DevOps and Data: Knowledge Functional

Brandon Klein, Cody Tyler, Sam Fields

Overview

- Motivation

- Knowledge Functional
  - Association
  - Taxonomy
  - Applied

- Closing Remarks

Motivation

Motivation

The Pulsed Power facilities at Sandia have increased in complexity and usage demands resulting in scientific and operations difficulties, especially at the Z Machine. Desire to research and develop reusable data-driven methodologies and solutions coupled with DevOps to enable and empower science and operations with knowledge for quick, efficient information-driven decisions: faster-time-to-knowledge

Since 2019, our research and development efforts to support robust data analytics and data science at the Pulsed Power facilities in Sandia National Laboratories have helped uncover an efficient, reusable methodology and reference framework of applying DevOps to data — named the Knowledge Functional, abbreviated as KF(n)

The Knowledge Functional, takes in data and outputs insight, answers, predictions through the following principles guided by the DevOps methodology:

- Knowledge discovery
- Data architecture
- Data management
- Data engineering

- Data ontology
- Ontological data science
- Ontological data analytics

Knowledge Functional KF(n) Association

KF(n) Association – Reynold's Number

- In fluid mechanics Reynold's Number is a nondimensional value used to help predict if a fluid flow is laminar or turbulent and is calculated with the following equation:

$$Re = \frac{\rho u L}{\mu}$$

- [Re] is Reynold's Number, [ρ] is density, [u] is velocity, [L] is characteristic length, and [μ] is dynamic viscosity

- This is a simplified way of applying qualitative labels to quantitative data and helps in establishing what the current state of a system is so that you know what changes need to be made to move closer to the desired state

- We have adapted Reynold's Number for data / information theory as a generalized way to measure Chaos of a given system and to give insight in how to reduce said chaos

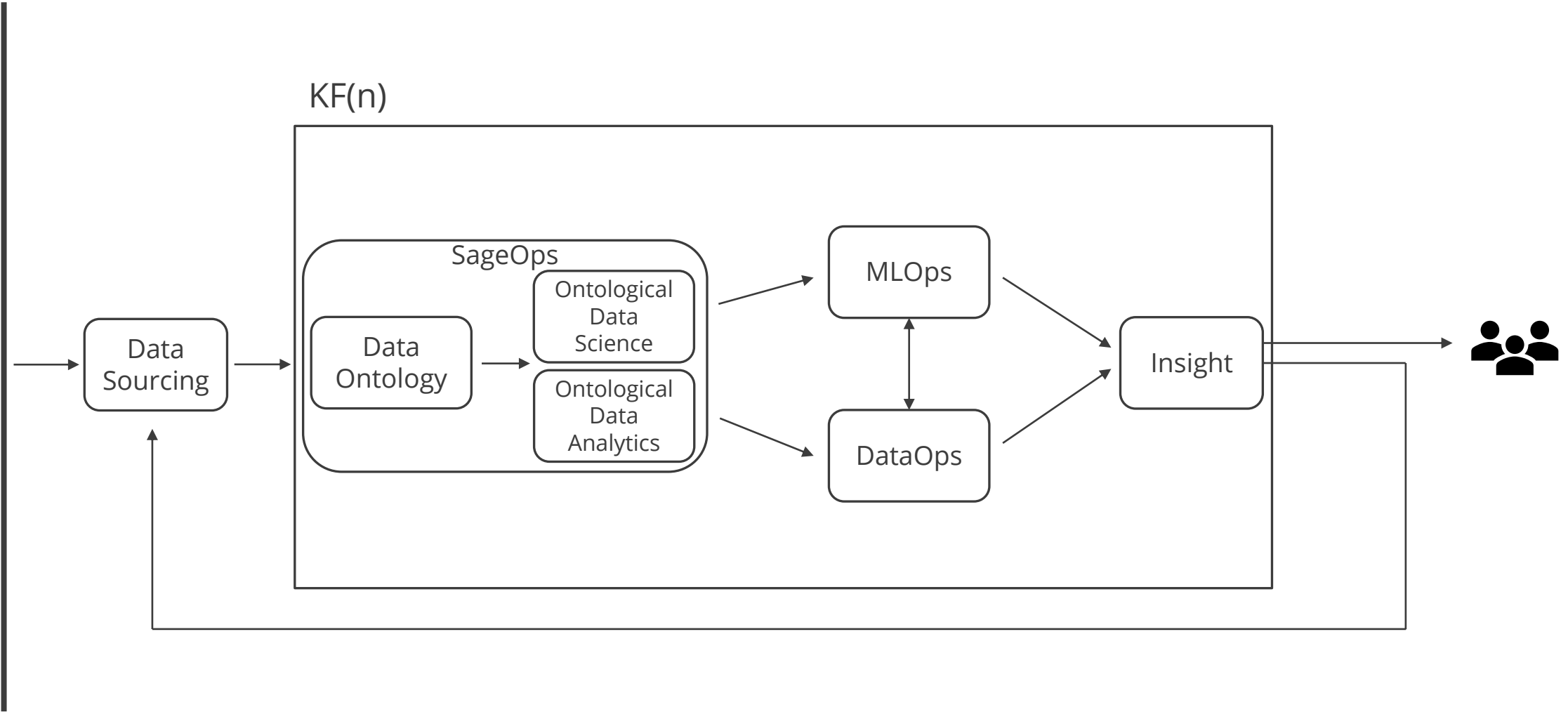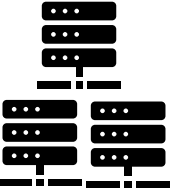$$KF(n) \approx \frac{dca}{u}$$

- [KF(n)] is the Knowledge Functional, [d] is data sources, [c] is data connections, [a] is applications, and [u] is users

