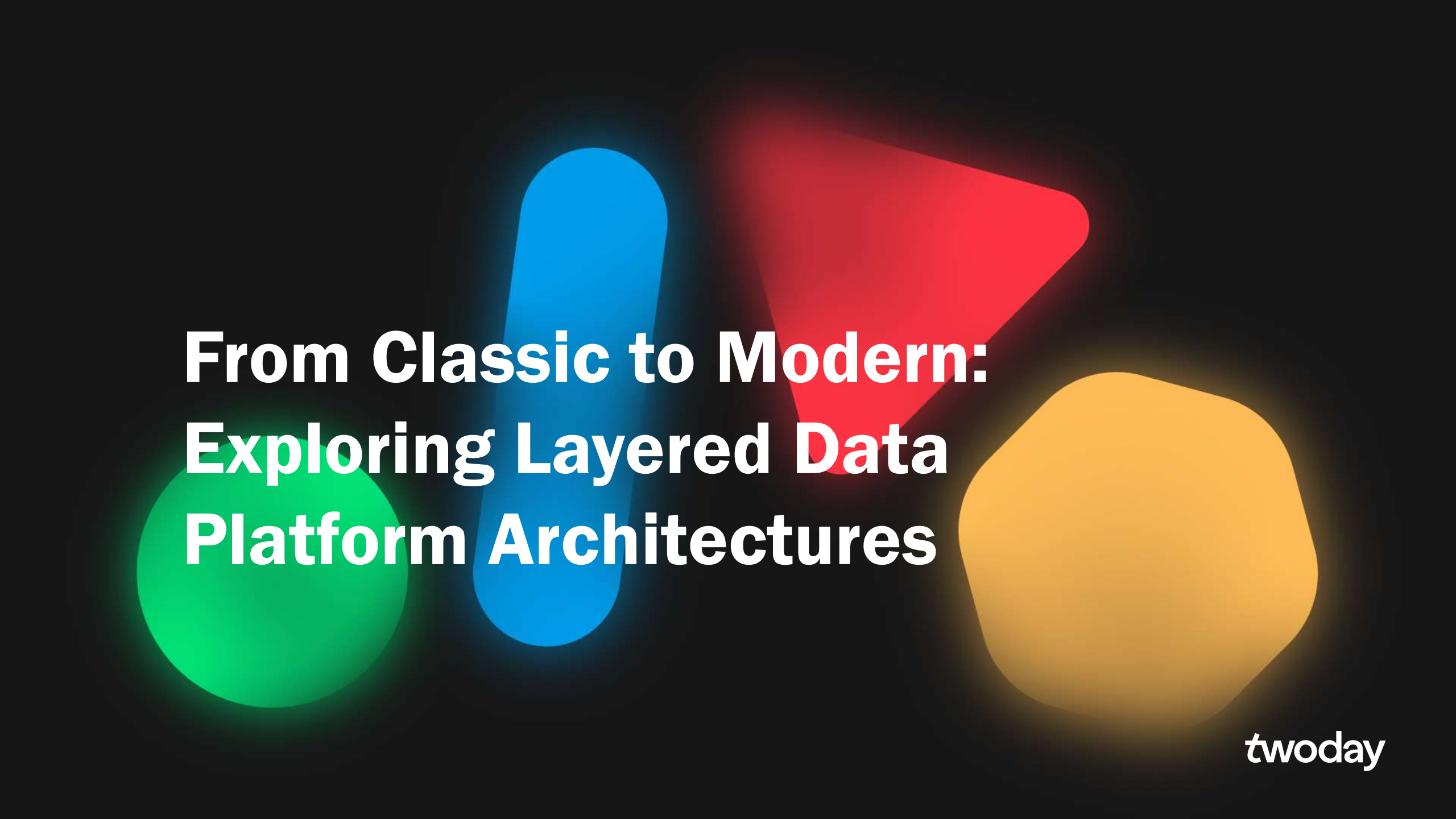


Norway	500 colleagues
Denmark	1,100 colleagues
Sweden	500 colleagues
Finland	700 colleagues
Lithuania	200 colleagues
UAE	15 colleagues

What is this session all about?

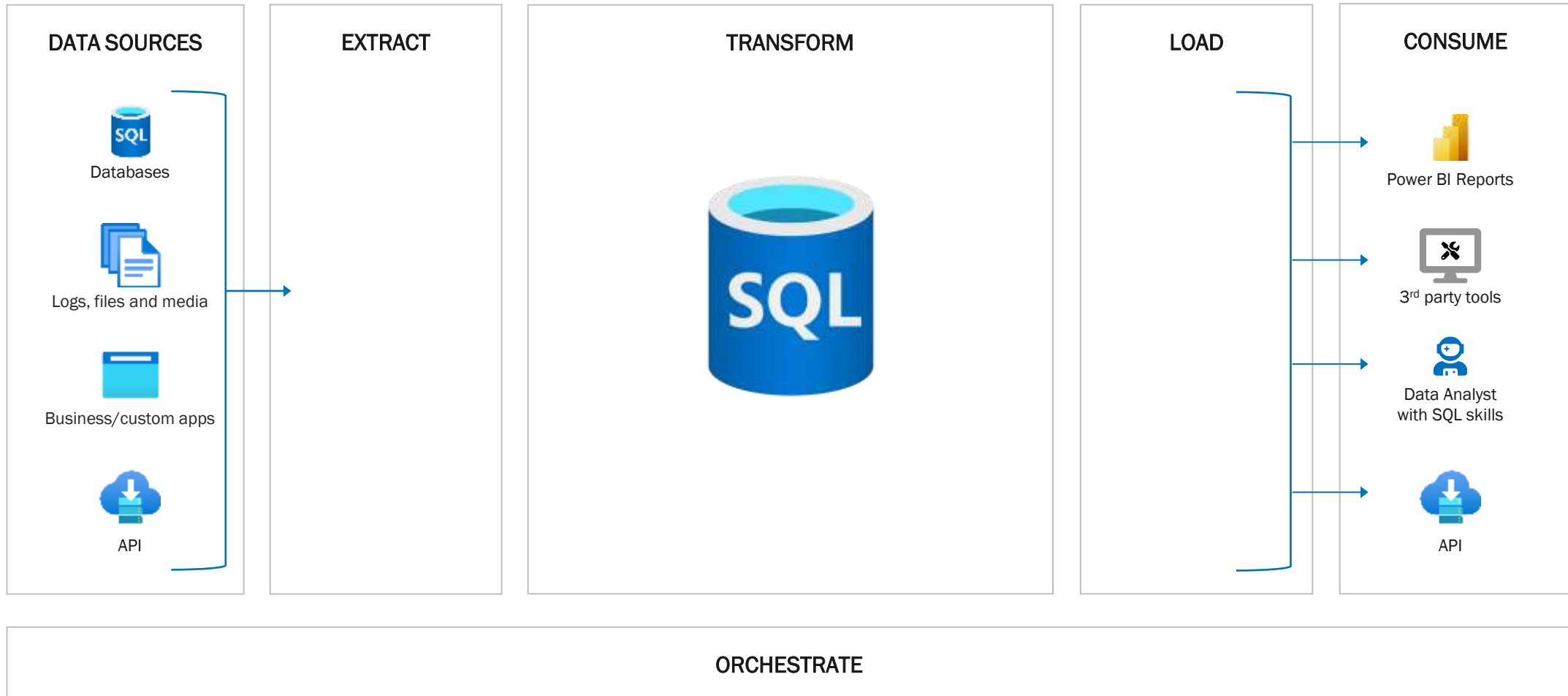
- What is this Data Lakehouse that everyone is talking about now?
- Why Bronze/Silver/Gold now?
- Why decoupling storage and compute now?
- Do we all need to learn Python?
- What about good old SQL?



