Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Descartes
research

Towards Splitting Monolithic Workflows into Serverless Functions and Estimating Their Run-Time in the Earth Observation Domain

Our Vision for Earth Observation Analysis in Serverless Computing

M.Sc. Dennis Kaiser & M.Sc. Bohdan Dovhan

10.11.2021

https://se.informatik.uni-wuerzburg.de

The Problem with EO

MOTIVATION

Motivation – The Problem with EO

- Earth observation [EO] projects usually work independently
- They set up individual and varying execution environments
- Their (legacy) workflows are often monolithic and not optimized for cloud or serverless computing

➔ No prevalent standard or framework exists that helps these scientists to set up, optimize & execute their analysis using a cloud provider or with provisioning cloud hardware

Background

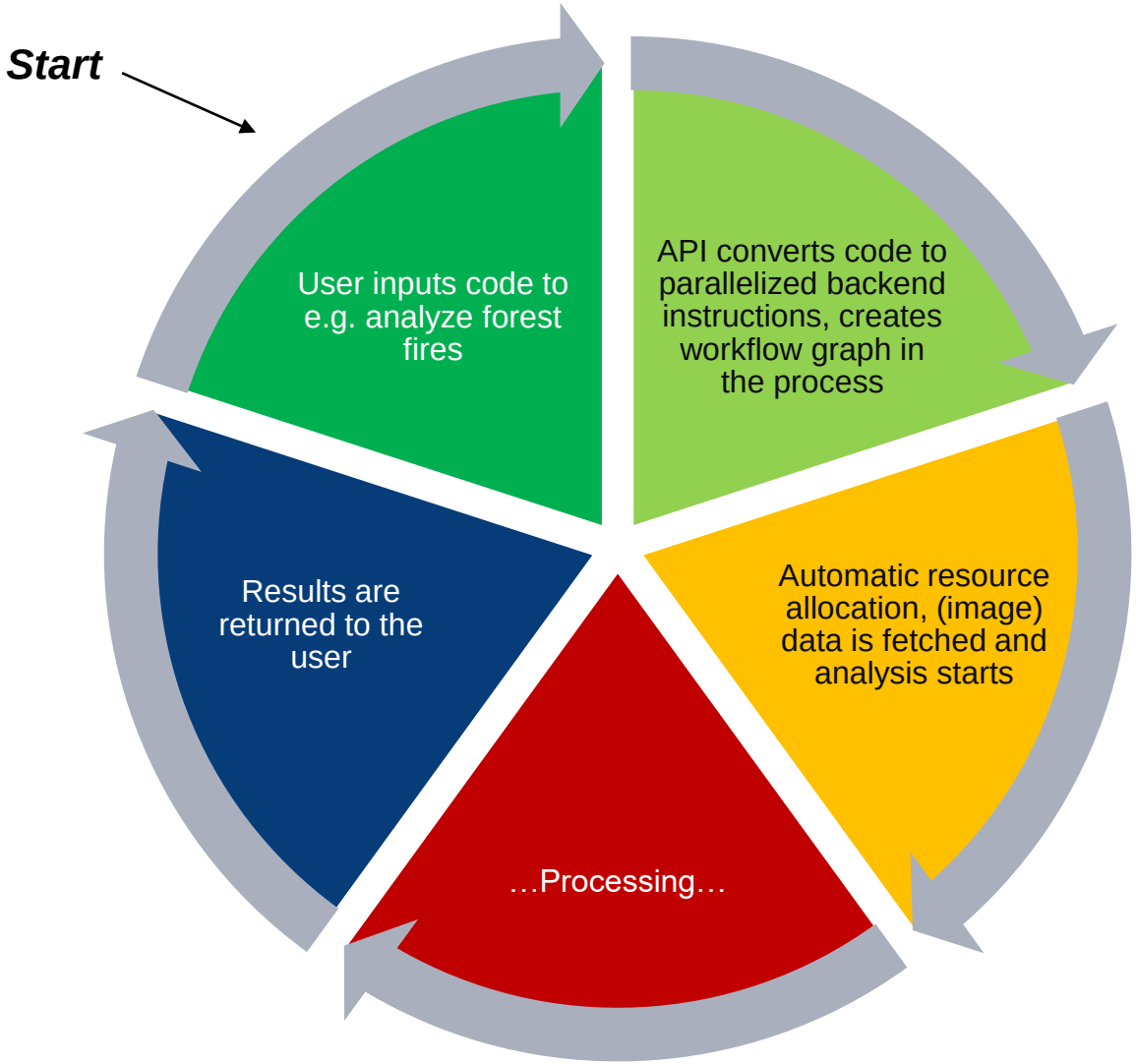
OUR MAIN OBJECTIVE

Our Main Objective

- Create a platform to enable scientists to easily analyze (earth observation) data
  - Knowledge of the backend is not required for users
  - Scientists can concentrate on their personal objective and research questions (e.g. atmospheric or image time series analysis)
