WIKIPEDIA

Dataflow

In computing, **dataflow** is a broad concept, which has various meanings depending on the application and context. In the context of software architecture, data flow relates to stream processing or reactive programming.

Contents

Software architecture

Dataflow computing is a software paradigm based on the idea of representing computations as a directed graph, where nodes are computations and data flow along the edges.[1] Dataflow can also be called stream processing or reactive programming.[2]

There have been multiple data-flow/stream processing languages of various forms (see Stream processing). Data-flow hardware (see Dataflow architecture) is an alternative to the classic von Neumann architecture. The most obvious example of data-flow programming is the subset known as reactive programming with spreadsheets. As a user enters new values, they are instantly transmitted to the next logical "actor" or formula for calculation.

Distributed data flows have also been proposed as a programming abstraction that captures the dynamics of distributed multi-protocols. The data-centric perspective characteristic of data flow programming promotes high-level functional specifications and simplifies formal reasoning about system components.

Hardware architecture

Hardware architectures for dataflow was a major topic in computer architecture research in the 1970s and early 1980s. Jack Dennis of the Massachusetts Institute of Technology (MIT) pioneered the field of static dataflow architectures. Designs that use conventional memory addresses as data dependency tags are called static dataflow machines. These machines did not allow multiple instances of the same routines to be executed simultaneously because the simple tags could not differentiate between them. Designs that use content-addressable memory are called dynamic dataflow machines by Arvind. They use tags in memory to facilitate

parallelism. Data flows around the computer through the components of the computer. It gets entered from the input devices and can leave through output devices (printer etc.).

Concurrency

A dataflow network is a network of concurrently executing processes or automata that can communicate by sending data over *channels* (see message passing.)

In Kahn process networks, named after Gilles Kahn, the processes are *determinate*. This implies that each determinate process computes a continuous function from input streams to output streams, and that a network of determinate processes is itself determinate, thus computing a continuous function. This implies that the behavior of such networks can be described by a set of recursive equations, which can be solved using fixed point theory. The movement and transformation of the data is represented by a series of shapes and lines.

Other meanings

Dataflow can also refer to:

- Power BI Dataflow, a Power Query implementation in the cloud used for transforming source data into cleansed Power BI Datasets to be used by Power BI report developers through the Microsoft Dataverse (formerly called Microsoft Common Data Service).
- Google Cloud Dataflow, a fully managed service for executing Apache Beam pipelines within the Google Cloud Platform ecosystem.

See also

- Binary Modular Dataflow Machine (BMDFM)
- Communicating sequential processes
- Complex event processing
- Data flow diagram
- Data-flow analysis, a type of program analysis
- Data stream
- Dataflow programming (a programming language paradigm)
- Erlang (programming language)
- Flow-based programming (FBP)
- Flow control (data)
- Functional reactive programming
- Lazy evaluation
- Lucid (programming language)
- Oz (programming language)
- Packet flow
- Pipeline (computing)
- Pure Data

- TensorFlow
- Theano

References

1. Schwarzkopf, Malte (7 March 2020). "The Remarkable Utility of Dataflow Computing" (https://www.sigops.org/2020/the-remarkable-utility-of-dataflow-computing/). *ACM SIGOPS*. Retrieved 31 July 2022.
2. A Short Intro to Stream Processing (http://www.jonathanbeard.io/blog/2015/09/19/streaming-and-dataflow.html)

The dictionary definition of *dataflow* at Wiktionary