



Optimizing Analytics Platforms with ClickHouse

March 19, 2025



👋 **Hello**



Gary Lin
Co-Founder & CEO at Explo



What is Explo?

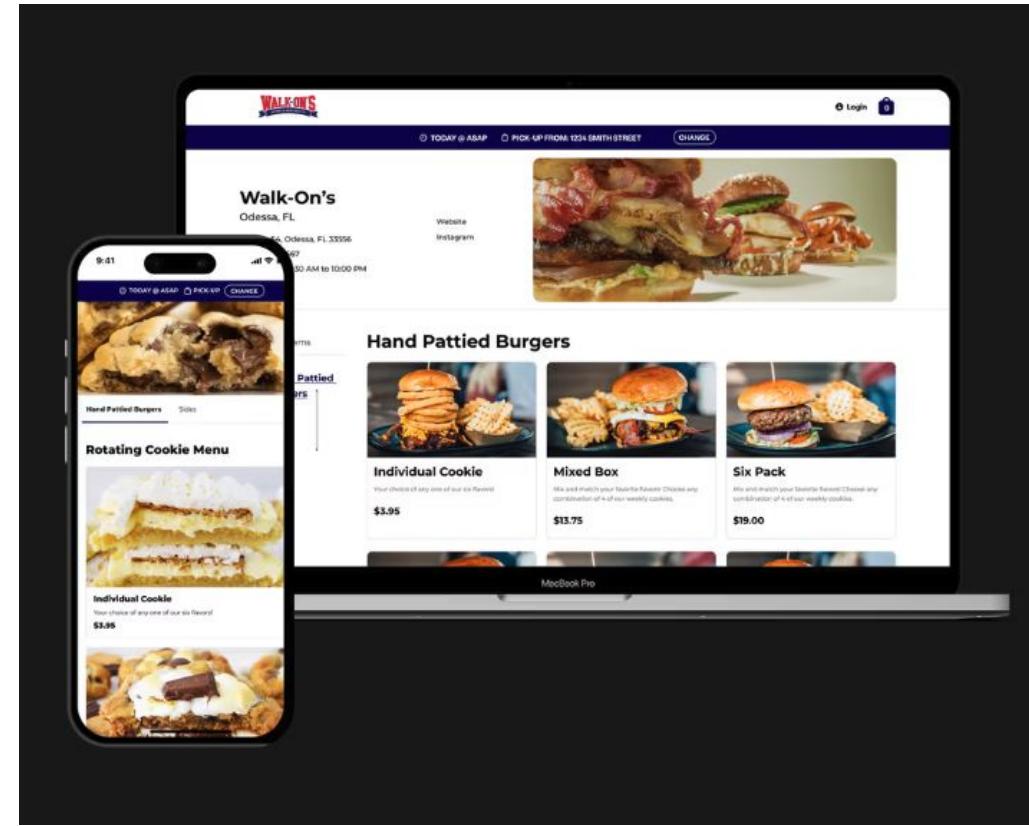


Quick Example

LUNCHBOX

Enterprise Order Management

- app and web ordering
- catering management
- etc



CHOPT

TORCHYS



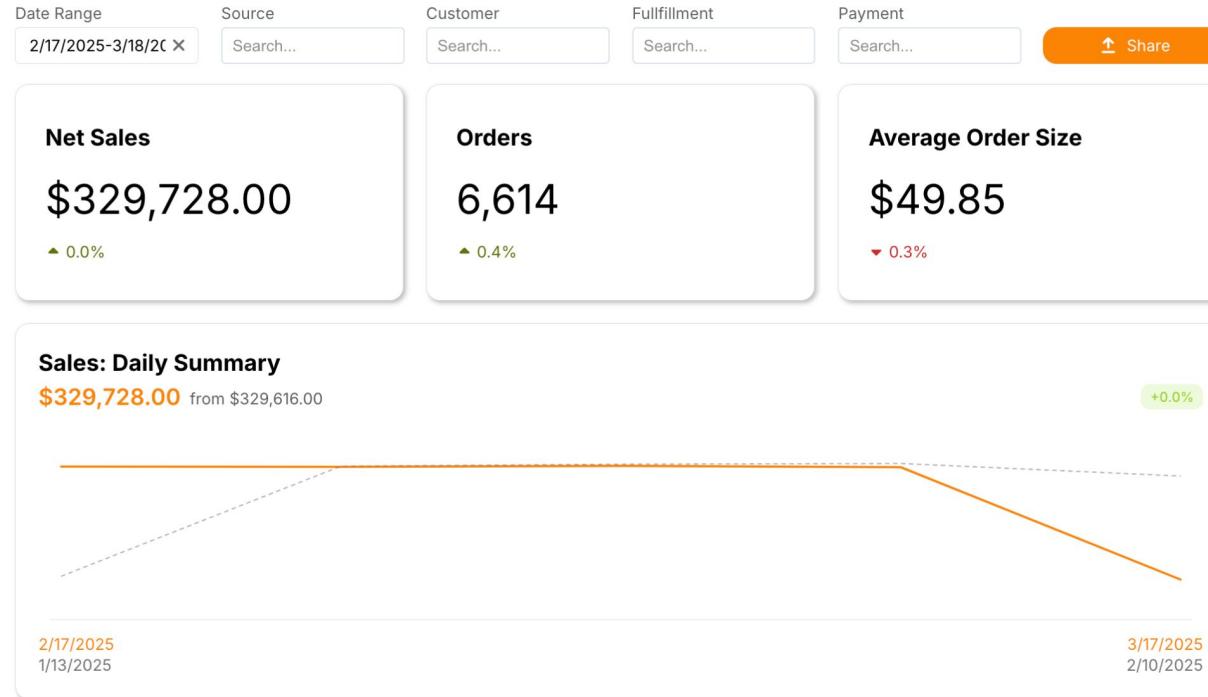
PARIS BAGUETTE

Friendly's

FIREHOUSE SUBS



Quick Example

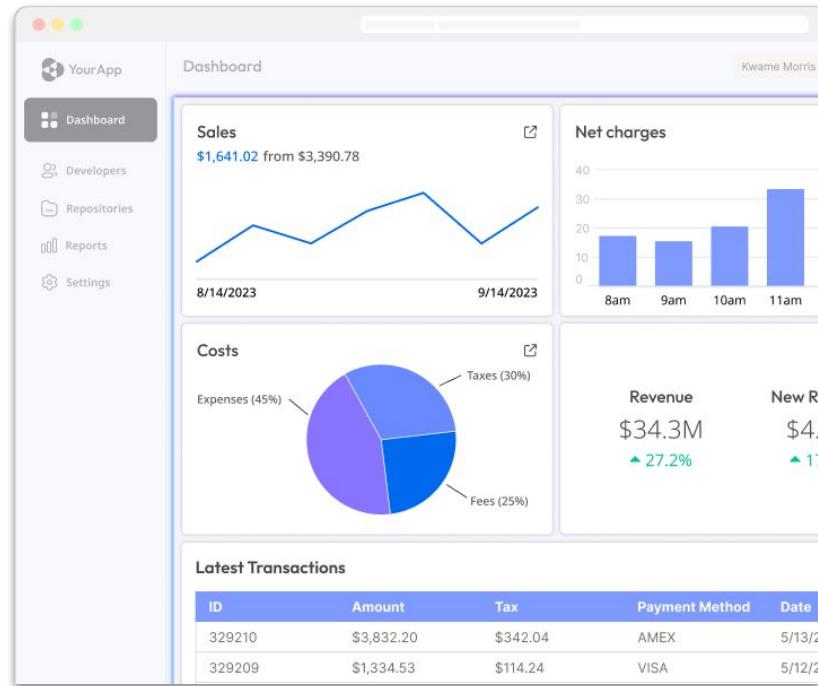


LUNCHBOX



Our Mission

Explo is the **customer-facing analytics platform**, changing the way companies think about **sharing data** with their customers (or end-users).





Let's define some terms



Terms

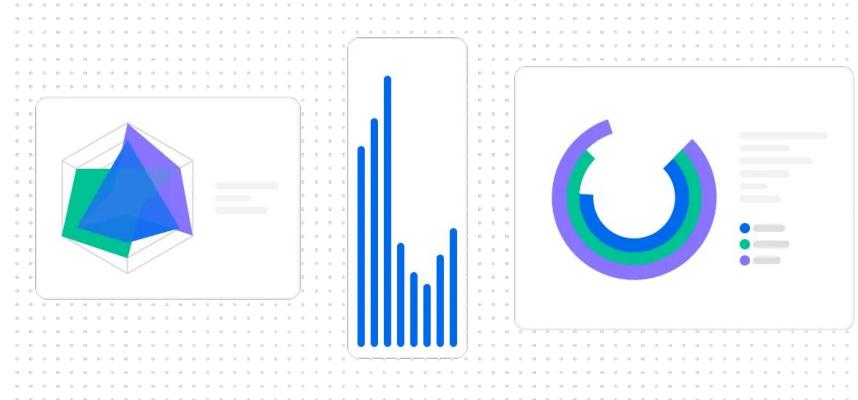
Data Orchestration

How the querying of data is performed and how that data is passed through an architecture to the desired destination.

Analytics Platform

A platform used for internal or external data analysis, designed to make it easier to derive insights from data.

For the purposes of this discussion, we'll use **Explo** as our analytics platform.

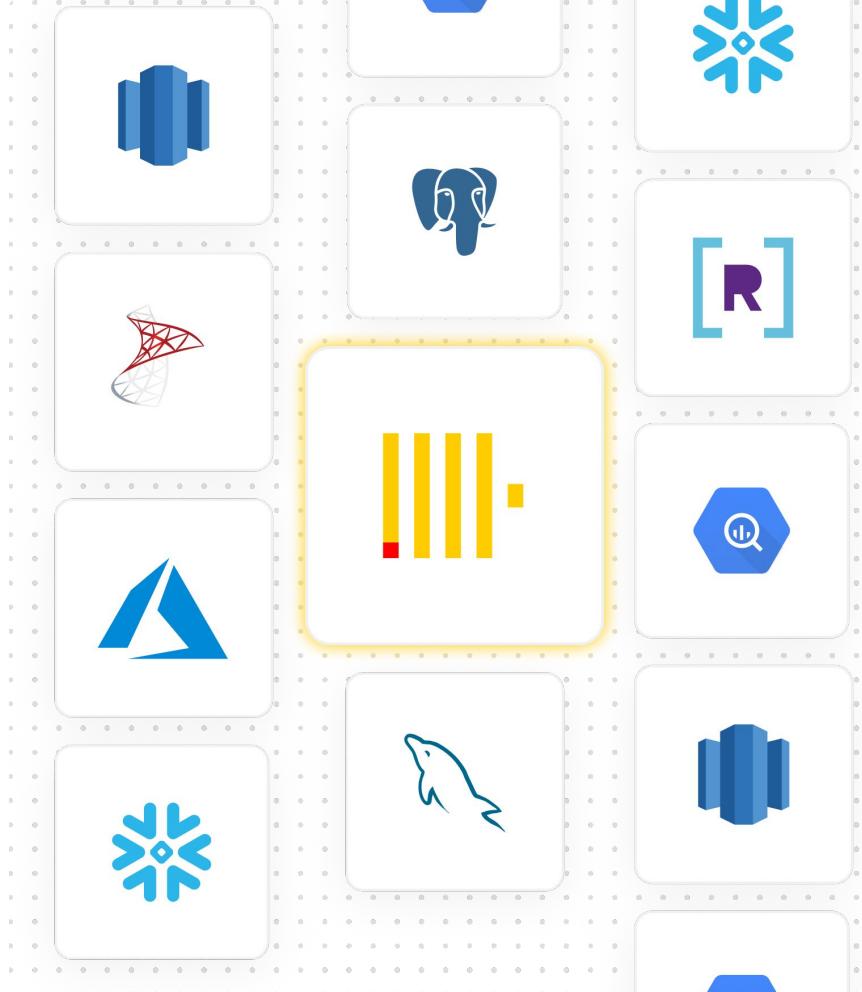




How does Explo work?