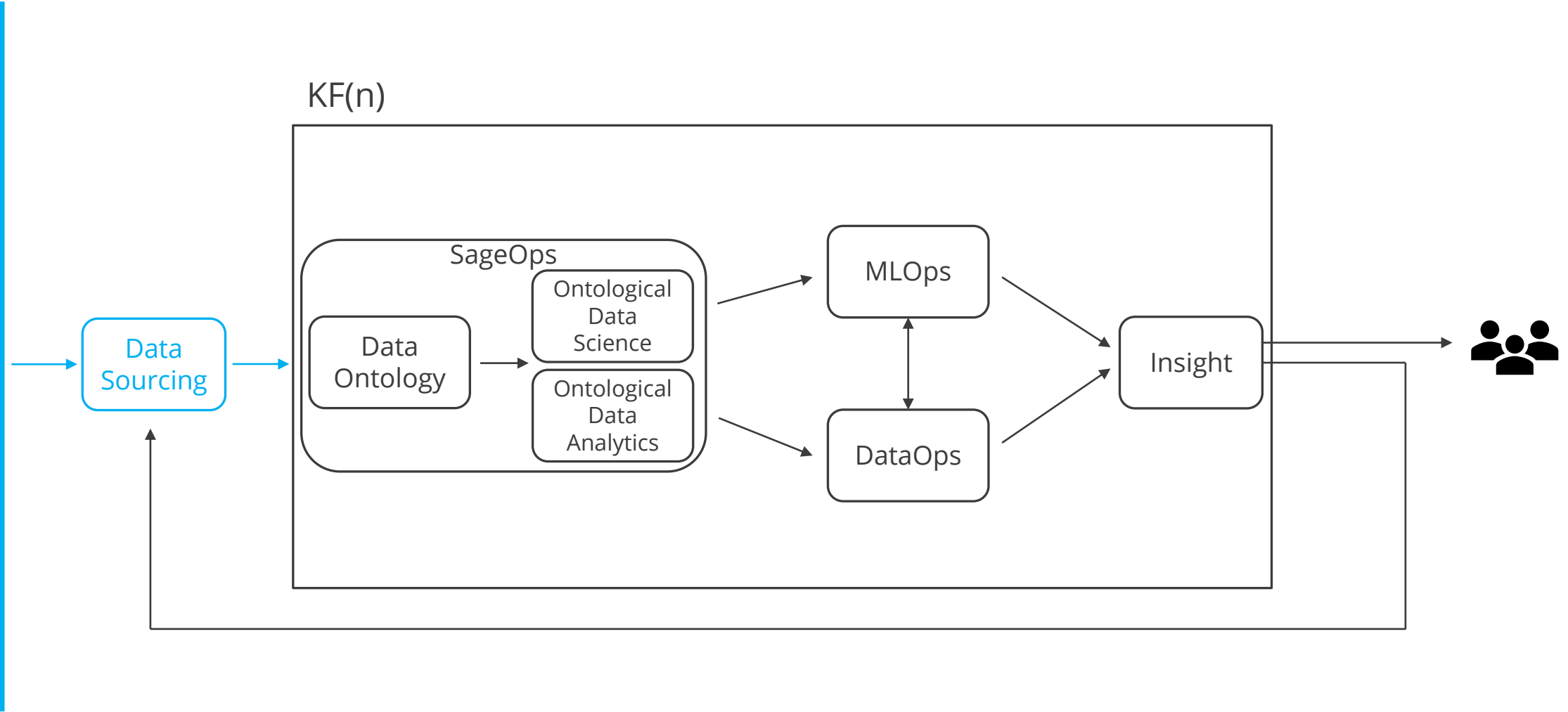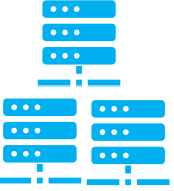# Knowledge Functional KF(n) Taxonomy

KF(n) Taxonomy

KF(n) Taxonomy – Data Sourcing
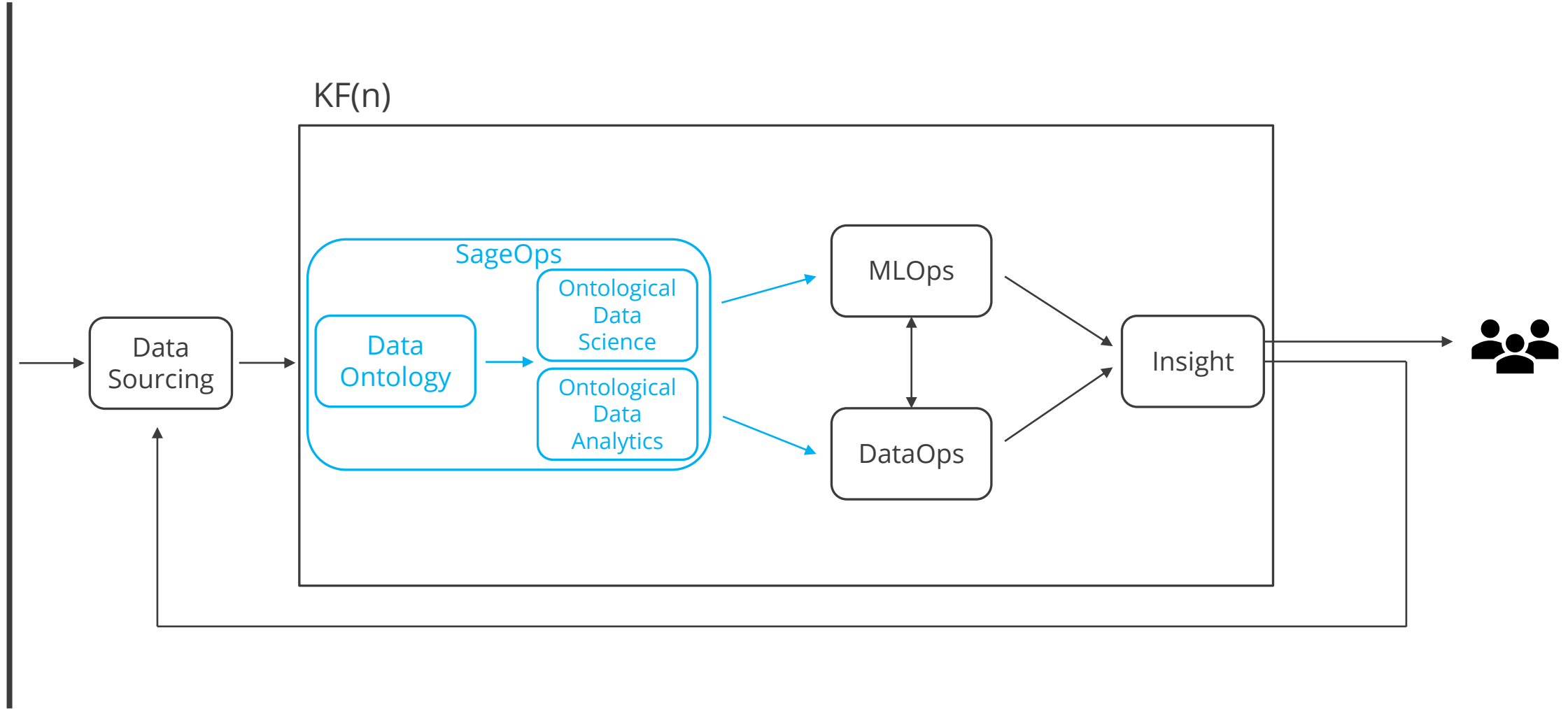
KF(n) Taxonomy – Data Sourcing

