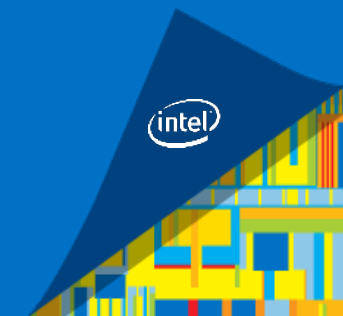
HIVE: Scalable, Cross Platform Graph Analytics Framework in Python

Vincent Cavé - Intel

Stanley Seibert - Anaconda

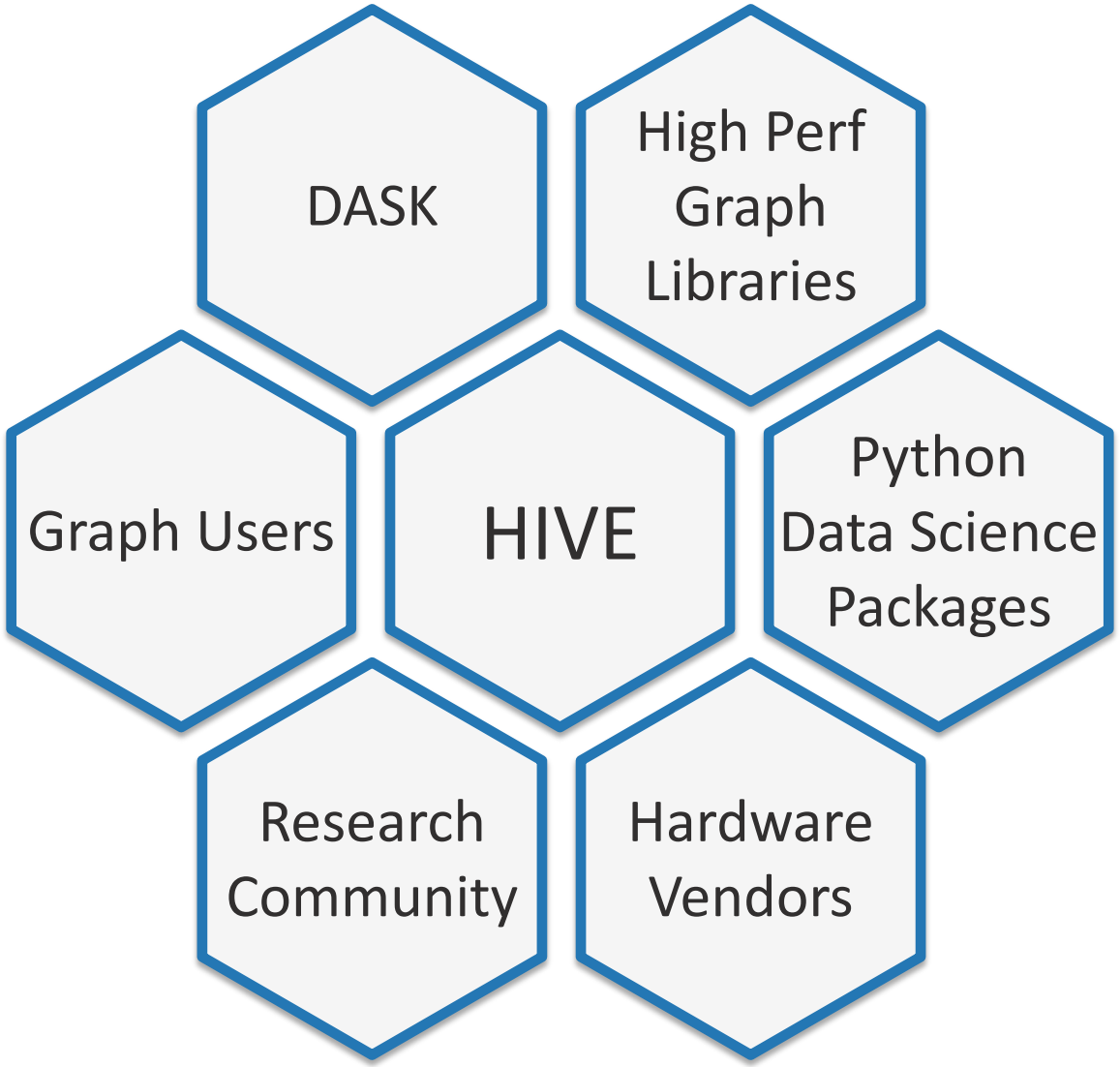