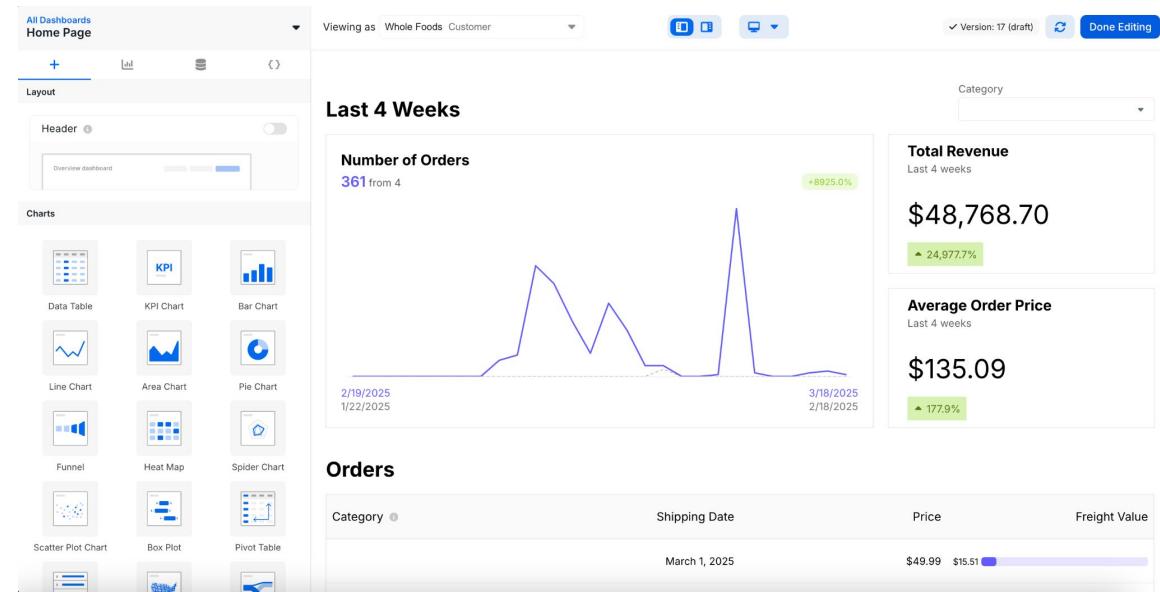
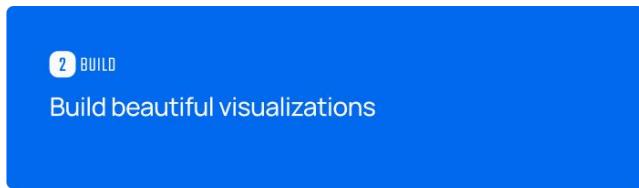


How it works - Connect





How it works - Build





How it works - Style

1 CONNECT
Connect your databases

2 BUILD
Build beautiful visualizations

3 STYLE
Style components to match any app

Styling

Viewing Home Page X as Whole Foods

Last 4 Weeks

Number of Orders
7,636 from 6,500 +17.5%

11/11/2023 12/9/2023

Orders

Category	Shipping Date	Price	Freight Value
Garden Tools	November 26, 2023	\$99.00	\$33.08
Garden Tools	November 12, 2023	\$99.00	\$45.99

Total Revenue
Last 4 weeks \$870,865.24 +12.7%

Average Order Price
Last 4 weeks \$114.05 ▼4.0%

Styling Overview

Theme to edit: default

Base
Number of dashboard columns: 12

Color
Background: #FFFFFF Actions: #635BFF
Custom interaction colors: #635BFF

Spacing
20px Narrow

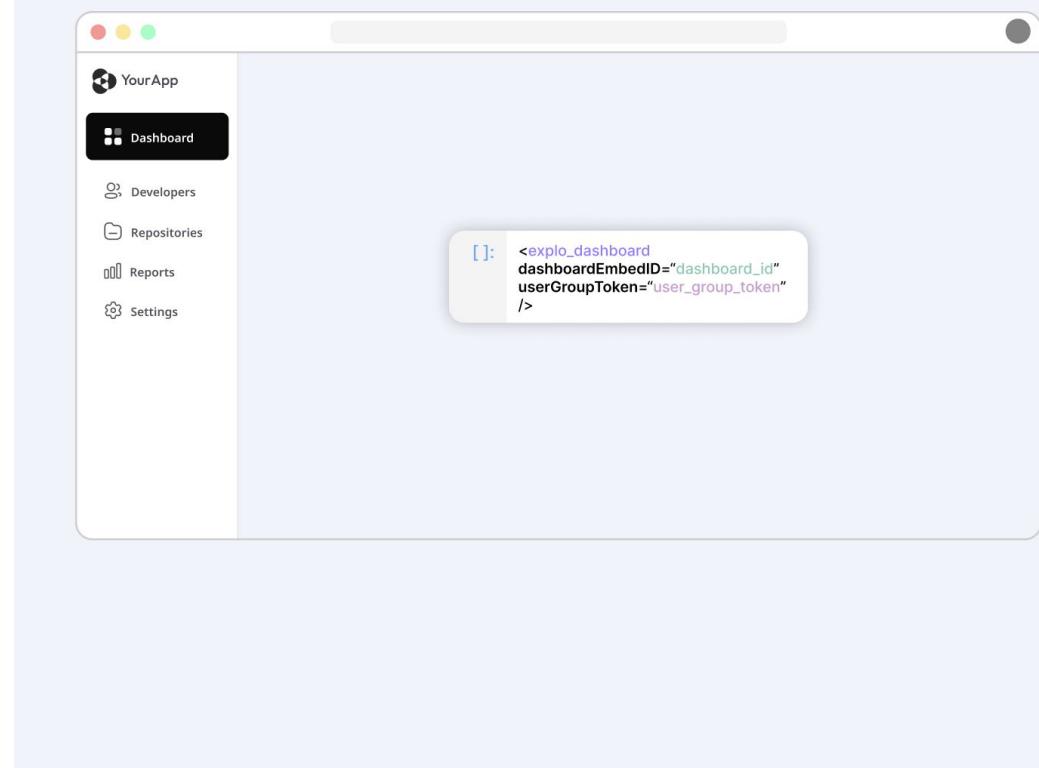
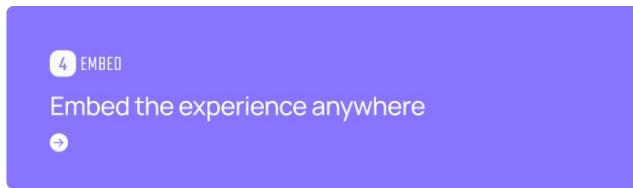
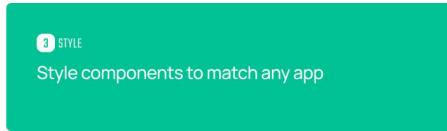
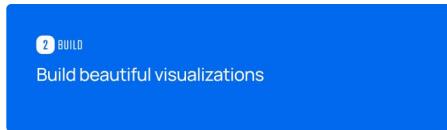
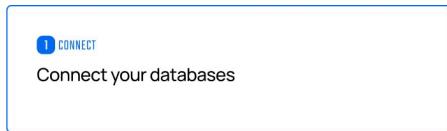
Cards
Fill and border: #FFFFFF

Actions
#635BFF

Reset Save



How it works - Embed





Some Examples

✓

Data	Documents	Status	Value (R\$)
06/07/2021	N/A	Active	90,00
07/07/2021		Inactive	620,00
07/07/2021	N/A	Inactive	200,050,00
07/09/2021		Active	620,00
07/10/2021		Inactive	500,00
08/07/2021	N/A	Active	40,00

Novo pedido

Overview Finance Marketing Tags

Date Range: Last 50 days Merchant: Search...

Revenue: \$0.51 m AOV: \$40.87 Orders: 12.57 k Blended CPA: 424.70 Blended ROAS: 1.89 x Total spend: \$7.7 k

Revenue: +8.3% (\$513,634.06 Jul 31 - Aug 30) Orders: +26.8% (12,566 Jul 31 - Aug 30) Total spend: -8.3% (\$7,577.78 Jul 31 - Aug 30)

Revenue: \$431,878.89 Jun 30 - Jul 30 Orders: 9,903 Jun 30 - Jul 30 Total spend: \$46,830.92 Jun 30 - Jul 30

Revenue: \$431,878.89 Jun 30 - Jul 30 Orders: 9,903 Jun 30 - Jul 30 Total spend: \$46,830.92 Jun 30 - Jul 30

EXAMPLES

Monitor Devices Cameras Host customers VCF Explo Dashboards Date Dog Device VPN

Total event counts: 36k, 24k, 12k, 0 (Frame events from 8/27 to 9/2)

Vehicle and Ped. Counts: (line chart showing vehicle and pedestrian counts over time)

Analytics

Spels Search Spels Create

Unique Viewers: Daily, Weekly, Monthly (Jan 2022 to Jan 2023)

Top Viewers:

User Name	Total Views
Justin Time	1,921
Garett Gutierrez	1,846
Valerie Lambert	1,842
Harper Sanders	1,787
Owen Harris	1,636

Dashboard Patients Appointments COVID-19 waitlist Flu shot waitlist Reports

Date grouping: Last 7 days Start date: 01 / 01 / 2022 End date: 01 / 26 / 2022 Share

Number of patients for all Appointment types

Number of Appointments

app_type

- Asymptomatic COV...
- Asymptomatic COV...
- COVID-19 Private P...
- COVID-19 Vaccine...
- COVID-19 Vaccine...
- COVID-19 Vaccine...
- MedCheck Annual...
- MedCheck for Dia...
- MedsCheck Follow...
- PCR Test for Travel...
- Pediatric COVID-19...
- Pediatric COVID-19...

Number of patients for all Appointment types

Appointment Type	Count
Asymptomatic COV...	4
Asymptomatic COV...	4
COVID-19 Private P...	3
COVID-19 Vaccine...	3

Home Calendar Schedule changes Recurring services Requests Ticket dashboard Locations Users Reporting Invoices Rate cards Help center Contact support My account Logout

Report type: Current month dashboard

Location:

Service line: Service type:

Open Requests

Type	Count	Percentage
corporate review	1	12%
market	83	100%
corporate bid revl...	5	6%
scheduling	7	8.4%
scheduled	21	25.3%