- This state is agnostic

- Non-exhaustive list of typical data sources users may encounter:
  - Filesystems
  - Databases
  - Workstations
  - Paper documents
  - Code Version managers (e.g. Git)

$$KF(n) \approx \frac{dca}{u}$$

- [KF(n)] is the Knowledge Functional, **[d] is data sources**, [c] is data connections, [a] is applications, and [u] is users

KF(n) Taxonomy – SageOps
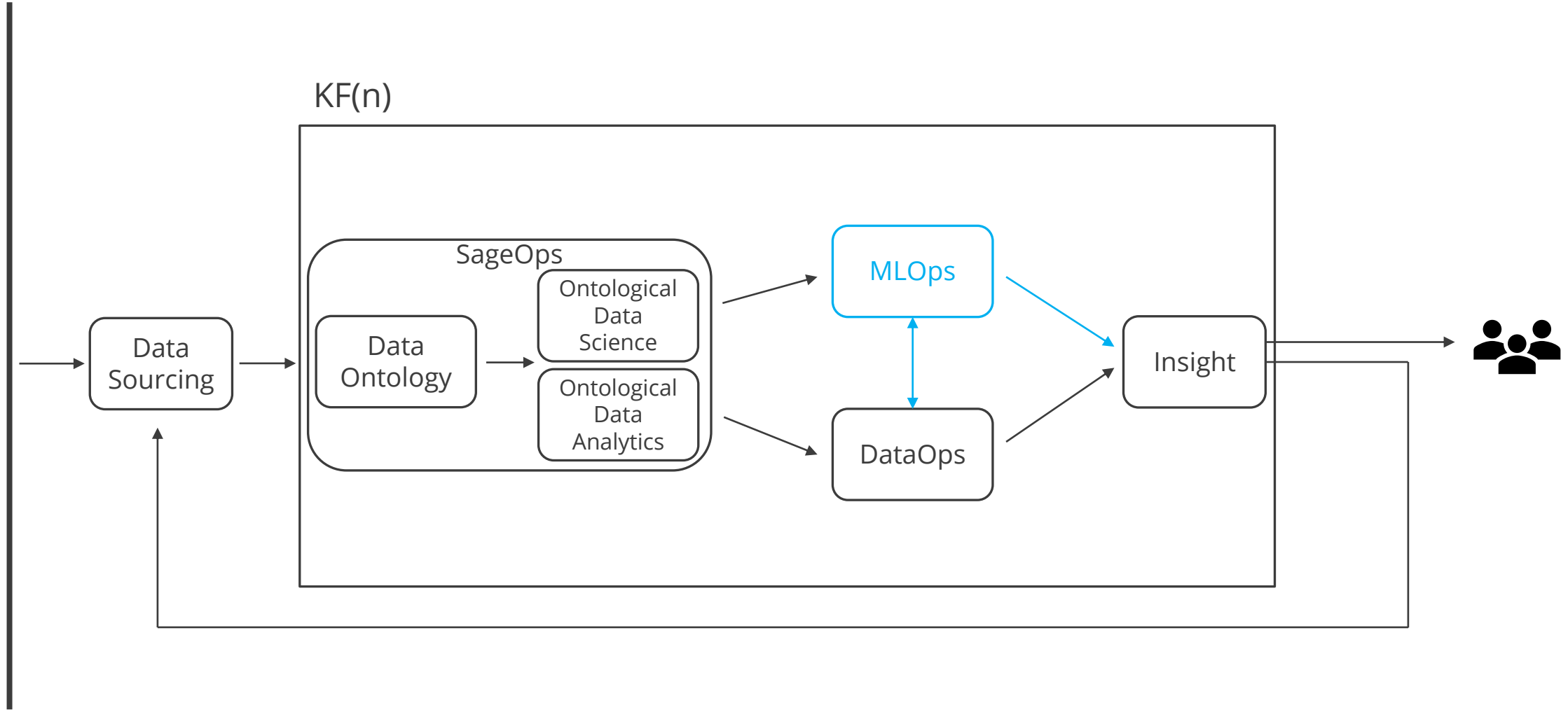
KF(n) Taxonomy – SageOps

- We postulated data ontology with DevOps would prove to be useful and we labeled the practice SageOps
  - SageOps as a synergistical coupling of data ontology and DevOps, for quick and efficient knowledge discovery of taxonomic relationships amongst data to support sagacious decisions

- SageOps looks to foster and obtain knowledge discovery sooner for ontological data science and ontological data analytics; immediate feedback can be cyclically and intrinsically reinvested

- Any patterns, relationships, or correlations uncovered through the initial ontological data mining, findings could be passed along for, and between, ontological data science and ontological data analytics to provide an accelerated advantage

$$KF(n) \approx \frac{d\boldsymbol{c}a}{u}$$

- [KF(n)] is the Knowledge Functional, [d] is data sources, **[c] is data connections**, [a] is applications, and [u] is users

