

The Data Infra Behind Zillow's 3x Growth in Experiment Volume



Aaron Wroblewski

Sr. Mgr, Experimentation



Zillow the Housing Super App

“Zillow”

Is more searched
than “Real Estate”
on Google

200M+

Avg. monthly
unique users



CURRENT HOME



63M

Tours
facilitated

4M

Buyers used
Zillow in 2021



NEW HOME



Accelerating Zillow's transition with Data Driven Innovation

Experimentation Challenges

Lack of infrastructure for scaling

Numerous 3rd party or custom-built tools

Inefficient processes leading to slow results
and lack of trust

Data Infrastructure
the backbone of
innovation
experimentation
personalization



Infrastructure Challenges

Limited support for self-service deployments

1000s of applications on shared infrastructure

Inefficient management of resources and scaling

Hindered productivity and velocity



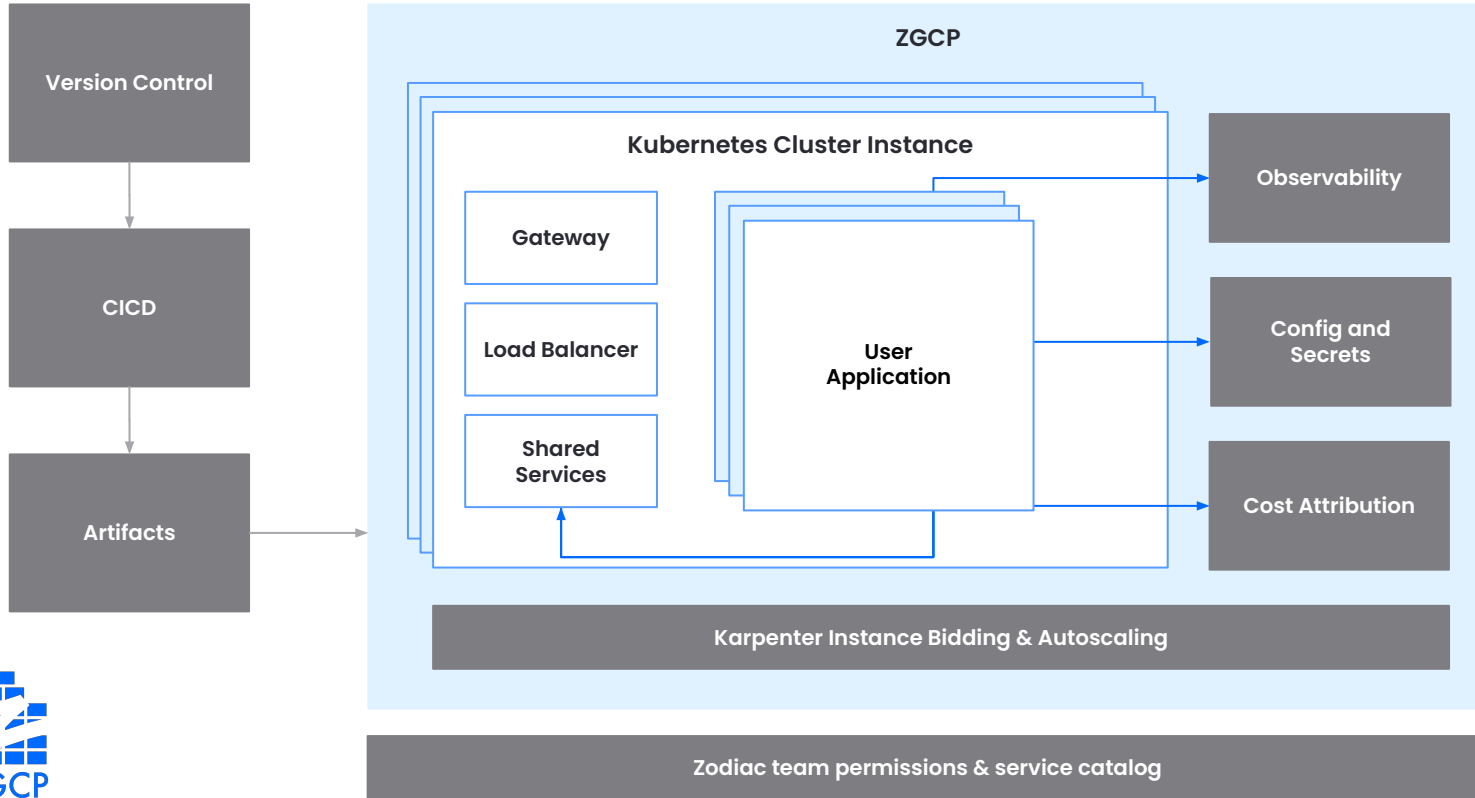
Kubernetes-based infrastructure

Dramatically increases productivity and velocity

Supports critical features: CI/CD, observability, config & secrets, autoscaling, team-level cost attribution

