



# EUROPEAN **MiCROSOFT** **FABRiC**

Community Conference

STOCKHOLM 24-27 SEPTEMBER 2024

JOIN THE CONVERSATION

#FABCONEUROPE



# Demystifying the Data Lakehouse in Fabric

**Just Blindbæk**

Principal Architect and MVP, twoday, Denmark

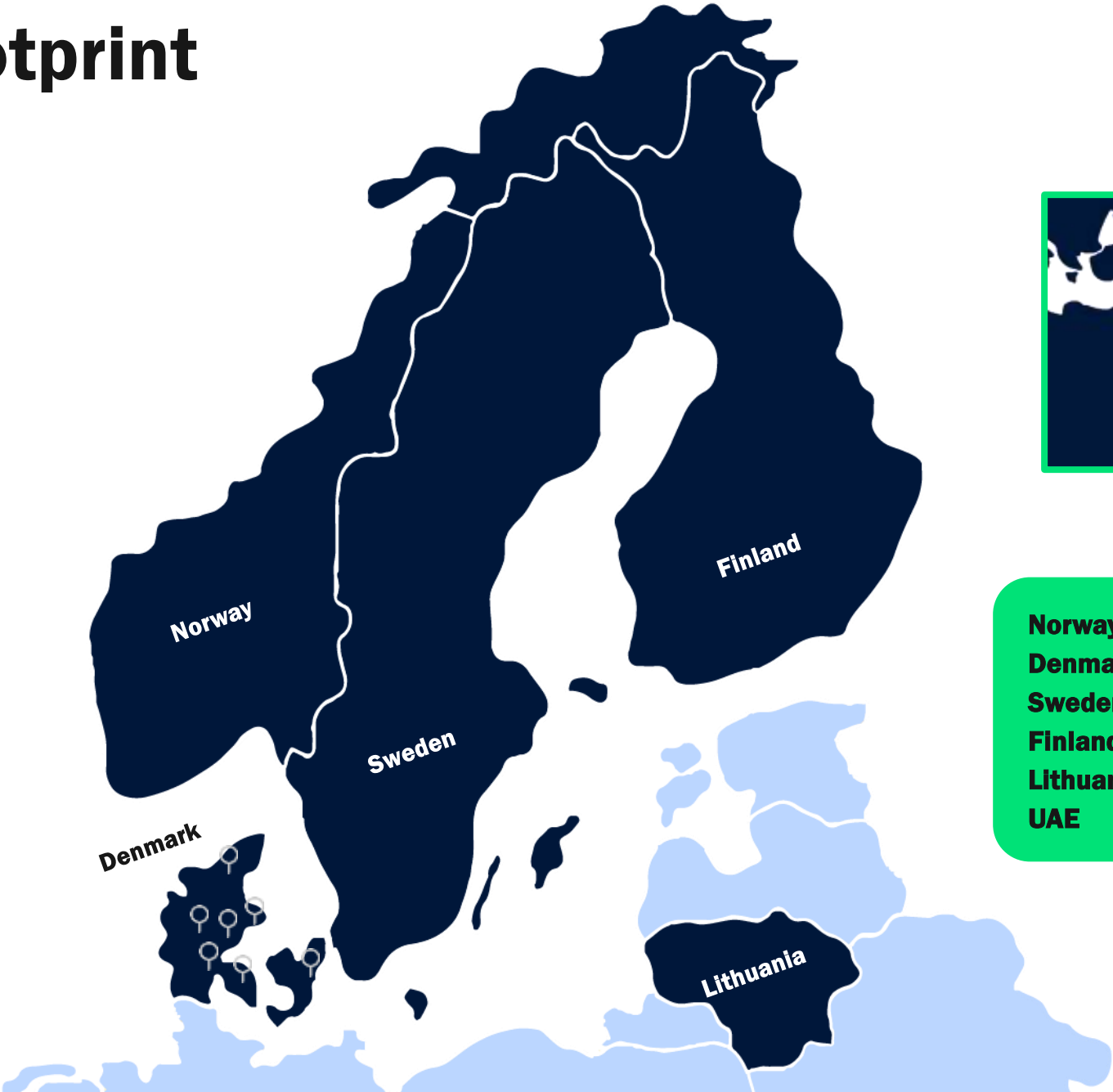
# twodays footprint

**3,000**  
Colleagues

**6**  
Countries

**35**  
Offices

**40 %**  
Co-owners



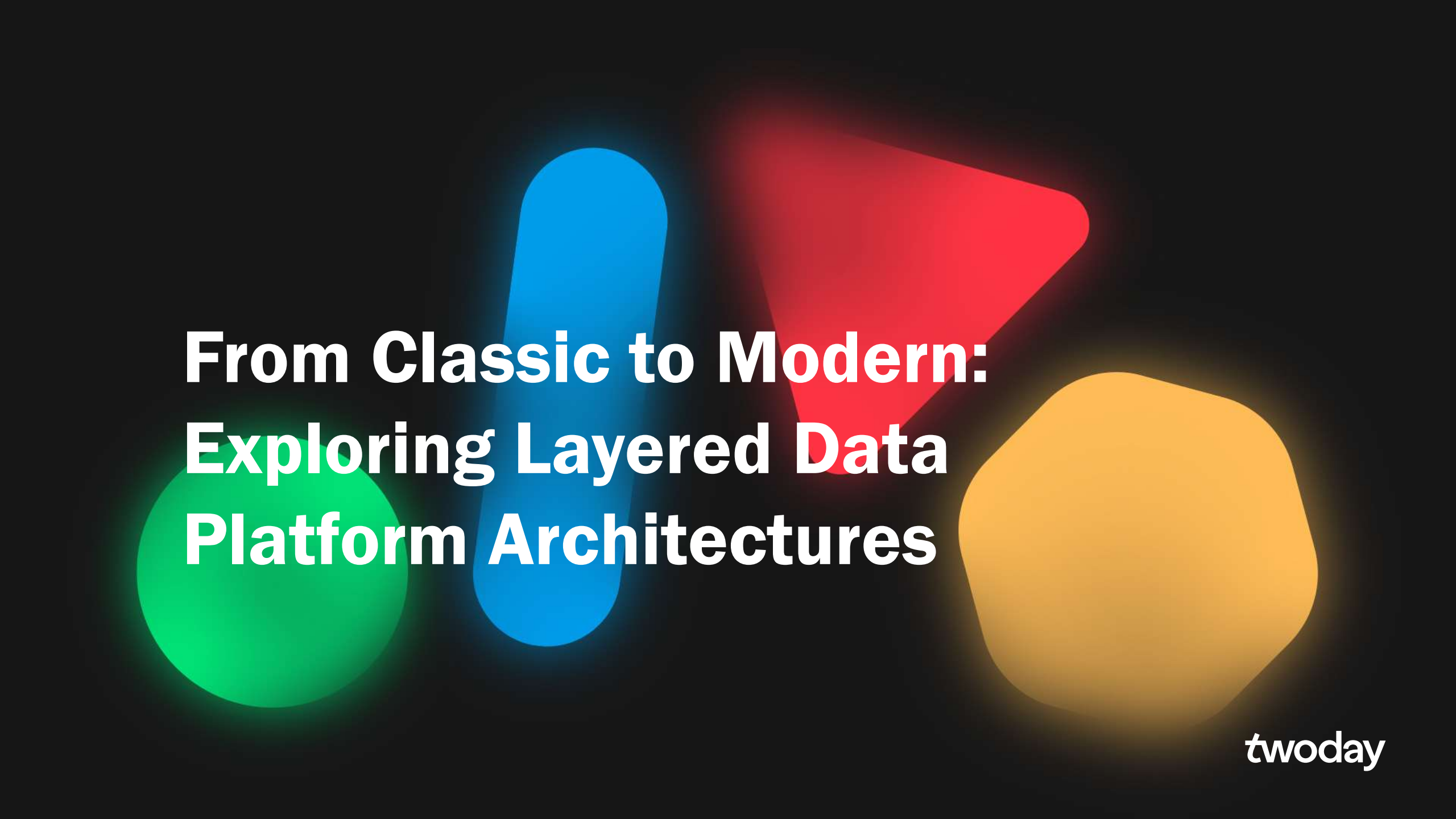
Norway	500 colleagues
Denmark	1,100 colleagues
Sweden	500 colleagues
Finland	700 colleagues
Lithuania	200 colleagues
UAE	15 colleagues

