



FEATURE STORE SUMMIT

12-13 OCTOBER | 08:30 AM - 4:00 PM PT

ORGANIZED BY HOPSWORKS



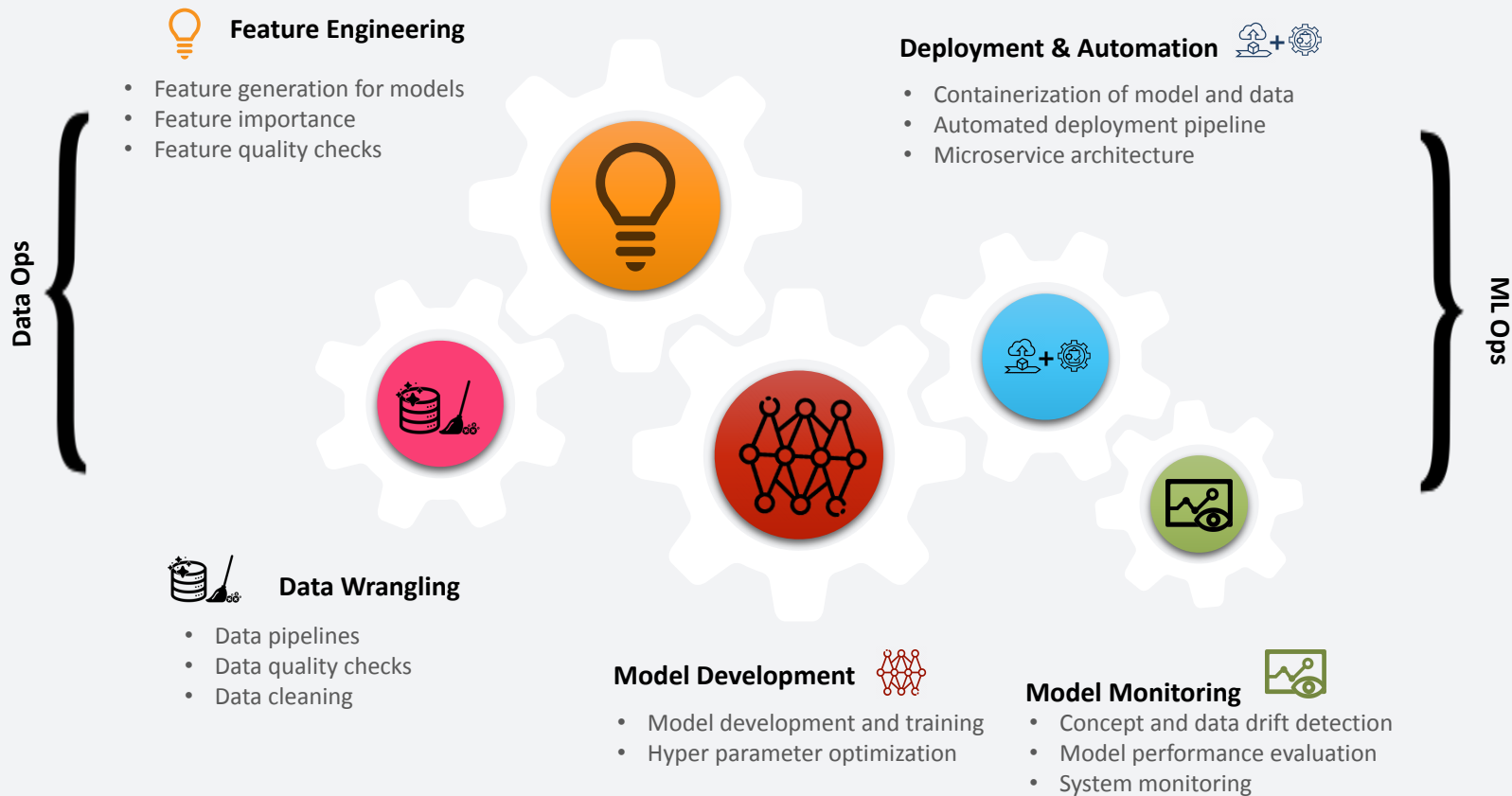
Feature Store

The Heart of ML Ops

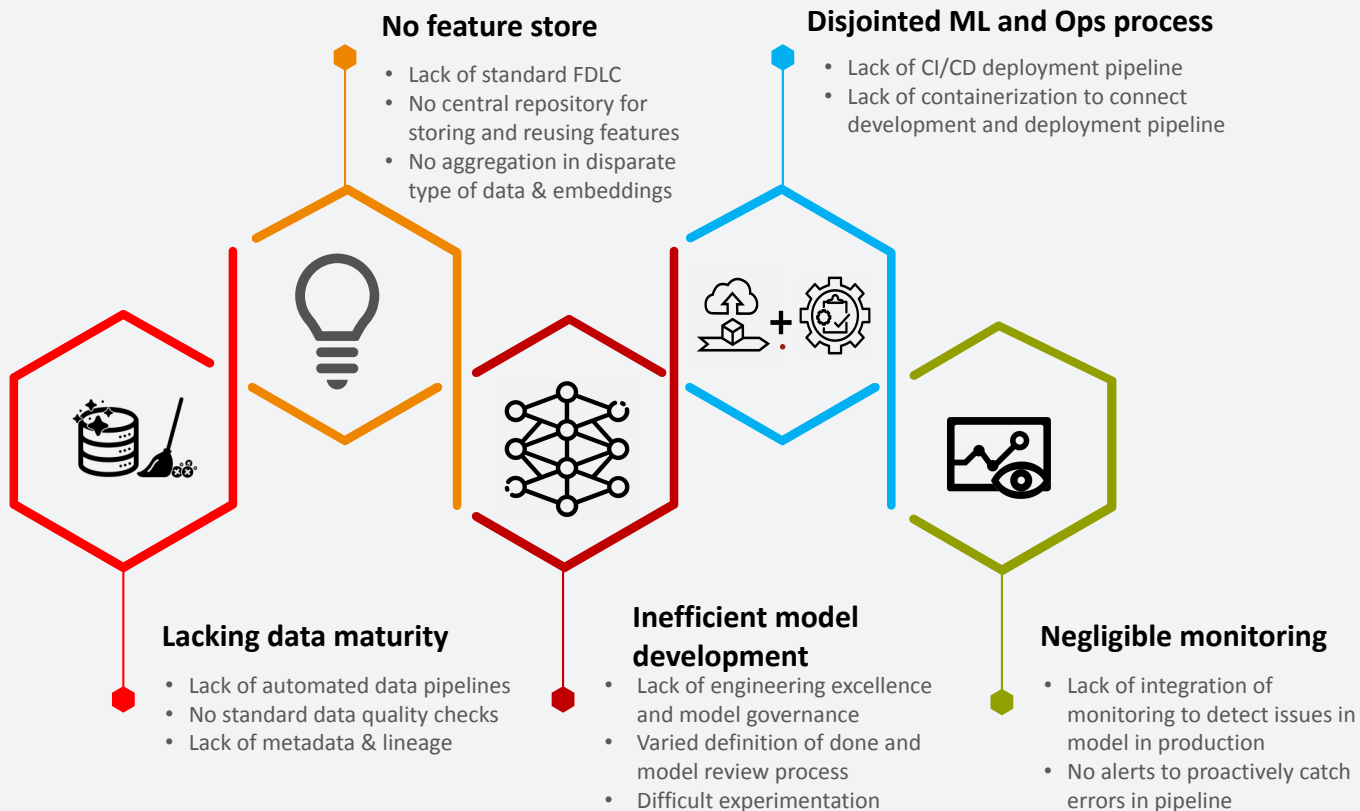


Richa Sachdev
Head of Machine
Learning Engineering
Vanguard

Model Ops – Bringing model development and deployment closer



Challenges we face to achieve an ideal process



Rubik's Cube – Our feature store



SEARCH AND REUSE OF DATA

- Connected to centralized metadata repository
- Sharing of data for shorter cycle time to train in several teams