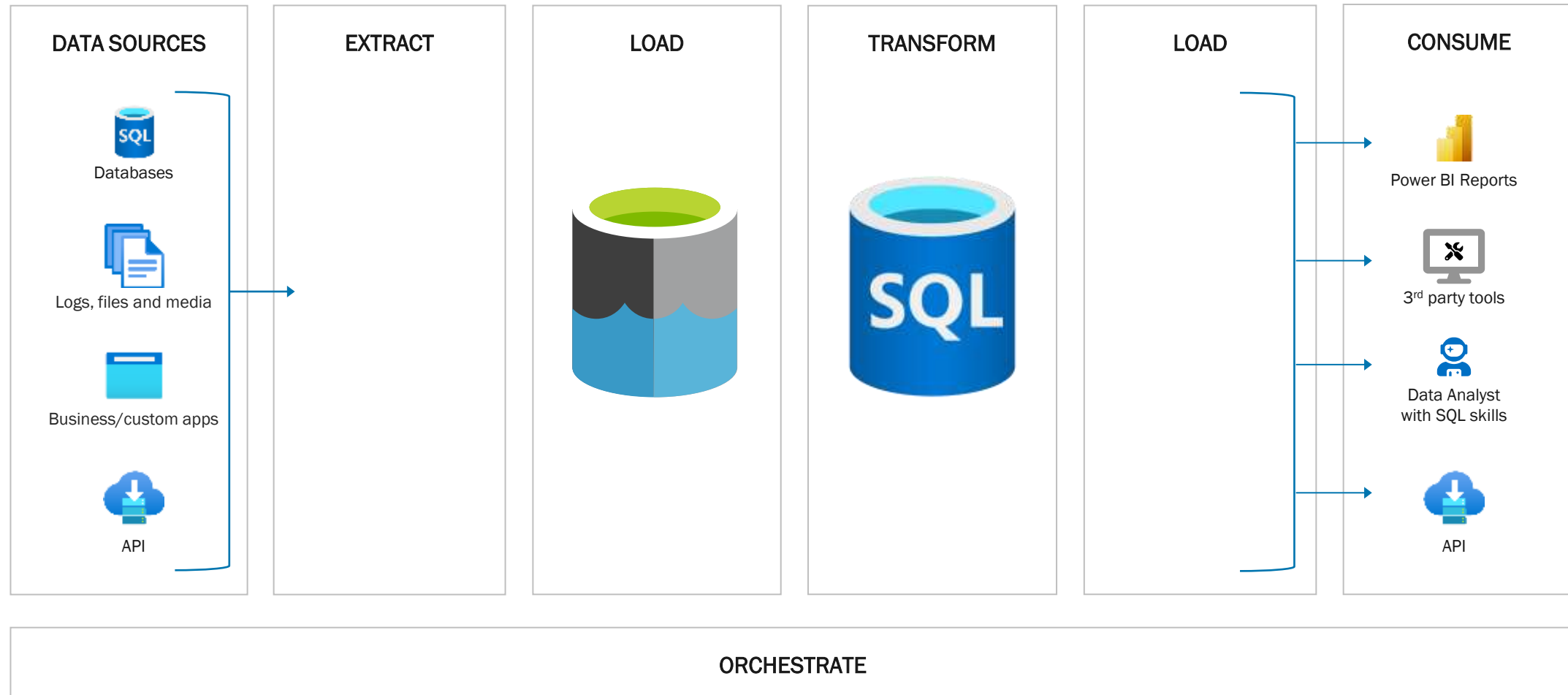
The background is dark with several glowing, semi-transparent shapes: a blue vertical pill shape, a red triangle pointing right, a green circle, and a yellow irregular blob on the right.