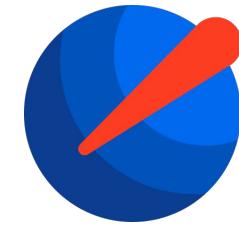
Some Optimizing Constraints



Multiple Levers to Pull



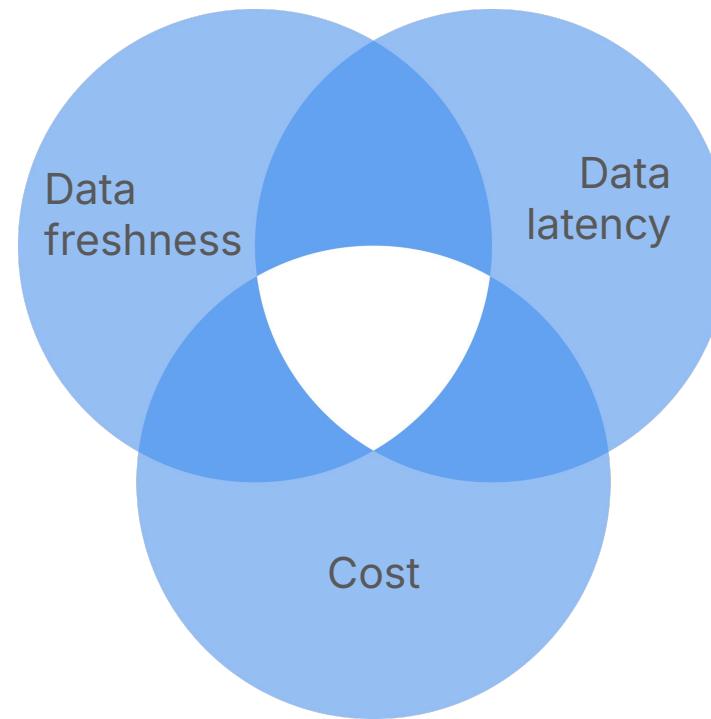
Company



Analytics Platform



Can't Have it All: Fresh vs. Cheap vs. Fast





Steps to Optimize Analytics Platforms

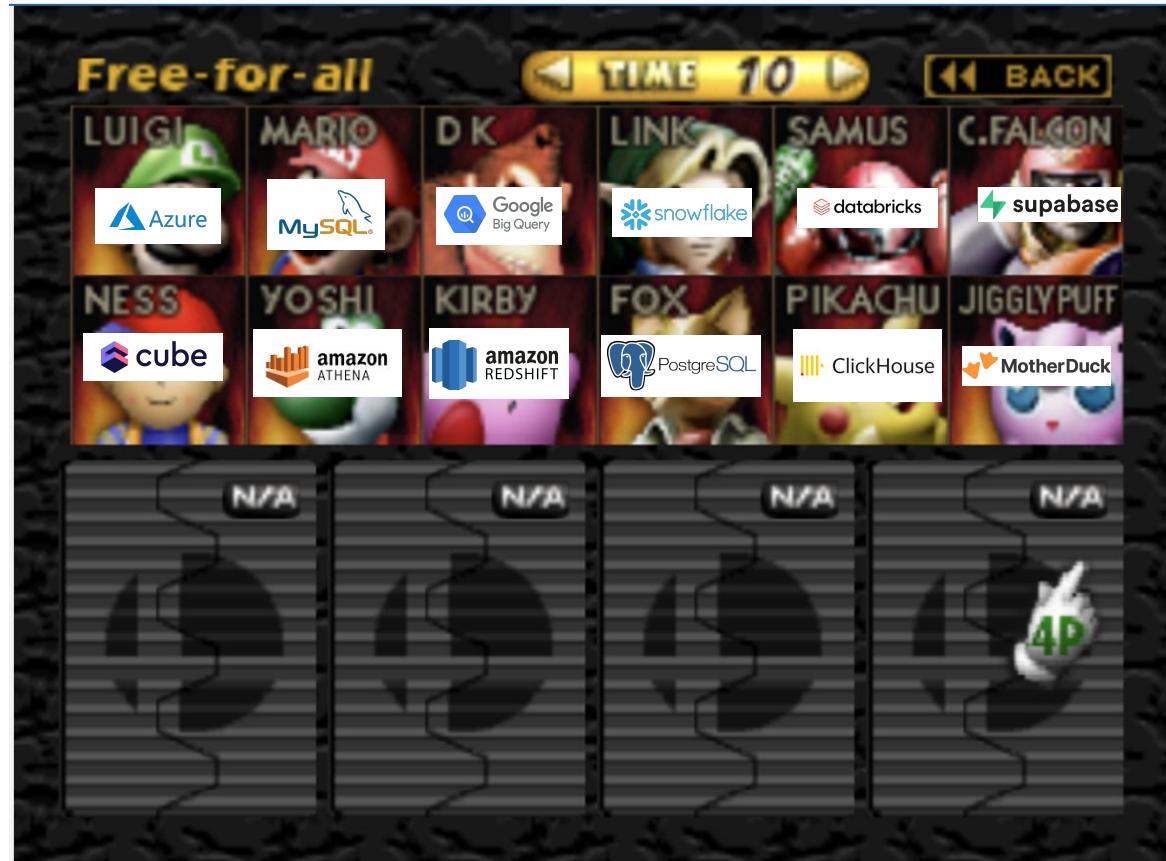


1

(You) Choose the Right DB



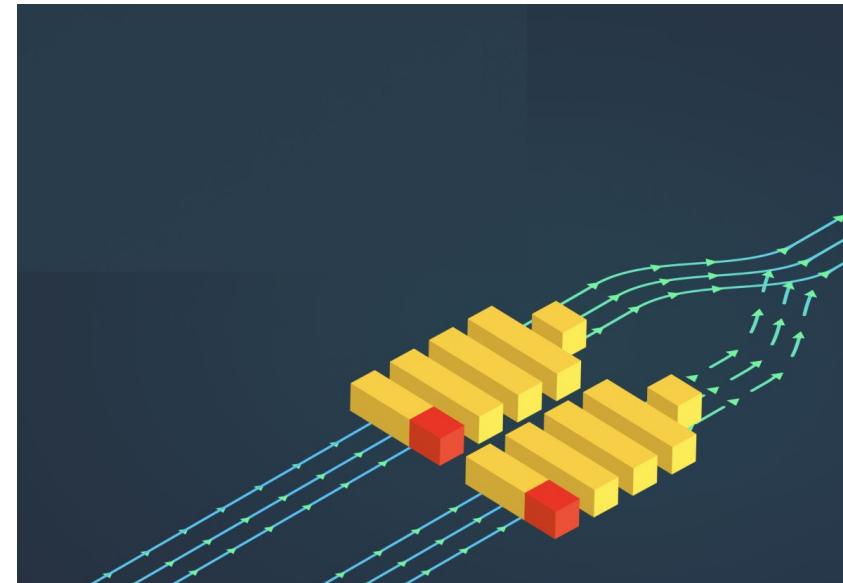
Choose your Character





Why ClickHouse

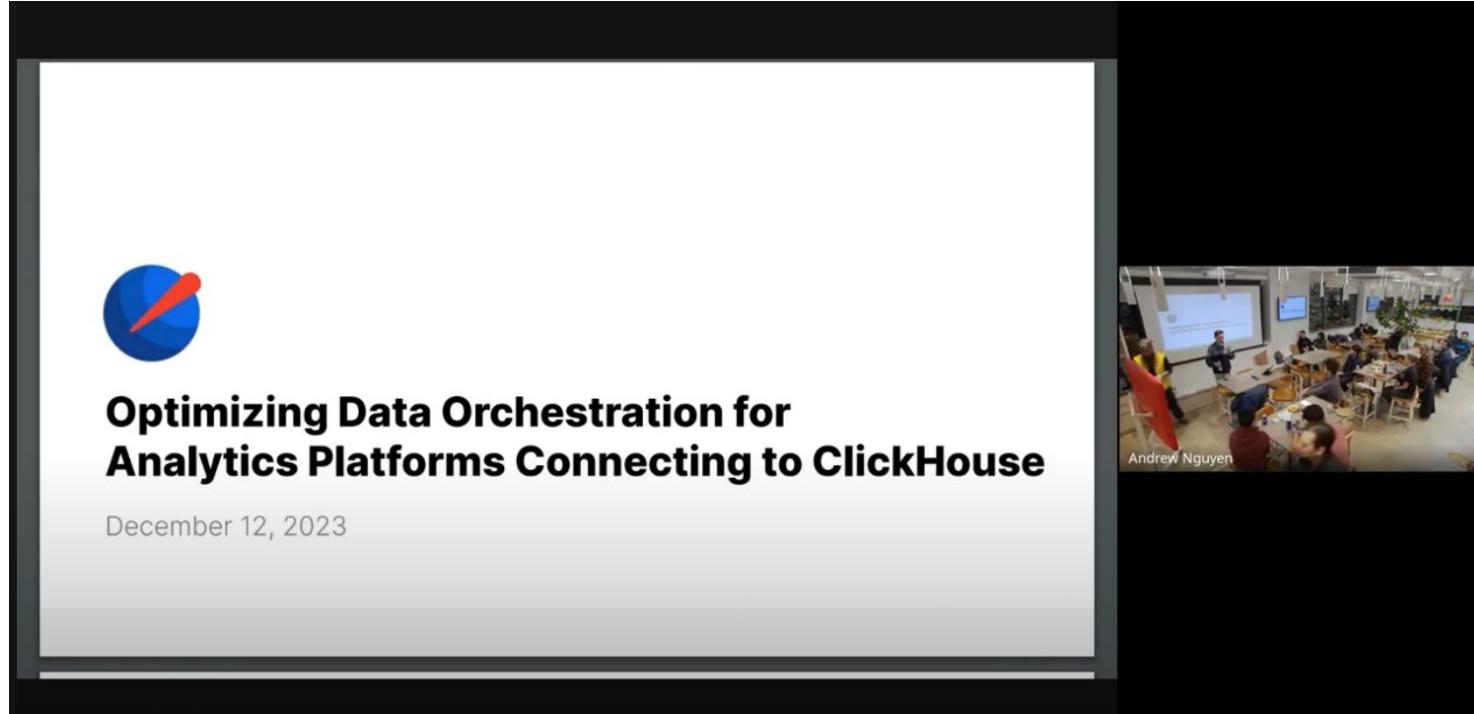
- Speed & Performance
 - Columnar Storage
 - Data Compression
- Scalability & Concurrency
 - OLAP (Online Analytical Processing)
 - Distributed Architecture
- Indexing & Query Optimization
 - Sparse Indexes
- Handling Large Datasets
- Cost Efficiency
 - Lower Storage Costs
 - Faster Query Execution





(We) Optimize Data Orchestration

✓ Flashback to Dec 2023 (in NYC)



The image shows a presentation slide on the left and a photograph of a presentation on the right.

Slide Content:

- A blue circular logo with a red compass-like symbol.
- Optimizing Data Orchestration for Analytics Platforms Connecting to ClickHouse**
- December 12, 2023

Photograph Content:

- A photograph of a presentation in progress in a room with several people seated at tables and a person standing near a whiteboard.
- Caption: Andrew Nguyen



Upgrading Explo's Data Orchestration





Upgrading Explo's Data Orchestration





What's a Backronym

F I D O
— — — —



What's a Backronym

F I D O

A N A R
S T T C
T E A H
R O A E
O P S E
R E S T
A B R A T
L E R O R



What's a Backronym

R O V E R
— — — — —



What's a Backronym

R O V E R

— — — — —

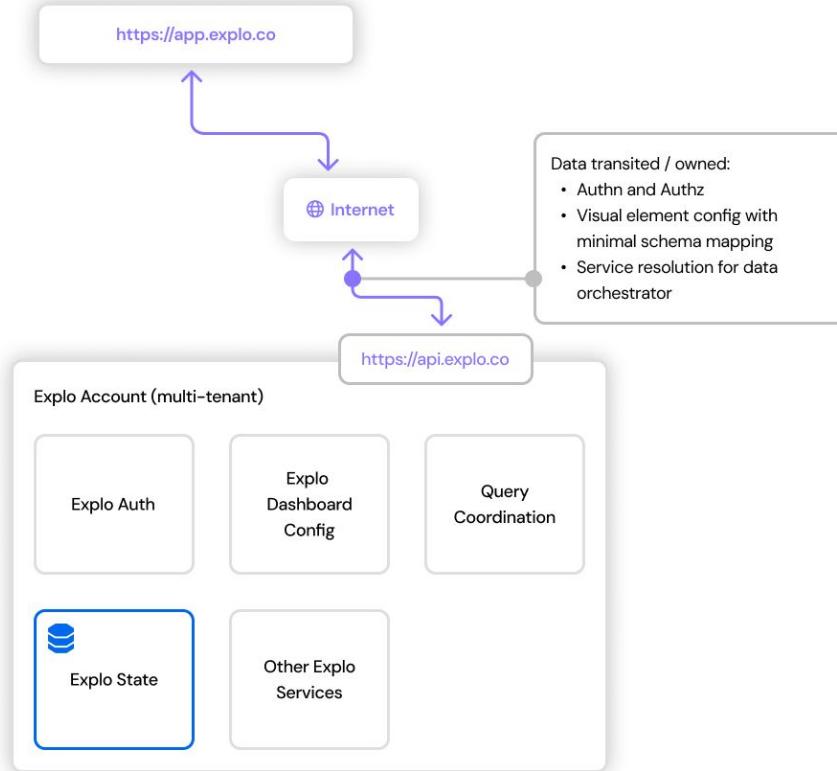
E H I M A E P
L C I L I O R O R T
I H C L I L I O R T S
A E S T R A T I O N
B R C H E S T R A T I O N
L A B L E (F O R)
E



Let's dive in!