AGGREGATION OF DISPARATE DATA SOURCES

- Features from structured and unstructured data all in one platform
- Easy querying of features across multiple data sources



FEATURE STANDARDIZATION

- Feature creation by agreed upon business logic to ensure one gold copy
- One format/nomenclature to store and access features across departments
- Standard data quality framework



EASY ACCESS AND SCALABILITY



- Features as a service available via API and querying against one platform
- Low latency in querying data by choosing smart DB to store features and via scalable API

ALERTS AND CONSUMPTION



- Alerts to multiple teams regarding feature refreshes and failures
- Monitoring in place to catch problems with feature pipelines, data quality and model performance, drifts etc.
- Reports to show consumption to measure usability

GOVERN, ACCESS CONTROL AND VERSION CONTROL



- Version control enforced to track changes
- Access control and security set up to grant correct level of access
- Proper governance process integrated to ensure correct usage of data for model deployment and consumption

Solving ML Ops challenges with a holistic feature store design

DATA MANAGEMENT/OPS

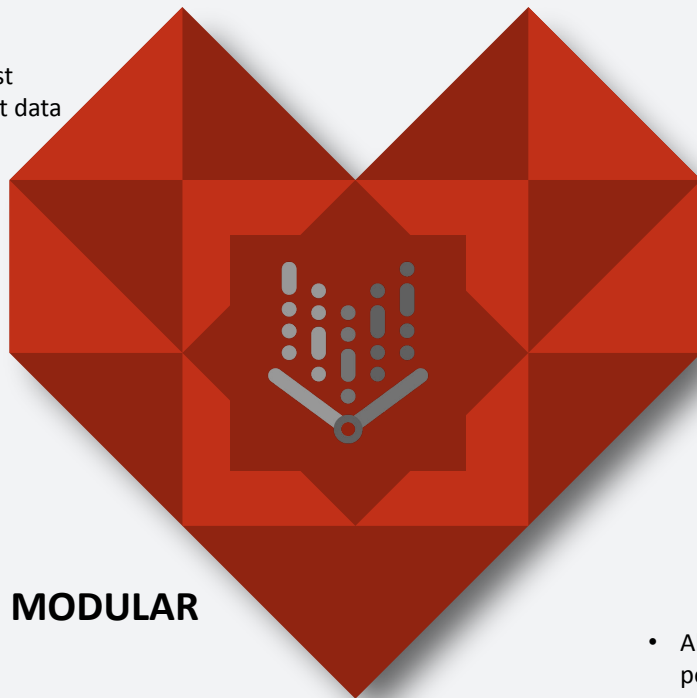
- Standard definition of gold copy of data is a must
- Automation and workflow management of input data
- Automated data quality checks and reports
- Mandatory metadata and lineage

FEATURE STANDARDIZATION

- What classifies as a feature
- Automated feature creation pipeline
- Feature drift and quality verification
- Embeddings and time series data

REUSABLE AND MODULAR CODE

- Publishing and using standard coding templates
- Centralized repo of code connected to model development platform
- Reusable features from a centralized repo



CI/CD BATCH AND API PIPELINES

- Containerization of features and model
- One click deploy and change record
- Blue/Green deployment
- Workflow orchestrator and scheduler to handle interdependencies

MODEL, DATA, RISK MONITORING

- Model registry for up-to-date model metadata
- Mandatory review as per enterprise model and risk governance

ALERTS AND USAGE

- Alerts to detect drift, bias, performance deterioration in models
- Tracking of end user/application consumption of model insights



Thank you!

Do you have any questions?

Speaker Info



<https://www.linkedin.com/in/richasachdev/>