## 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?

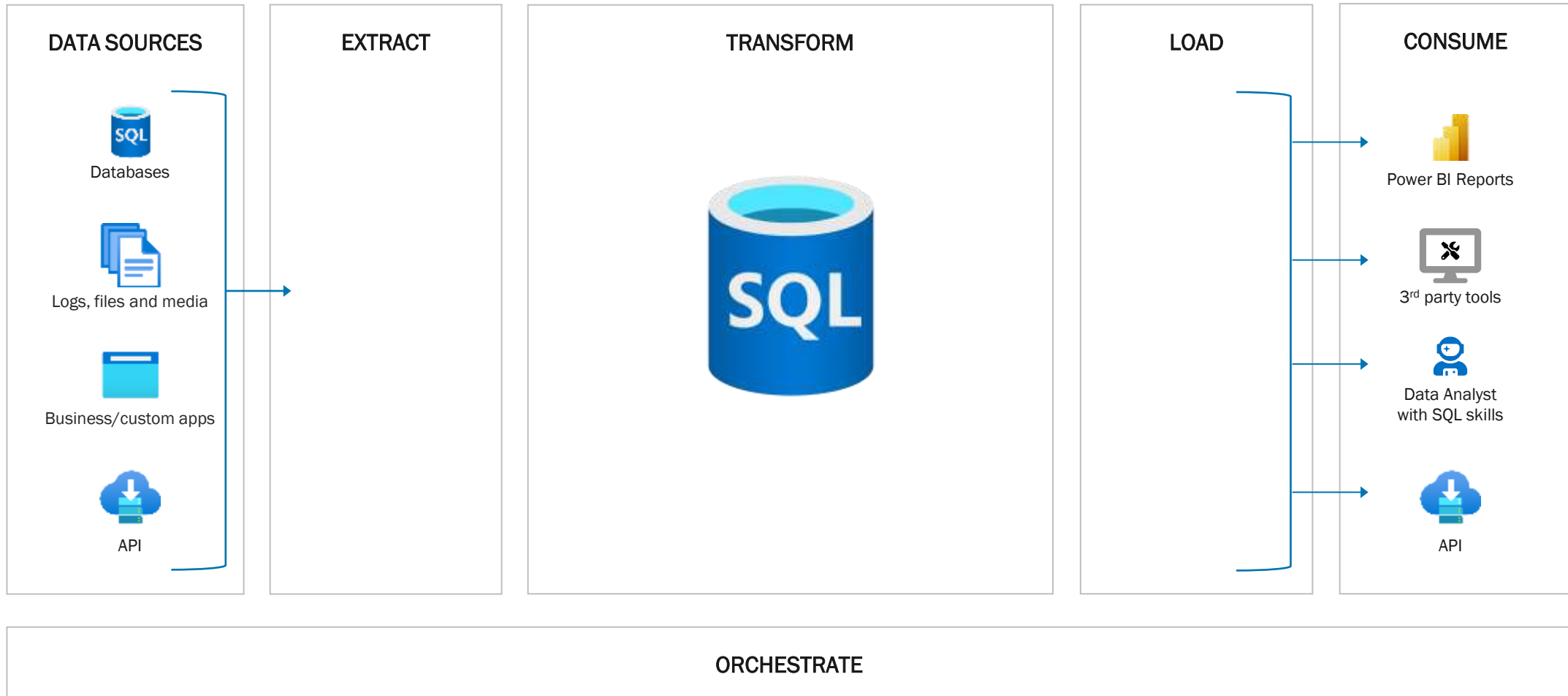




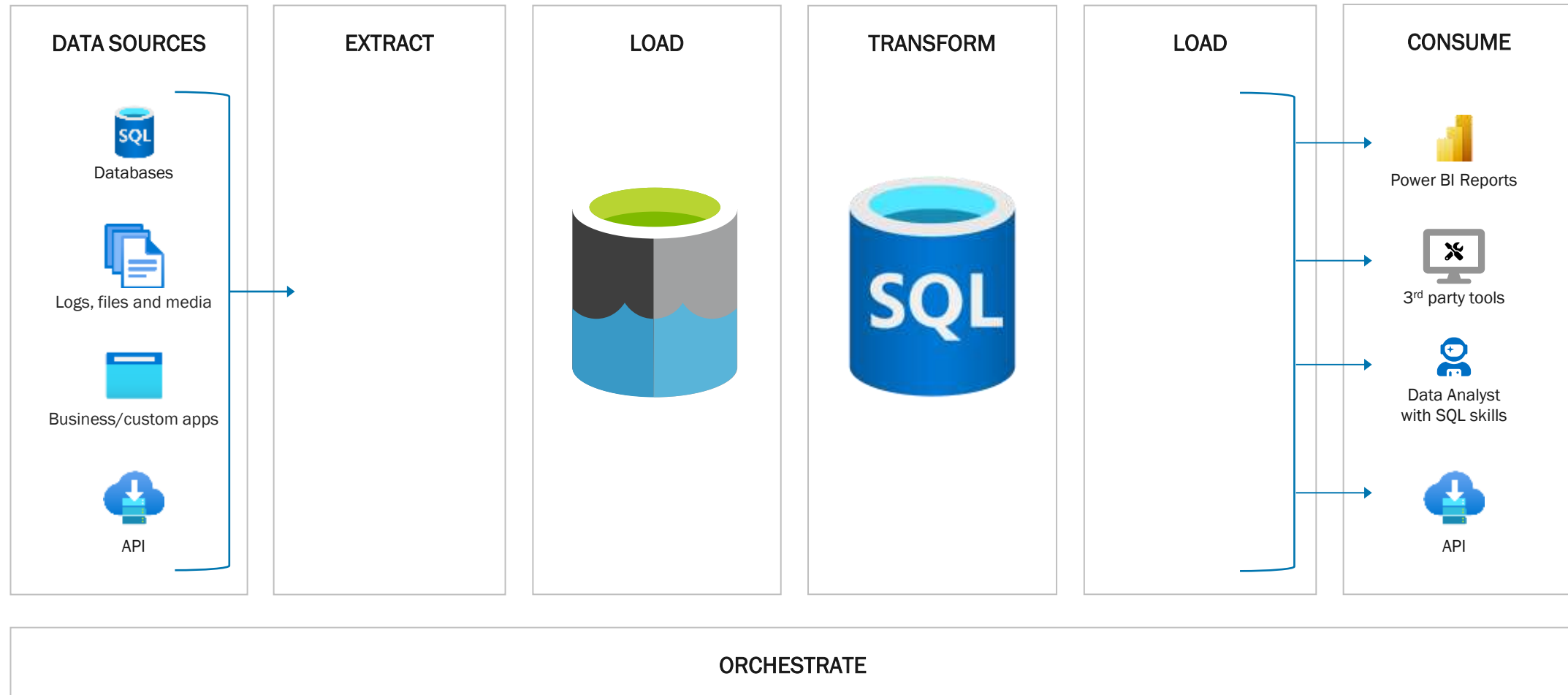
The background is dark with several glowing, semi-transparent shapes: a blue vertical pill shape, a red triangle pointing right, a green circle, and a yellow irregular blob on the right.