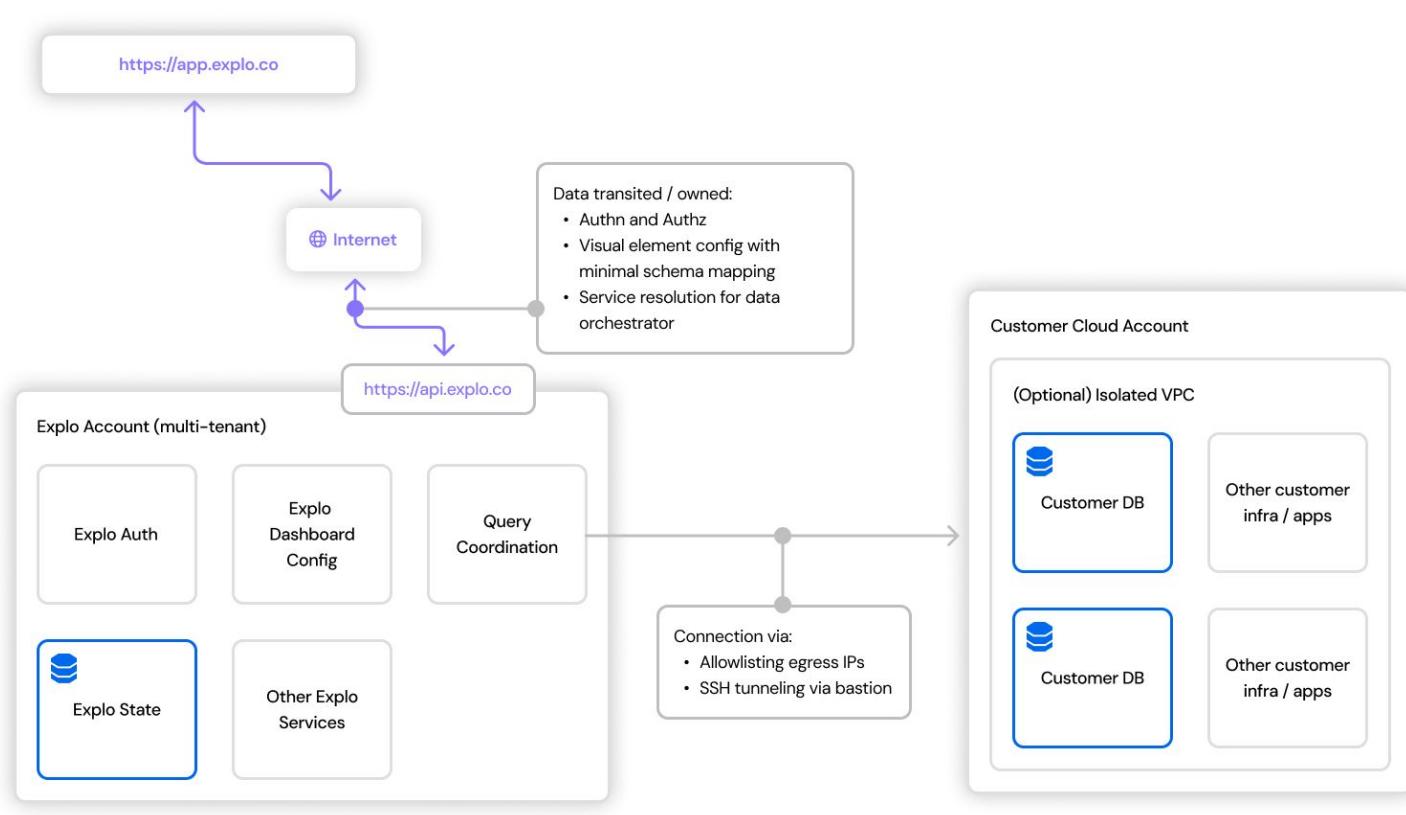


Data Orchestration v1



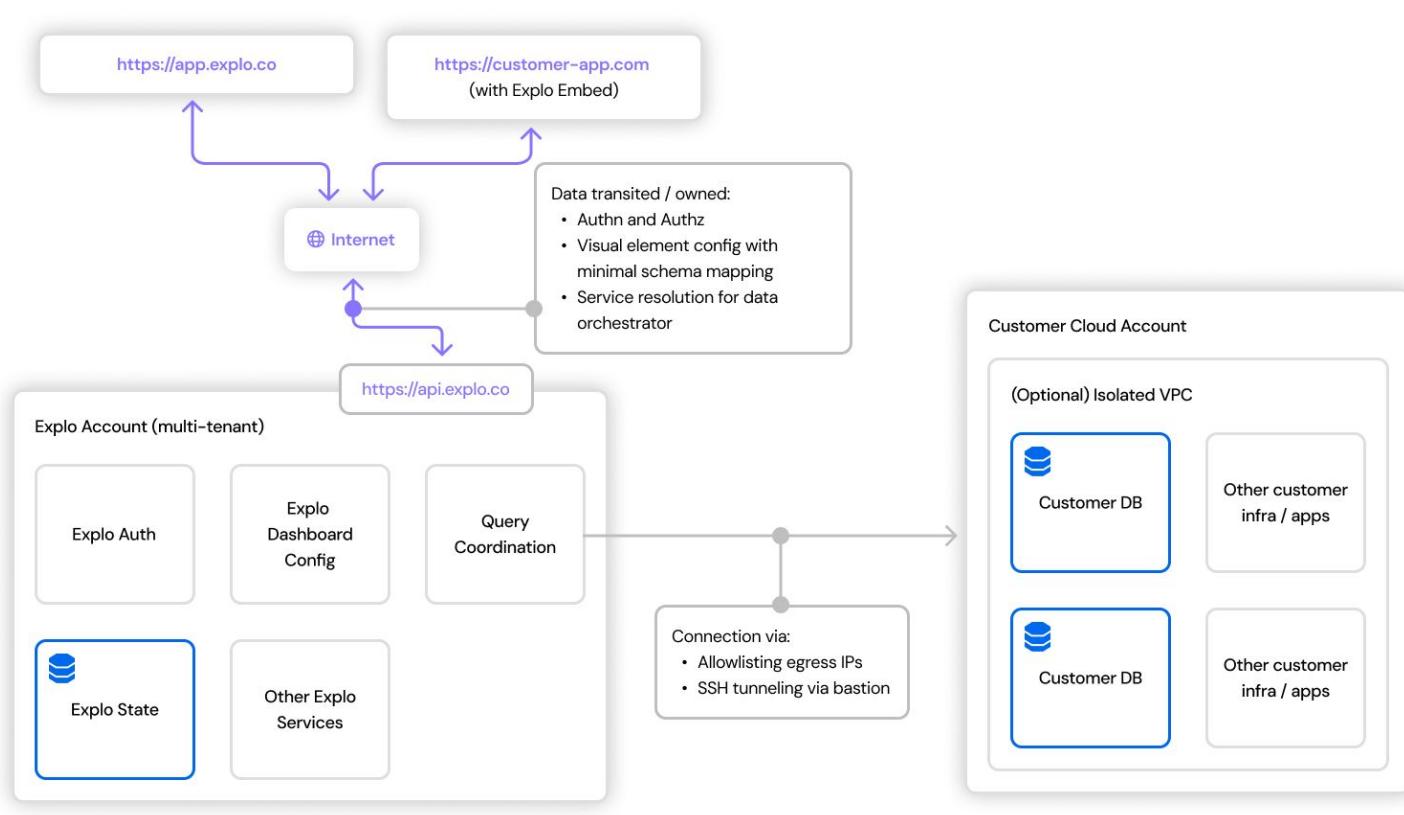


Data Orchestration v1





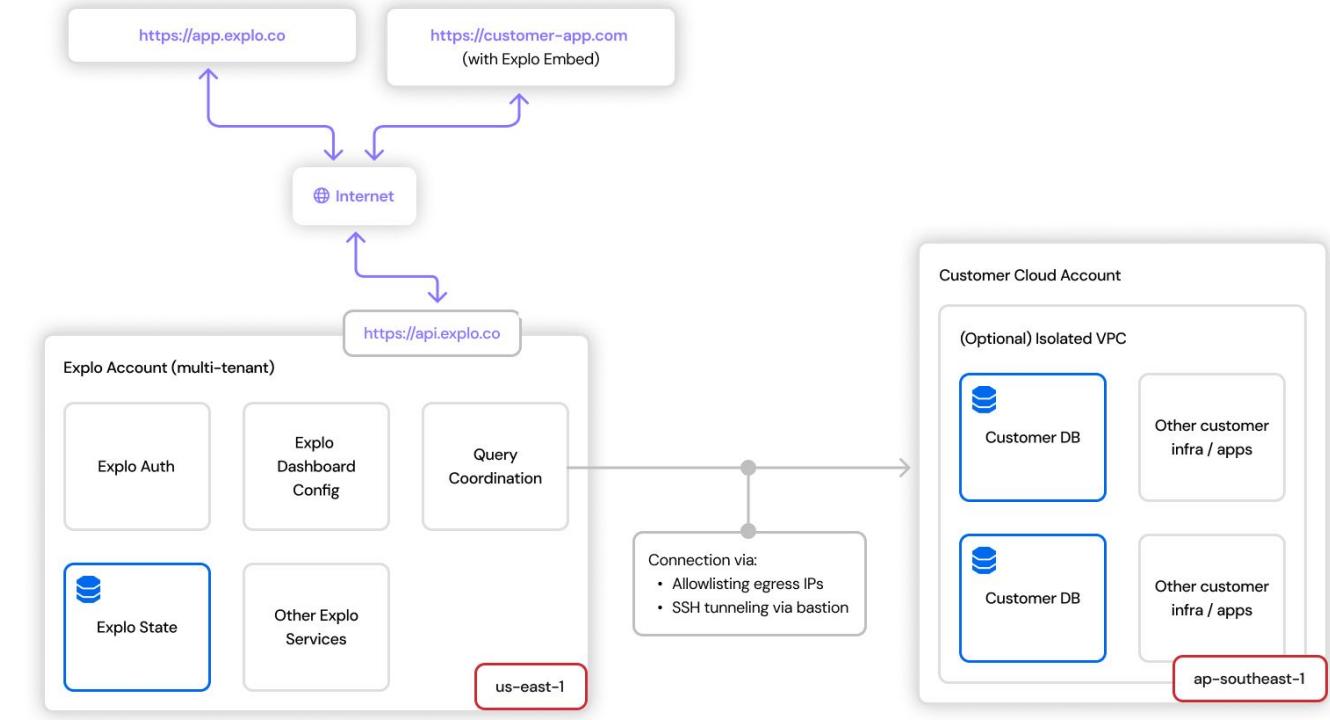
Data Orchestration v1





Issues with this Architecture

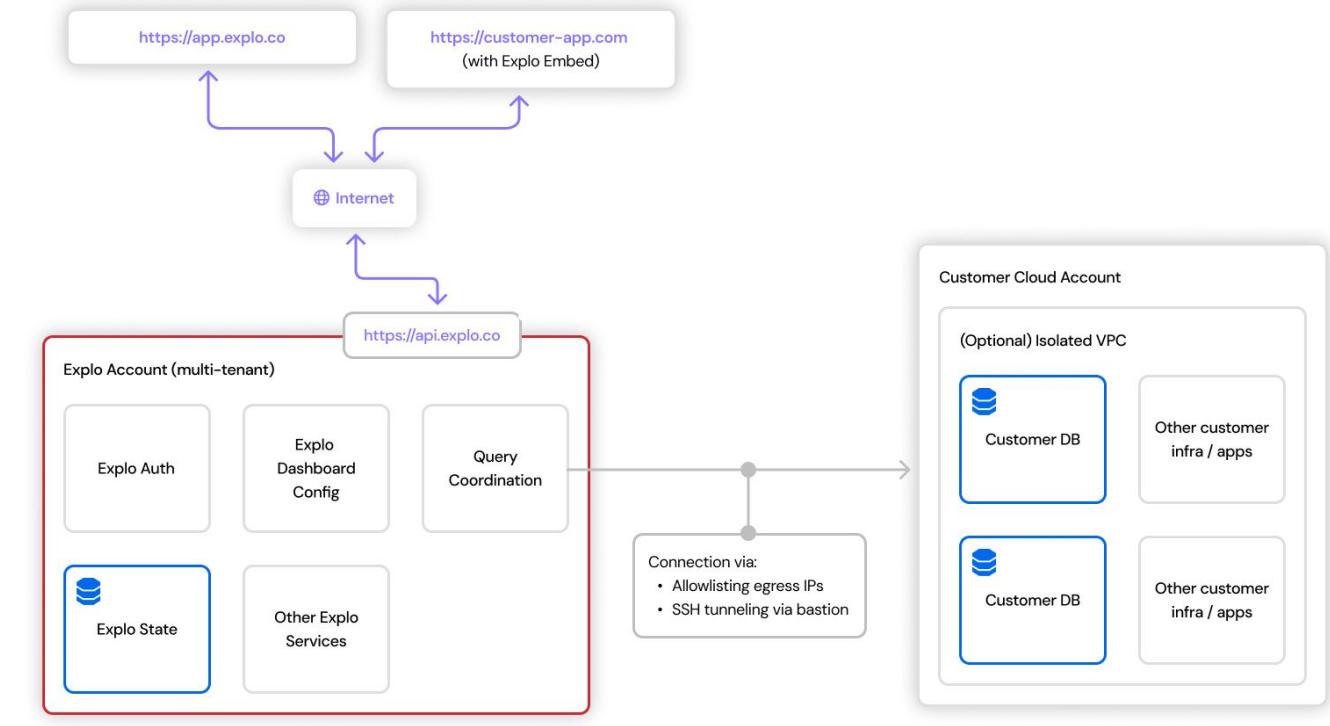
- Query coordination hosted in us-east-1
 - Latency
 - Region specific data requirements (GDPR)