FOSDEM 2020



Outline

- What is HIVE?
- Architecture
- Interfaces
- Extensibility
- Summary

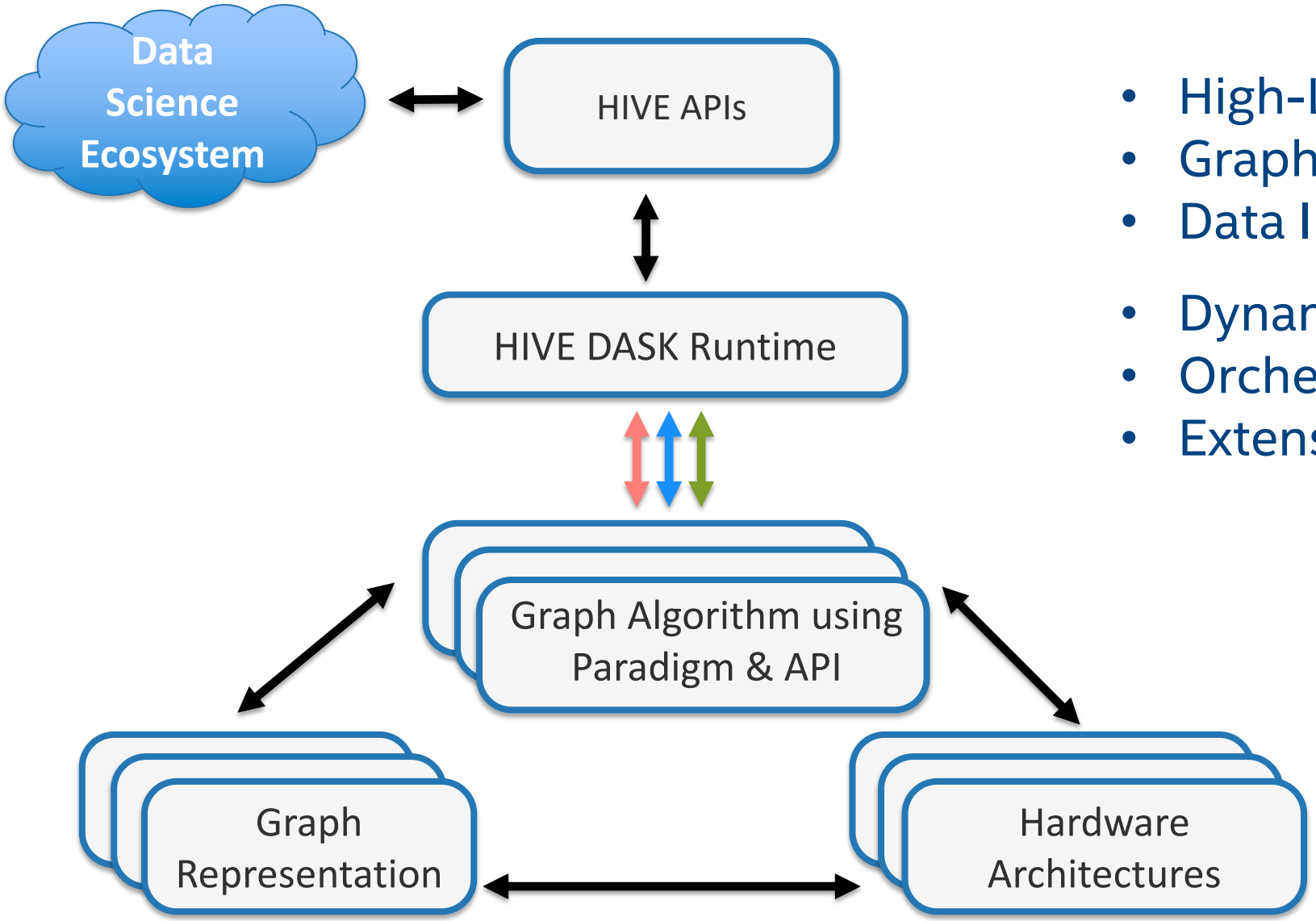
HIVE: A Bridge Between Graphs and Data Science