# **From Classic to Modern: Exploring Layered Data Platform Architectures**

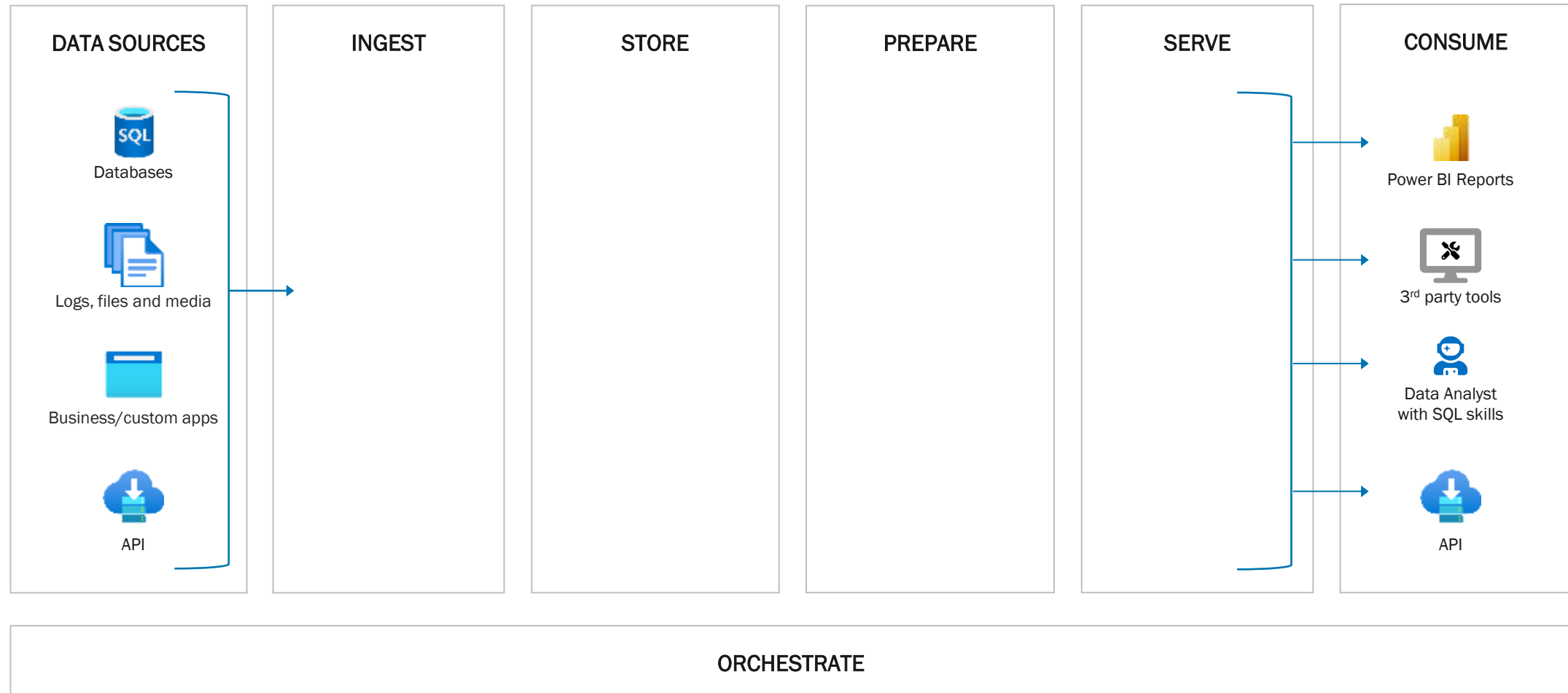
# Dataflow in a Classic Data Warehouse



# Dataflow in a Modern Data Warehouse

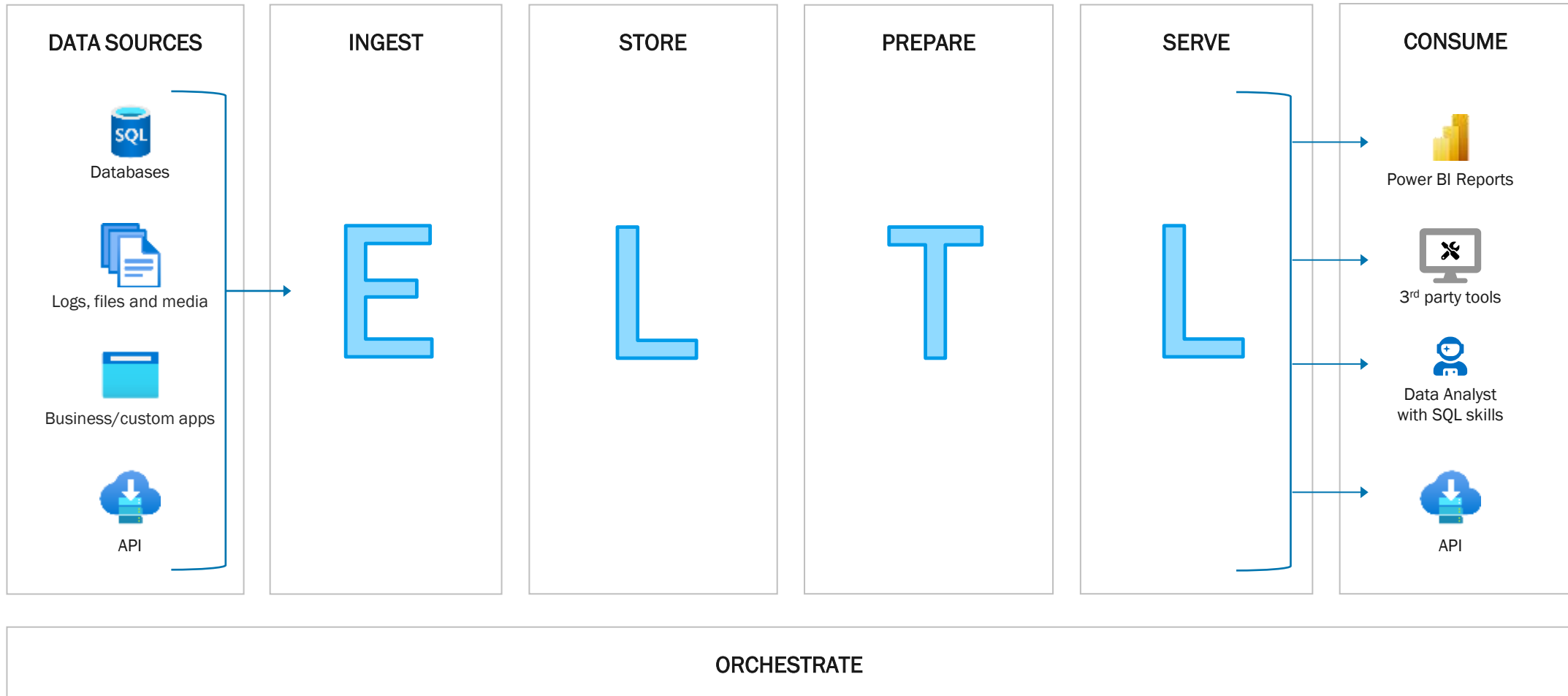


# Dataflow in a Modern Data Warehouse





# Dataflow in a Modern Data Warehouse

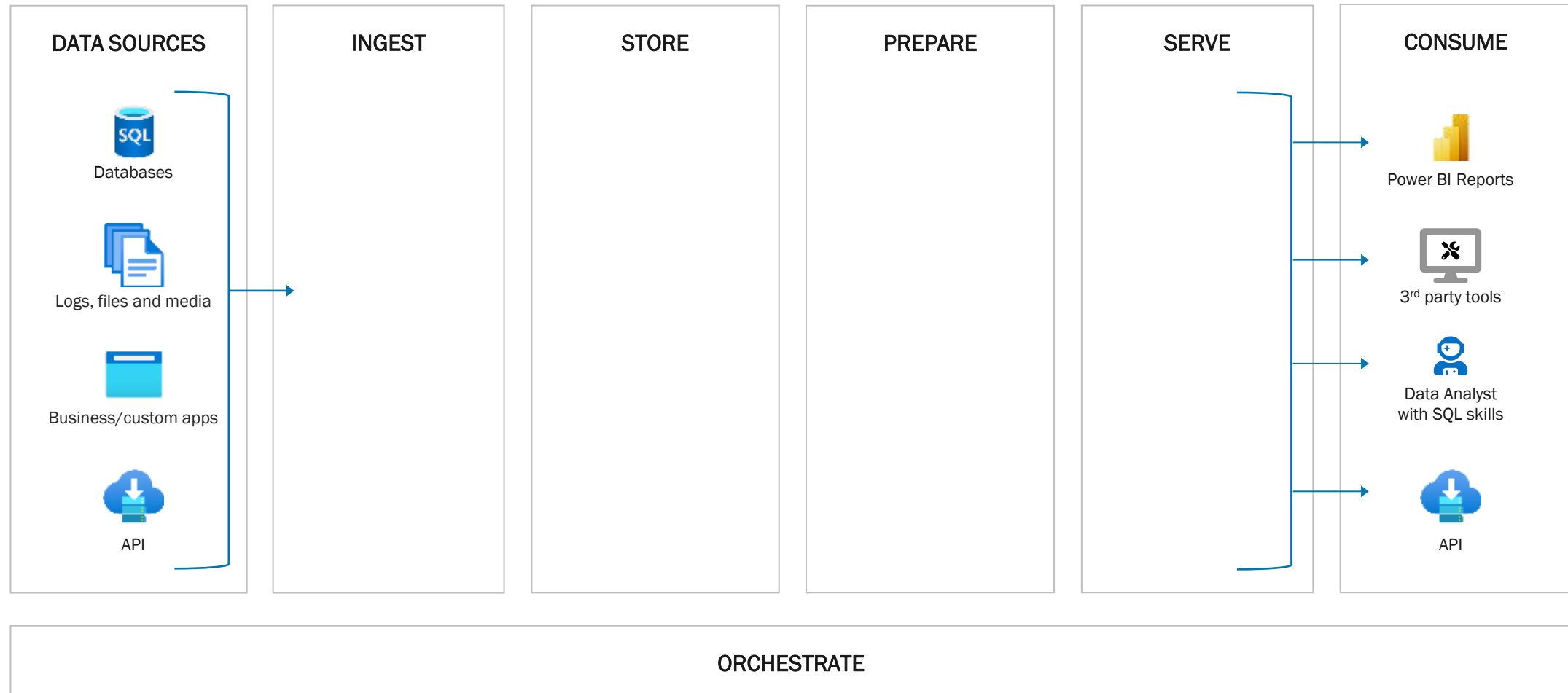


The background is dark with four glowing, semi-transparent shapes: a green circle on the left, a blue vertical pill shape in the center, a red triangle pointing right in the upper right, and a yellow irregular shape in the lower right.