  - This abstracts users from the inner workings of the platform (e.g., parallelization, optimizations, server allocation, …)

- Working together with the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt [DLR])

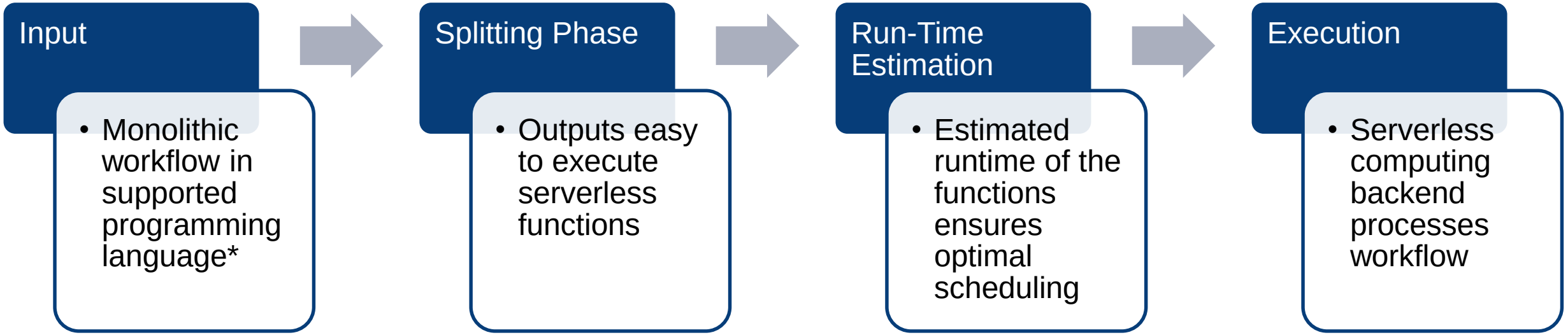
Inside the Blackbox – Shortened Iterative Workflow Example



Monolithic Workflows

OUR VISION

Our Vision for Monolithic Workflows



*Planned: At least Python & Javascript

Towards Serverless Functions

SPLITTING MONOLITHIC WORKFLOWS

Splitting Monolithic Workflows

Currently reviewing and researching:

- How to improve scalability, parallelization and runtime by:
  - Shrinking the percentage of serial program code
  - Assessing optimizations for the remainder
  - Allowing users to annotate their code with e.g. specific tags that enable additional options

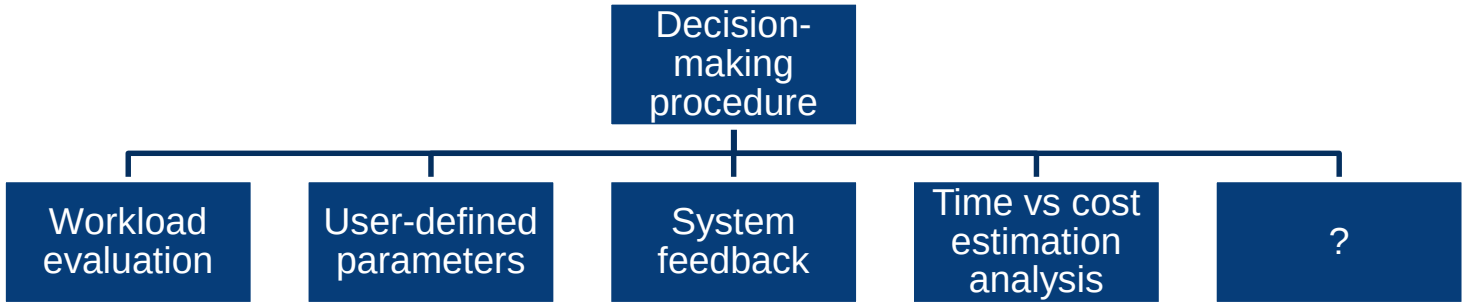
- What a process might look like for each specific workflow, for example:
  1. Construct a graph representation of the workflow
  2. Split off nodes by using a predefined ruleset
  3. Map theses resulting nodes to (serverless) functions that can be executed