- Graph Analytics in Python
- Data-science Inter-Operability
- High Performance
- Transparent Orchestration
- Community Driven
- Hardware Agnostic

- In development, to be open sourced in 2020

One Indirection to target them all



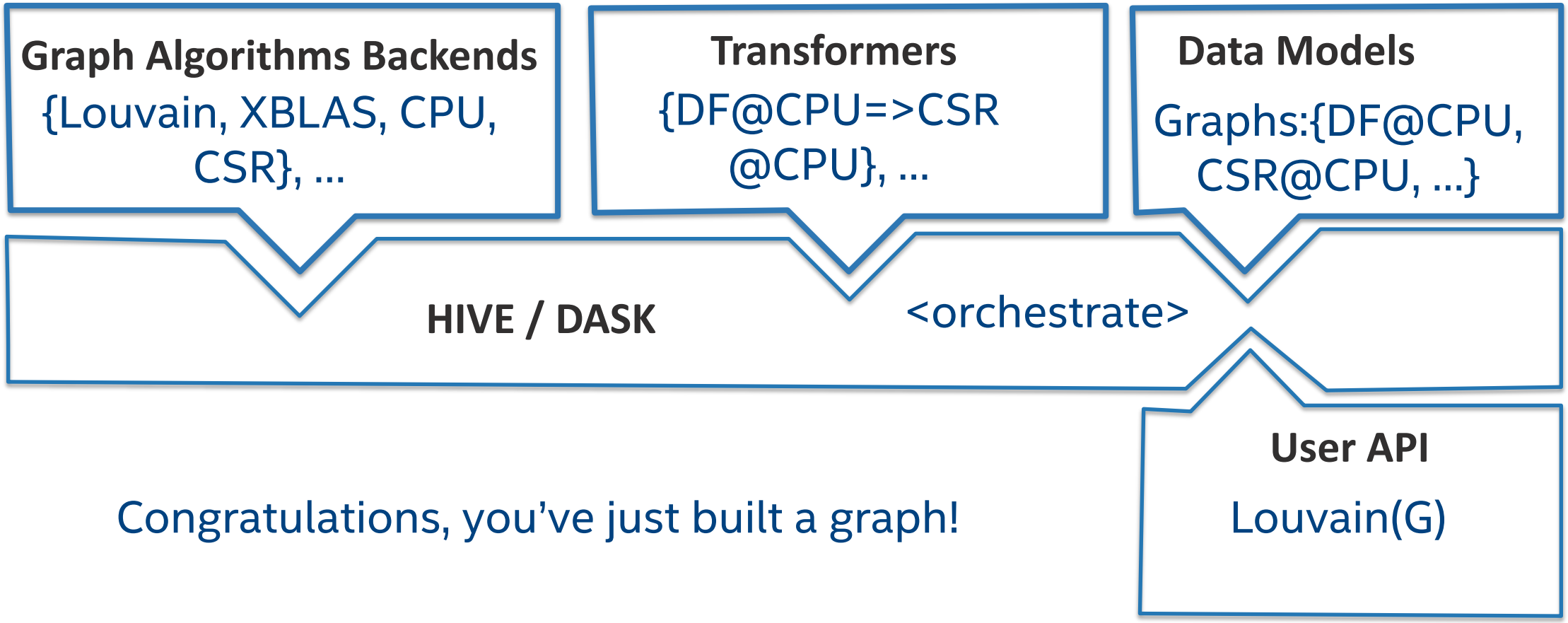
- High-Level Graph API
- Graph Query API with Numba
- Data Inter-Operability