Issues with this Architecture

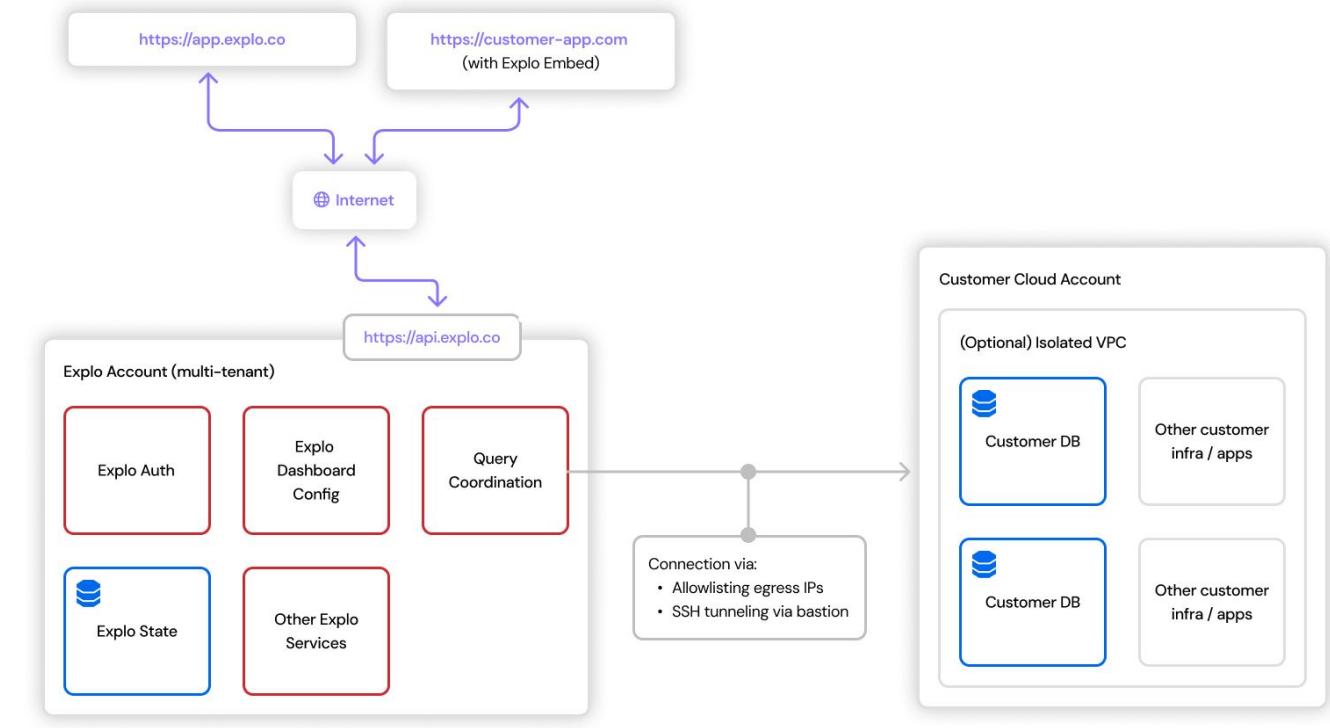
- QC hosted in us-east-1
 - Latency
 - Region specific data requirements (GDPR)
- **Very expensive On-prem**





Issues with this Architecture

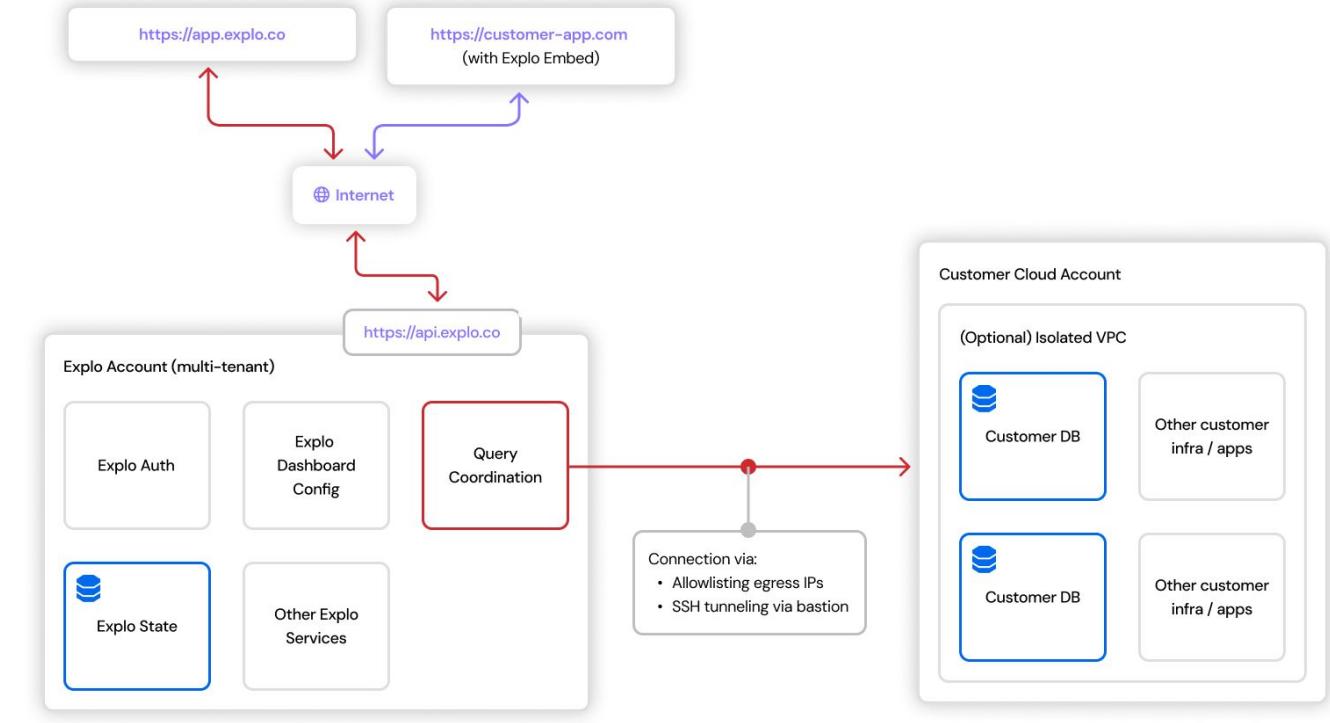
- QC hosted in us-east-1
 - Latency
 - Region specific data requirements (GDPR)
- Very expensive On-prem
- QC competes for resources





Issues with this Architecture

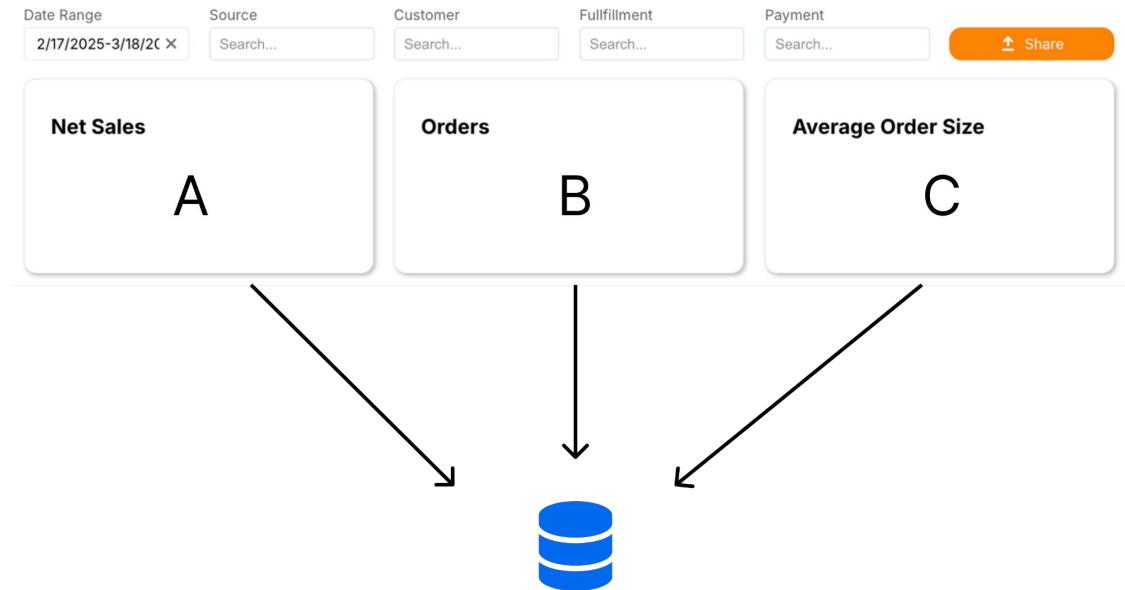
- QC hosted in us-east-1
 - Latency
 - Region specific data requirements (GDPR)
- Very expensive On-prem
- QC competes for resources
- **Non-persistent connections**





Issues with this Architecture

- QC hosted in us-east-1
 - Latency
 - Region specific data requirements (GDPR)
- Very expensive On-prem
- QC competes for resources
- Non-persistent connections



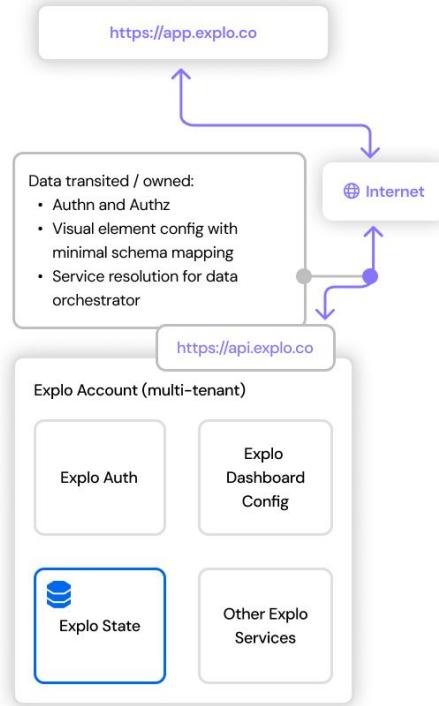


Enter: FIDO

(Fast Interoperable Data Orchestrator)

