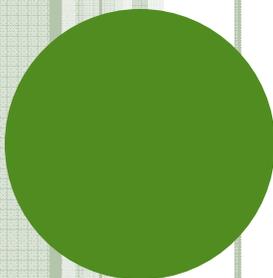
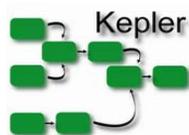


FLUXO DE DADOS EM WORKFLOWS CIENTÍFICOS: O SISTEMA KEPLER

Aluno: William Tankiti Yamamoto
Orientador: João Eduardo Ferreira

Departamento de Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo



CONTEÚDO

- Introdução
 - Definição de Workflow
 - O que é Workflow Científico?
 - Workflow Científico vs. Workflow de Negócio
- Motivação
- Sobre o Kepler
- Fluxo de Dados no KEPLER

DEFINIÇÃO DE WORKFLOW

- É a automação do processo de negócio onde documentos, informações ou tarefas são passadas de um participante para o outro de acordo com um conjunto de regras de procedimentos para se atingir um objetivo de negócio [*Workflow Management Coalition, 1995*]

O QUE É WORKFLOW CIENTÍFICO?

- Aplicação do conceito de *workflow* em ambientes científicos. Diferentemente dos ambientes de negócio, os ambientes científicos visam a análise e a manipulação de grande volume de ***dados***

DIFERENÇAS

Workflow Científico



Controle de fluxo de dados

Orientado a dado

Grande volume dados

Workflow de Negócio



Controle de fluxo de atividades

Orientado à atividade

Processos complexos

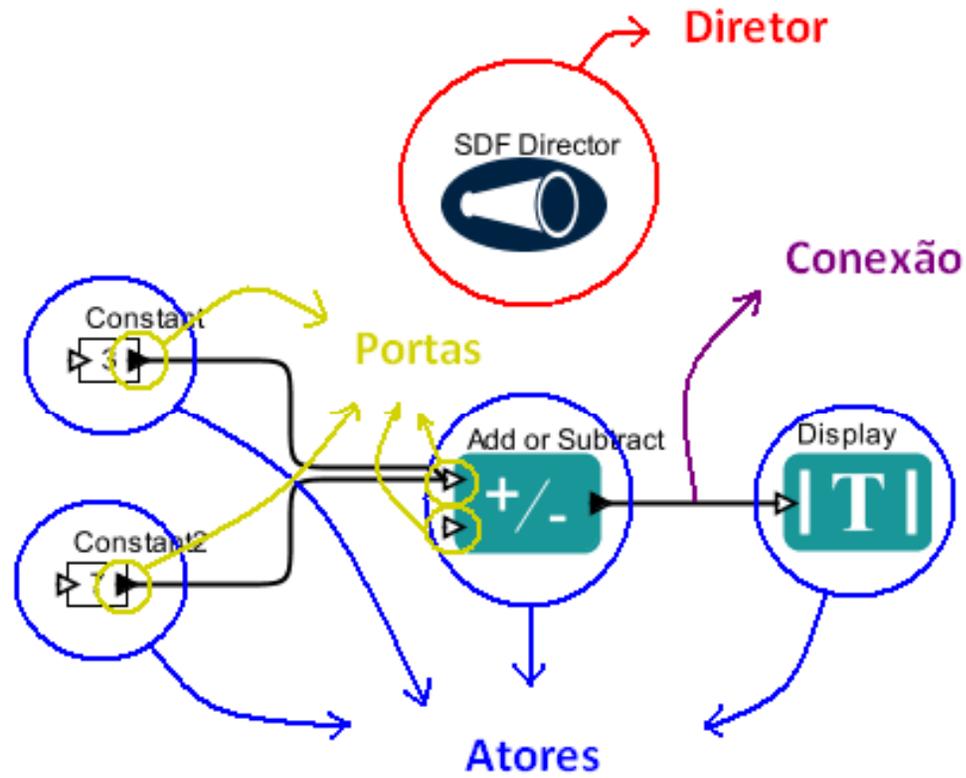
MOTIVAÇÃO

- Combinar o controle de fluxo de dados com o fluxo de atividades
- *Controle de fluxo de atividades* em Workflows de Negócios bastante difundido, utilizado e estudado
- Funcionamento do *fluxo de dados* em Workflows Científicos ainda precisa ser explorado
- Entender o *dataflow* é fundamental para integração