- Dynamic Task Graph
- Orchestrate compute & data
- Extensible via plugins

Graph Frameworks

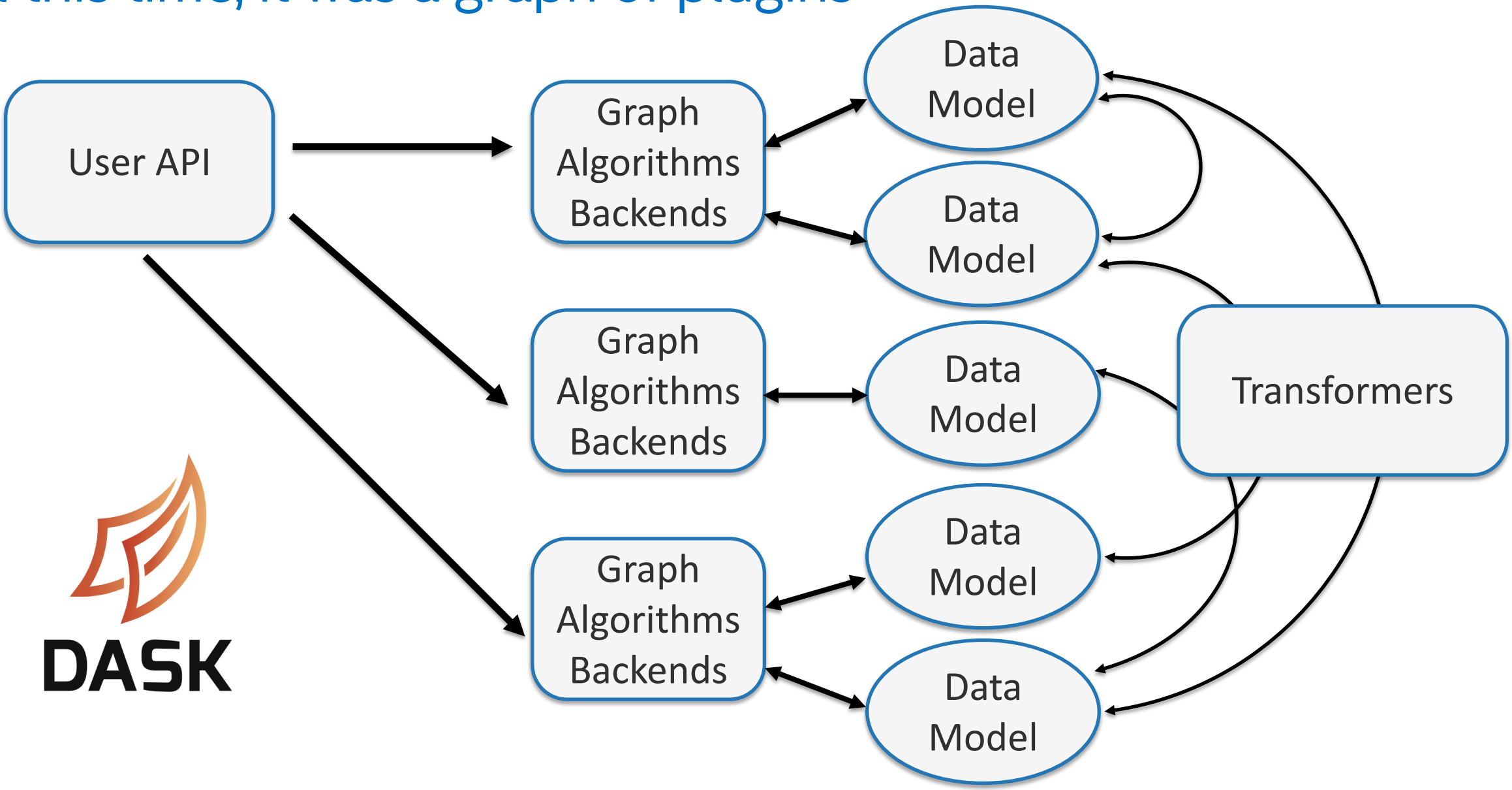
- SuiteSparse
- Galois
- GraphIt
- Gunrock
- …