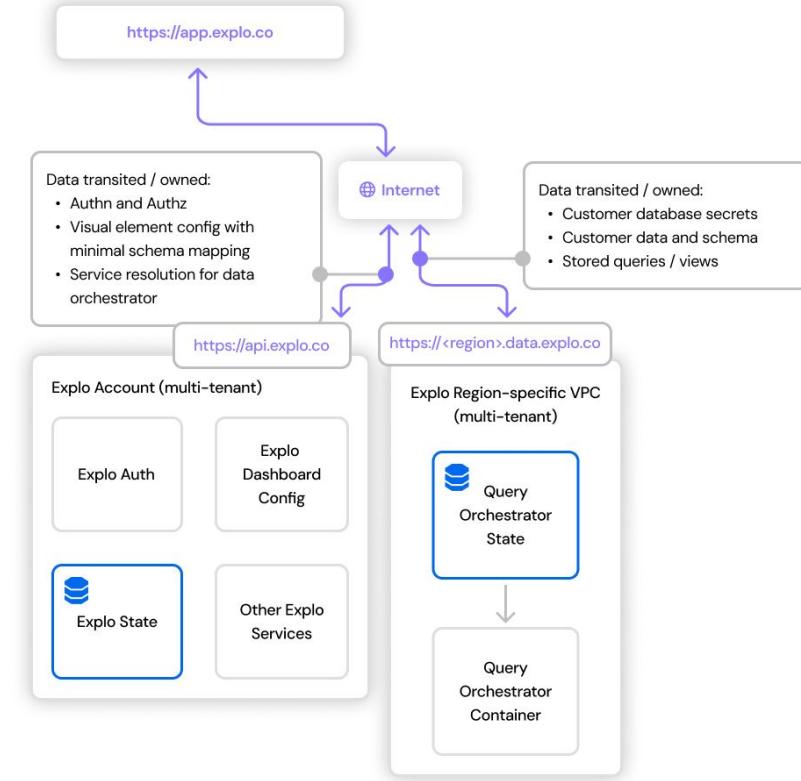


Data Orchestration v2



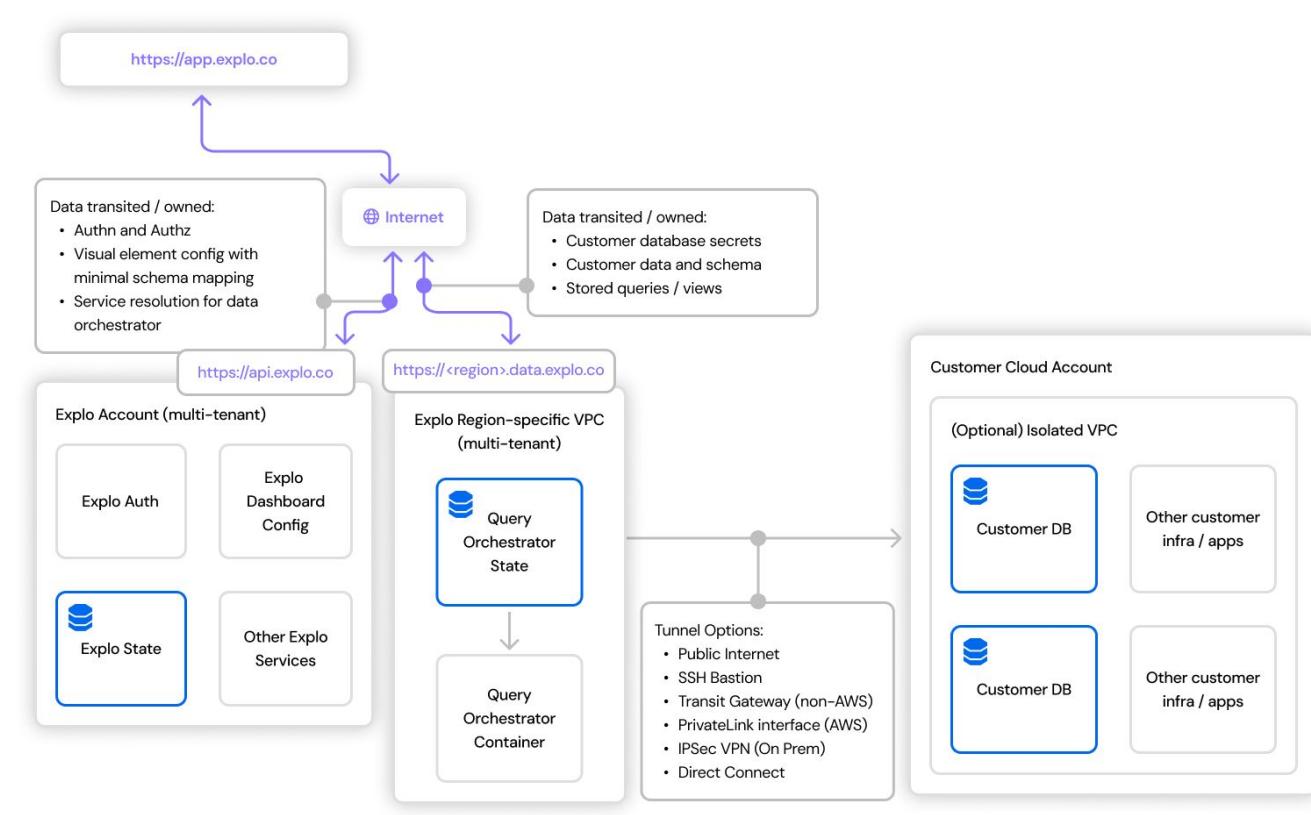


Data Orchestration v2



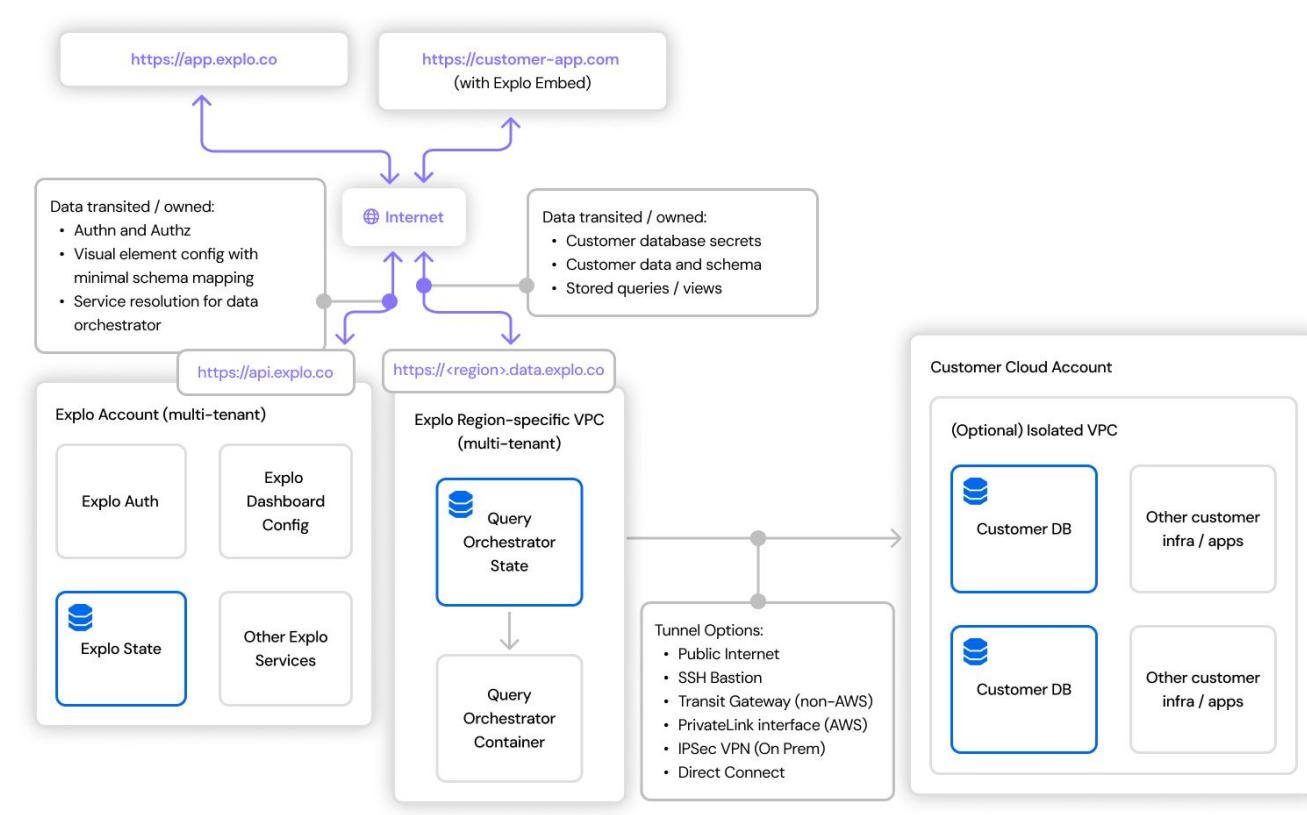


Data Orchestration v2



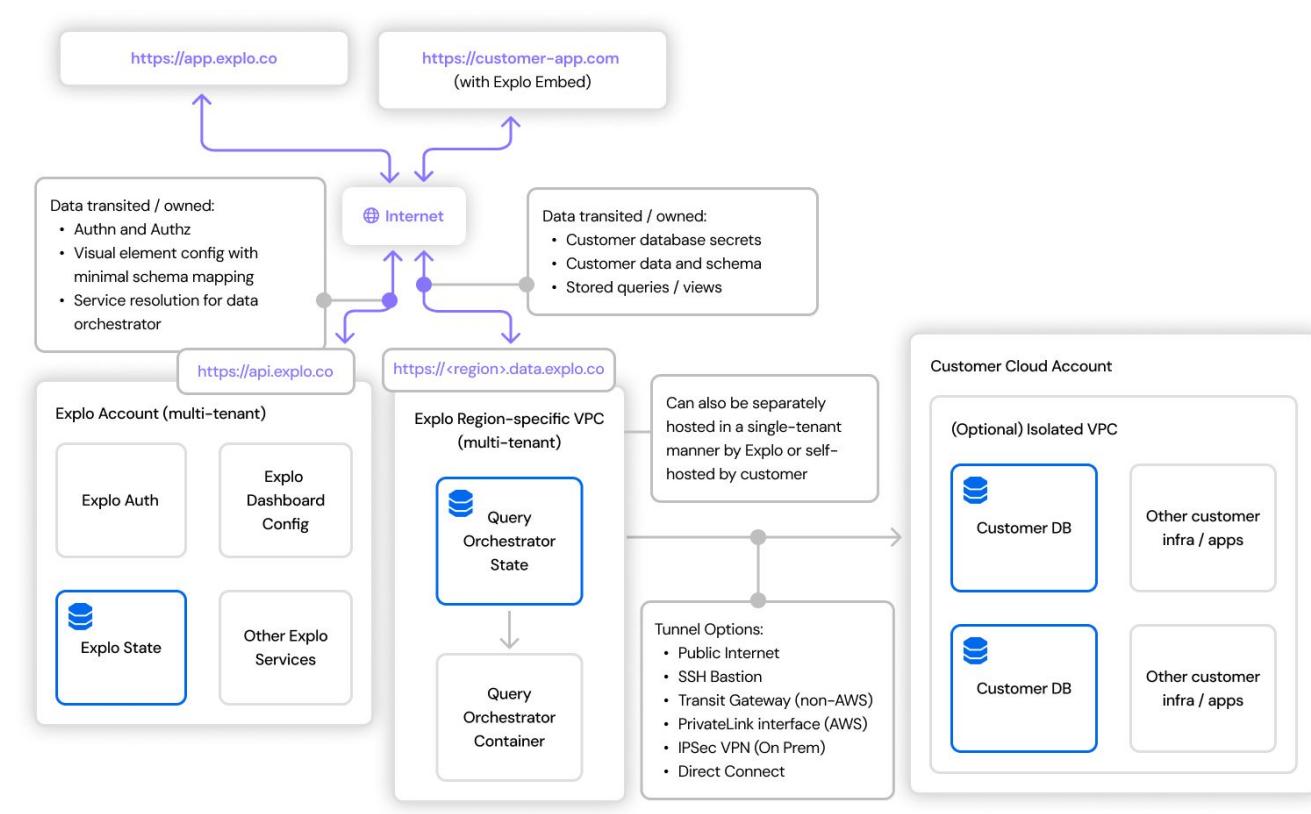


Data Orchestration v2





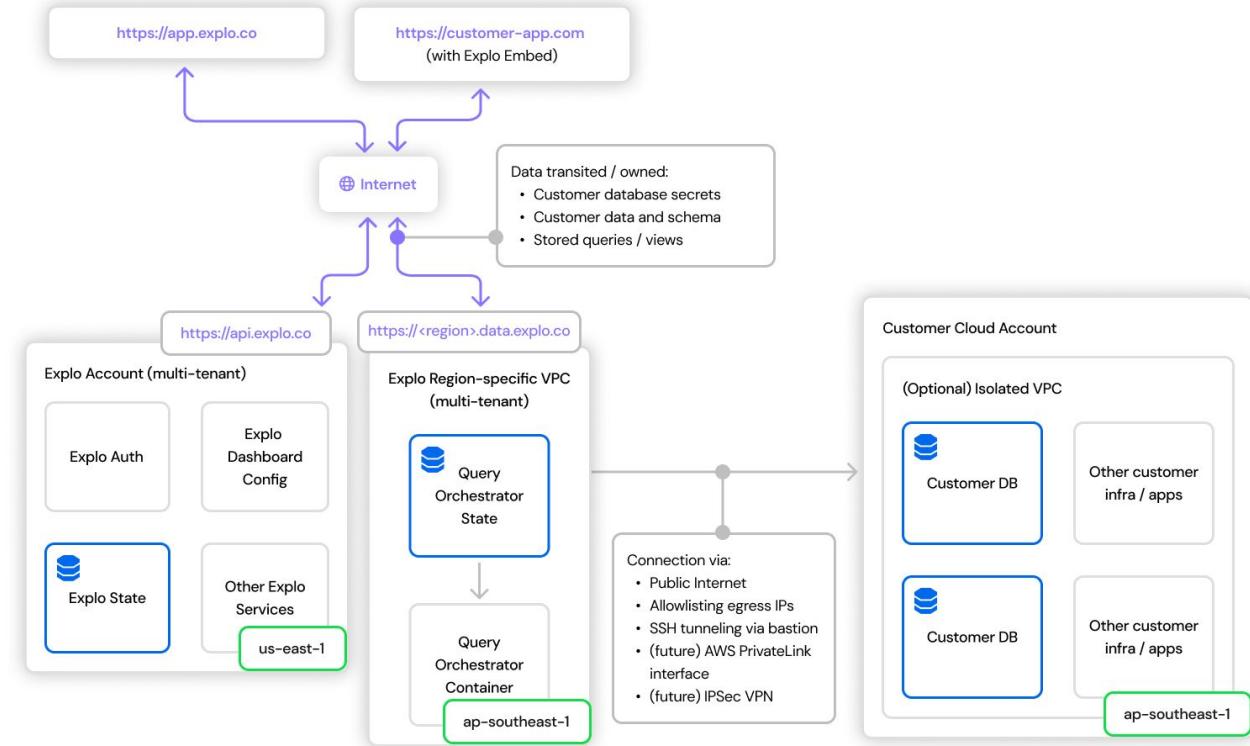
Data Orchestration v2





Benefits of this Architecture

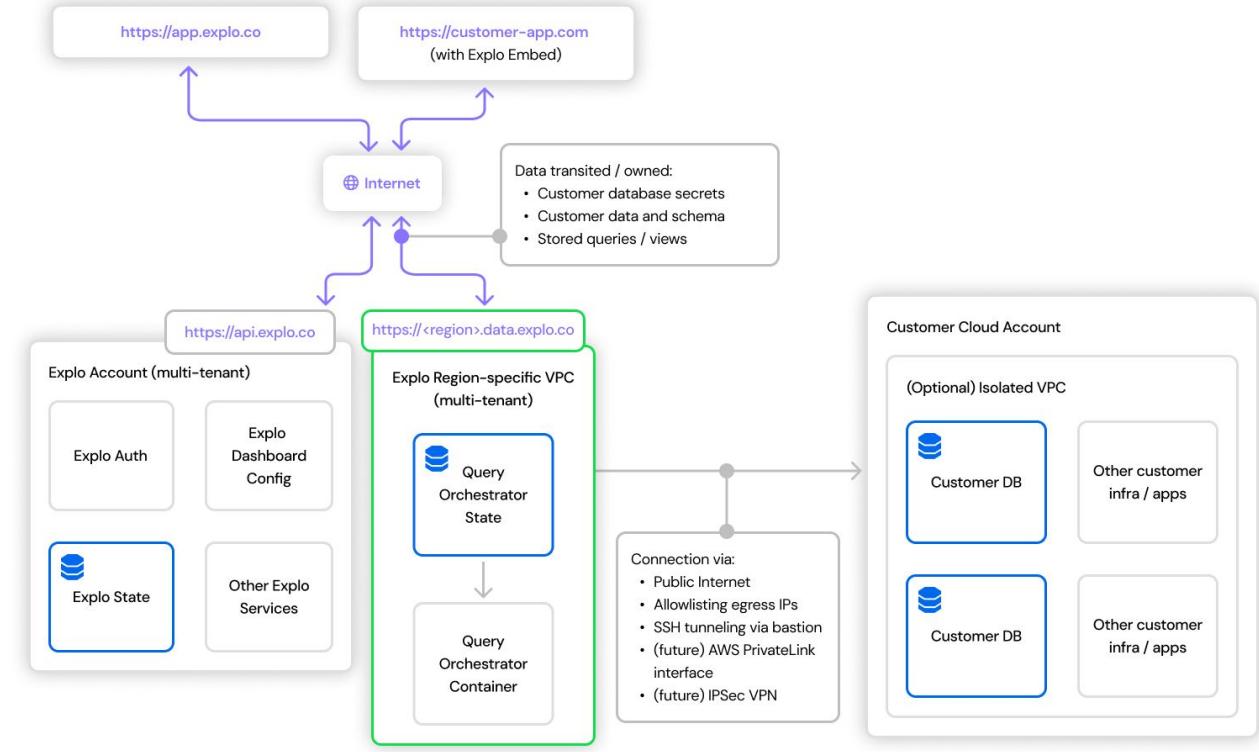
- Query service hosted in any region
 - Latency
 - Region specific data requirements (GDPR)