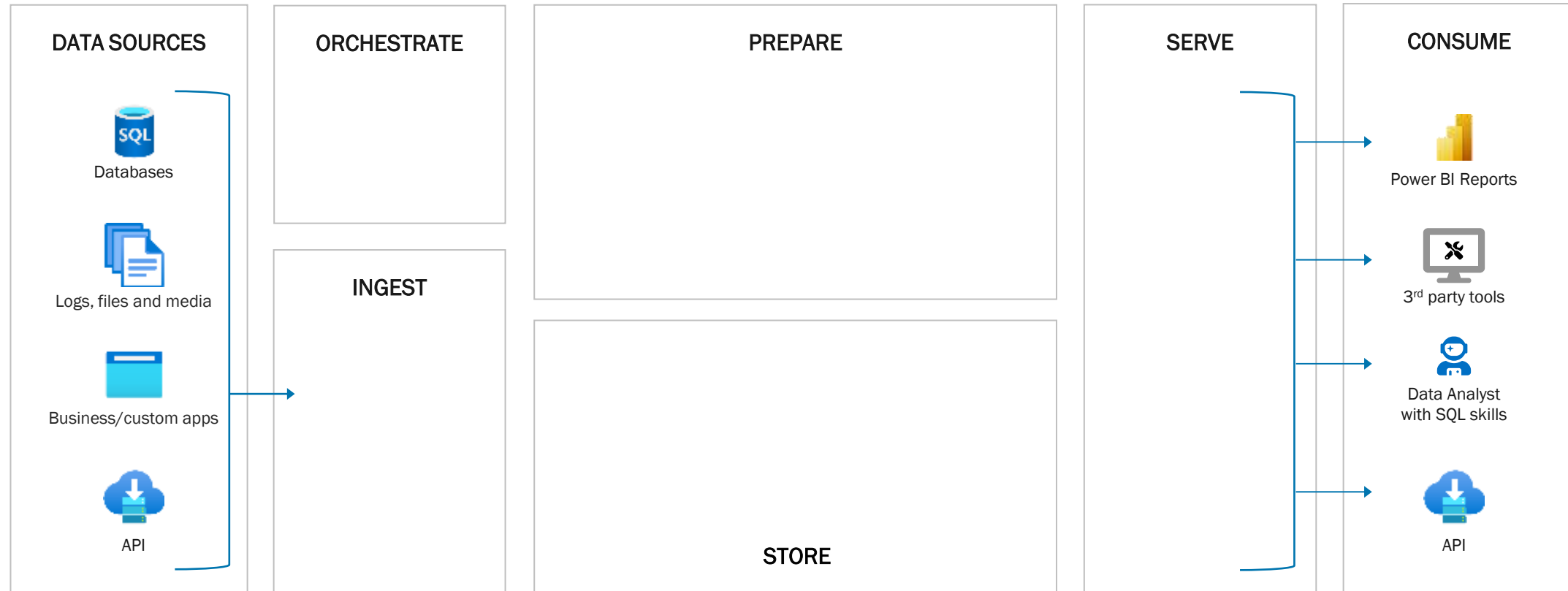
# Unleashing the Power of Decoupled Storage and Compute

*two*day

# Dataflow in a Modern Data Warehouse



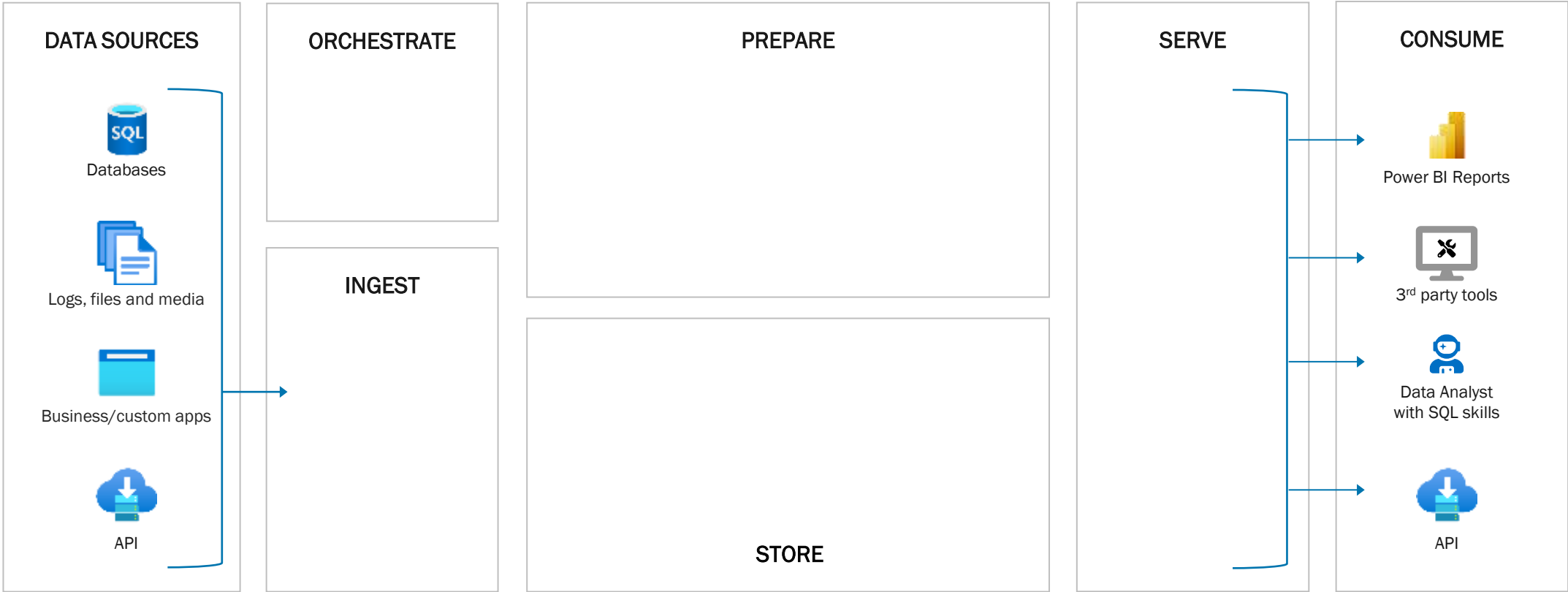
# Dataflow in a Data Lakehouse architecture