HIVE Framework Interfaces



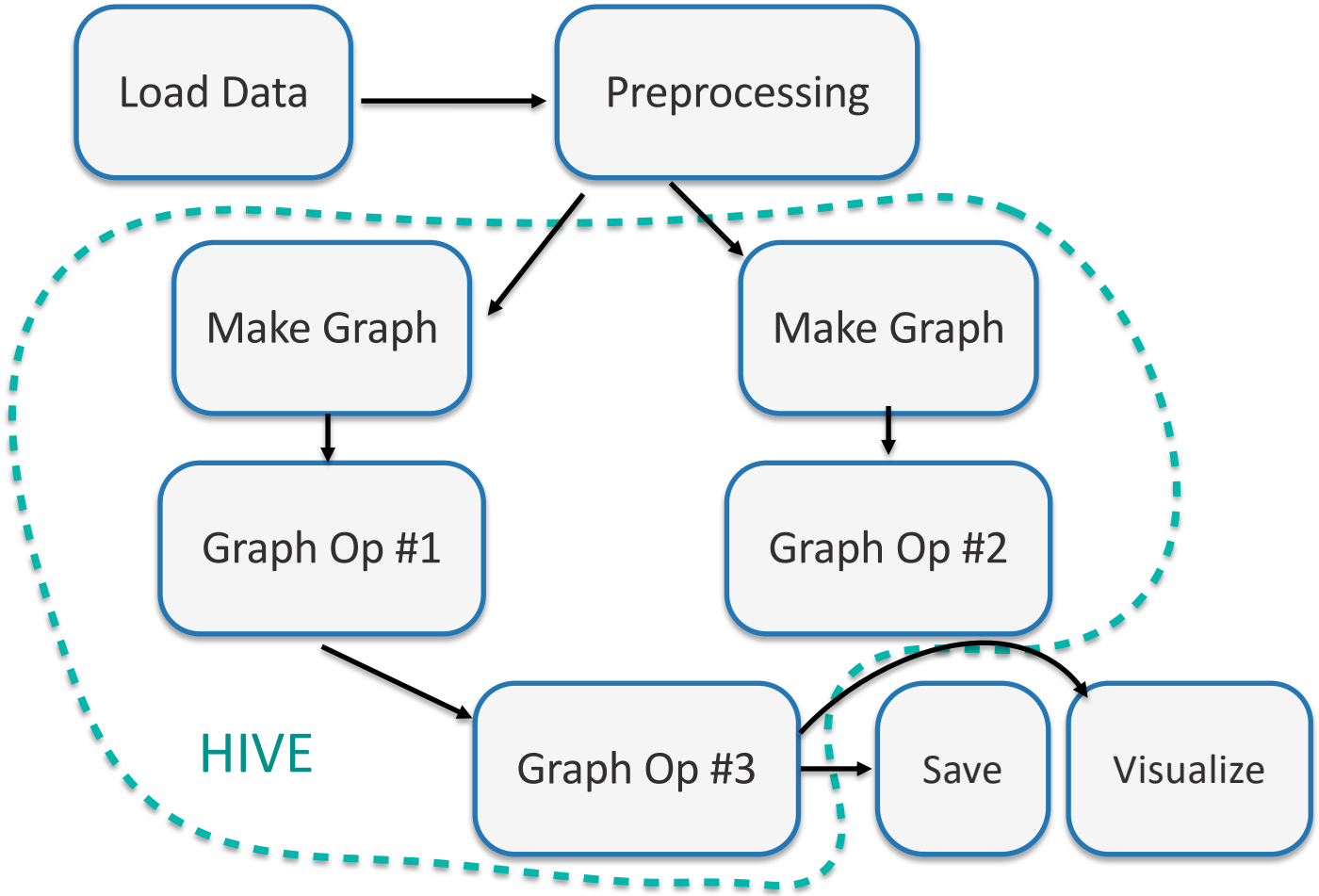
Congratulations, you've just built a graph!