KF(n) Taxonomy – MLOps
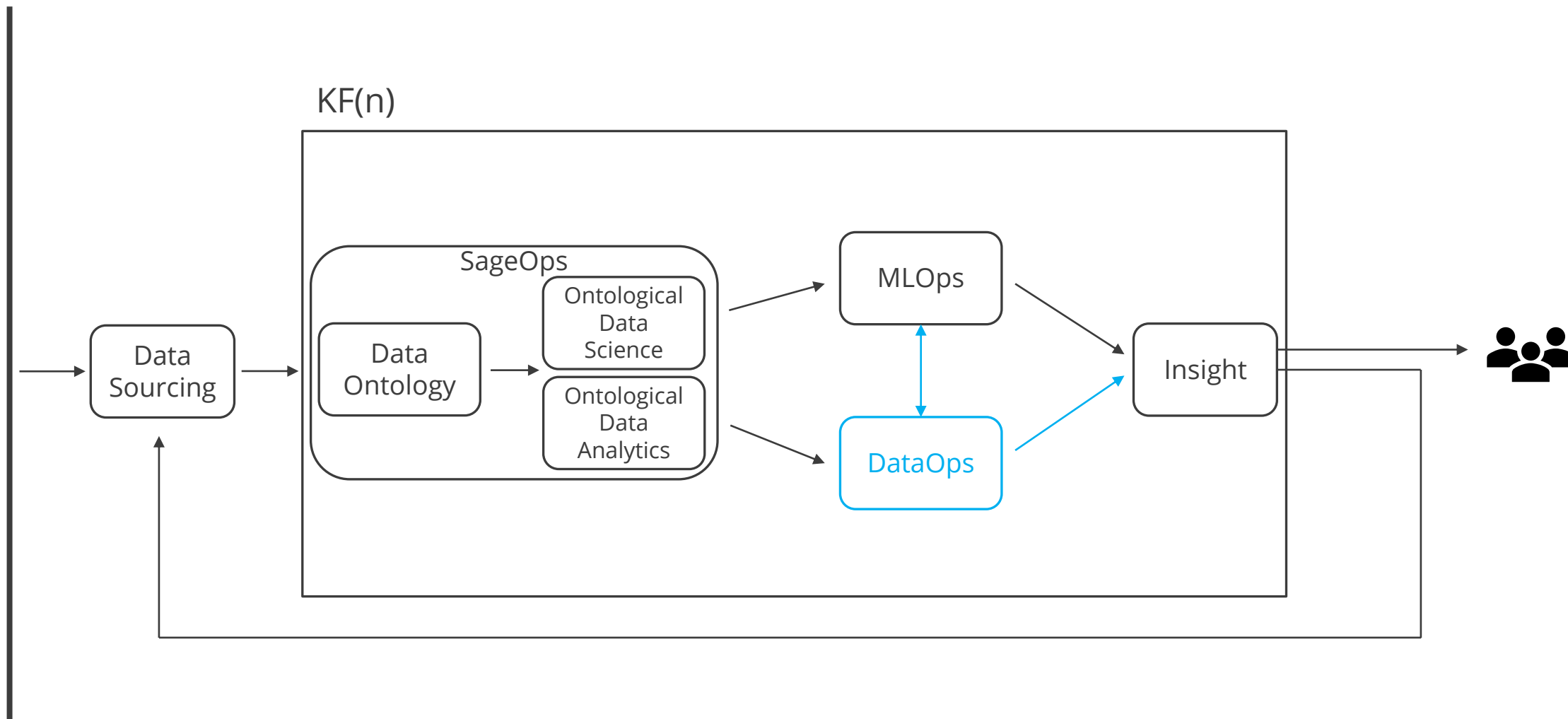
KF(n) Taxonomy – MLOps

- The MLOps state was identified as a desired approach to offer efficient, quality assured data science solutions
  - MLOps provides a set of practices to deploy and maintain ML models for production use, efficiently and reliably

- ML outputs from SageOps are designed to be efficiently available for use through established DevOps pipelines intended for external sources to efficiently consume/integrate the ML outputs

- The sooner identified data models are discovered through SageOps, the faster said data models can be shared via MLOps to the community for consumption, critique, improvement, and faster-time-to-knowledge
  - Deployed ML solutions could provide insight into discovered patterns in practices faster than classical ML classification methodologies

- MLOps interplays with DataOps to support efficient quality-assured ML models and/or solutions

$$KF(n) \approx \frac{dc\boldsymbol{a}}{u}$$

- [KF(n)] is the Knowledge Functional, [d] is data sources, [c] is data connections, **[a] is applications**, and [u] is users

KF(n) Taxonomy – DataOps
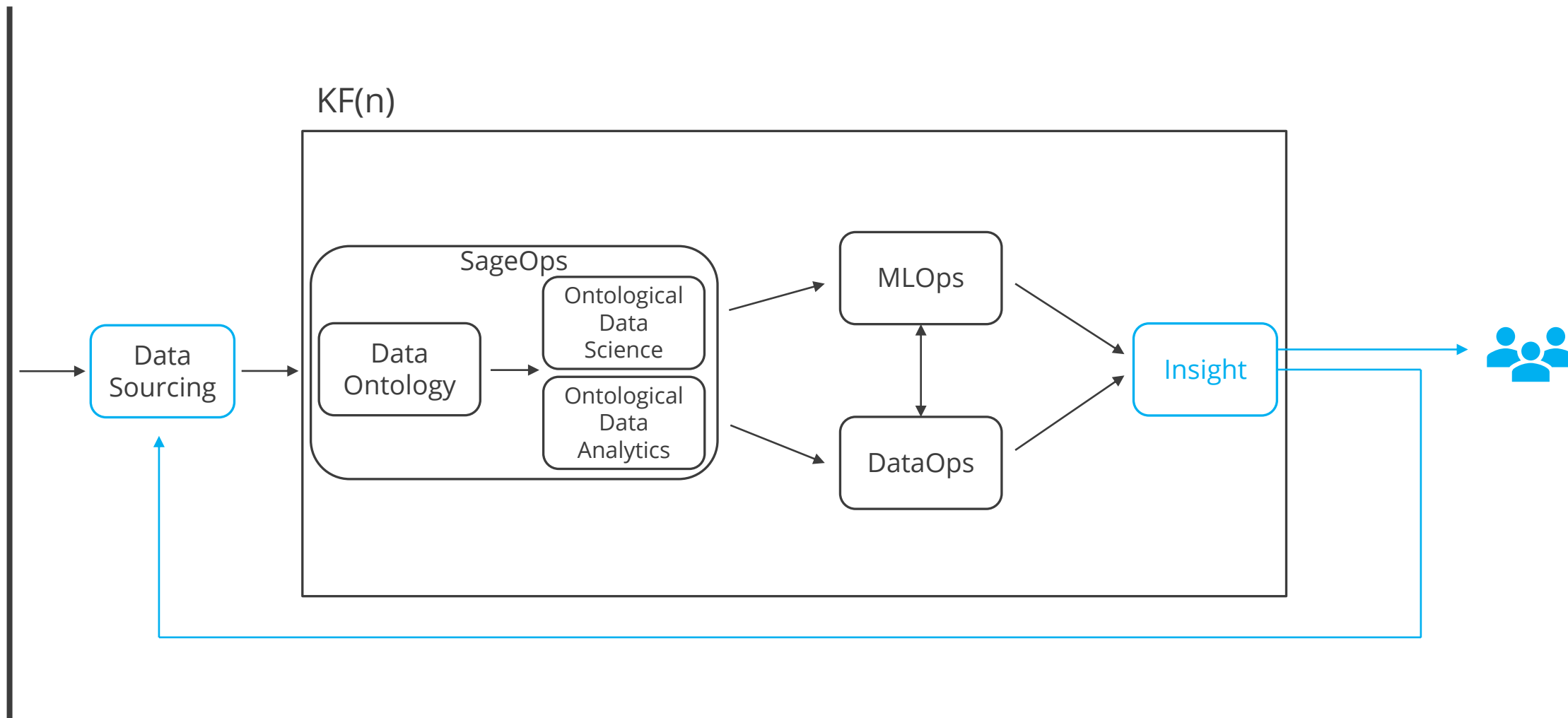
KF(n) Taxonomy – DataOps