# Fabric as an end-to-end unified analytical platform

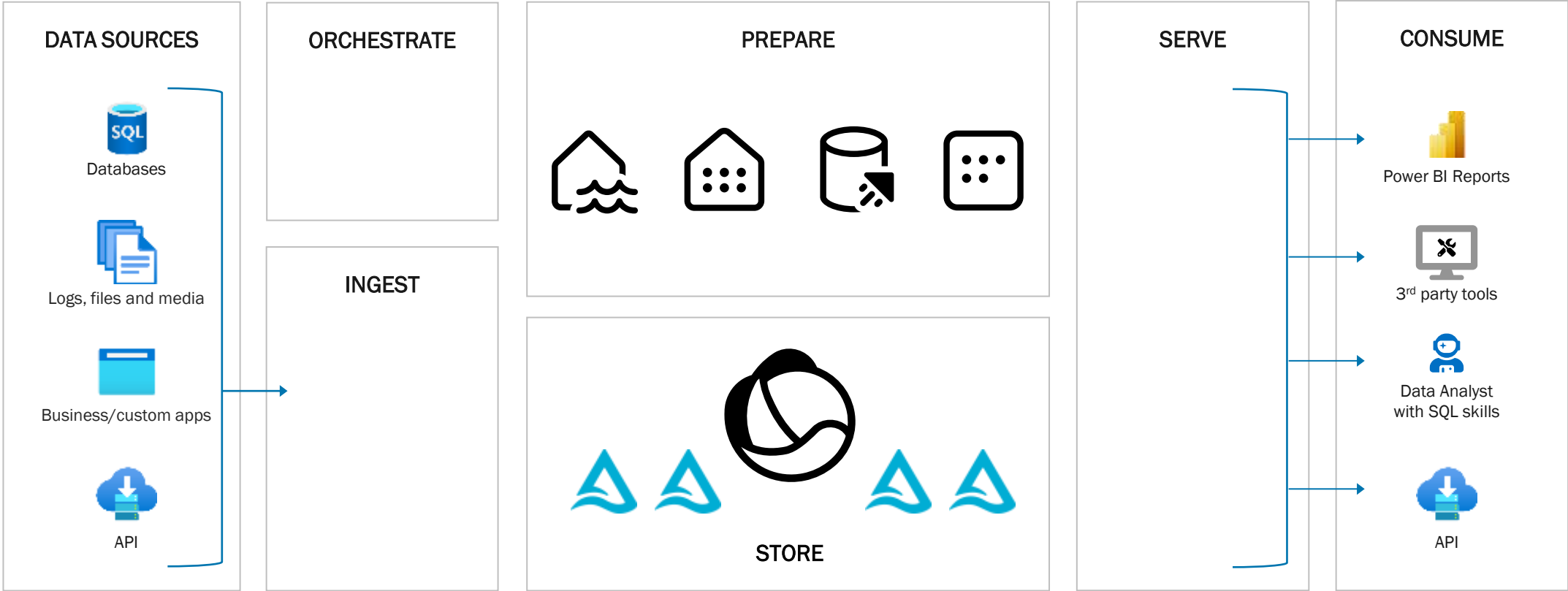


# Separation of Compute and Storage





# 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**

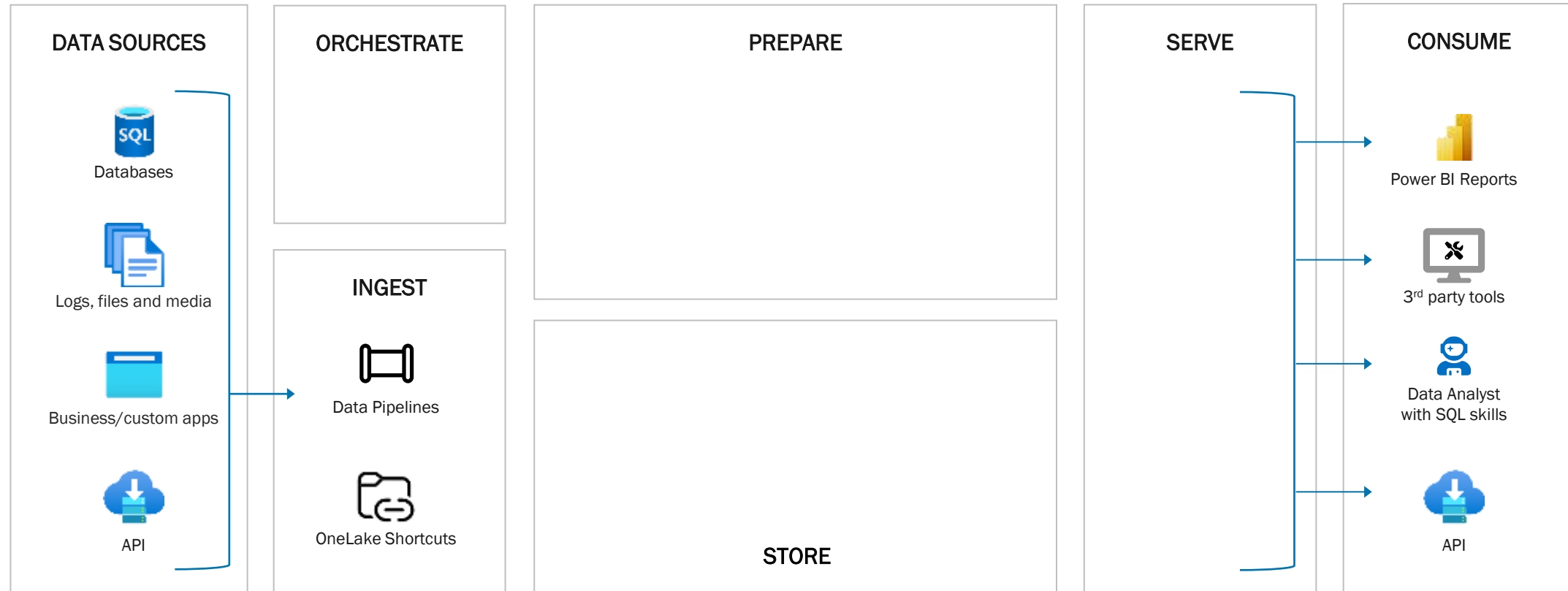
*twoday*

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 on the bottom right.

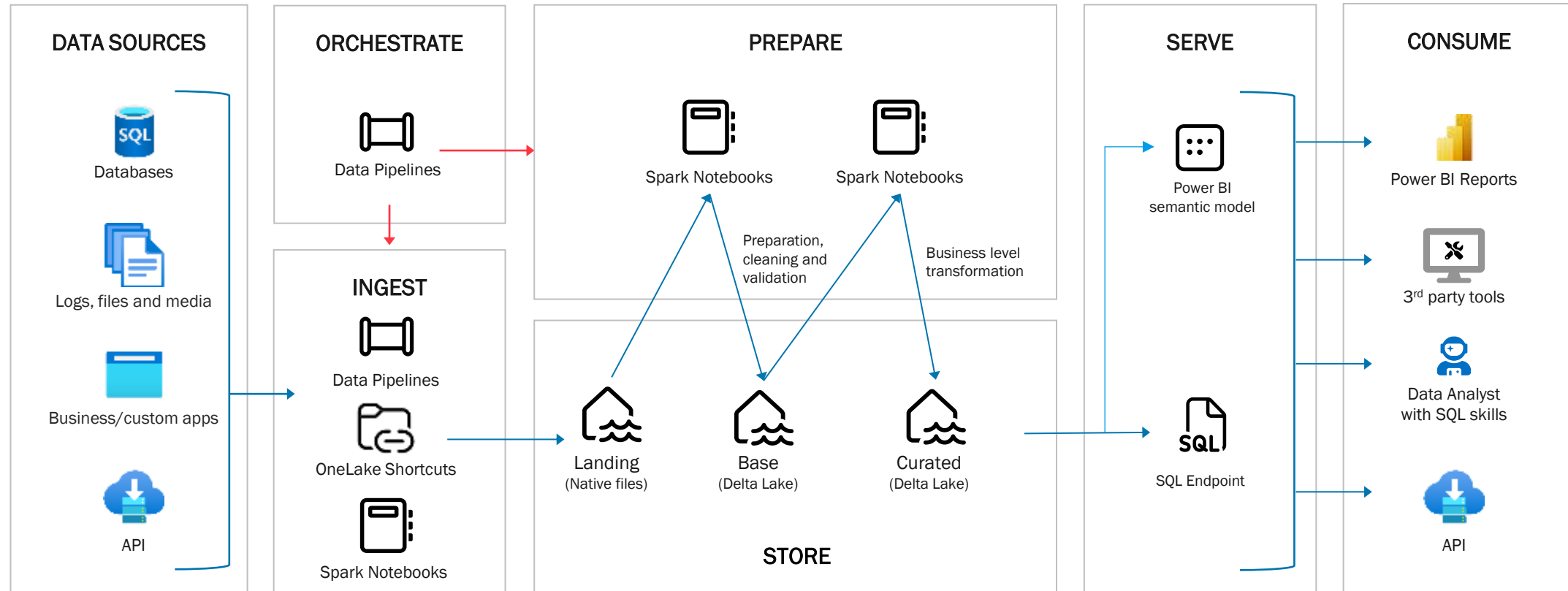
# **When PySpark Steals the Show: Key Scenarios for Success**



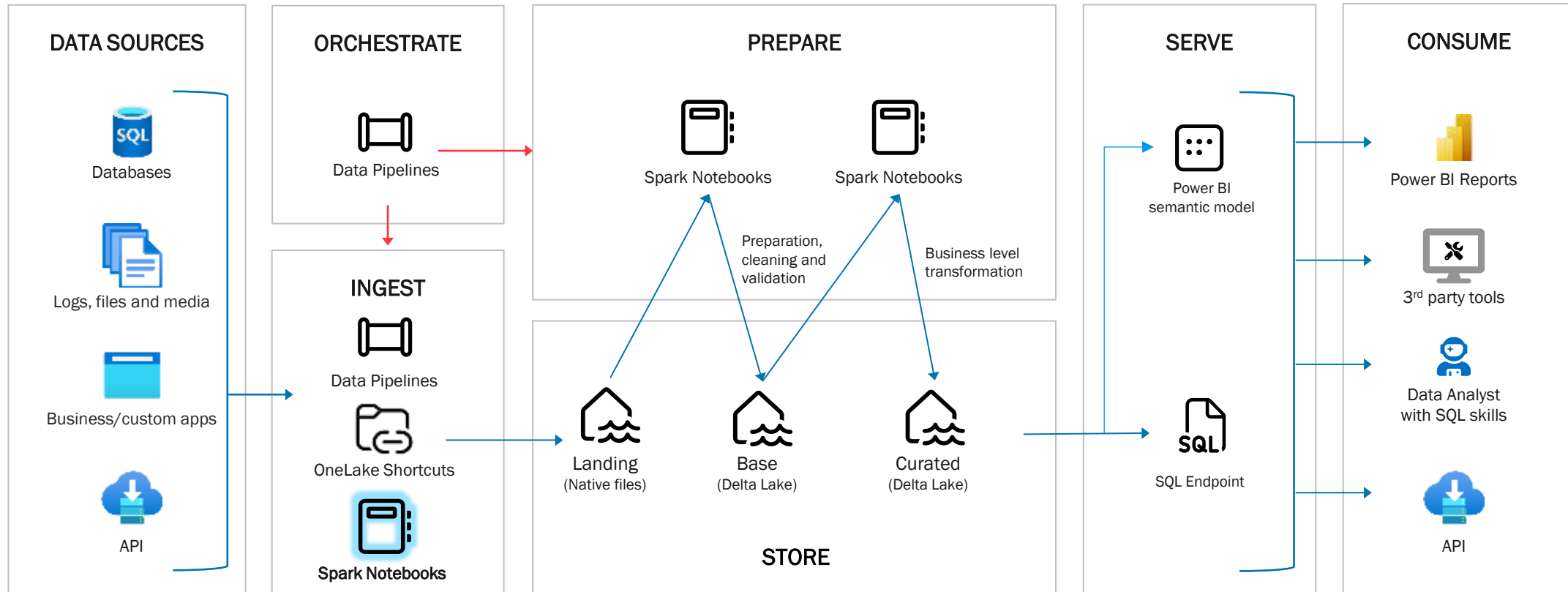
# Dataflow in a Data Lakehouse Architecture



