

Palette Metadata Journey

Paarth Chothani, Staff Software Eng, Uber

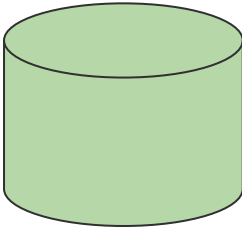
Nicholas Marcott, Staff Software Eng TLM, Uber

Michelangelo Palette

Feature Preparation
Batch + Streaming ETLs



Palette Feature Store
Online serving
Offline serving



Feature Discovery
Sharing across Models
Automatic feature selection



Monitoring
Pipeline & Data Quality



Michelangelo Transformer
Model specific feature xforms



History

Shared
Repository

Online
Serving

Offline
Serving

Scalability

Dispersals
Optimized Serving

Near real-time features

**Automatic
Feature
Selection**

Data Quality

Feature APIs

Embeddings

Reliability

2016

2023

Uber

Palette at Scale

Online Serving

10s of Millions of QPS at P99 < 10ms

Offline Serving

100s of TBs per training run

Pipeline Scale

1000s of Feature Pipelines

Metadata Changes

1000s of Metadata related changes

Problems with Metadata

Unreliable

Bad metadata crashing live services

Slow developer velocity

Difficult and slow to manage

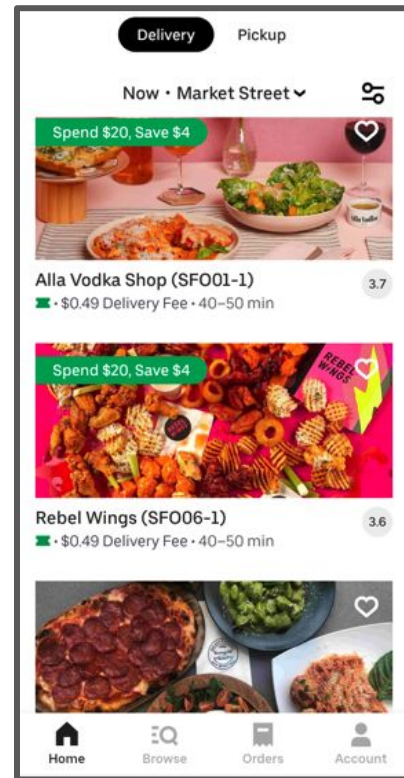
Customer dissatisfaction

User confusion

Problems with Metadata

System misbehavior

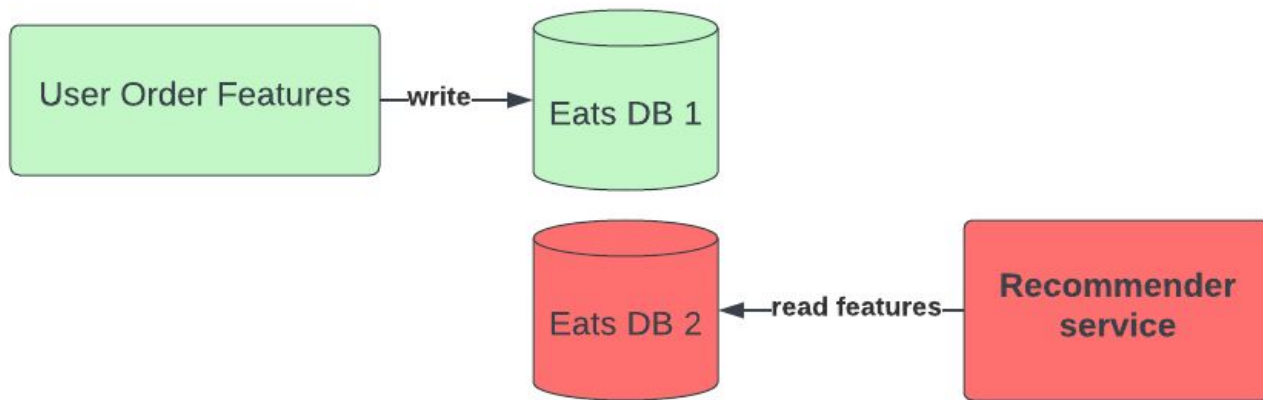
Bad metadata affecting downstream services