All this time, it was a graph of plugins



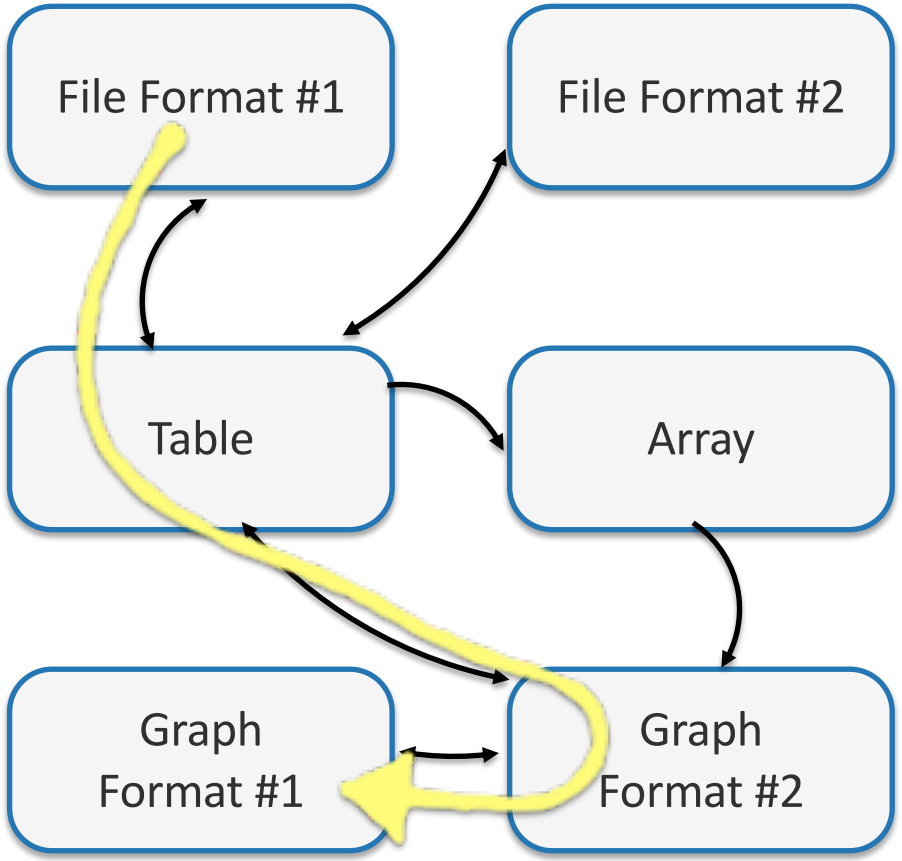
Doing Graph Analytics With The Help of Graphs

Workflow Task Graphs



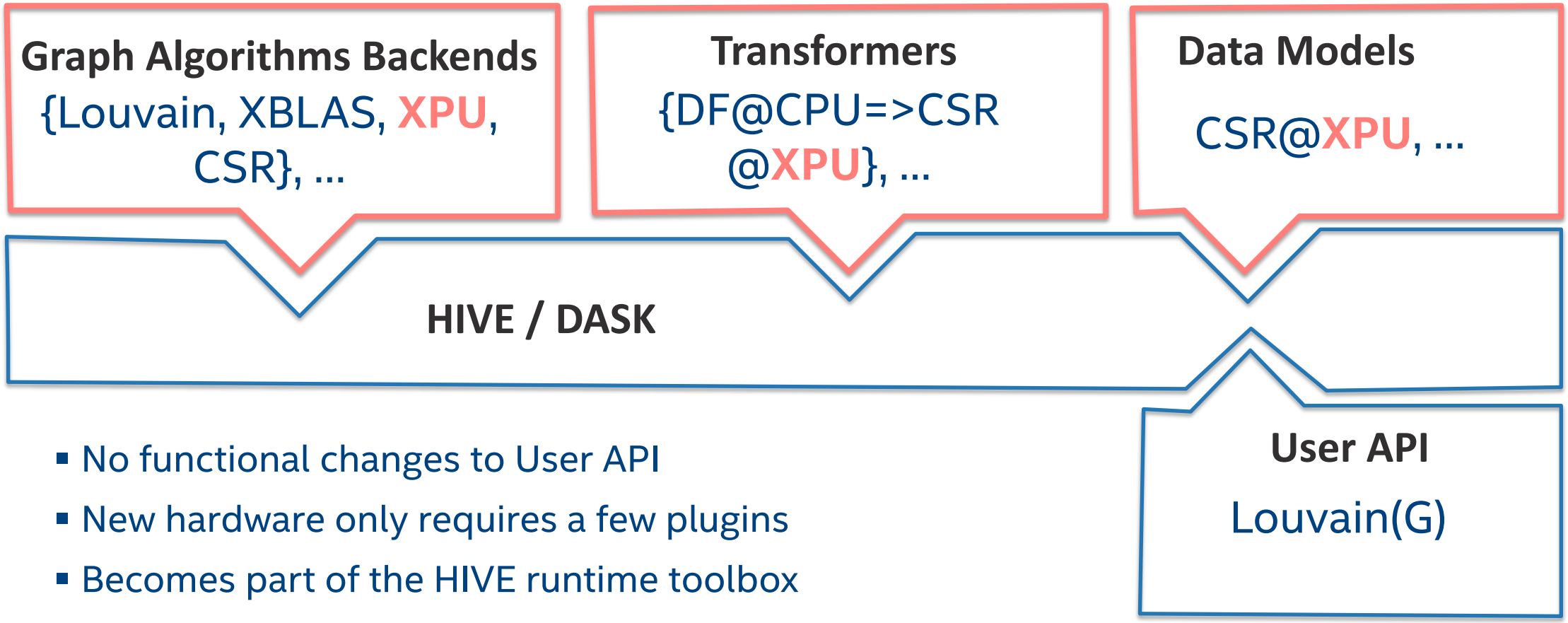
Orchestrate HW backend selection & data movement