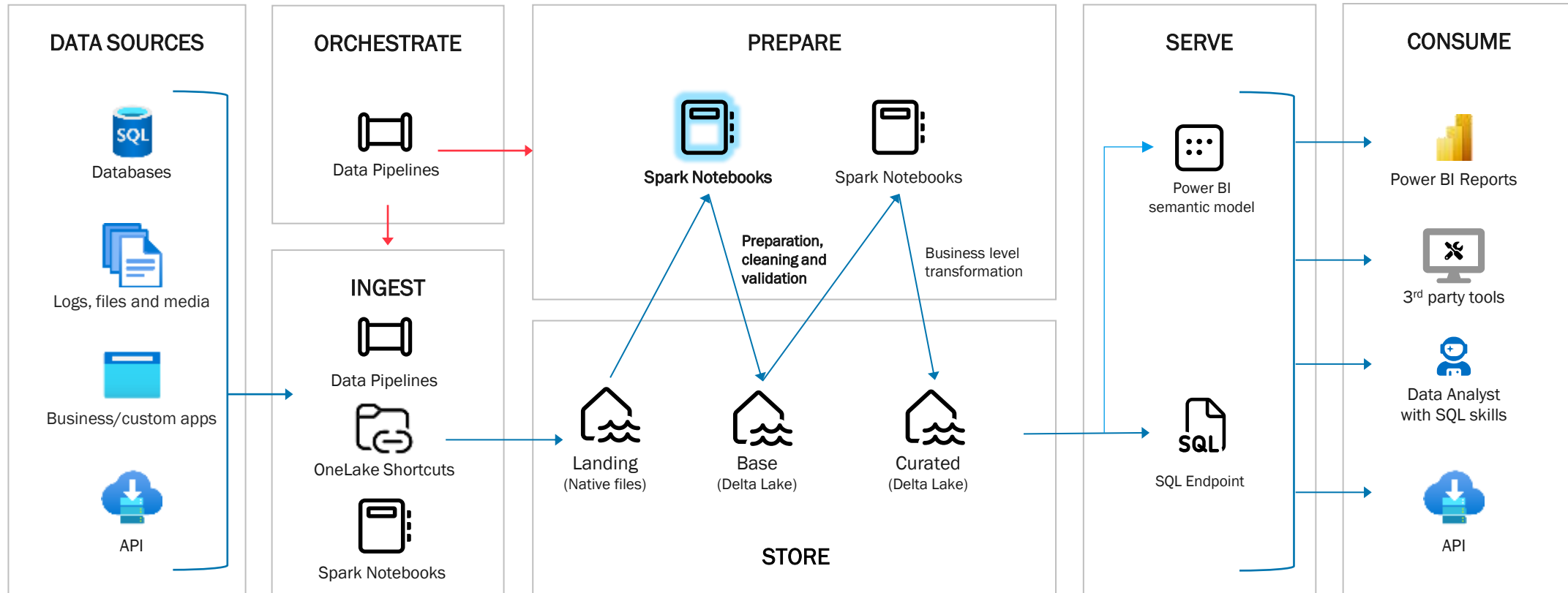
# Dataflow in a Data Lakehouse Architecture