From Classic to Modern: Exploring Layered Data Platform Architectures

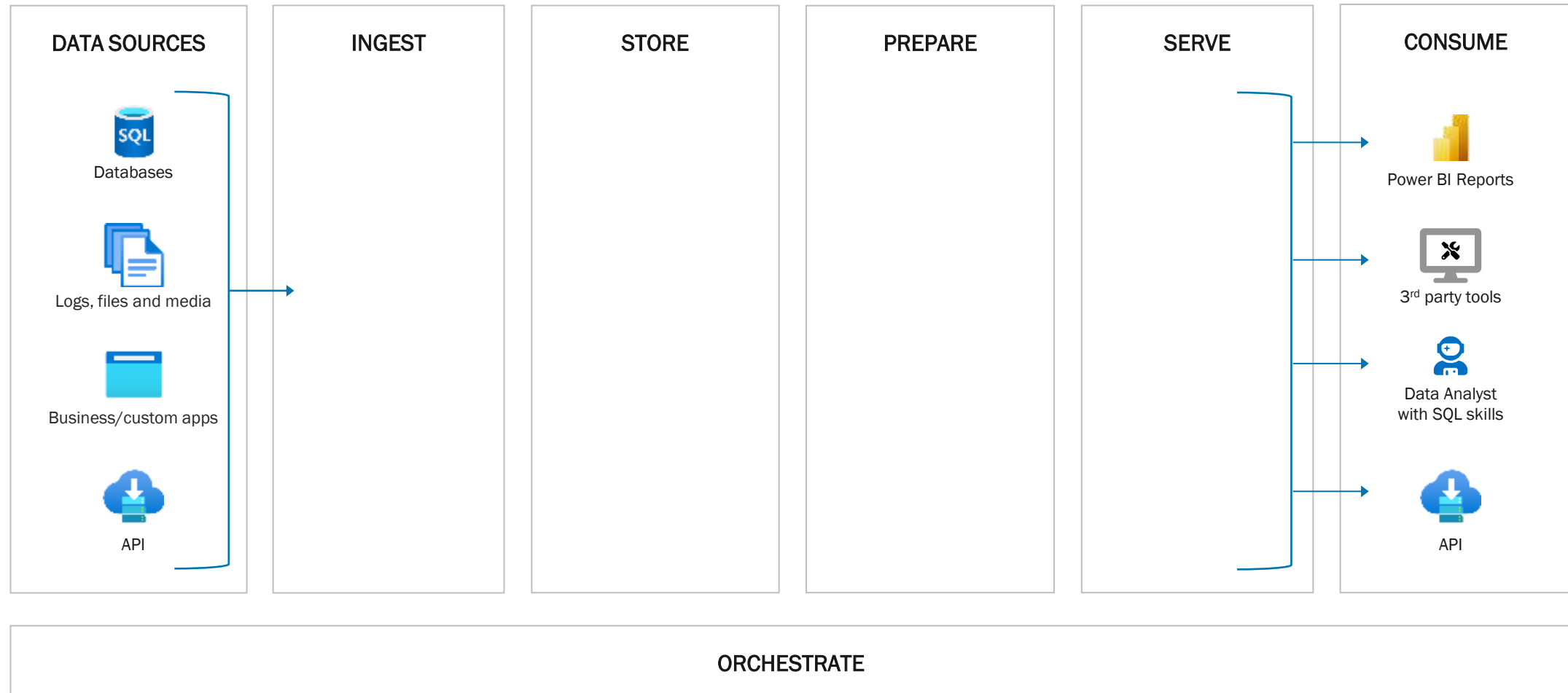
Dataflow in a Classic Data Warehouse



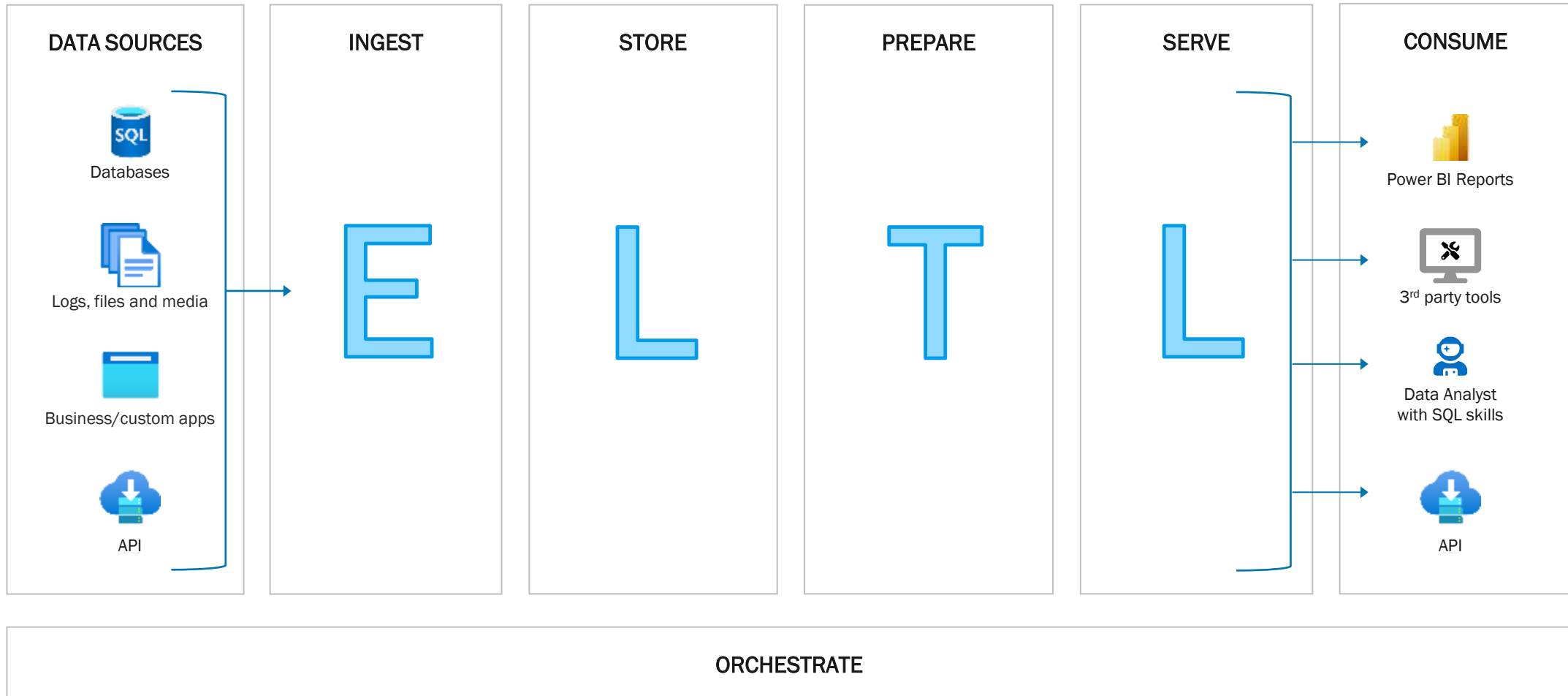
Dataflow in a Modern Data Warehouse



Dataflow in a Modern Data Warehouse



Dataflow in a Modern Data Warehouse

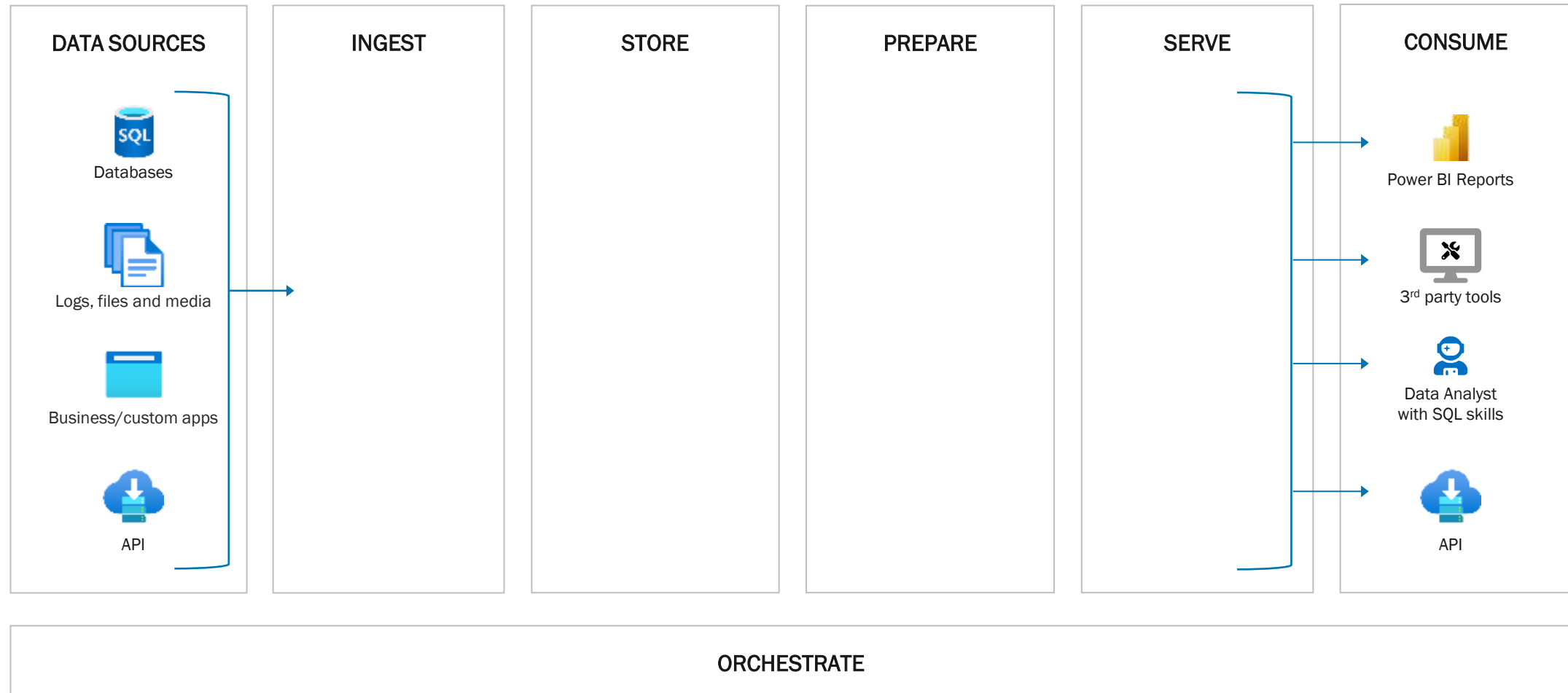


The background is dark with four glowing, semi-transparent shapes: a green circle on the left, a blue vertical pill shape in the center, a red triangle pointing right in the upper right, and a yellow irregular shape in the lower right.

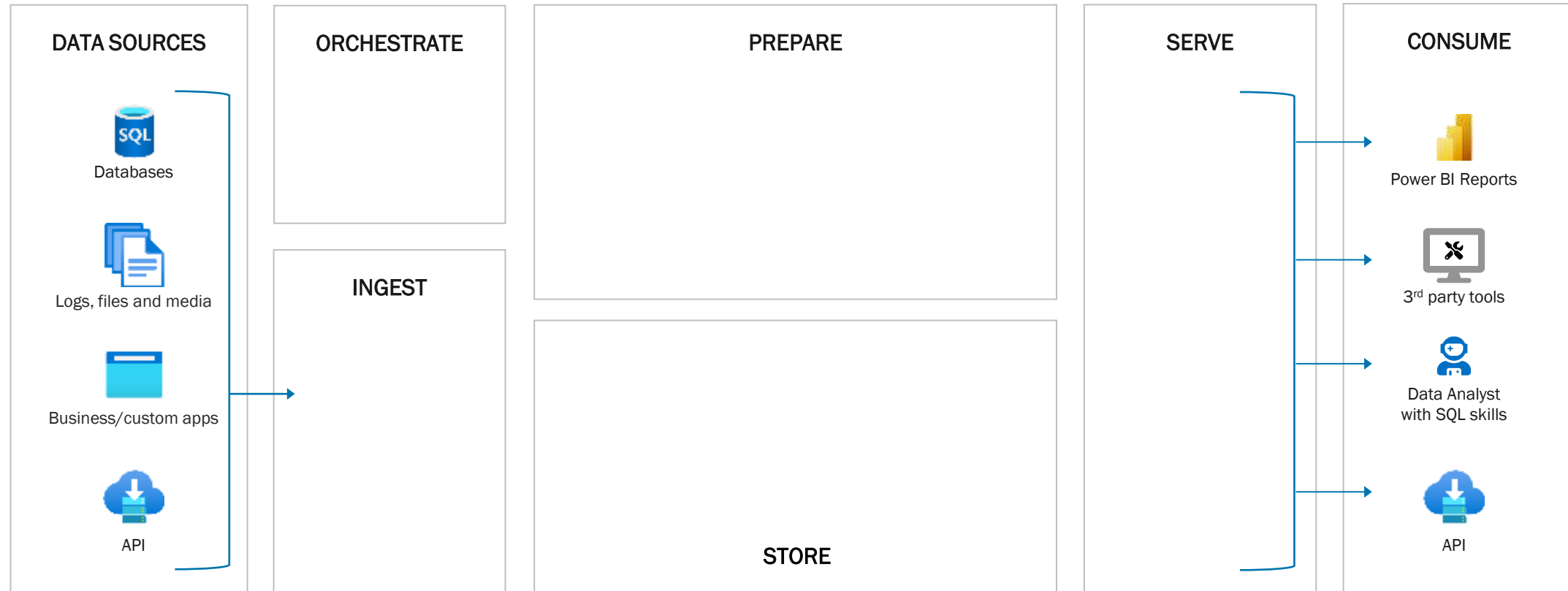
Unleashing the Power of Decoupled Storage and Compute