Benefits of this Architecture

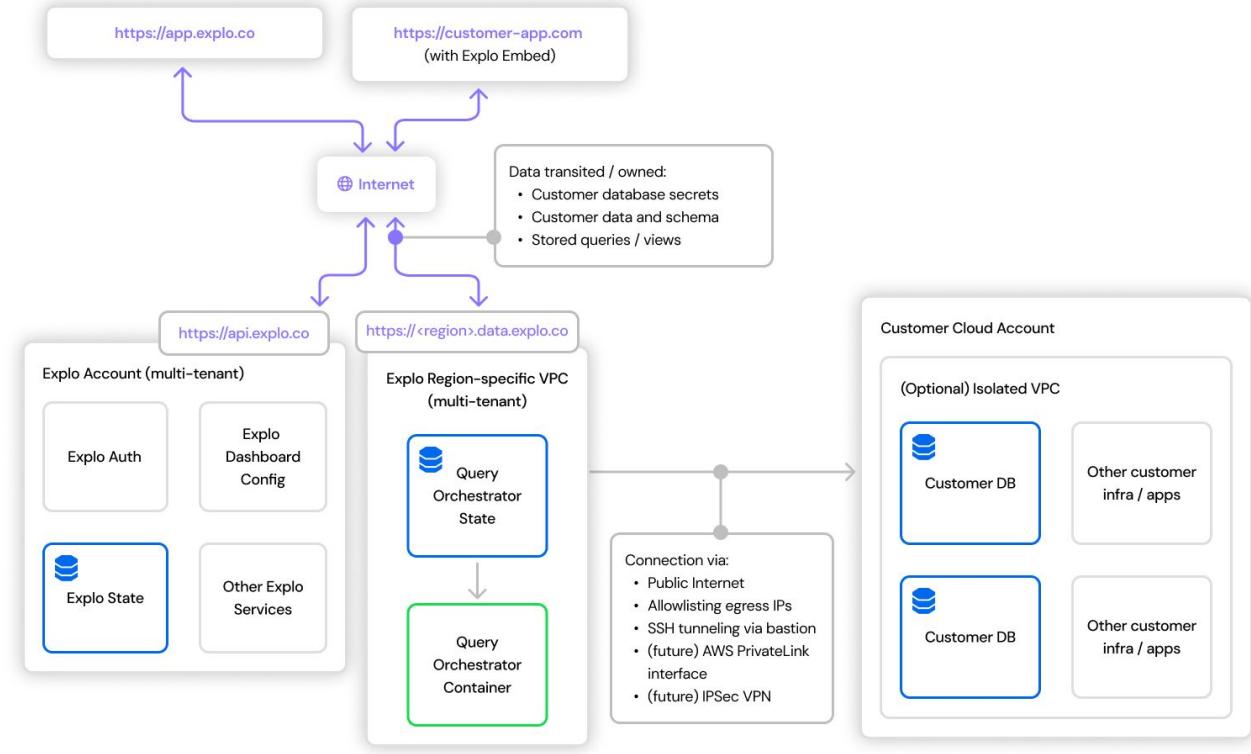
- Query service hosted in any region
 - Latency
 - Region specific data requirements (GDPR)
- On-prem query service





Benefits of this Architecture

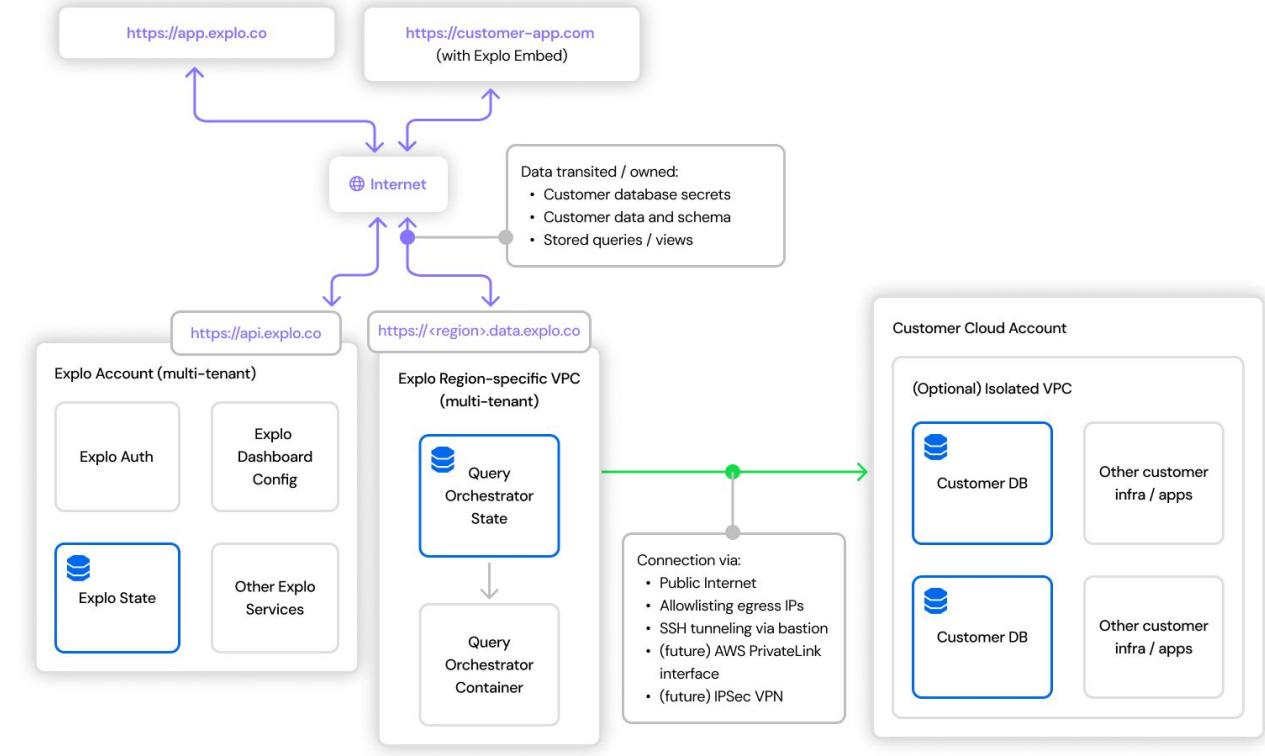
- Query service hosted in any region
 - Latency
 - Region specific data requirements (GDPR)
- On-prem query service
- Dedicated query service resources





Benefits of this Architecture

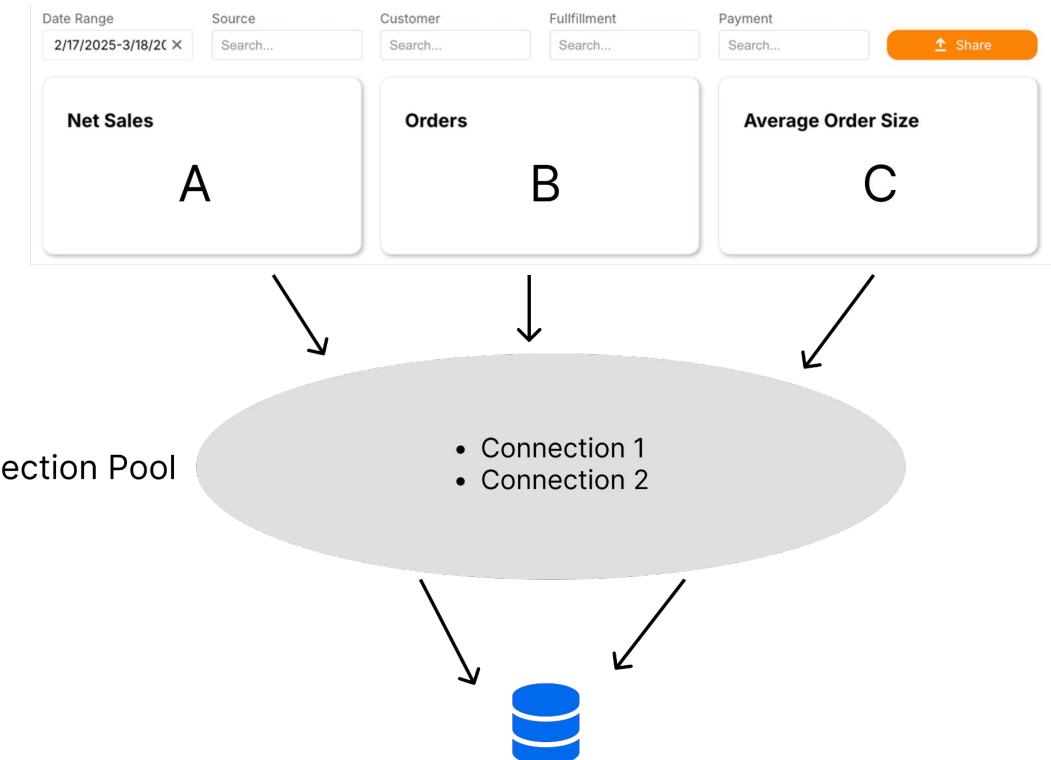
- Query service hosted in any region
 - Latency
 - Region specific data requirements (GDPR)
- On-prem query service
- Dedicated query service resources
- **Persistent connections**
 - Connection Pooling