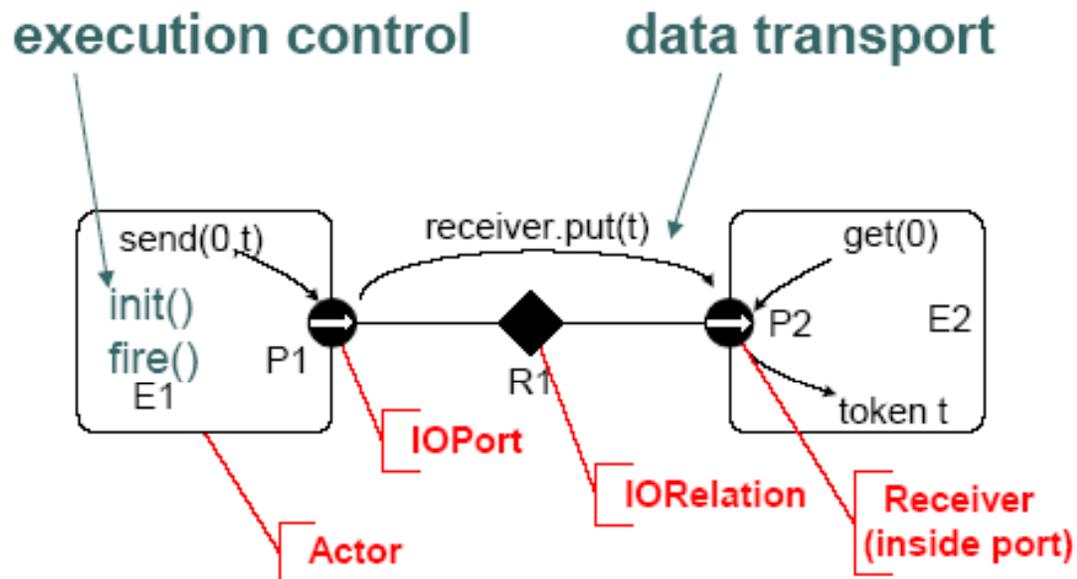
SOBRE O KEPLER

- Ambiente para *workflows científicos*
- Utilizado em áreas como a bioinformática, ecoinformática e geoinformática
- Modelo orientado a atores

COMPONENTES



FLUXO DE DADOS NO KEPLER



* Desenho retirado de
<http://embedded.eecs.berkeley.edu/concurrency/lectures/Lee05.pdf>

FLUXO DE DADOS NO KEPLER

- Comunicação entre os atores
 - Transporte de “mensagens” por meio de *portas* (single/multi)
 - Tipos de dados polimórficos (números, strings, arrays...)
 - Comportamentos polimórficos (process network, time-triggered, synchronous dataflow...)
- Receptores
 - Contêineres que armazenam os dados
 - *Portas* contêm *receptores*
 - Criados pelo *diretor*
 - Implementam a comunicação

REFERÊNCIAS BIBLIOGRÁFICAS

- Edward A. Lee and Stephen Neuendorffer, "Concurrent Models of Computation for Embedded Software", *Technical Memorandum UCB/ERL M04/26*, University of California, Berkeley, CA 94720, July 22, 2004
- Grace A. P. Borges, "Controle de dados em Processos de Negócio e Workflows Científicos", *Exame de Qualificação de Mestrado*, 2007
- S. Bowers and B. Ludaescher. Actor-oriented design of scientific workflows. In *International Conference on Conceptual Modeling (ER'2005)*, 2005
- Workflow Management Coalition, 1995