*two*day

Dataflow in a Modern Data Warehouse



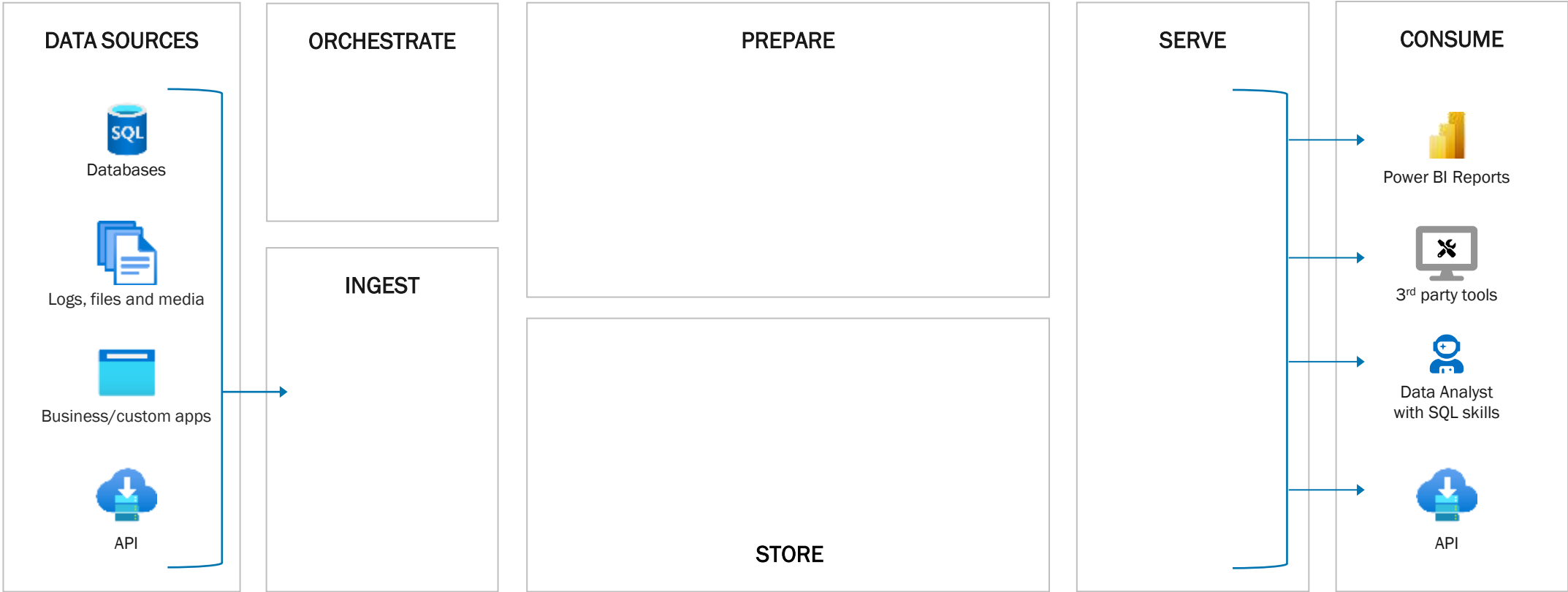
Dataflow in a Data Lakehouse architecture



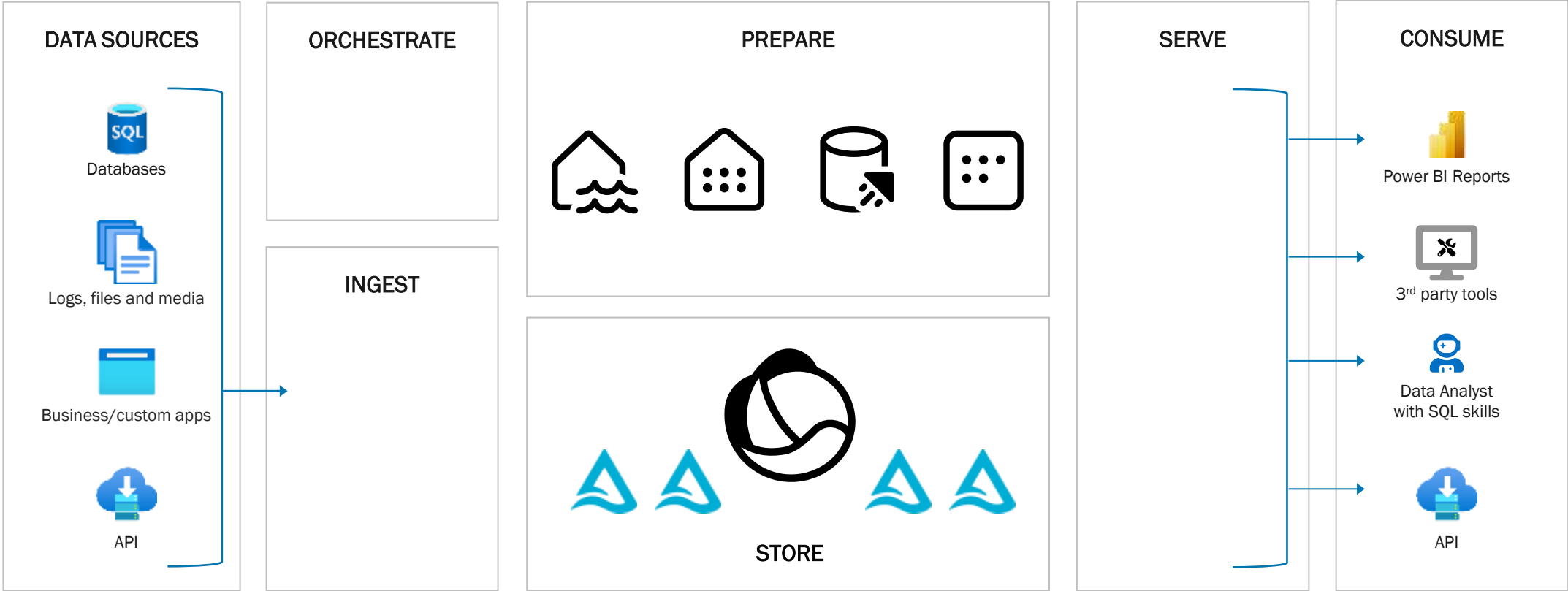
Fabric as an end-to-end unified analytical platform



Separation of Compute and Storage



Separation of Compute and Storage



Separation of Compute and Storage

1. Storage is Cheap, Compute is Expensive

- Decouple to scale compute only when needed, reducing costs by avoiding over-provisioning.

2. Scalability Across Multiple Compute Engines

- Handle big data processing with Spark, analyze real-time events with Eventhouse, and scale each engine independently for maximum efficiency.

3. Tailor Compute Resources to Workloads

- Optimize performance by assigning the right compute engine—Spark, Data Warehouse, Eventhouse, or Analysis Services—to the appropriate tasks.

4. Always-On Storage with OneLake

- Keep your data accessible even when compute is inactive, ensuring persistent storage with minimal downtime.

5. Run Compute Where it Makes Sense

- Leverage compute engines in different environments while keeping your data centralized in OneLake, enabling seamless cross-cloud operations.

Delta Lake: The Magic Behind OneLake

- **The Backbone of Fabric's Lakehouse Architecture:** Delta Lake is the open-source storage layer that powers the Lakehouse, seamlessly integrating with engines like Spark for ultimate flexibility.
- **Default Storage Format for All Workloads in Fabric:** No matter what you're doing—streaming, batch processing, or analytics—Delta Lake is your go-to format.
- **ACID Transactions & Scalable Metadata Handling:** Go beyond Parquet! Delta Lake adds a transaction log for ACID compliance and robust metadata management, making it perfect for massive data environments.
- **Time Travel Made Easy:** Effortlessly access and revert to previous data versions for audits, rollbacks, or historical analysis. Delta Lake brings time travel to your data!
- **Turn Files into Relational Tables:** Files are no longer static! With Delta Lake, your files behave like relational tables, allowing for seamless querying and data manipulation.
- **"It's Parquet, but Better!"**



Delta Lake: The Magic Behind OneLake



Name ↓

📁	_delta_log
📄	02403940

Name

📄	000000000000
📄	000000000000

```
{  
  "add": {  
    "path": "4374f6f8-f668-42ac-9ef9-fe85478de719.parquet",  
    "partitionValues": {},  
    "size": 173993,  
    "modificationTime": 1696242397107,  
    "dataChange": true,  
    "tags": {}  
  }  
}  
}  
"commitInfo": {}  
}
```