Data Transformation Graphs



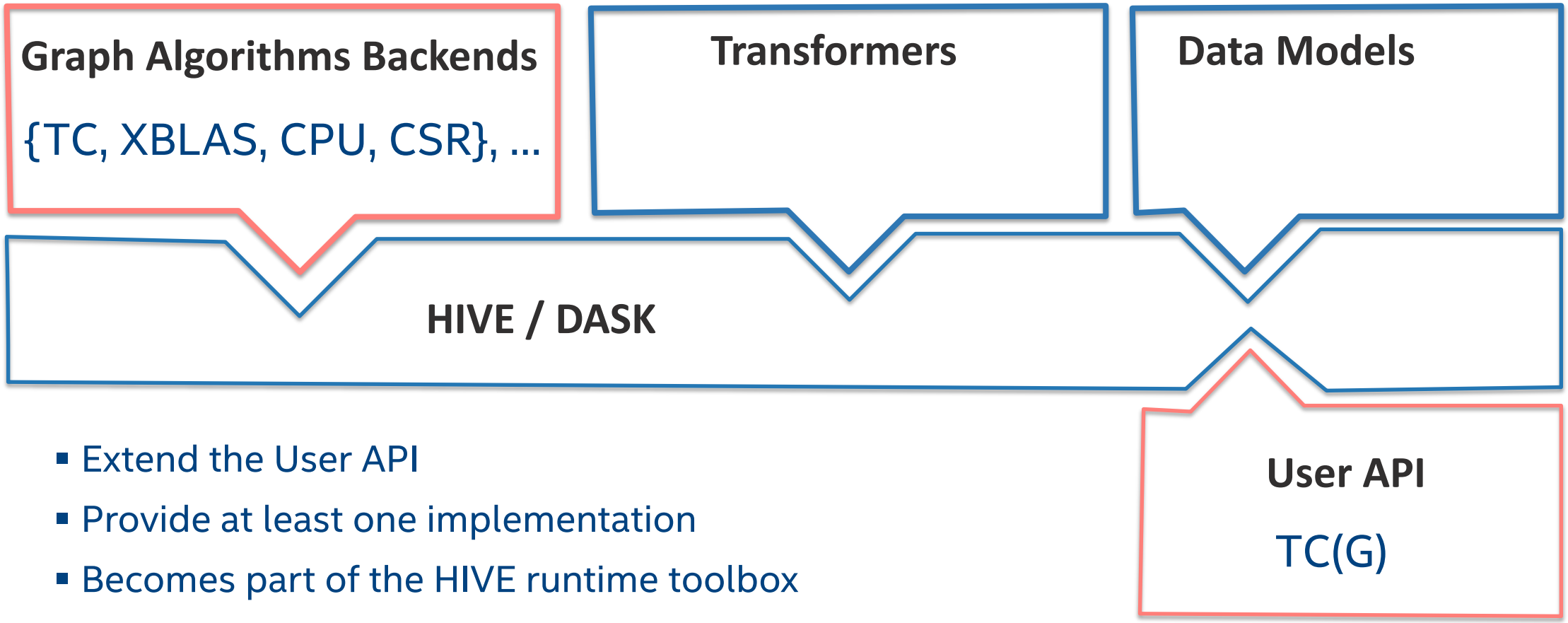
Automated data transformers selection

Extensibility: Supporting New Hardware



- No functional changes to User API
- New hardware only requires a few plugins
- Becomes part of the HIVE runtime toolbox
- Mixing between HW architectures is automatically supported

Extensibility: Supporting a new User API



- Extend the User API
- Provide at least one implementation
- Becomes part of the HIVE runtime toolbox