# Scenario 1: Integration with API-Based Sources



# Scenario 2: Automated Data Cleaning and Validation

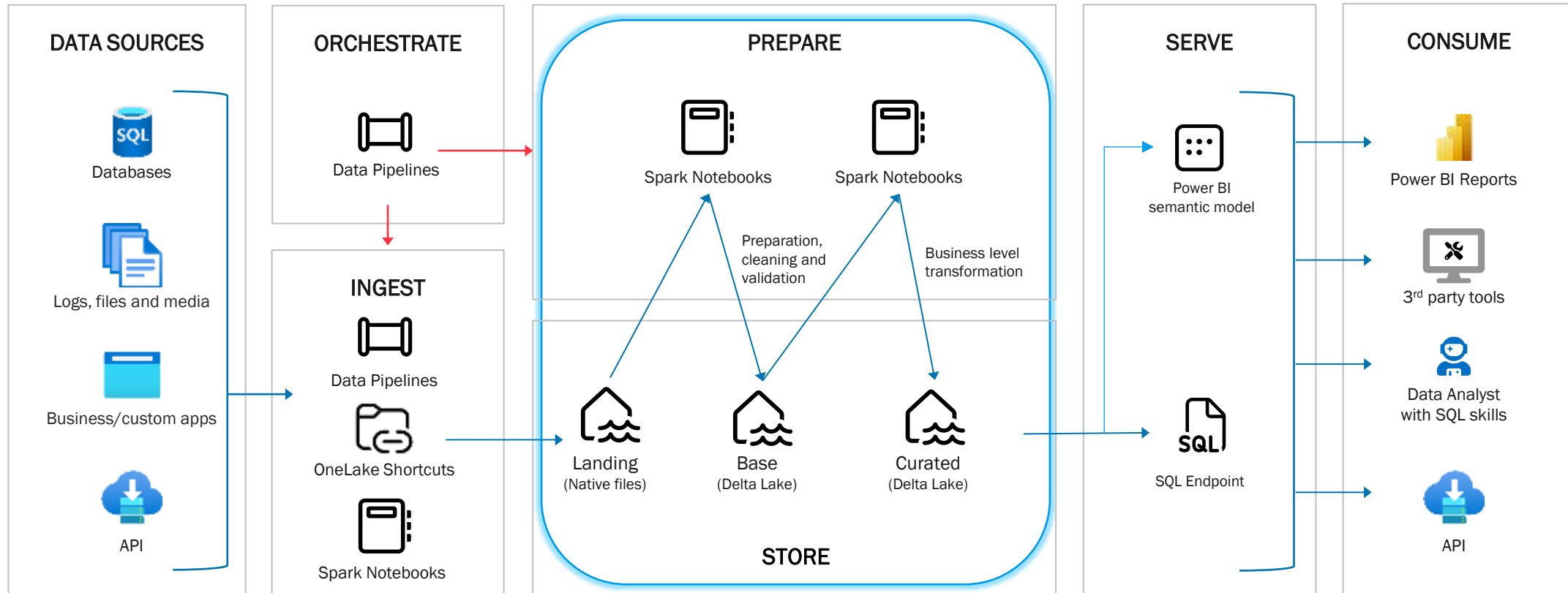




**DEMO TIME**

**twoday**

# Scenario 3: Breaking Free from SQL Constraints







**DEMO TIME**

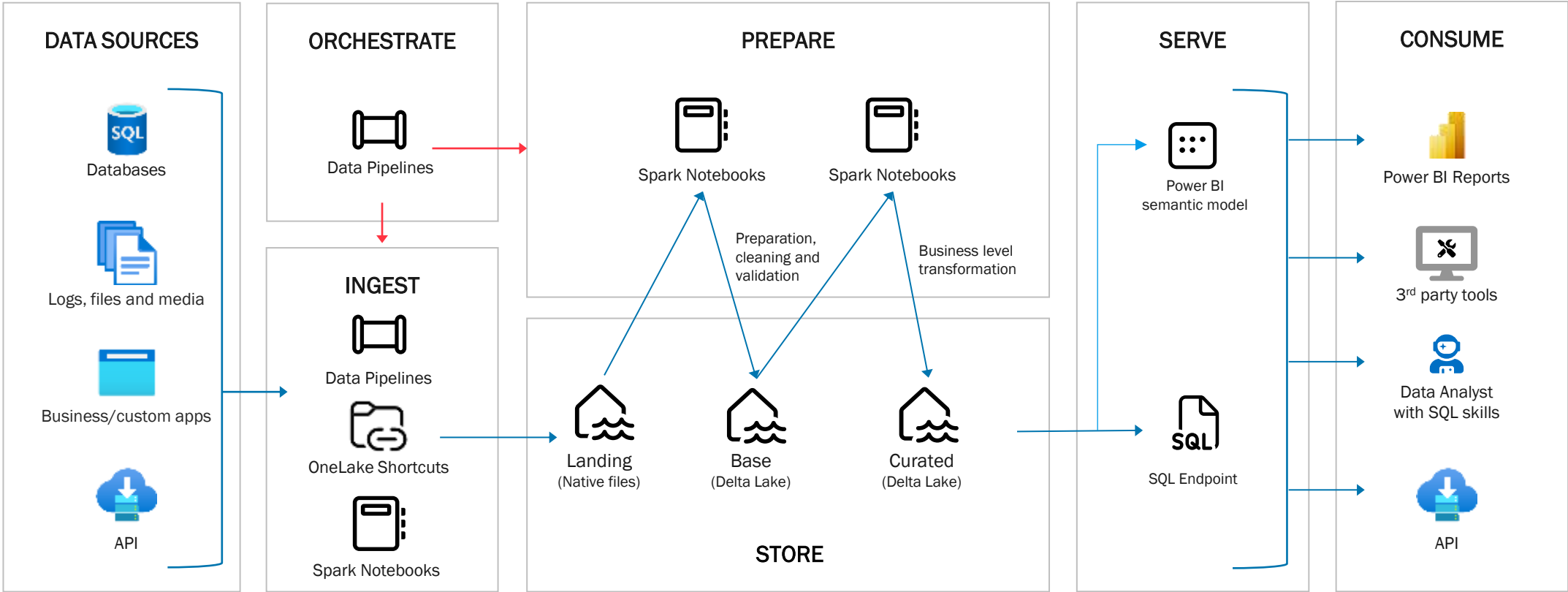
*twoday*



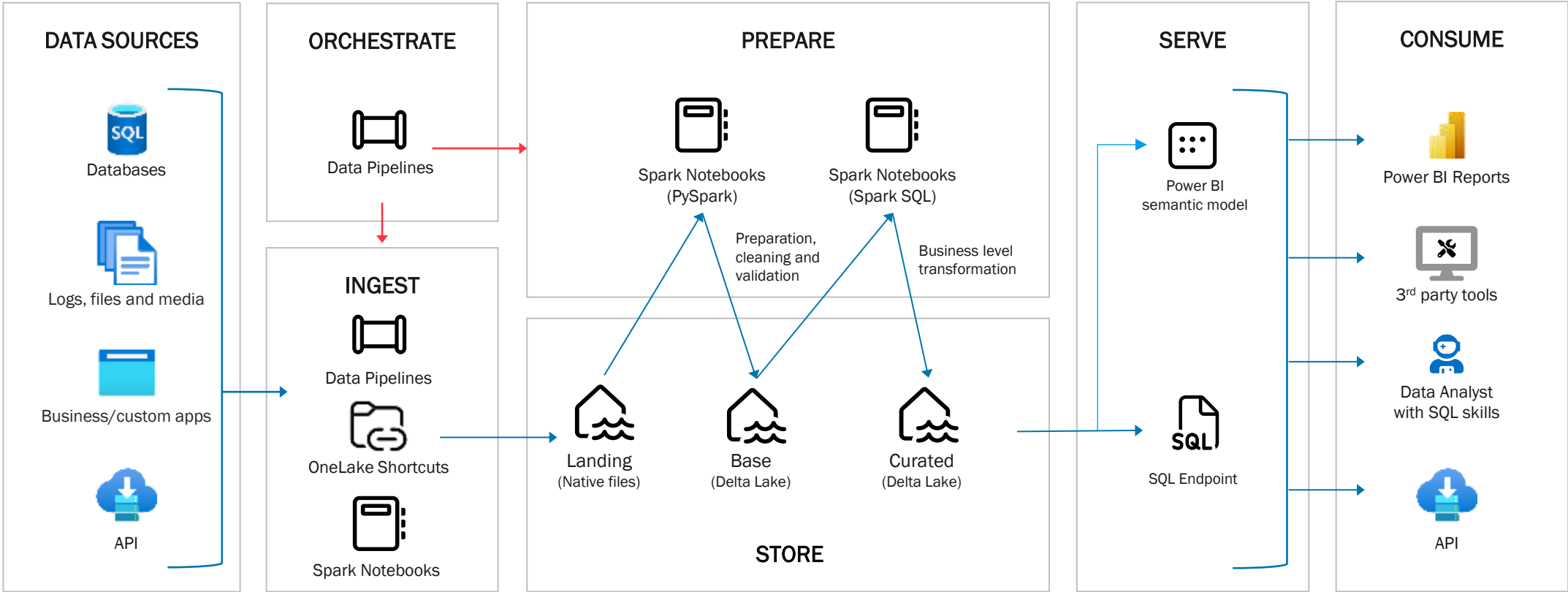
# **SQL is Still Cool: Why It's Here to Stay**