Archetypes for rapid service and application launch

Zillow Group Container Platform



Distributed Data Challenges

**Over 13,000 Spark jobs consuming 15% of
cloud budget**

**Teams spending significant time on cluster
management and scaling issues**

**Need for more efficient and seamless
workload & dependency management**

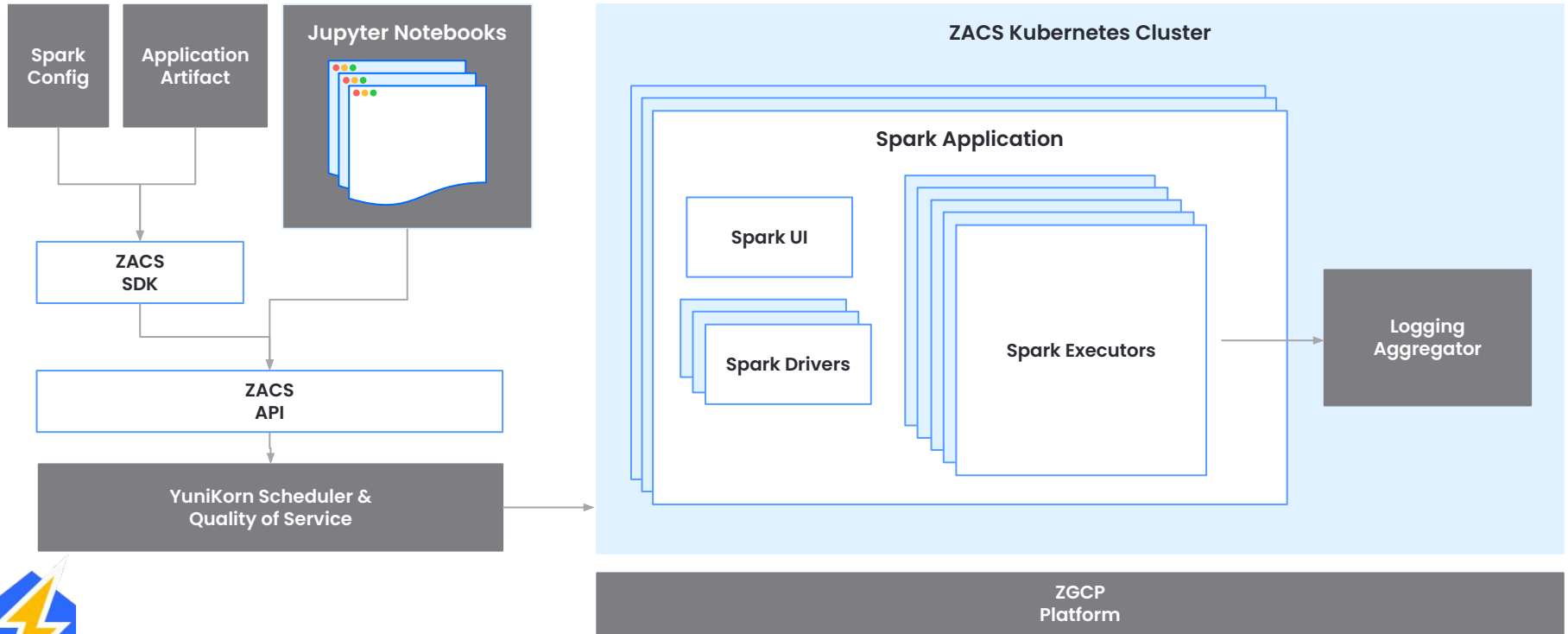


Convenient SDK and API for submitting workloads

YuniKorn scheduler for predictable capacity and cost management

JupyterLab support for ad-hoc exploration and interactive Spark jobs

Zillow Analytics Compute Services



Data Application Challenges

No streaming events support

**Lack of scalable & reliable solutions for event
ingestion and processing**

Weeks or Months to detect quality issues

**Data location, schema & quality were tribal
knowledge**



Data Portal



Self-service control center & data catalog



Schema registry



Contracts for defining tests over streaming and batch data



Data quality monitoring & anomaly detection

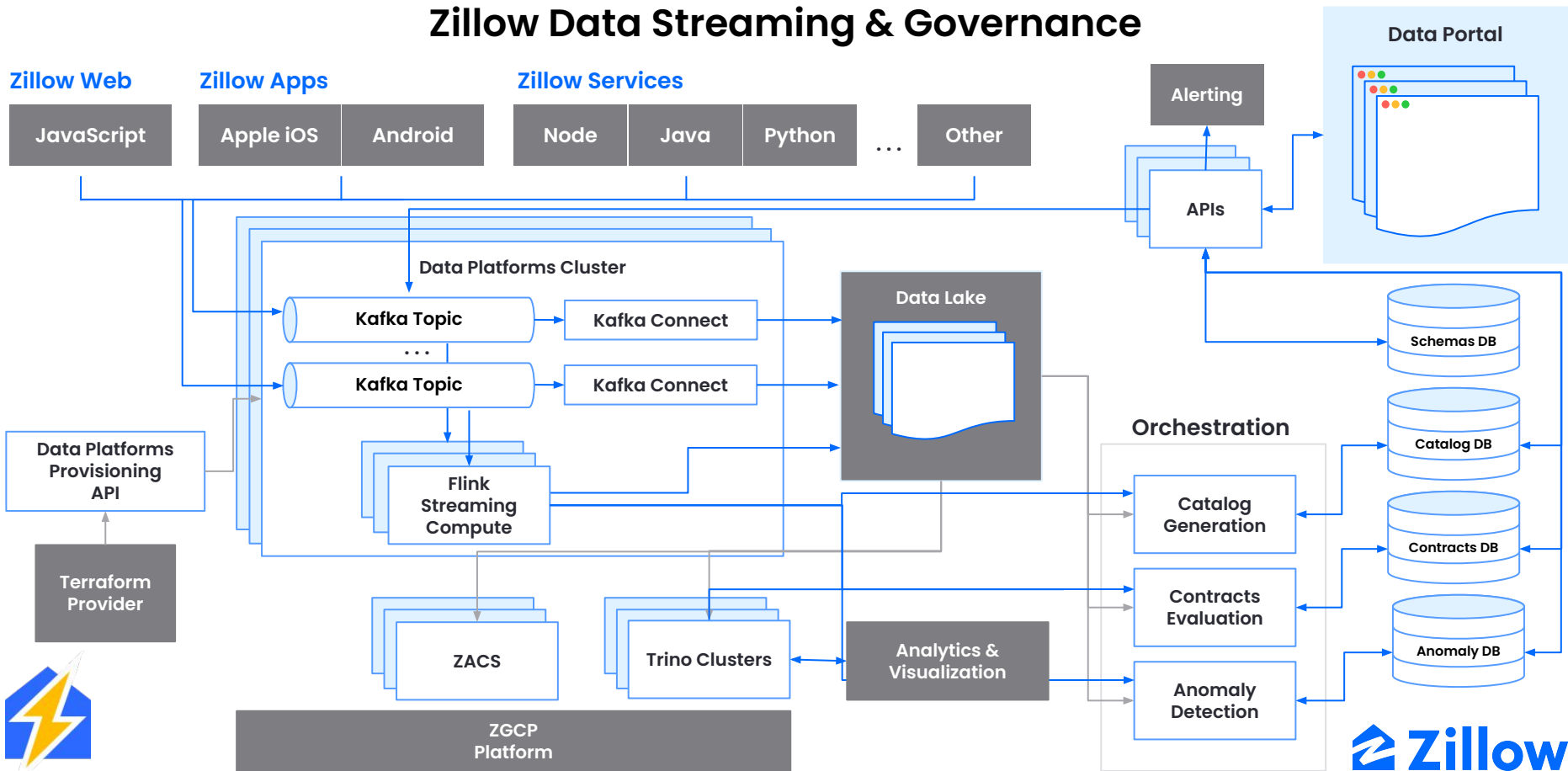


Easy provisioning of topics, tables, and streaming jobs

Schema validation

Efficient real-time & batch processing of validated data

Zillow Data Streaming & Governance



Accelerating Zillow's transition with Data Driven Innovation