twoday



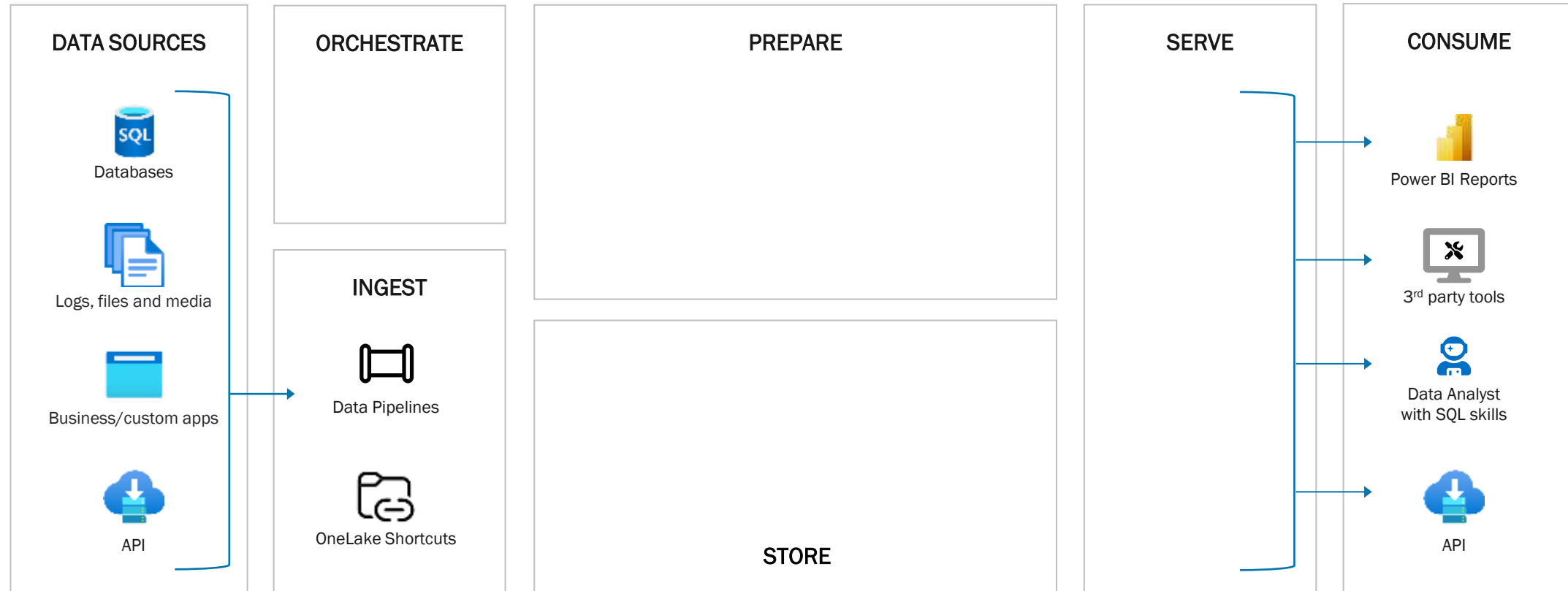
DEMO TIME

twoday

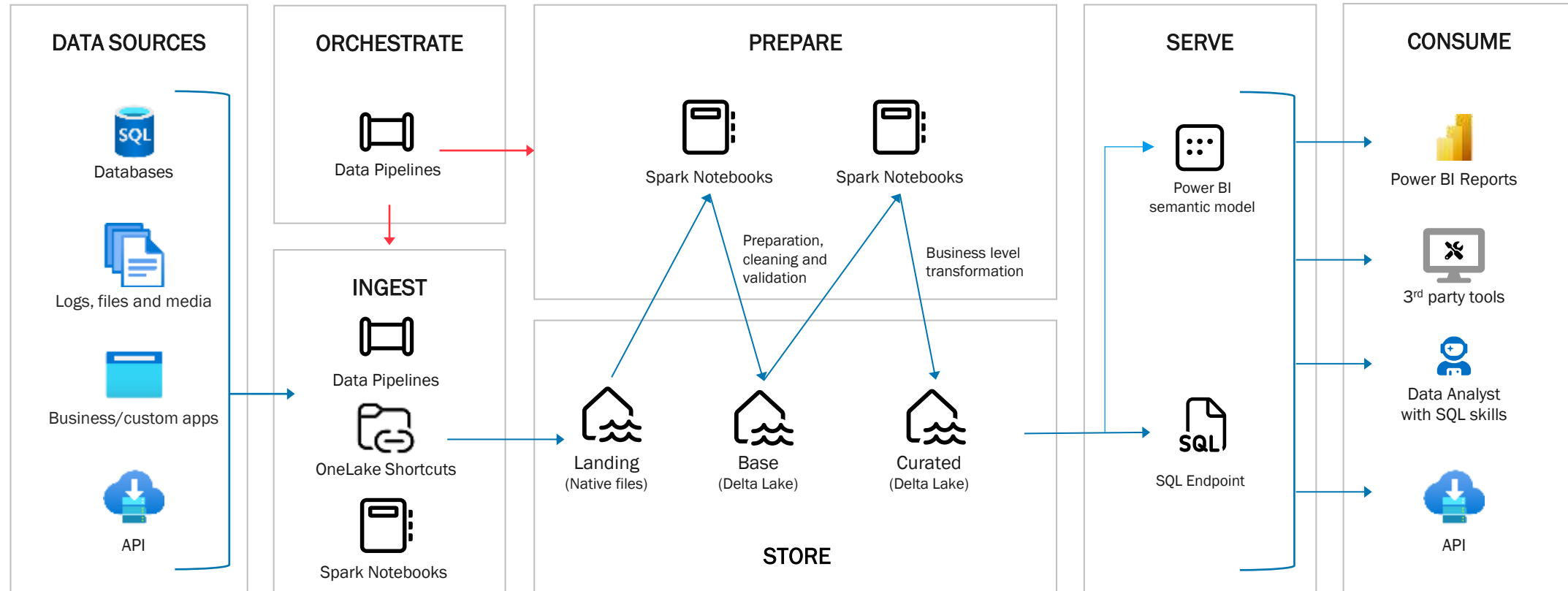
The background is dark with several glowing, semi-transparent shapes: a green circle on the left, a blue vertical pill shape in the center, a red triangle pointing right in the upper right, and a yellow irregular shape in the lower right.

When PySpark Steals the Show: Key Scenarios for Success

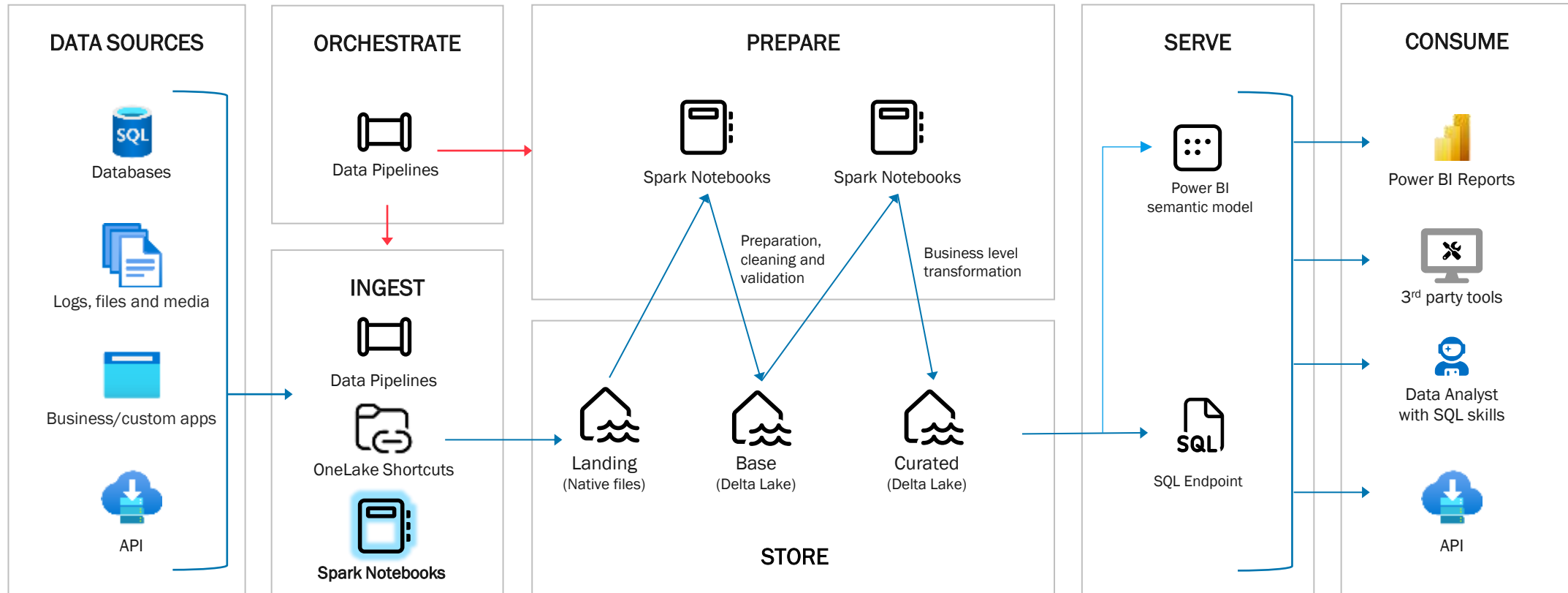
Dataflow in a Data Lakehouse Architecture



