Feeding The World One Plant At A Time

Presented by:

Fahad Khan, Head of Product, Data & CVML Platform Naveed Farooqui, Sr. Director, Data & CVML Platform

June 2023

Spraying the whole field?

"Would anyone want to do it if they had a choice?"







A CVML Accelerator Platform



Integrated Ecosystem of Infrastructure, Data, Tools, Applications and Vendor Workflows





Go-to Platform for Data and Computer-Vision Based ML





Who are our customers?



Our Users and their goals







Program Engineering

Enable Intelligent machinery by rapidly using advanced ML algorithms and vision technology

OHN DEERE

Platform Engineering

Provide a secure and scalable ecosystem for Computer Vision Machine Learning lifecycle

Business Owners

Track and align value delivery and cost effectiveness of program offerings



Technical product development challenges



TECHNOLOGY

How are we solving these problems for our users?



IN DEERE

Log Analytics

Playback messages to understand machine behavior / errors Re-play to verify software fixes Annual report to verify savings

> Model Evaluation / Release Management (models, datasets, code versions), Reproducibility, Release, Monitor Performance

Model Development

Creating datasets, training, testing (base, test, eval) against models, What/if scenarios



The New Platform Stack



Databricks: The Data Enablement Foundation



- Flexible Databricks Infrastructure Independent Progran Team Workspaces
- Platform Team manages the lifecycle of all CVML entities
- Data is stored in Platform-Managed Program Workspaces
- Goverance of data is via a shared, single Unity Catalog
- Infra as code takes care of managing AWS, Databricks and 3 rd party integrations
- Tracking dashboards provide necessary oversight





Use-Case Walkthrough





CVML Lifecycle







Machine Log Ingestion







Common Ingestion Pipeline



CVML Lifecycle

OHN DEERE





Image Curation and Labeling workflows







CVML Lifecycle







CVML Pipeline: Self Driving Use-Case







CVML Pipeline: Self Driving Use-Case







Distributed Data Processing and Training pipeline



Promising initial results, iterating and testing further

A streamlined data management strategy is the key to successful ML offerings. Databricks has already proven to help us significantly with its lakehouse offering in this regard

Today's AI systems are hybrid offerings with purposeful choices on compute. We have a compelling need to build a common orchestration layer between on-prem and Cloud for machine learning.

What was not possible a few years ago is becoming increasing possible! Ping us to share your success endeavours on a hybrid ML ecosystem

We would 💙 to hear from you!





Accelerating a flywheel for innovation at scale



Help us revolutionize Agriculture!







BLUE RIVER TECHNOLOGY