Self-service, comprehensive tools

Empowering data producers to own quality

Scalable compute and data production



Experimentation Challenges

~~Lack of infrastructure for scaling~~

Numerous 3rd party or custom-built tools

Lack of viz into of cross-LOB impact

Inefficient processes leading to slow results
and lack of trust



Increase velocity, trust, and transparency

Single system for all brands & experiences

Hybrid build & buy strategy

ZEXPy: Flexible, Trustworthy, & Reliable Analysis



Assignment, exposure, and metric data sources

Granular fault tolerance, checkpointing and automatic backfill

Self-service customization & open to contribution

FastAB: Single Source of Truth API for All Experiments

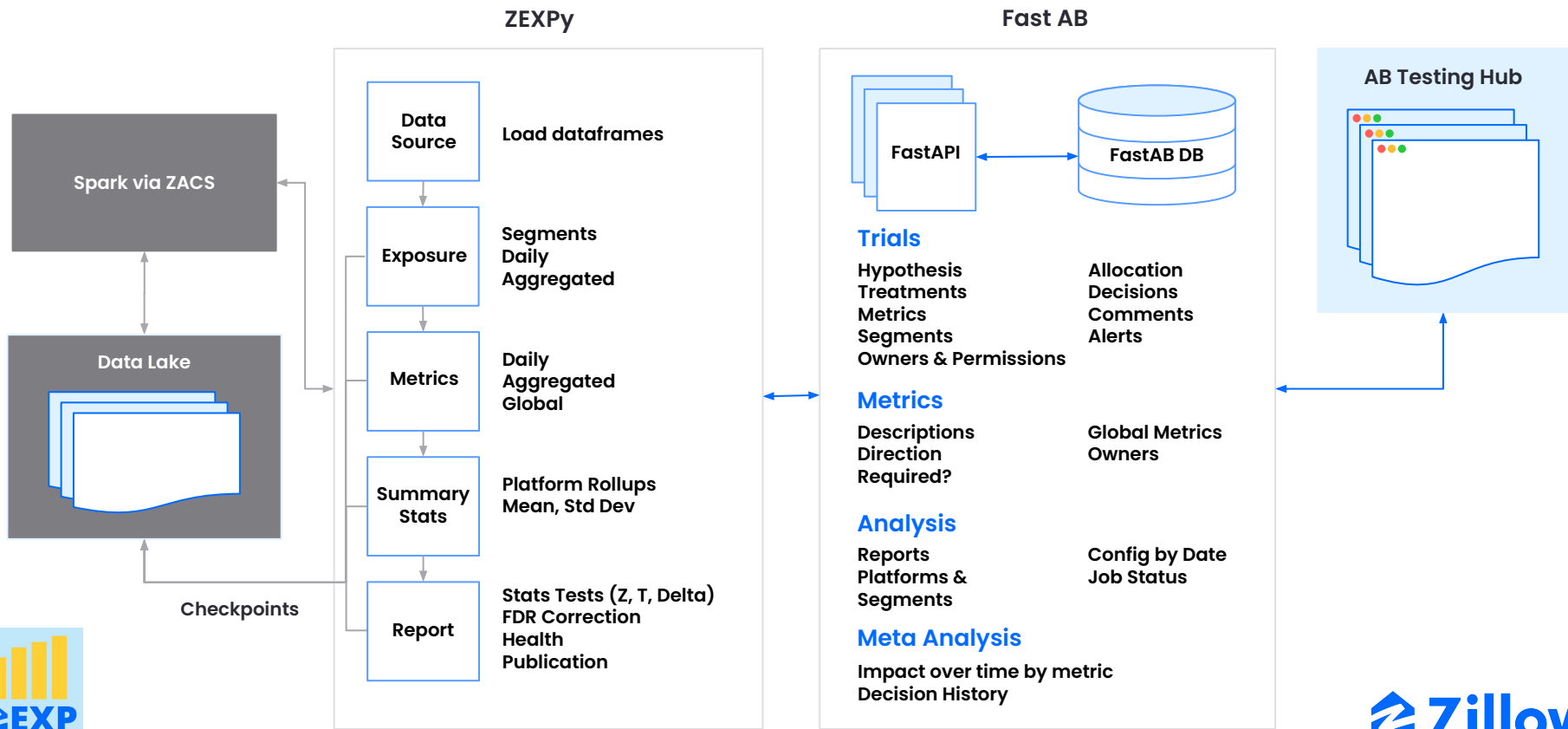


Enables experiment transparency and governance

Job configuration & status tracking for ZEXPy orchestration

Deployed as a ZGCP service for seamless integration

ZEXP Analysis



AB Testing Hub: Self-Service Experimentation



Create & configure experiments

Monitor experiment health, results, and alerts

Learn from past experiments: hypothesis, goal metrics, and outcomes

AB_DASHBOARD_AA_TEST - / x +

← → ↻

Data Portal Search all ZG data

☰

☺

< All AB Tests

AB_DASHBOARD_AA_TEST