Problems with Metadata

Split brain system

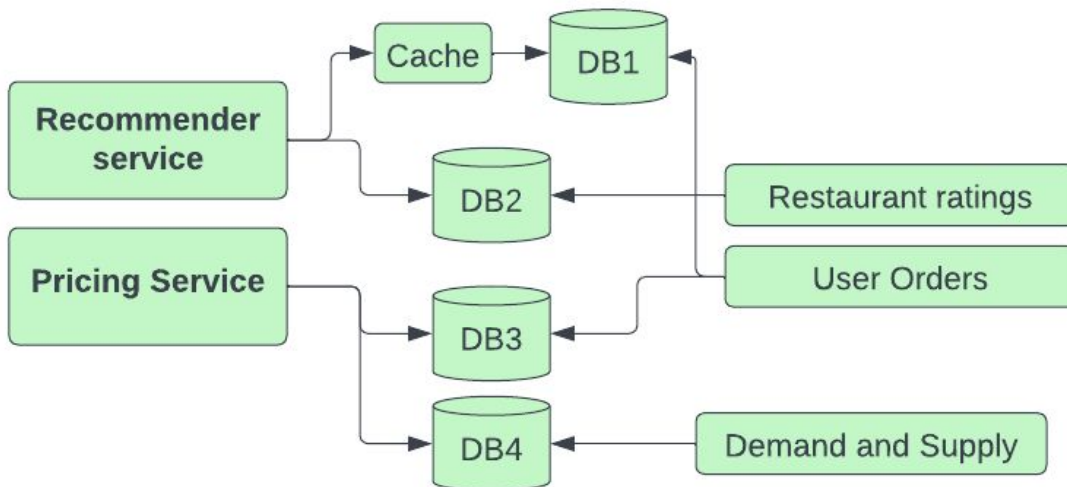
Disagreement between system metadata



Problems with Metadata

High operational complexity

Difficult and slow to manage



Solutions

Validation

Schema driven with additional validation to maintain system health

Unified

One source of truth to drive complex multi-component system

Dynamic and Automated