- The DataOps state was identified as a desired approach to offer efficient, quality assured data-driven solutions
  - DataOps provides a set of practices to deploy and maintain data models for production use, efficiently and reliably

- Data outputs are discovered through SageOps with the use of ontological data mining, ontological data science, and ontological data analytics

- Quality assured data can be shared via DevOps Pipelines to the community for consumption, critique, improvement, and faster-time-to-knowledge

- DataOps interplays with MLOps to support efficient quality-assured data models and/or solutions

$$KF(n) \approx \frac{dca}{u}$$

- [KF(n)] is the Knowledge Functional, [d] is data sources, [c] is data connections, **[a] is applications**, and [u] is users

KF(n) Taxonomy – Insight

KF(n) Taxonomy – Insight

- Provides knowledge discovery for end users to make information-based decisions derived from MLOps and DataOps solutions

- End users will be provided information that could help identify trends, patterns, and outliers to gain insight to maintain, improve, or deprecate existing processes

- By applying DevOps to data, data will be available more efficiently and translated into information sooner to enable faster-time-to-knowledge

$$KF(n) \approx \frac{dca}{\boldsymbol{u}}$$

- [KF(n)] is the Knowledge Functional, [d] is data sources, [c] is data connections, [a] is applications, and **[u] is users**
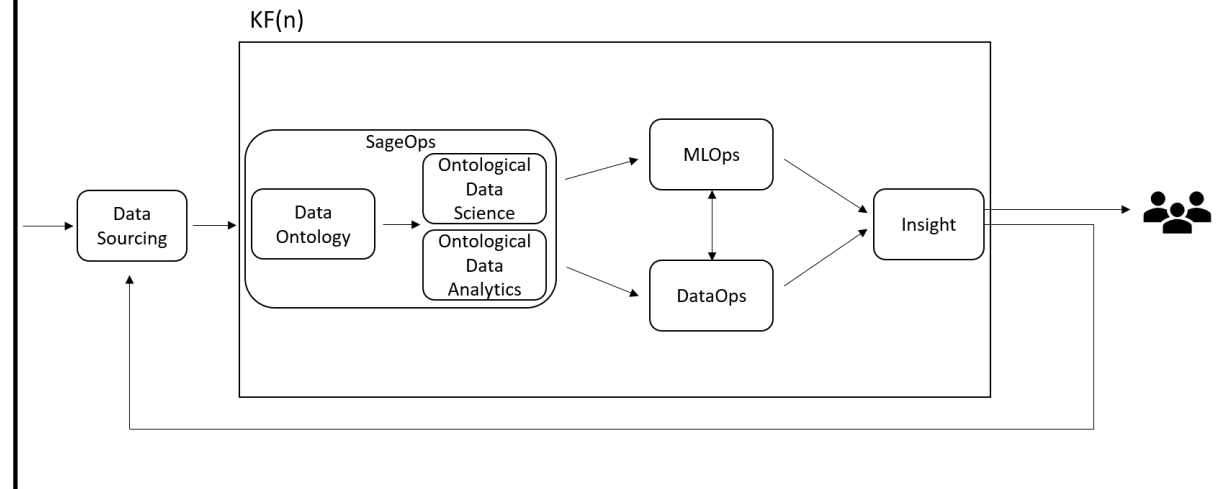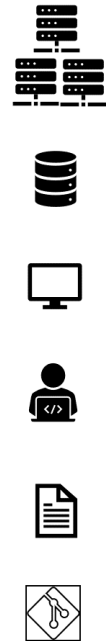
Knowledge Functional - Applied

KF(n) Applied

Day in the life of a Pulsed Power data set through KF(n):