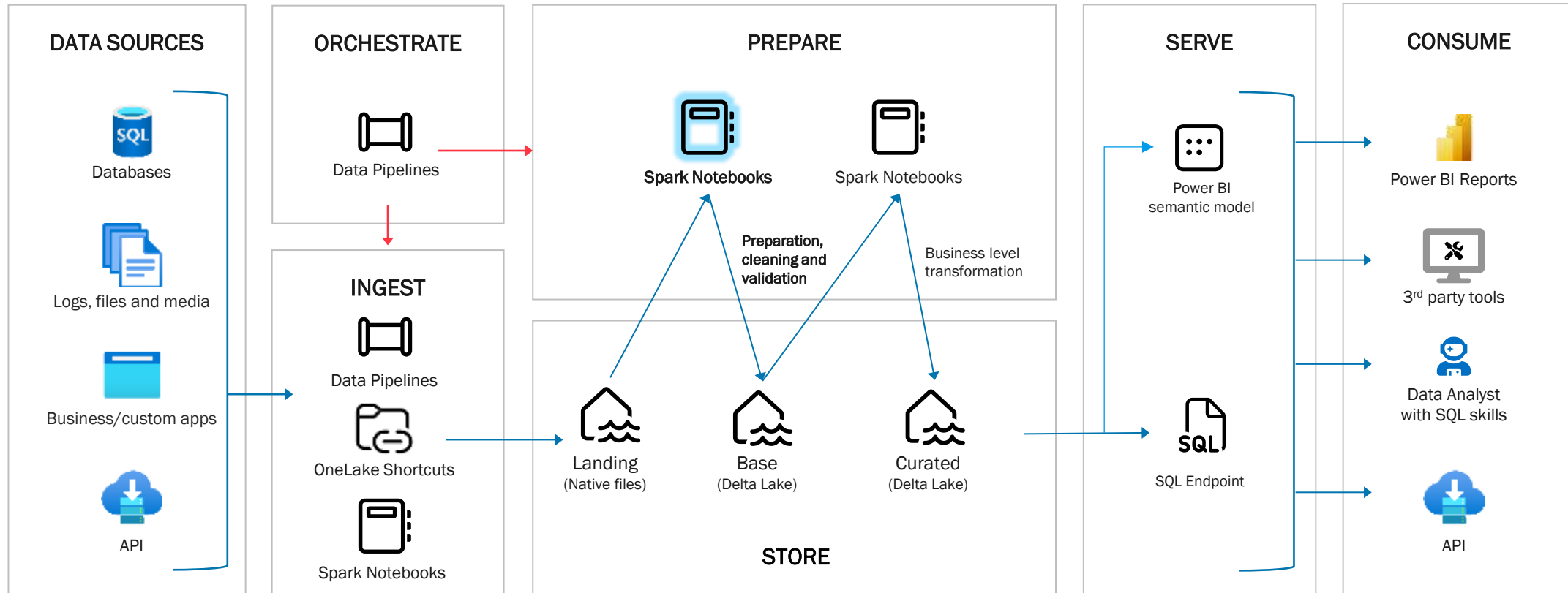
Dataflow in a Data Lakehouse Architecture