Reduce the number of operations and time to deploy

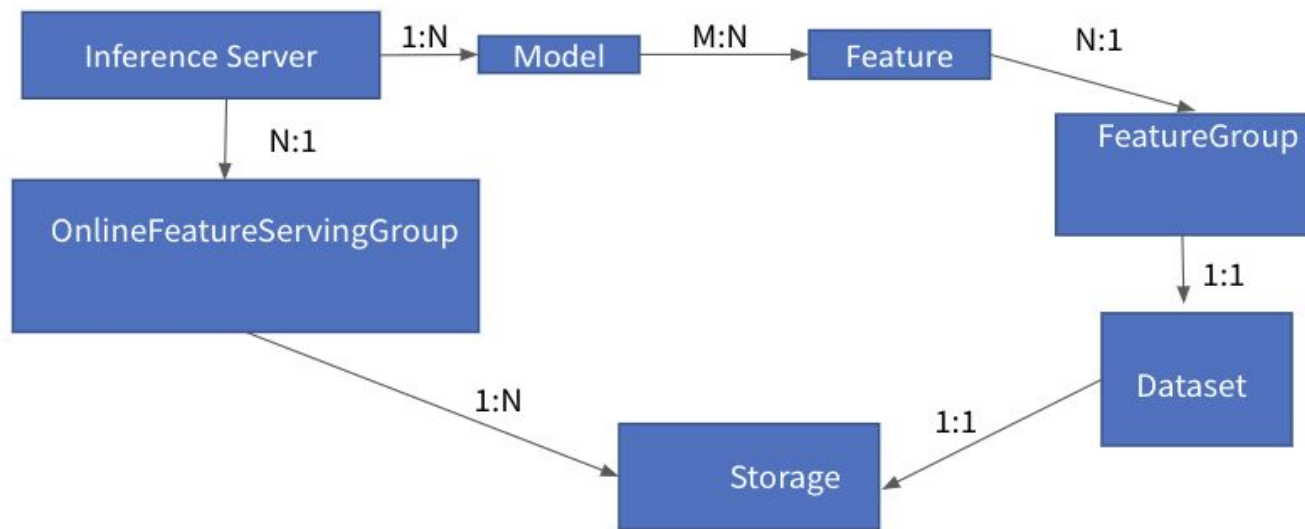
New Abstractions

Better primitives for modelling problem space

Solution: Feature Schema and APIs

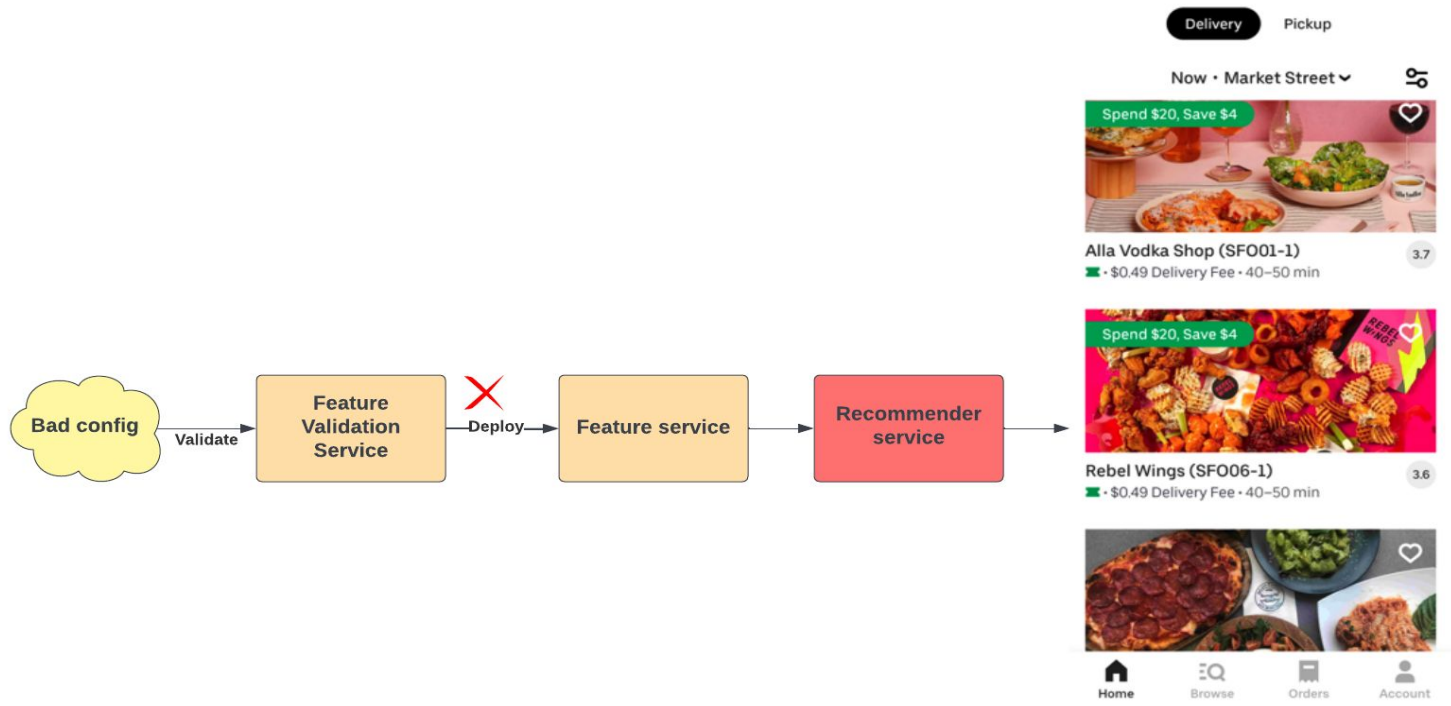
- APIs for First-Class Entities
- Support for both UI and code driven interactions
- Consistency and No fragmentation
- Extensibility for future