*twoday*

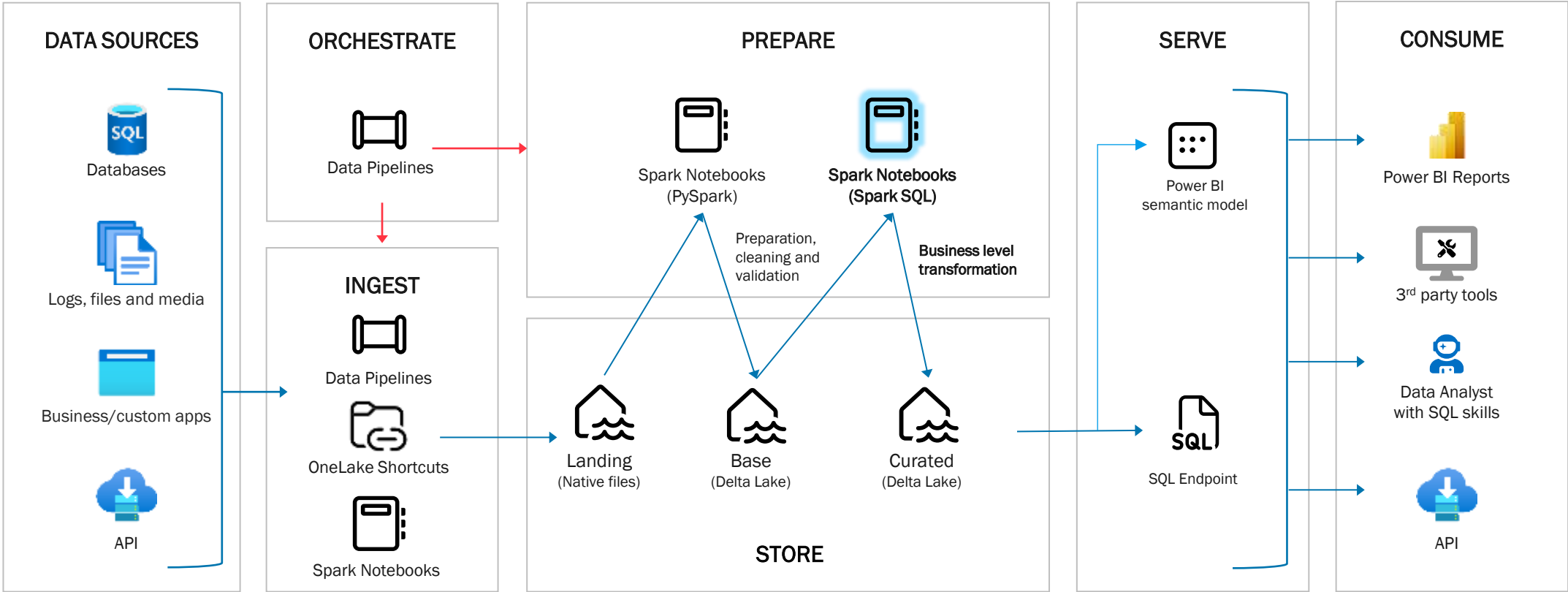
# SQL's Cool Factor



# SQL's Cool Factor



# SQL's Cool Factor





**DEMO TIME**

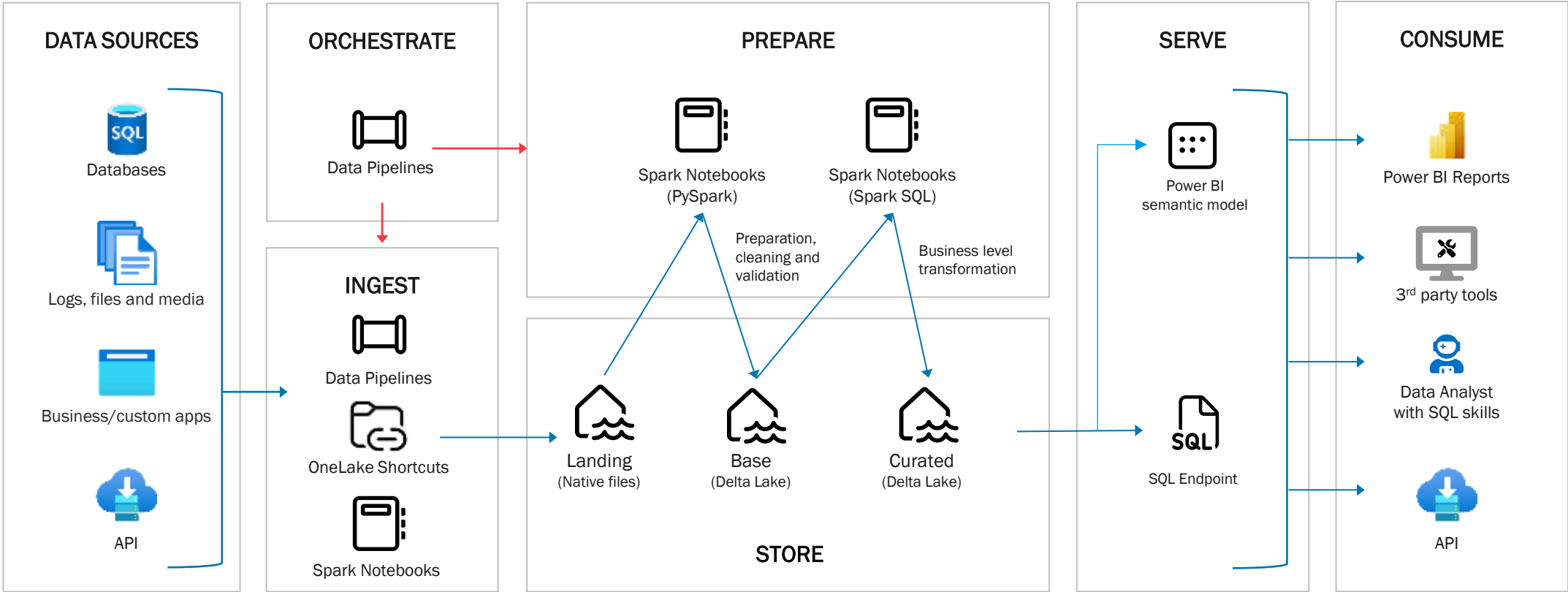
*twoday*



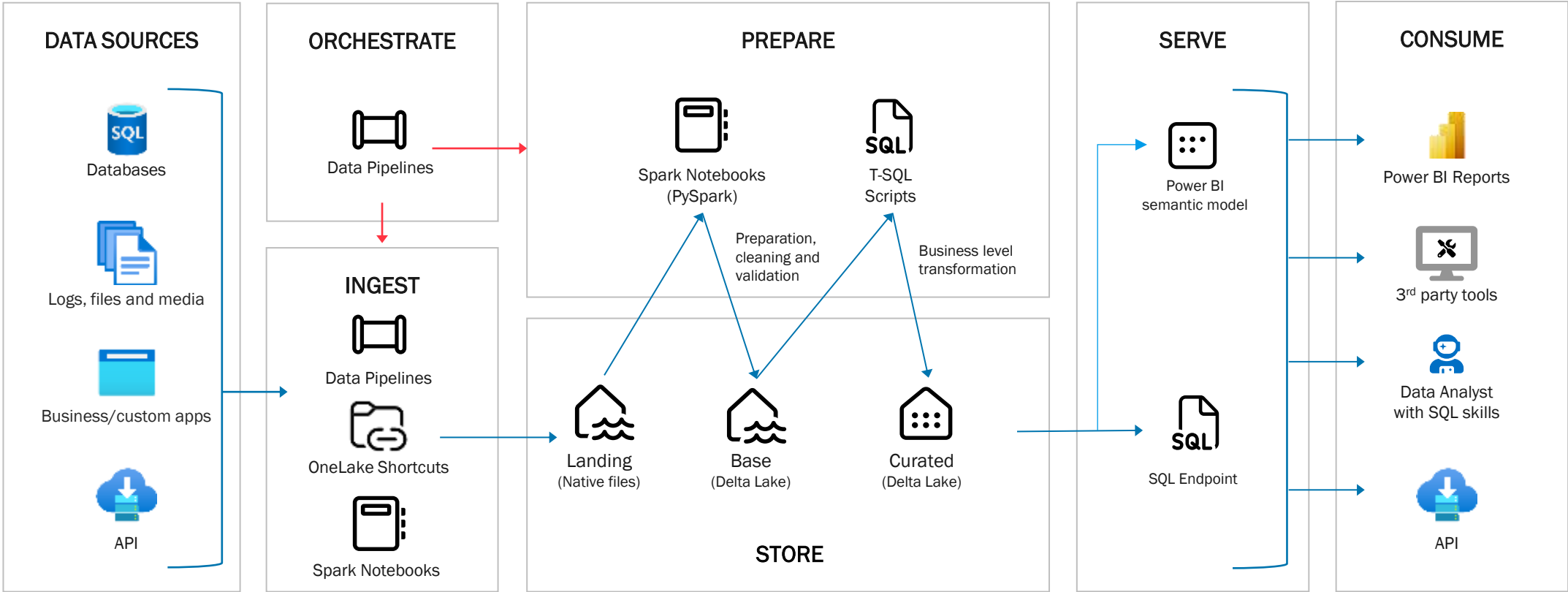
# SQL's Cool Factor

- **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.

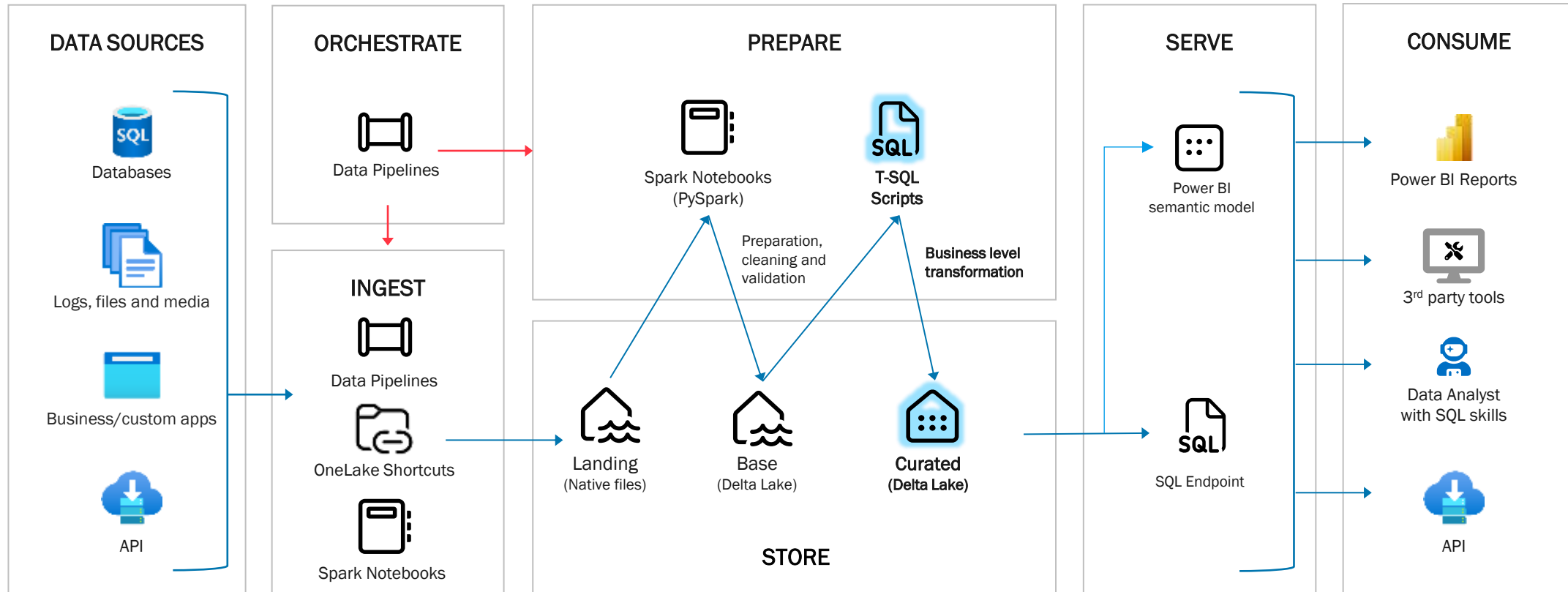
# SQL is Still Cool



# SQL is Still Cool



# SQL is Still Cool





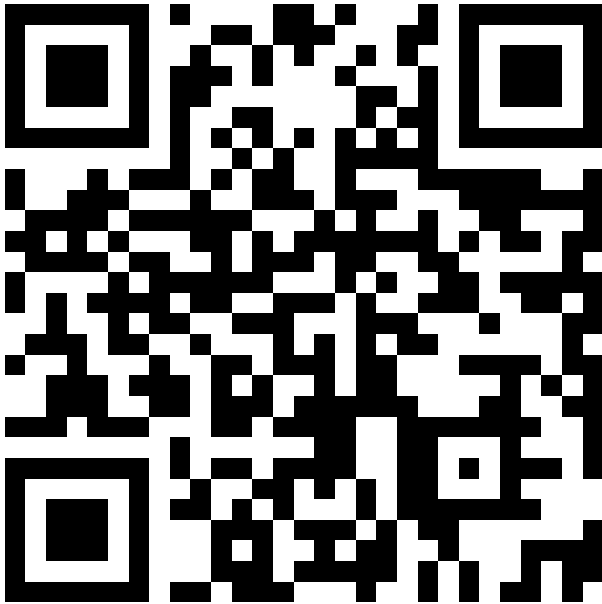
**DEMO TIME**

*twoday*

**Q&A**

**twoday**

# FREE Microsoft Fabric



## Certification Exam

Are you ready to get Fabric certified by the end of October?

Claim your **100% discount voucher** for Exam DP-600: Fabric Analytics Engineer.

Come find us in the Community Lounge under the Get Certified banner to learn more!

[aka.ms/FabCon24/IAmReady](https://aka.ms/FabCon24/IAmReady)



# Get Involved in the Fabric Community



## [aka.ms/FabricCommunity](https://aka.ms/FabricCommunity)

Connect with community members, ask questions, and learn more about Fabric



## [aka.ms/FabricUserGroups](https://aka.ms/FabricUserGroups)

Find a user group that matches your interests in your area or online



## [aka.ms/SuperUsers](https://aka.ms/SuperUsers)

Spread your Fabric knowledge, insights, and best practices with others



## [aka.ms/MVP](https://aka.ms/MVP)

Technology experts that share their knowledge and passion with the community





Please rate  
this session  
on the app



cvent





# Learn more about Microsoft Fabric

Unify your data to unlock AI innovation

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. The word 'twoday' is centered in white, lowercase, sans-serif font.

twoday

Where tomorrow is made