- Did not exist prior, organic data collection

- Extract, transform, and load to our data ontology tools

- Ontological data science and analytics

- Sent to MLOps and DataOps pipelines
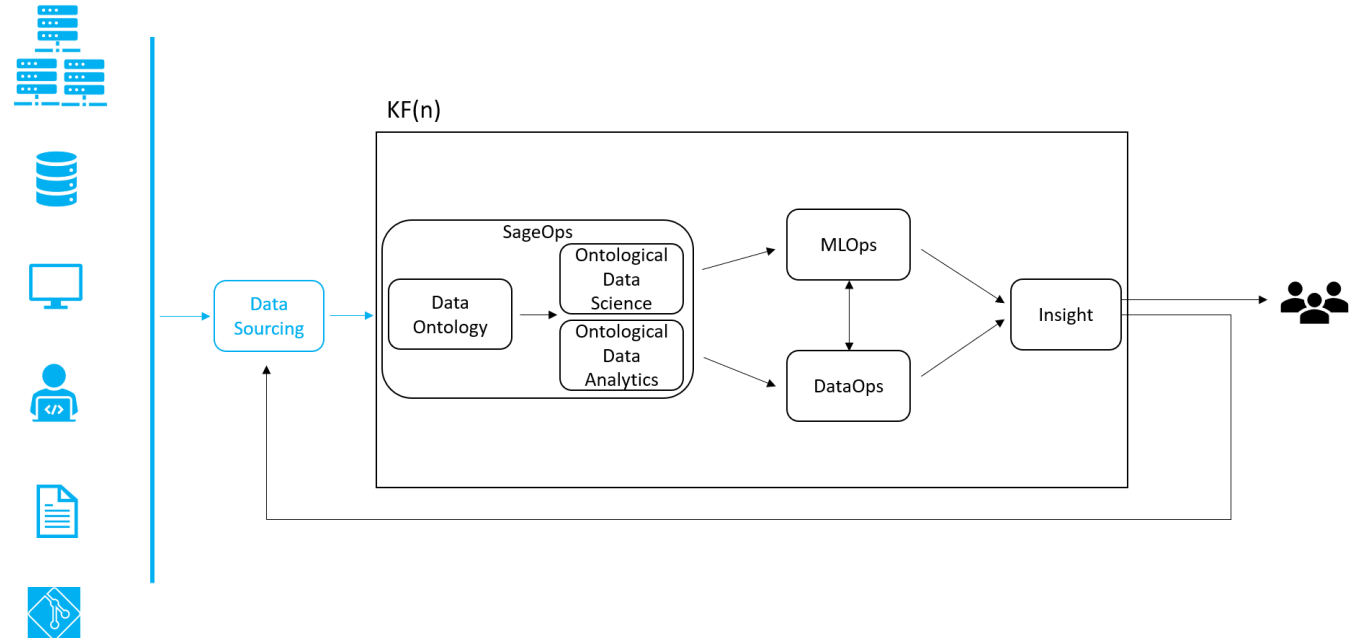
- Consumed by end-user for insight

KF(n) Applied

Day in the life of a Pulsed Power data set through KF(n):

- Did not exist prior, organic data collection

- Extract, transform, and load to our data ontology tools

- Ontological data science and analytics

- Sent to MLOps and DataOps pipelines

- Consumed by end-user for insight
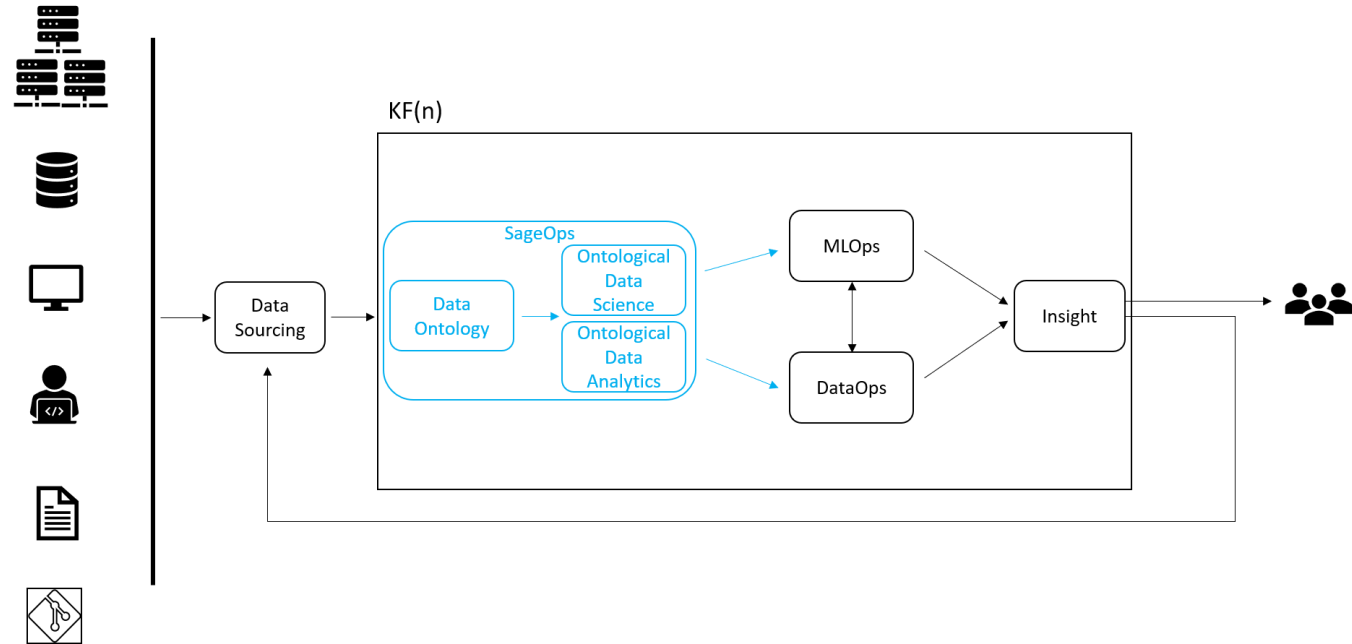
KF(n) Applied

Day in the life of a Pulsed Power data set through KF(n):

