

Federated Data Preparation, Learning, and Debugging in Apache SystemDS

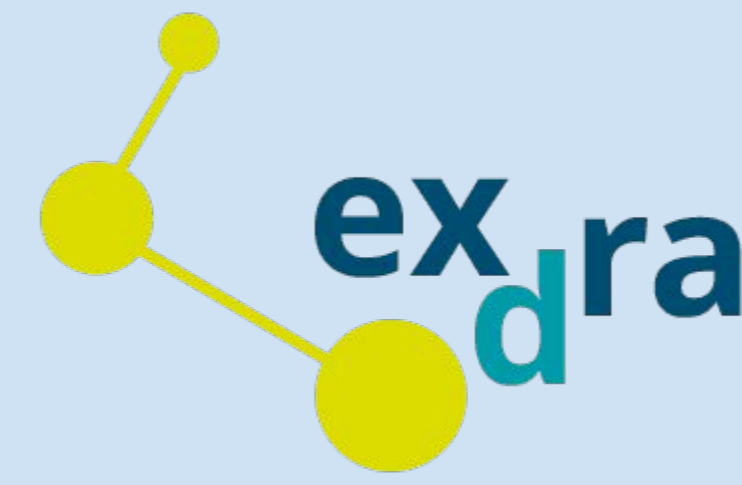
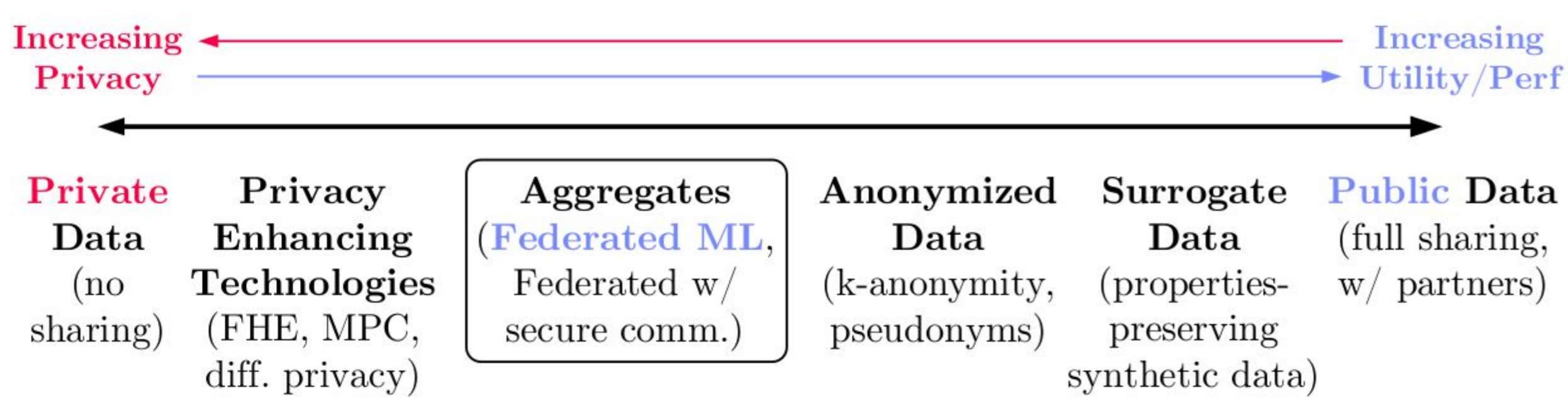
Sebastian Baunsgaard ^{1,*}, Matthias Boehm ^{1,2,*}, Kevin Innerebner ¹, Mito Kehayov ¹, Florian Lackner ¹, Olga Ovcharenko ¹, Arnab Phani ^{1,*}, Tobias Rieger ¹, David Weissteiner ¹, Sebastian Benjamin Wrede ^{1,2}

¹ Graz University of Technology; Graz, Austria; ² Know-Center GmbH; Graz, Austria; * now at TU Berlin; Berlin, Germany



Spectrum of Data Sharing

- Landscape of Techniques w/ Different Privacy, Utility and Performance Tradeoffs (new markets?)