Run-Time Estimation

ESTIMATING SERVERLESS FUNCTIONS

Estimating Serverless Functions

- System reliable performance requires run time estimation based on constant and holistic input data scheduling and validation
  - Review serverless functions run-time estimation approaches
  - Research and compare decision-making procedures and their components

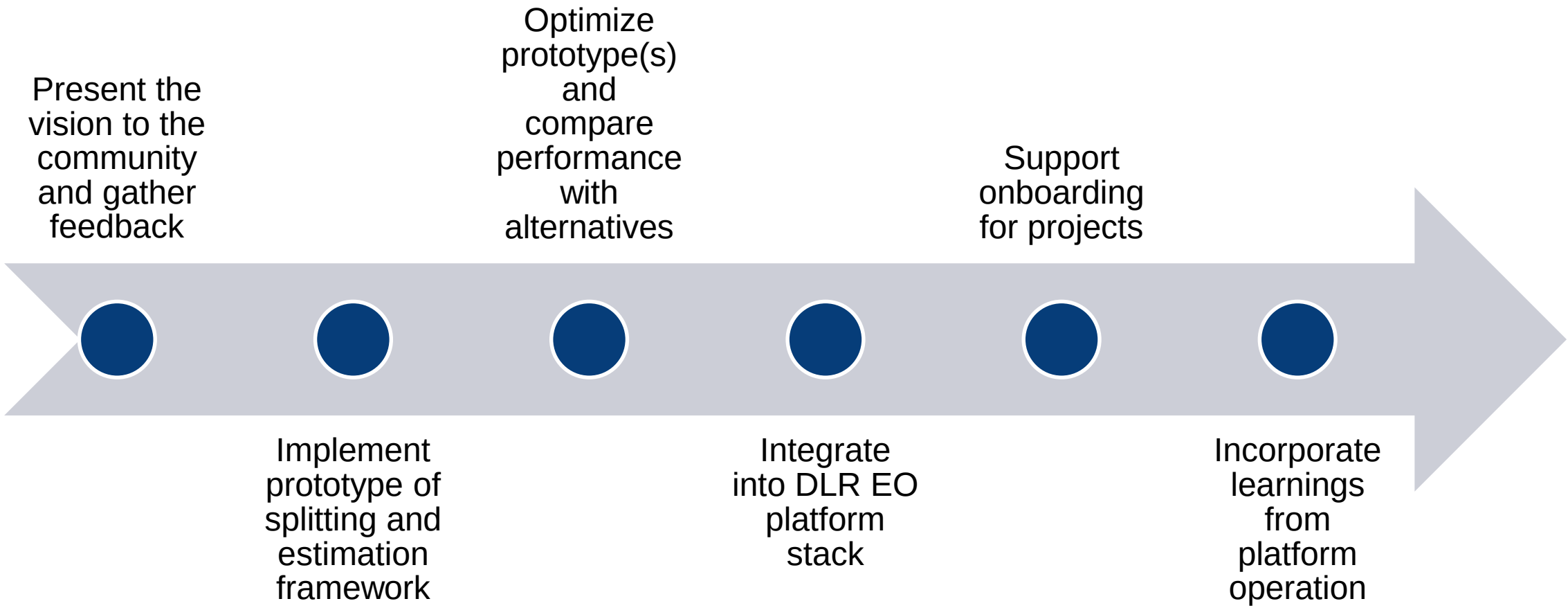


- Open research questions:
  - Feedback improvement
  - Third-party AI prediction engines and the initial bootstrap procedure
  - User-defined parameters in accordance with specific industry needs
  - The Estimation processes and their influence on the overall function execution time
  - Cost prediction of serverless computation and its importance for the business decision making

Future Work

NEXT STEPS

Planned Process from Now to Use on First Platform



The End (for now…)

Thank you all very much for your attention!

Dennis Kaiser

E-Mail:
dennis.kaiser@dlr.de

Site:
http://go.uniwue.de/kaiser

Bohdan Dovhan

E-Mail:
bohdan.dovhan@dlr.de