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: Graph Analytics Framework in Python – Vincent Cavé, Stanley Seibert – FOSDEM 2020
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**

1. Load Data
2. Preprocessing
3. Make Graph
   - Graph Op #1
   - Graph Op #2
   - Graph Op #3
4. Save
5. Visualize

**Data Transformation Graphs**

1. File Format #1
2. File Format #2
3. Table
4. Array
5. Graph Format #1
6. Graph Format #2

**Orchestrate HW backend selection & data movement**

**Automated data transformers selection**

HIVE: Graph Analytics Framework in Python – Vincent Cavé, Stanley Seibert – FOSDEM 2020
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

<table>
<thead>
<tr>
<th>Data Scientists</th>
<th>Graph Framework Developers</th>
<th>Researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Unified API for Graph Analytics</td>
<td>• Python frontend for algorithms</td>
<td>• Easy integration in workflows</td>
</tr>
<tr>
<td>• Python inter-operability</td>
<td>• Increased user base</td>
<td>• Easily extensible</td>
</tr>
<tr>
<td>• State of the art backends</td>
<td>• Performance feedback</td>
<td>• Performance monitoring &amp; optimization</td>
</tr>
<tr>
<td>• Transparent orchestration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Increased workflow portability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HIVE: A Bridge Between Graphs and Data Science

Questions?