Scenario 1: Integration with Data Sources



Scenario 2: Automated Data Cleaning and Validation

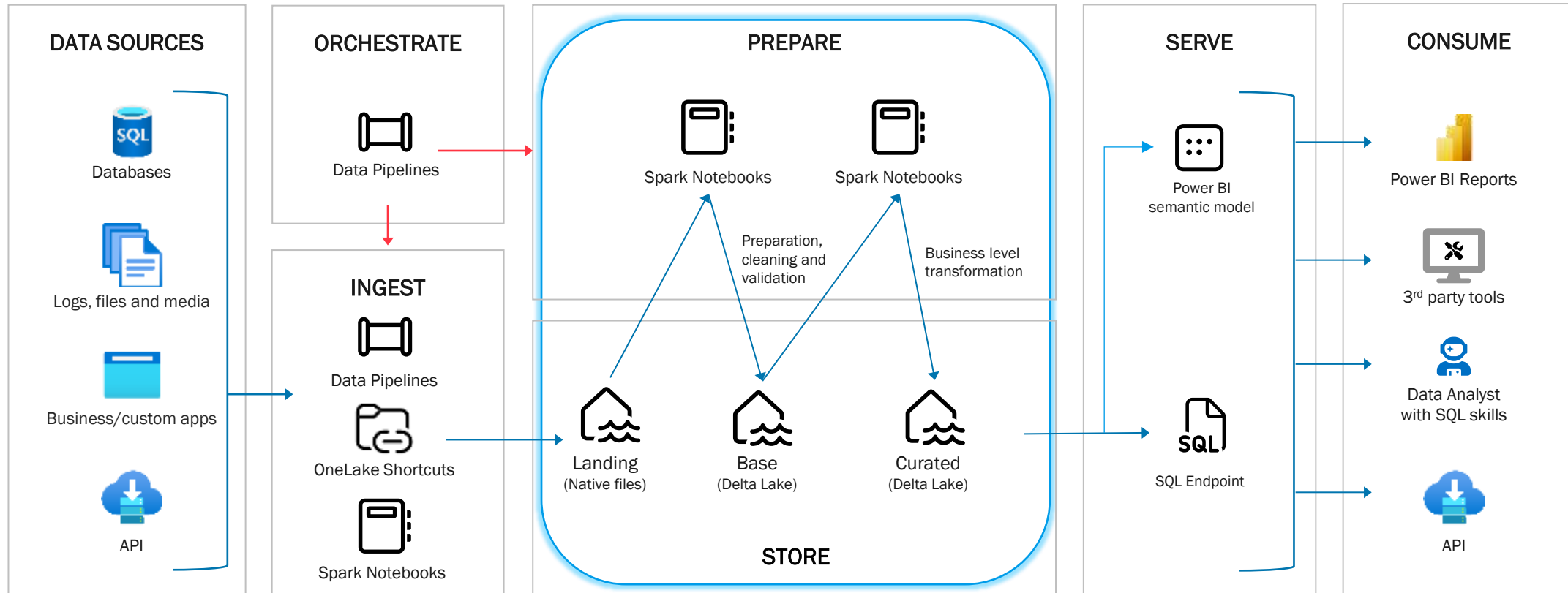




DEMO TIME

twoday

Scenario 3: Breaking Free from SQL Constraints





DEMO TIME

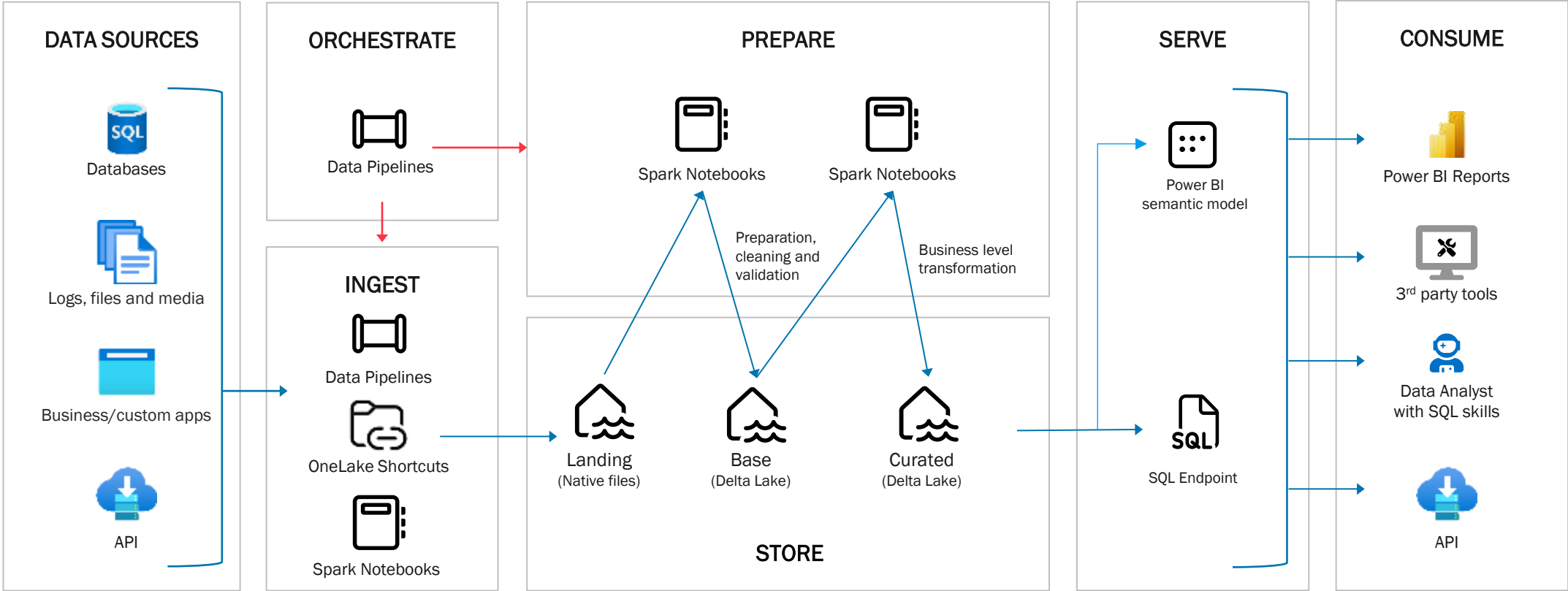
twoday



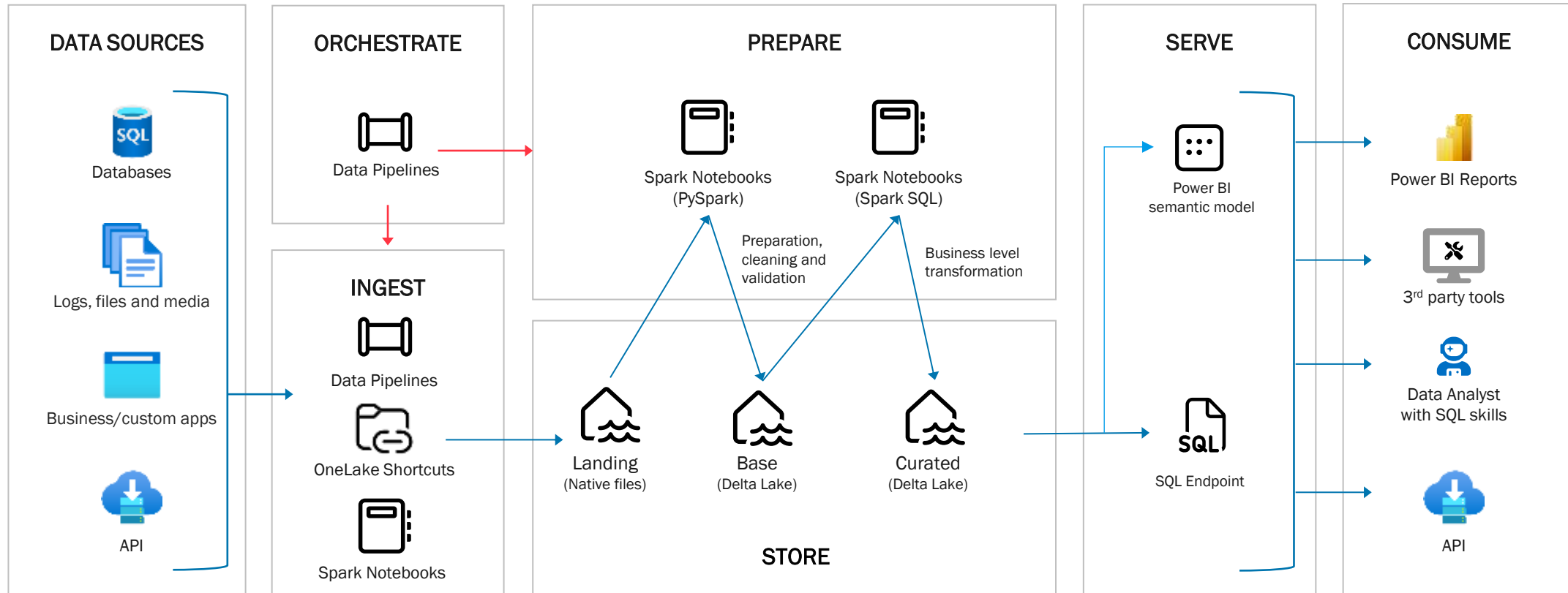
SQL is Still Cool: Why It's Here to Stay