- Did not exist prior, organic data collection

- Extract, transform, and load to our data ontology tools

- Ontological data science and analytics

- Sent to MLOps and DataOps pipelines
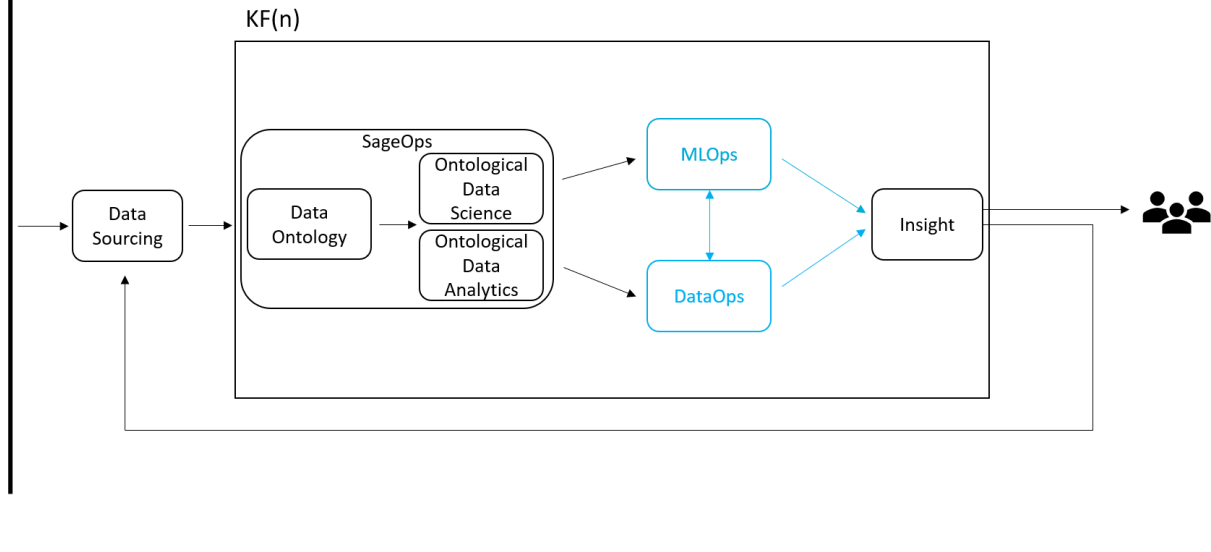
- Consumed by end-user for insight

KF(n) Applied

Day in the life of a Pulsed Power data set through KF(n):

- Did not exist prior, organic data collection

- Extract, transform, and load to our data ontology tools

- Ontological data science and analytics

- Sent to MLOps and DataOps pipelines
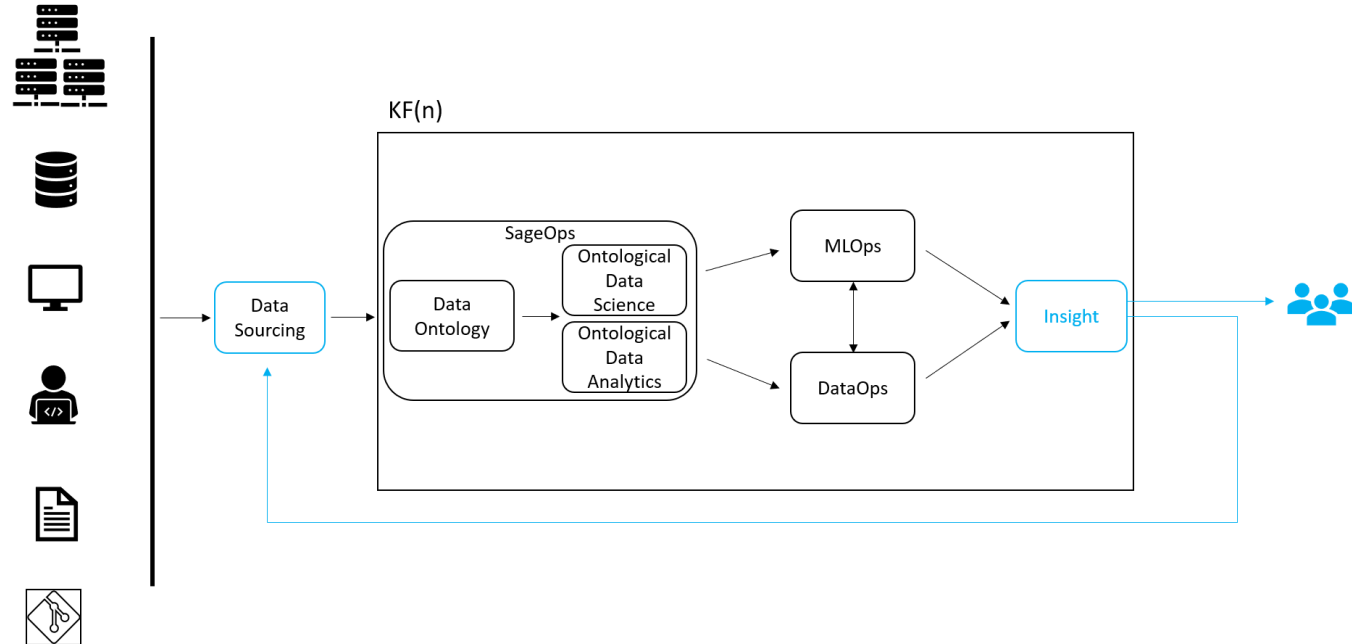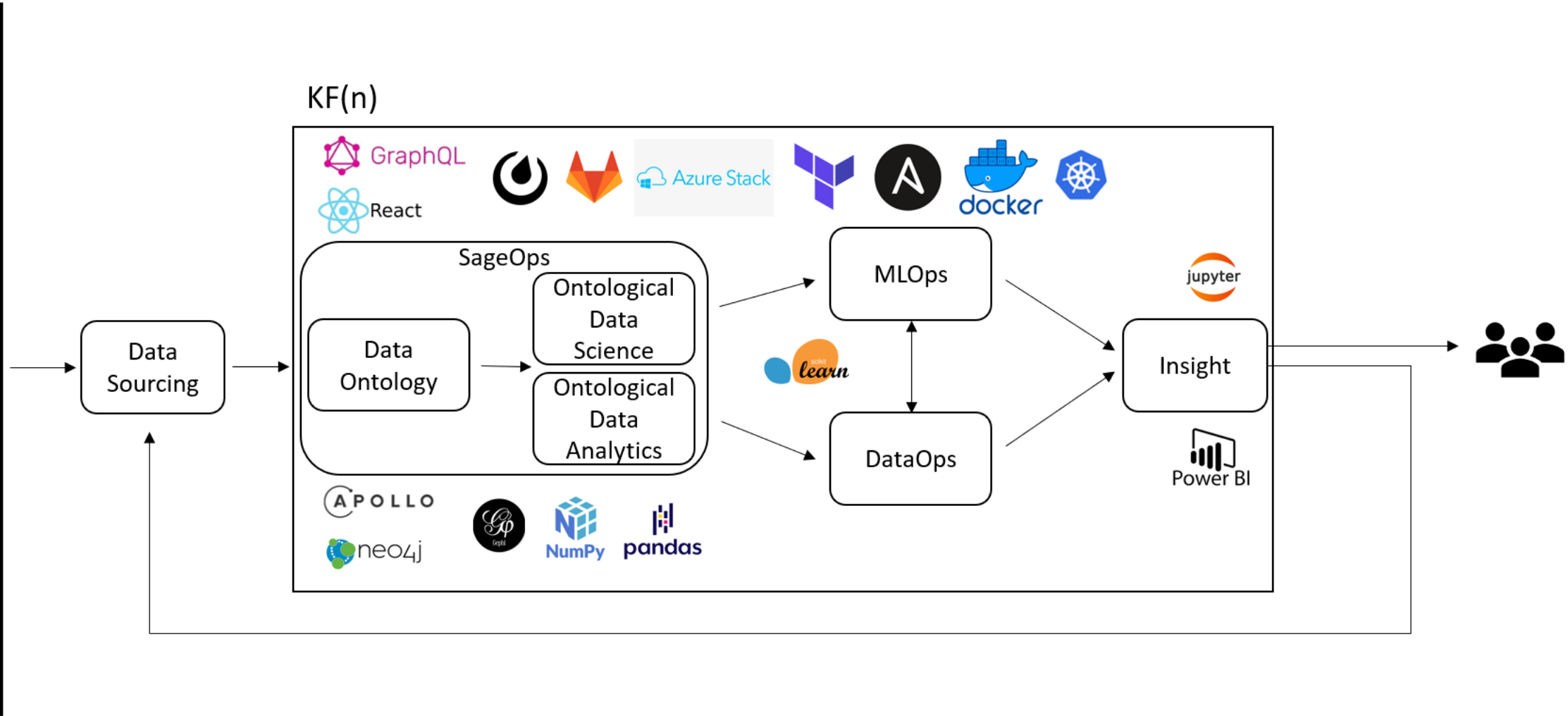
- Consumed by end-user for insight