Stakeholders View

Data Scientists

- Unified API for Graph Analytics
- Python inter-operability
- State of the art backends
- Transparent orchestration
- Increased workflow portability

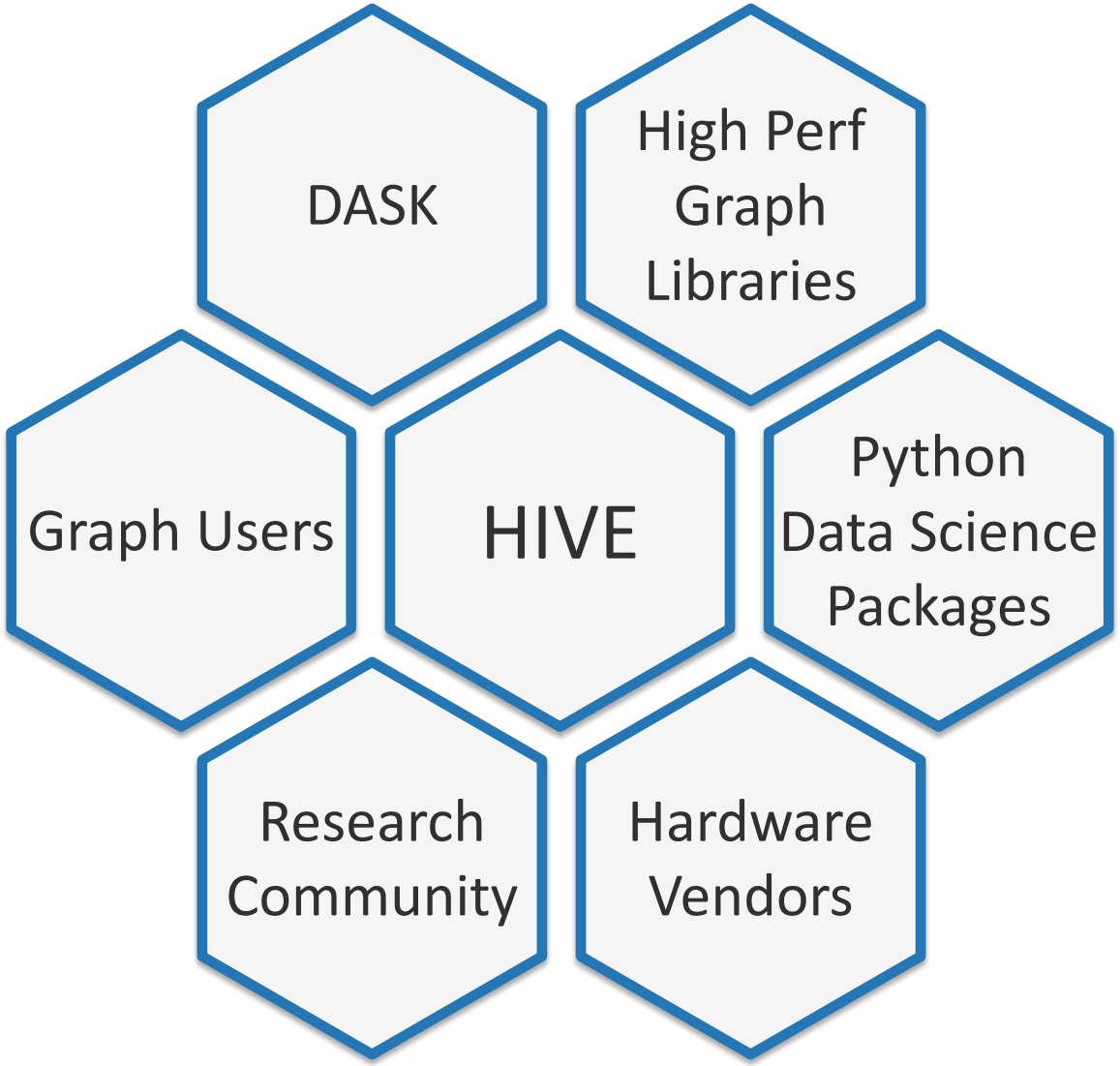
Graph Framework Developers

- Python frontend for algorithms
- Increased user base
- Performance feedback

Researchers

- Easy integration in workflows
- Easily extensible
- Performance monitoring & optimization

HIVE: A Bridge Between Graphs and Data Science



Questions?