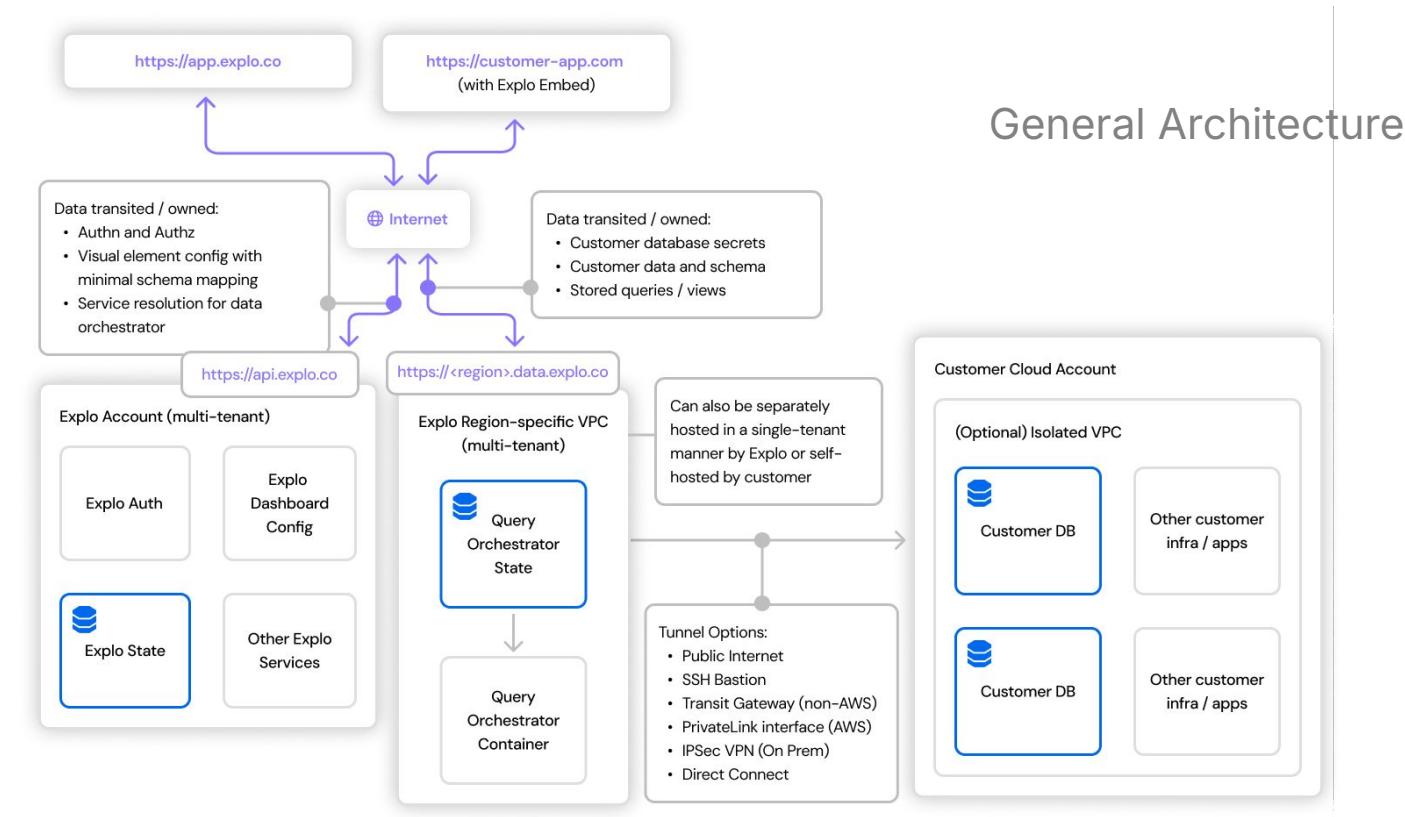
Benefits of this Architecture

- Query service hosted in any region
 - Latency
 - Region specific data requirements (GDPR)
- On-prem query service
- Dedicated query service resources
- **Persistent connections**
 - Connection Pooling



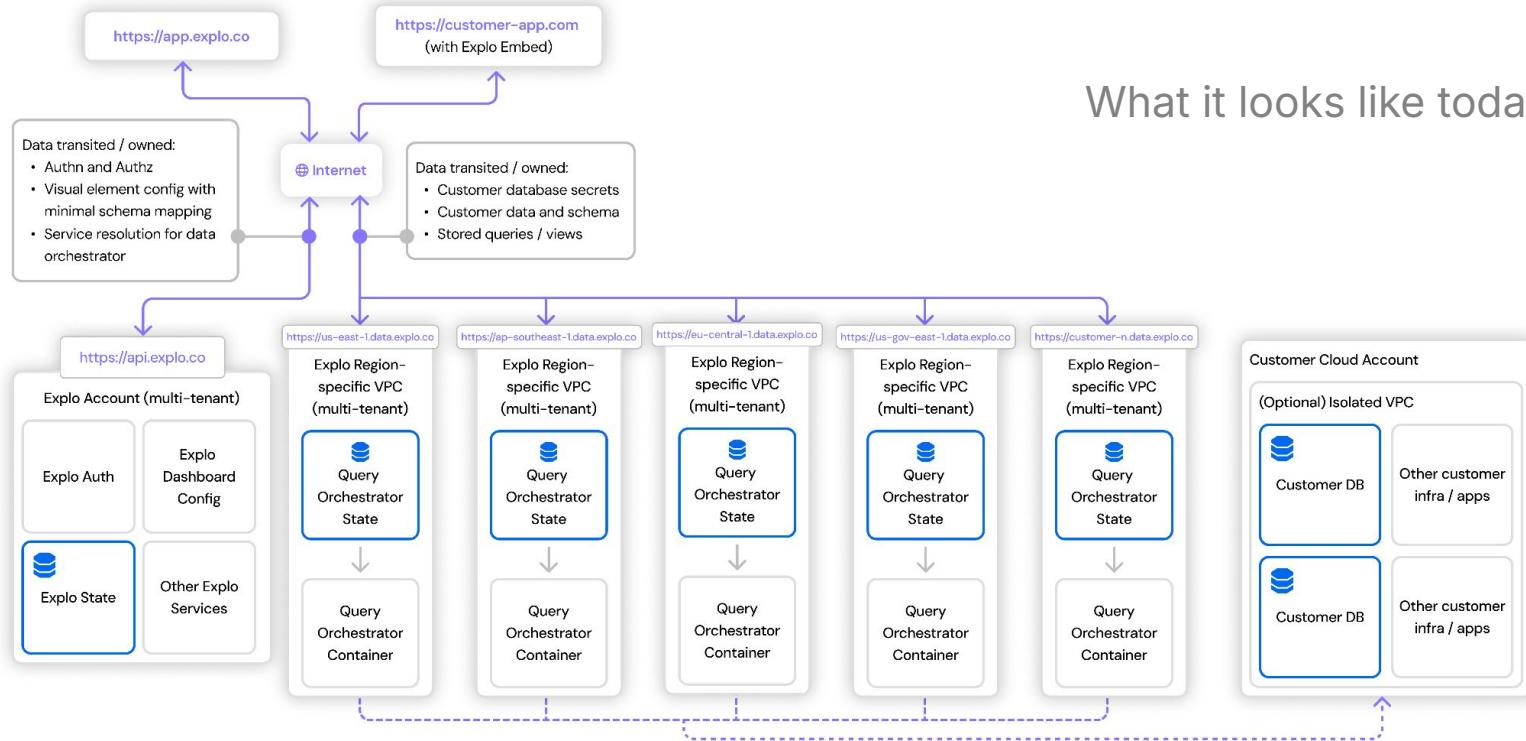


FIDO's International Coverage



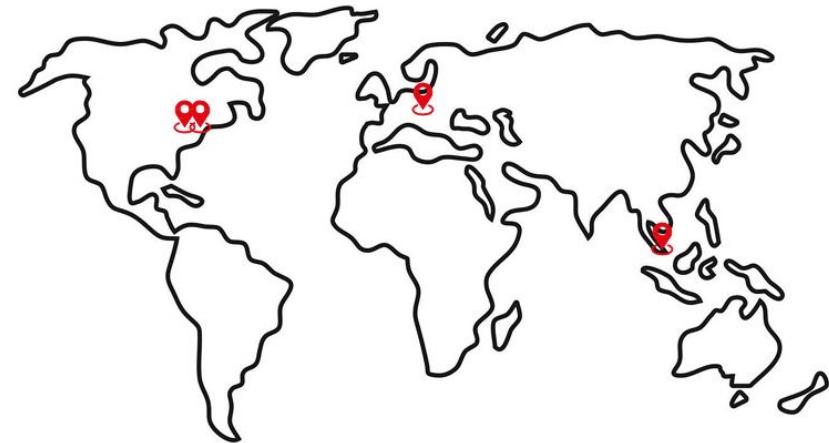
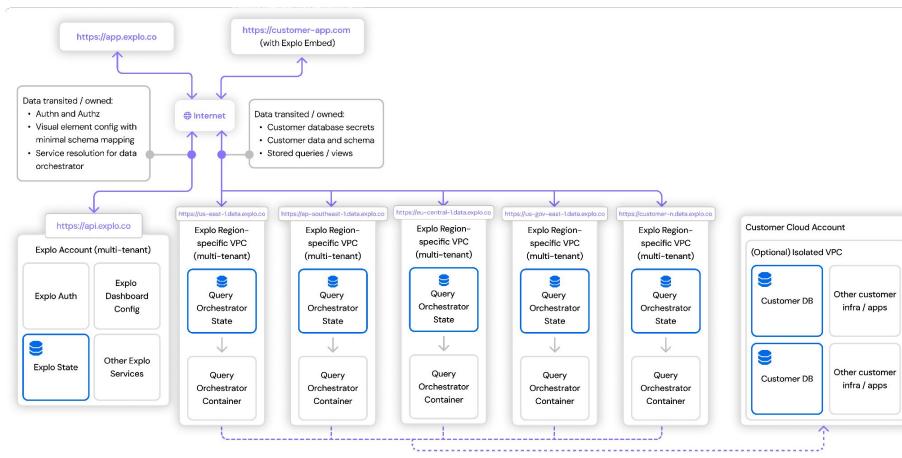


FIDO's International Coverage





FIDO's International Coverage





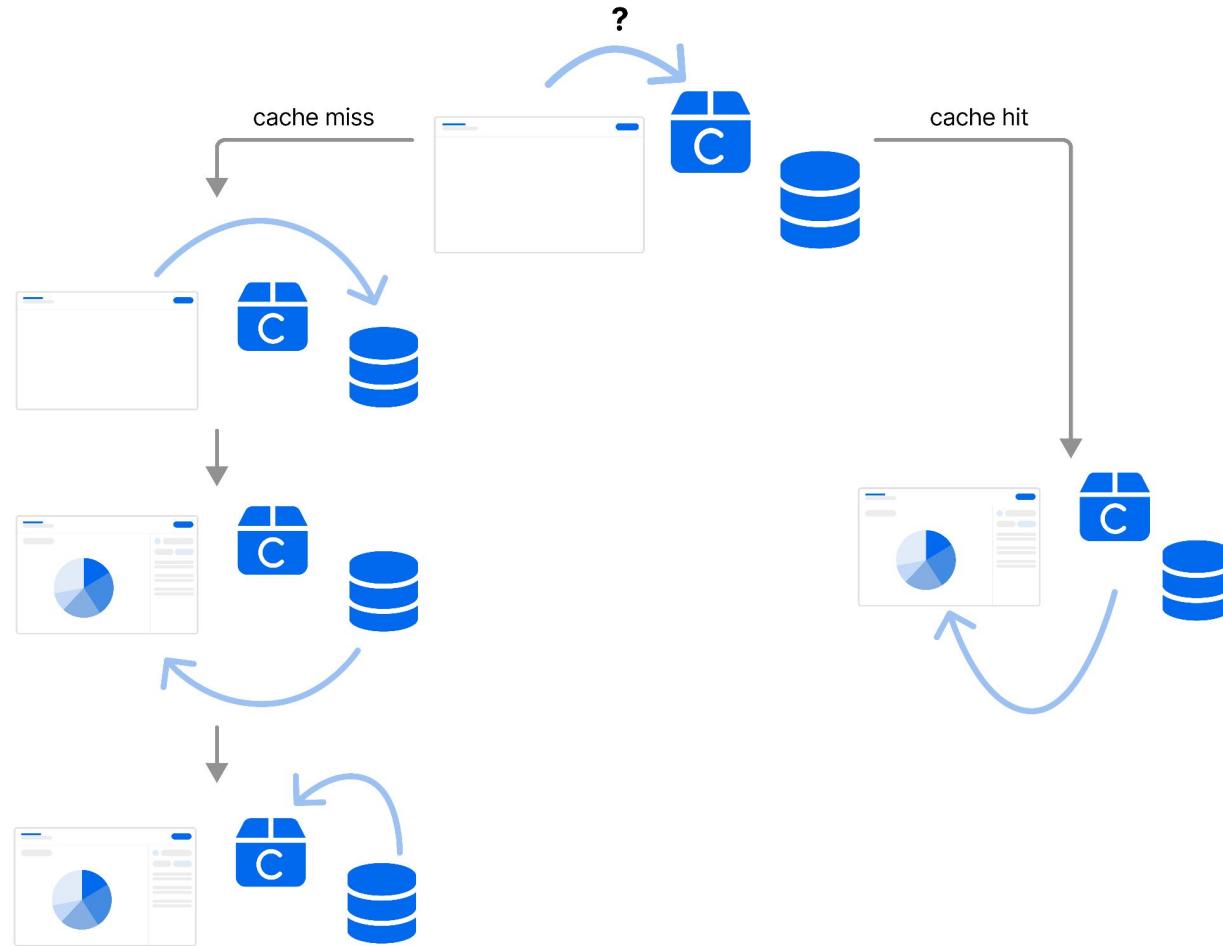
3

(We) Implement Caching



1) Caching Results

- Not too complicated
- Redis and Apache Ignite
- Queries are the cache keys
- Data is held in memory, never written to disk
- Default off (security reasons)
- Each FIDO instance handles its own caching (security)





2) Cache Warming



- Roadmapped!
- You want faster queries
 - Freshness matters within a range
- Pros
 - Fast with fresh (enough) data
- Cons
 - Expensive
 - Hard to warm everything
 - Relative time filters make this very hard

Customer 1



Customer 2



Customer n





TLDR



- ① Choose the Right DB (for you)
- ② Optimize Data Orchestration (asap)
- ③ Utilize Caching (correctly)



Questions?



Interested in trialing Explo?



email: gary@explo.co
twitter: @garymlin
website: explo.co