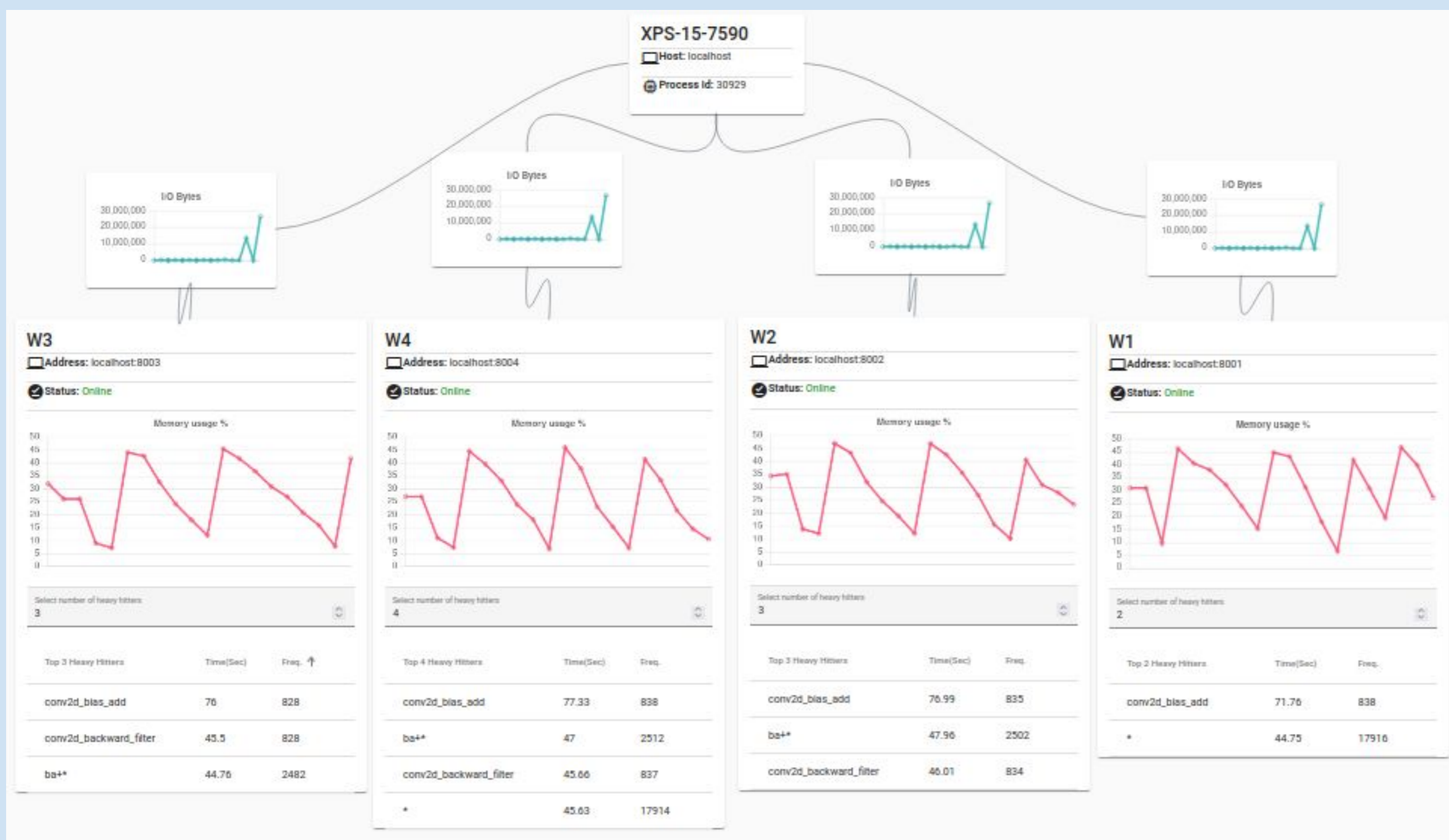


Data-centric Data Science Lifecycle

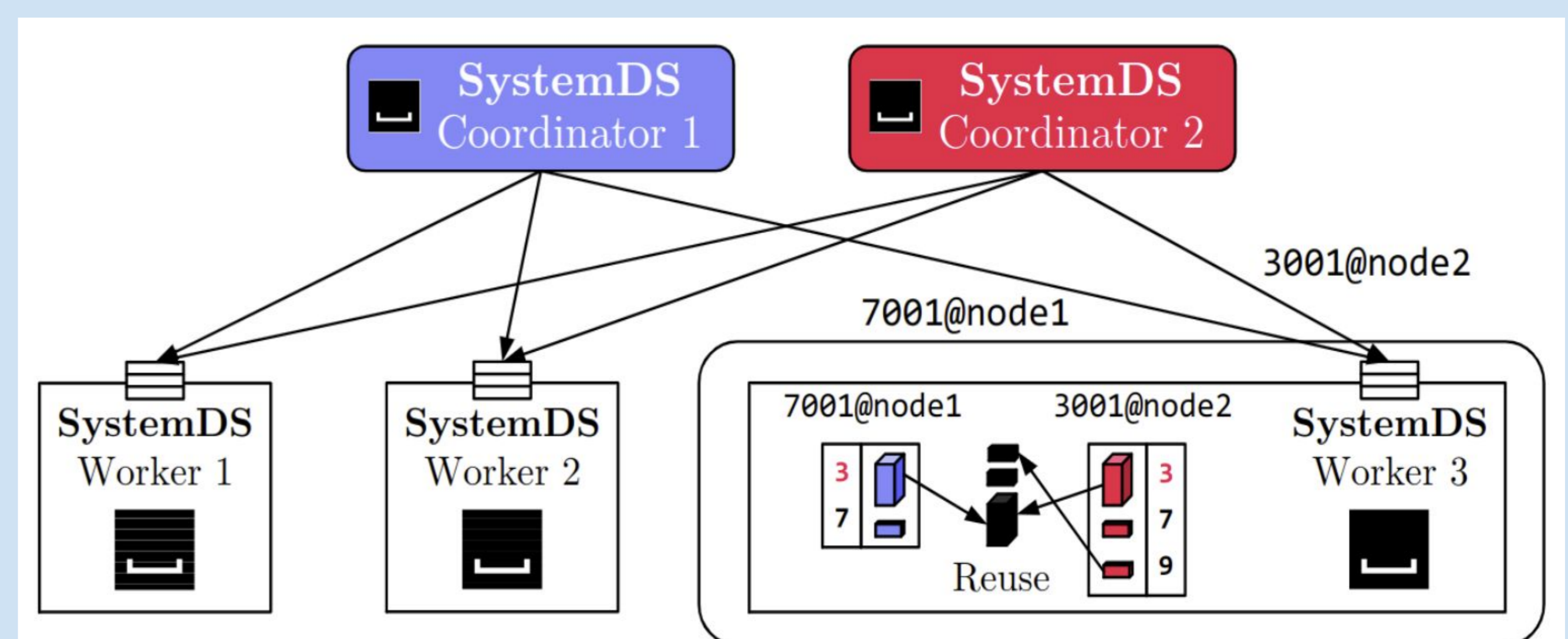
- Federated Linear Algebra & Parameter Servers
- Federated Data Preparation, Feature Transformations
- Model Training & Scoring, and Model Debugging
- Automatic Compilation of Federated Runtime Plans