twoday

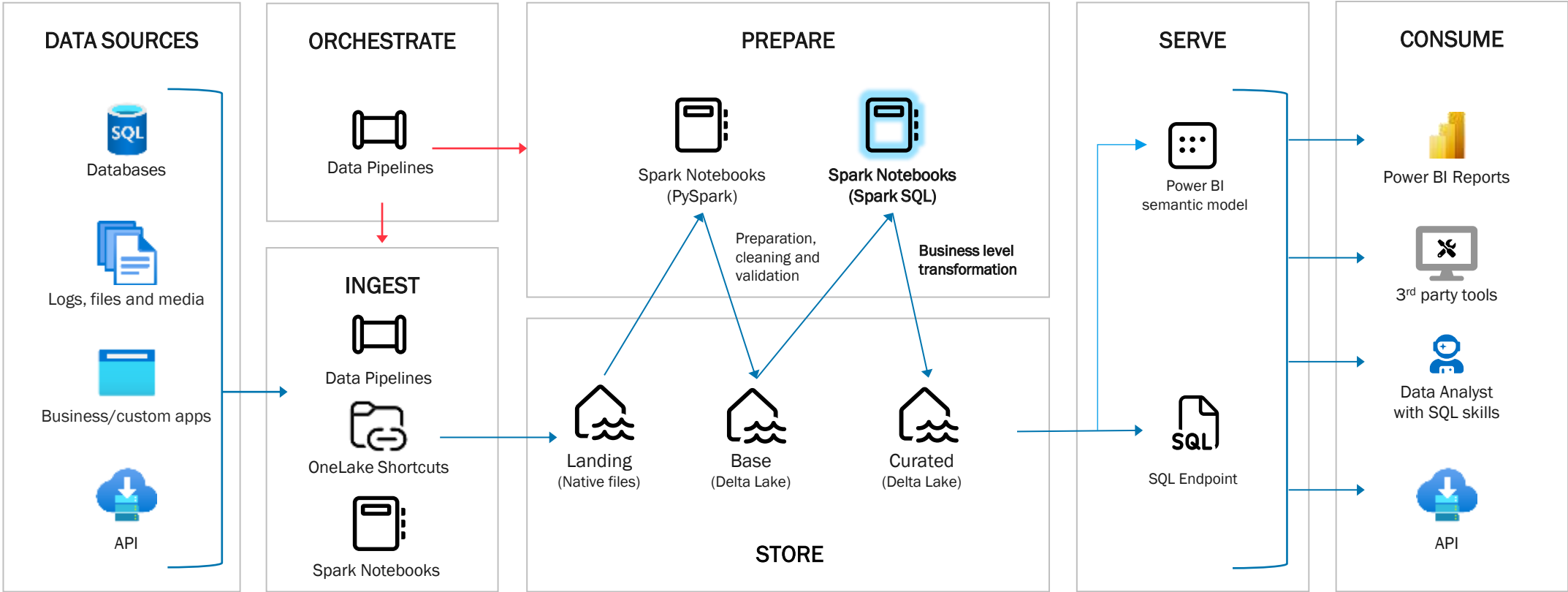
SQL's Cool Factor



SQL's Cool Factor



SQL's Cool Factor





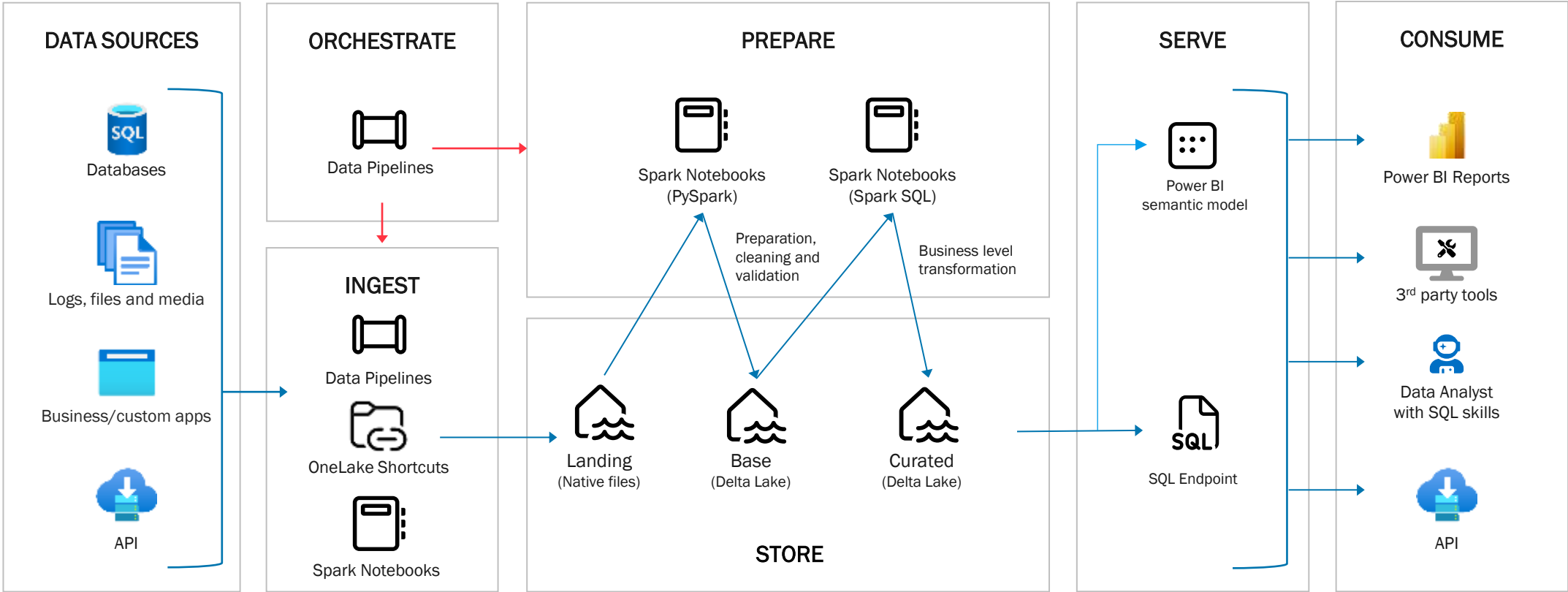
DEMO TIME

twoday

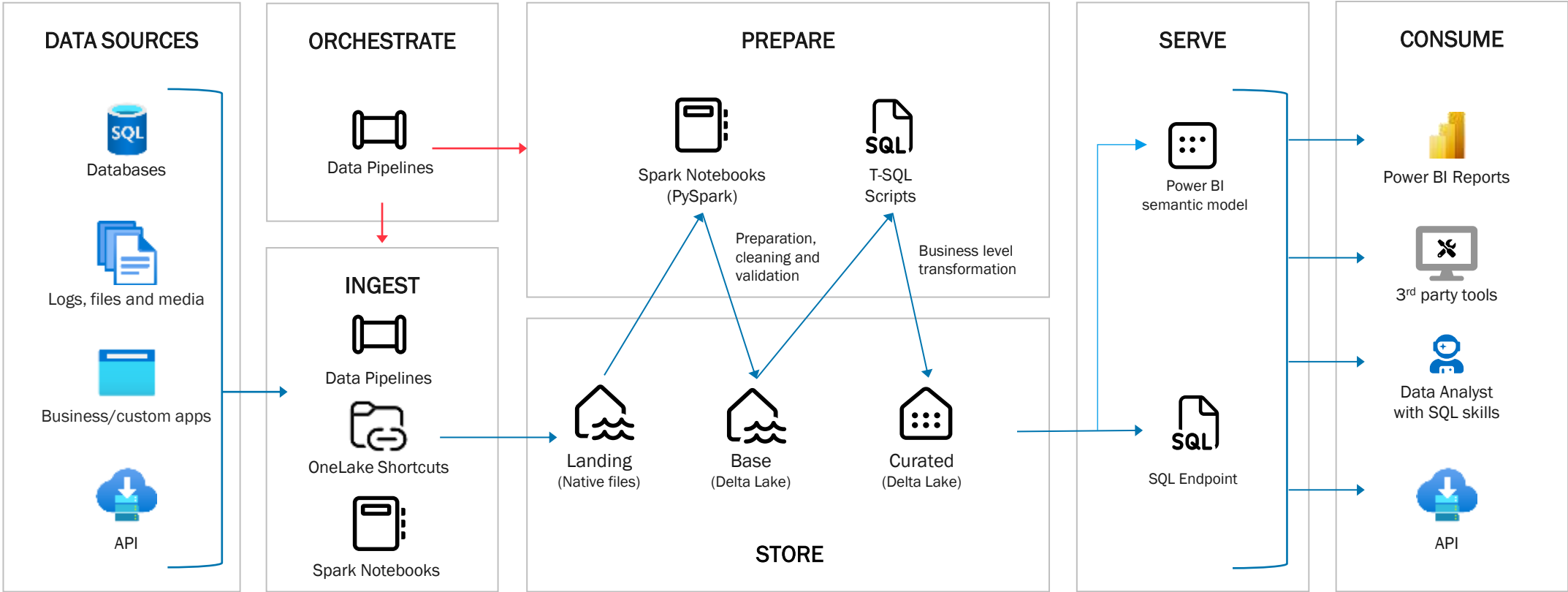
SQL's Cool Factor

- **Familiarity and Simplicity:** Leverage the power of a language you already know—SQL's declarative syntax is intuitive and widely adopted, making it easy for developers and analysts alike.
- **Seamless Transition to Modern Platforms:** SQL fits right into modern data ecosystems like Microsoft Fabric, allowing you to build on existing skills without the steep learning curve of new languages.
- **Powerful Performance with Spark SQL:** Execute SQL queries at scale with Spark SQL, merging the flexibility of SQL with the processing power of distributed systems like Spark.
- **Perfect for Business Logic:** SQL remains the gold standard for defining, refining, and executing business logic, ensuring clarity and precision in your data workflows.
- **Cross-Platform Portability:** SQL makes it simple to port business rules and queries across platforms, reducing the need for rework and ensuring consistency.
- **Integration with Modern Tools:** From Delta Tables to PySpark and Power BI, SQL plays well with today's most advanced tools, keeping it relevant and versatile.
- **SQL: The Glue for Data and Analytics:** SQL continues to bridge the gap between raw data and actionable insights, making it the connective tissue in any data architecture.

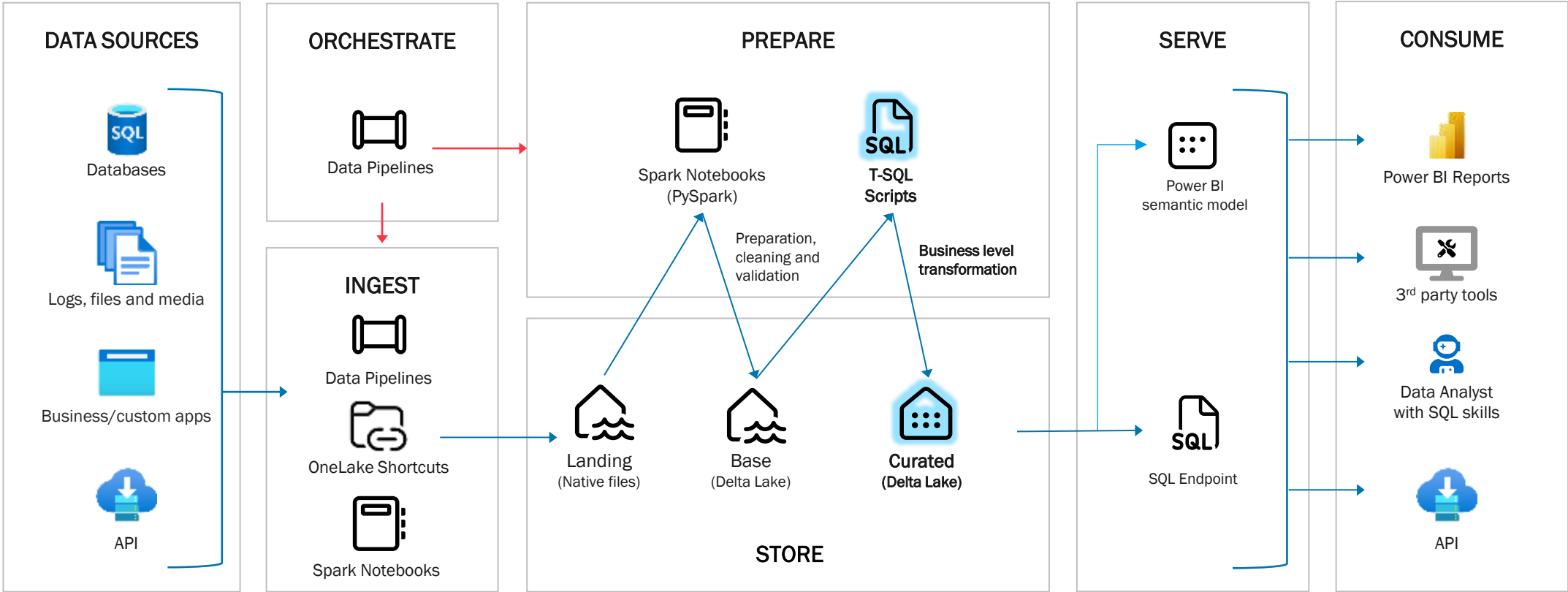
SQL is Still Cool



SQL is Still Cool



SQL is Still Cool





DEMO TIME

twoday

Q&A

twoday



twoday

Sponsors



 **Gold** 



SILVER 



BRONZE 



twoday