KF(n) Applied

Day in the life of a Pulsed Power data set through KF(n):

- Did not exist prior, organic data collection

- Extract, transform, and load to our data ontology tools

- Ontological data science and analytics

- Sent to MLOps and DataOps pipelines

- Consumed by end-user for insight

KF(n) Applied – Technologies

Closing Remarks

Closing Remarks

Things to Consider in Applying KF(n):

- Workforce must understand their existing data processes and data landscape in order to improve existing and design new data processes

- Workforce transformation must develop the skills and roles necessary to employ the advanced methodologies and technologies, such as:

  - DevOps Engineer
  - ML Engineer
  - Data Engineer
  - MLOps Engineer

  - DataOps Engineer
  - Data Scientist
  - Data Analyst
  - Data Architect

- Data used to derive information and form insight require sound data governance, data lineage, data provenance, data curation, data security, data fidelity, and data integrity

Also checkout!

Our paper on DevOps and Data

DevOps and Data: Faster-Time-to-Knowledge through SageOps, MLOps, and DataOps

Contact Information

Brandon Klein: btklein@sandia.gov

Cody Tyler: ctyler@sandia.gov
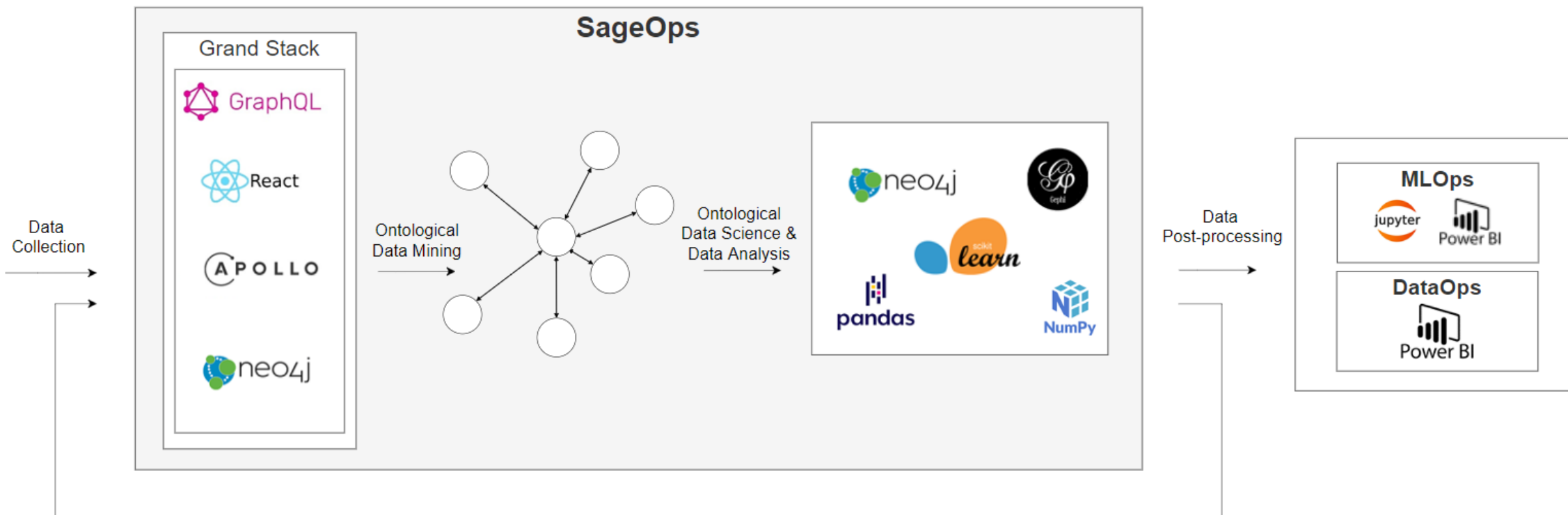
Sam Fields: sefield@sandia.gov

30

THANK YOU!

Extra Slides for Discussion

KF(n) Applied

KF(n) Applied –GitOps