A continually-running test to ensure that the entirety of the A/B pipeline is functioning correctly and current A/B results are valid

📄 AB Test Contract

✍ Complete Results Summary

👤 zexp

Owner
Aaron Wroblewski
Senior Manager, Machine Learning

Reviewer
Ryan Thomas Williamson
Senior Data Scientist

Slot: 10
Custom Dimension: 50

CONTROL 50%

ON 50%

📄 Change Allocations

📄 Make Decision

🚨 Report a Problem

Results | Comments 9 | Configuration | Decisions | History | Health | Alerts

Treatment: ON | Platform: zillow_all | Segment: all | Report Date Range: 2023-03-06 - 2023-03-13 7 day(s) | Report Version: BDO | Columns: Control +4 | Duration & Allocation: Show Widget

Multi Select

Only Significant

All Metrics 5

Core Metric

Type to filter metrics in selected groups (for example "PaidNCHDPPVs|Visits", or "Paid PV")

Metric Name	Control	Treatment	Weekly Impact	Global Lift	Relative Lift CI
					-1% -0.5% 0% 0.5% 1%

☺ ☺ ☺ ☺ ☺



Impact Dashboard - AB Testing x +

Data Portal Search all ZG data

Impact Dashboard

Metric: HomeSaves mean Channel: onsite Platform: zillow_all Segment: all Owner: Select Owners Report Date Range: 2023-02-12 - 2023-03-14

Report Version: ZEXPy_v2 Columns: Start Date +5 Include SRM: [toggle] Max P Value: 0.05 Max FDR P Value: [input] Concluded Only: [toggle] Shipped Only: [toggle] Min Lift: [input] % Max Lift: [input] %