Solution: Feature Schema and APIs

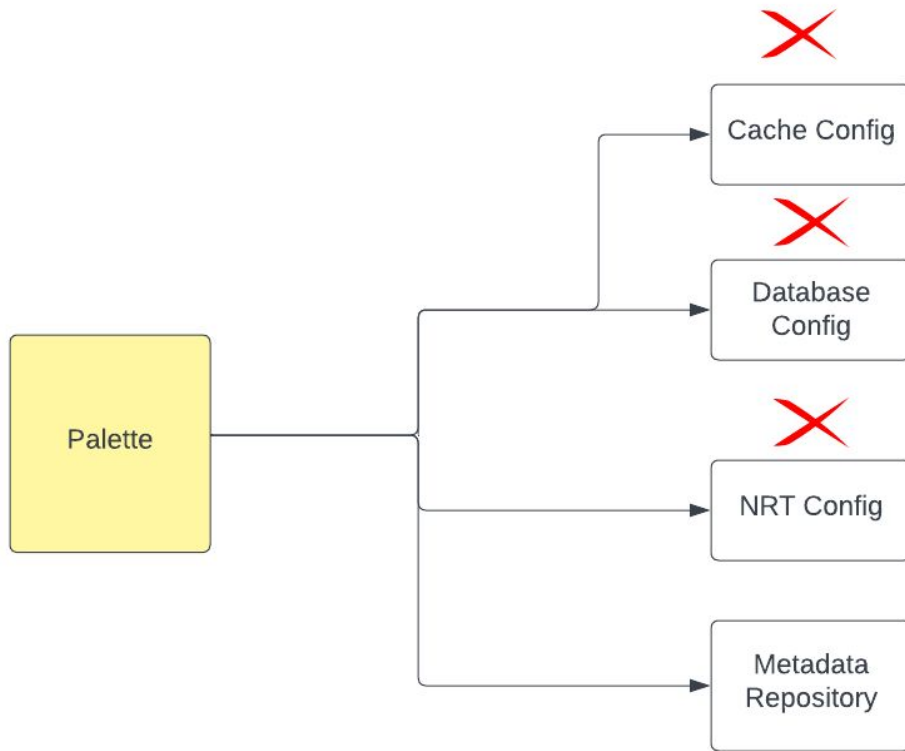


1:1*: Commonly 1:1 but in some situations 1:2+

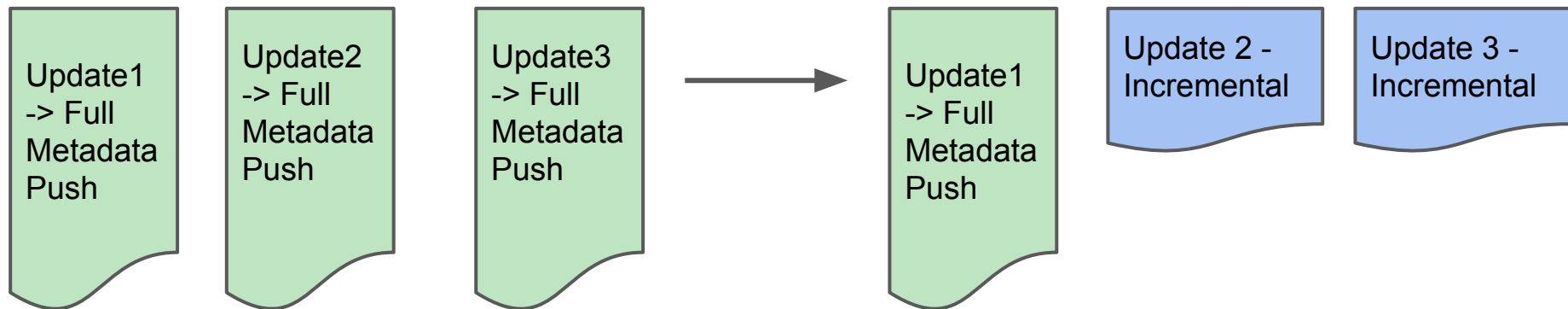
Solution: Validation



Solution: Unified

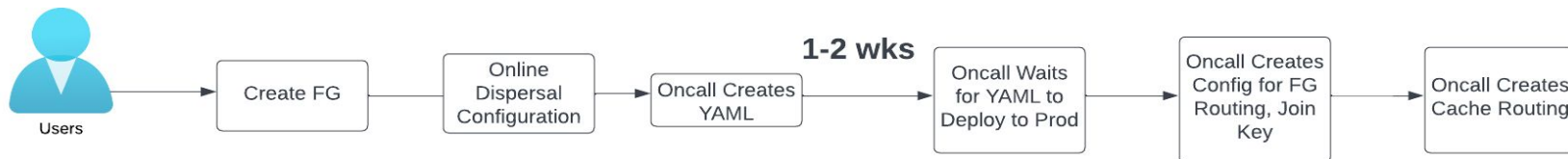


Solution: Dynamic and Incremental



Solution: Automated

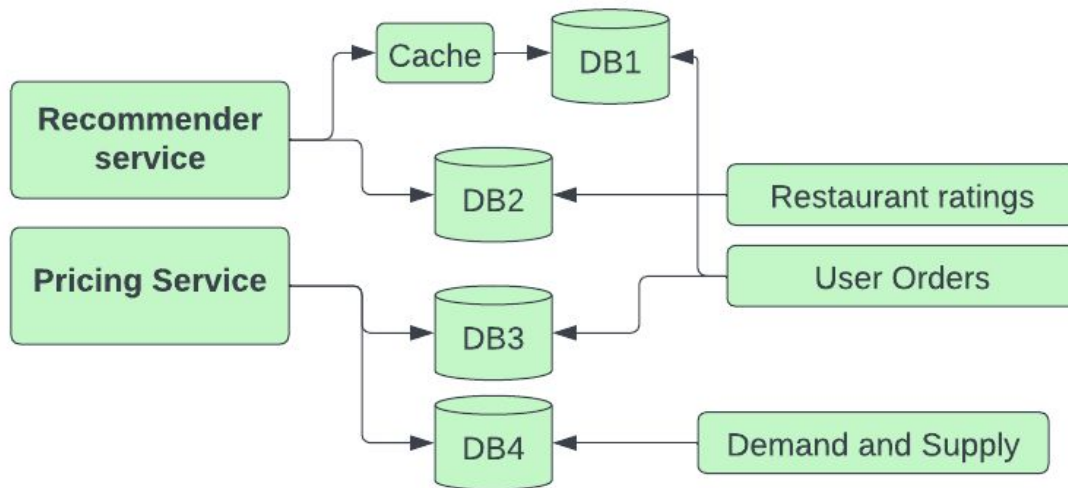
Old Flow



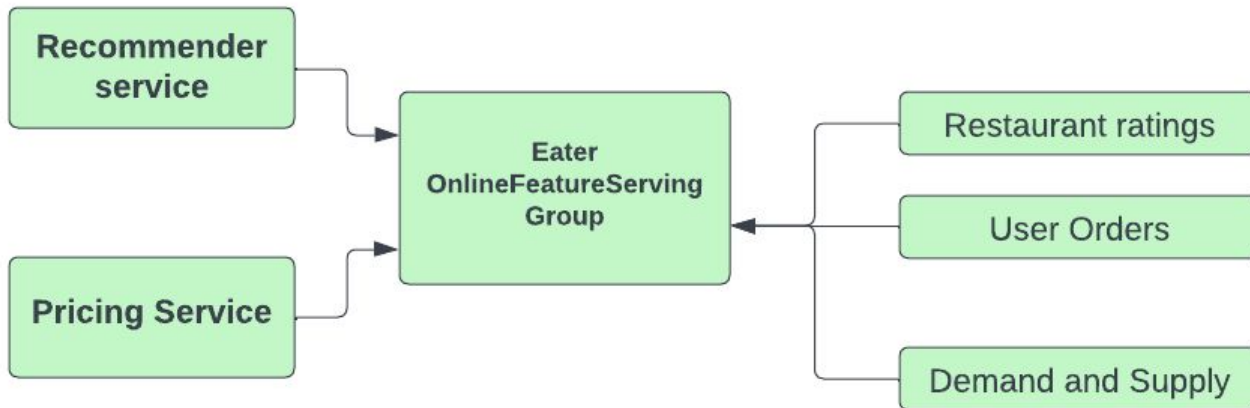
New Flow



No OnlineFeatureServingGroup