Demo: Monitoring Tool

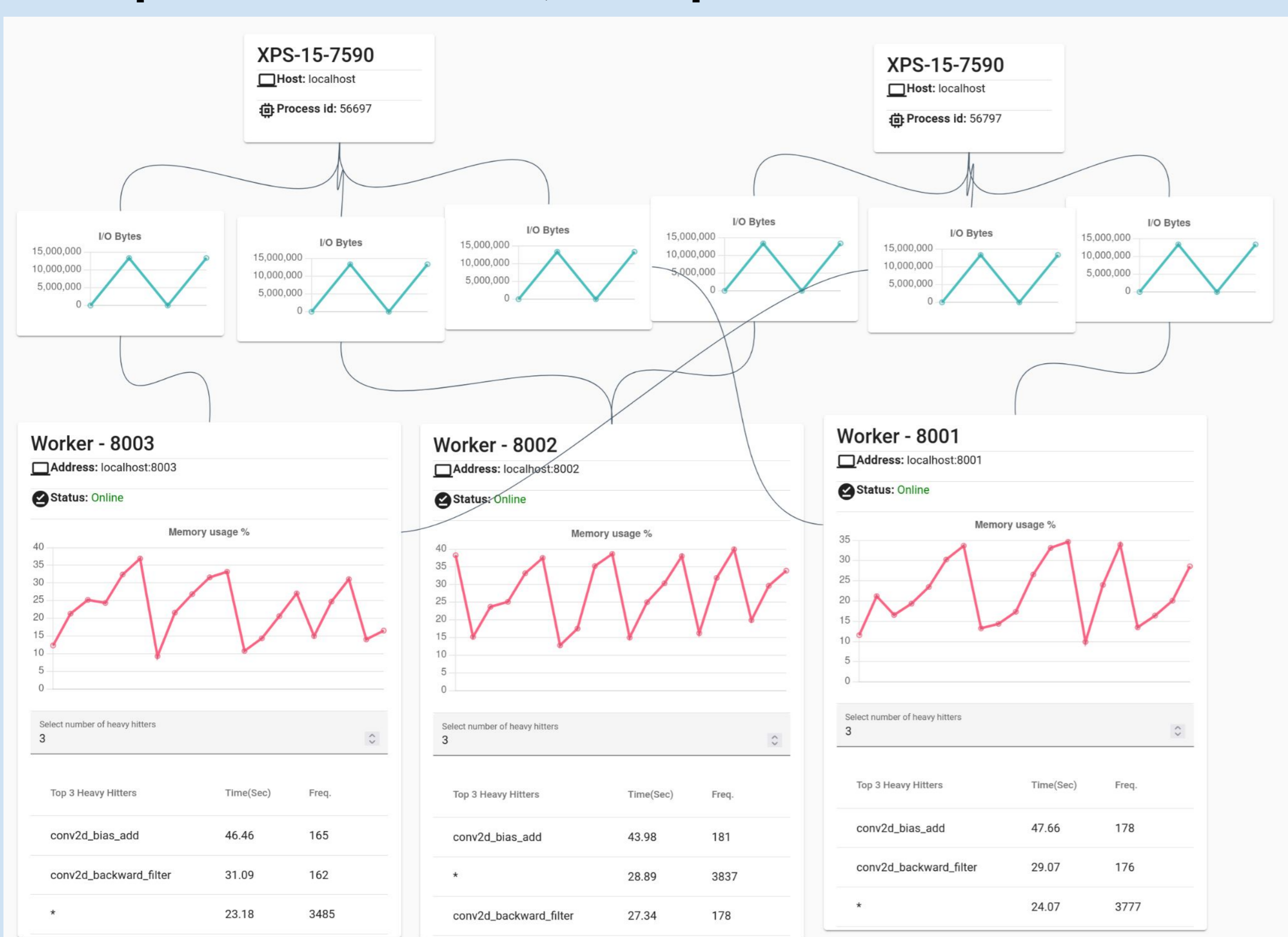
One Coordinator, Multiple Workers:



Multi-tenant Federated Learning



Multiple Coordinators, Multiple Workers:



Apache SystemDS



Source Code



Federated Tutorial



SIGMOD 2021 Paper

Experiments