HomeSaves: Home Saves

Trial Name	Treatment Name	Start Date	End Date	Control	Treatment	Global Lift	Relative Lift CI
							-5% -4% -2% 0% 2% 4% 5%



Home - Data Portal

Data Portal Search all ZG data

Job Statuses

ab-onsite-zexpy-v2 AA_TEST Exclude Start Date End Date State Rerun All Visible

Include Details Auto Refresh

Name	Run Date	Updated Date	State	Mark For Rerun
> report BDO AB_DASHBOARD_AA_TEST zillow_app_all all	2023-03-19	2023-03-20 7:07:55 pm	success	
> report BDO AB_DASHBOARD_AA_TEST zillow_web_all all	2023-03-19	2023-03-20 7:07:44 pm	success	
> report BDO AB_DASHBOARD_AA_TEST zillow_all all	2023-03-19	2023-03-20 7:07:21 pm	success	
> report BDO AB_DASHBOARD_AA_TEST zillow_android all	2023-03-19	2023-03-20 7:07:09 pm	success	

Future Work

Culture

Trial Duration

Inner Source

Metrics

Meta Analysis



Accelerating Zillow's transition with Data Driven Innovation

Zillow's Compute & Data Platforms

- ZGCP for scalable compute
- ZACS for scalable distributed data applications
- Data Portal for self-service quality & discovery
- Streamz for scalable streaming

ZEXP

- ZEXPy for scalable analysis
- FastAB as the single source of truth
- AB Testing Hub Self-service

Thank you!

**Find me online
@aaronwro**