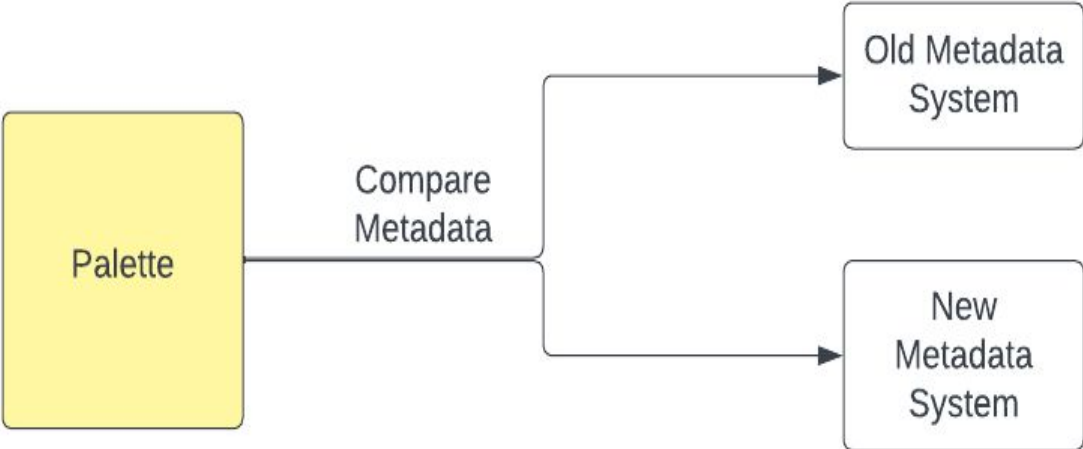


Solution: New Abstractions



Migration Challenges

- Data Integrity
- Metrics Comparison
- Thorough Testing
- Rollback Plan
- Automated Backfills
- Clear Onboarding Docs



Conclusion

- Ease of Use
- Reduces maintenance
- Reduces Operational costs
- Reduces Deployment time
- Customer onboarding time

Thank you

And much thanks to all the Palette team members:

- Xiyuan Feng
- Dehua Lai
- Divya Nagar
- Victoria Wu
- Chunhao Zhang