WESTERN SYDNEY UNIVERSITY

FOG AND EDGE COMPUTING: CHALLENGES AND EMERGING TRENDS

Dr Rodrigo N. Calheiros

School of Computer, Data and Mathematical Sciences
Western Sydney University, Australia

Seminar – IEEE Dhofar University Student Branch, 04-Nov 2021

WESTERN SYDNEY
UNIVERSITY

CONTENTS

01.
Clouds, fog, edge...

Let's time travel

The New York Times

Google to Acquire YouTube for $1.65 Billion

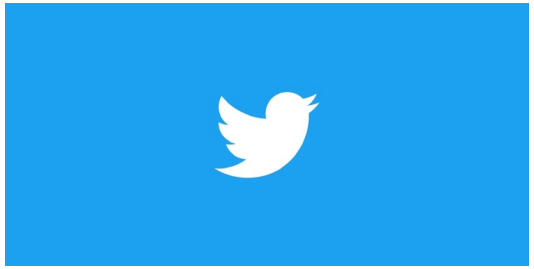
By Andrew Ross Sorkin and Jeremy W. Peters

Google announced this afternoon that it would buy YouTube, the popular video-sharing Web site, for stock that it valued at $1.65 billion.

Google beat out a number of other YouTube suitors, including Microsoft, Viacom, Yahoo and the News Corporation. By successfully negotiating the deal, Google has once again proved that it is the Internet's dominant player.

Under the terms of the deal, YouTube will retain much of its identity and will keep its name and its office in San Bruno, Calif., more than 25 miles from Google's headquarters in Mountain View.

Chad Hurley, YouTube's co-founder and chief executive, has repeatedly said he would prefer for his company to remain independent. Asked about such comments in a conference call with Wall Street analysts and investors held late this afternoon, Mr. Hurley said his company did want to stay independent, adding that "by working with Google, that's still the case."



Let's time travel

The New York Times

Google to Acquire YouTube for $1.65 Billion

By Andrew Ross Sorkin and Jeremy W. Peters

Google annou
popular video
billion.

Google beat o
Microsoft, Via
successfully n
that it is the Internet's dominant player.

Under the terms of the deal, YouTube will retain much of its identity and will keep its name and its office in San Bruno, Calif., more than 25 miles from Google's headquarters in Mountain View.

Chad Hurley, YouTube's co-founder and chief executive, has repeatedly said he would prefer for his company to remain independent. Asked about such comments in a conference call with Wall Street analysts and investors held late this afternoon, Mr. Hurley said his company did want to stay independent, adding that "by working with Google, that's still the case."

Trivia:
What these things have in common?

Let's time travel

The New York Times

Google to Acquire YouTube for $1.65 Billion

By Andrew Ross Sorkin and Jeremy W. Peters

Google annou
popular video
billion.

Google beat o
Microsoft, Via
successfully n
that it is the Internet's dominant player.

Under the terms of the deal, YouTube will retain much of its identity and will keep its name and its office in San Bruno, Calif., more than 25 miles from Google's headquarters in Mountain View.

Chad Hurley, YouTube's co-founder and chief executive, has repeatedly said he would prefer for his company to remain independent. Asked about such comments in a conference call with Wall Street a
Hurley said
"by working

Trivia:
What these things have in common?

Answer:
They all took place in 2006!

What happened in 2006 that is relevant to this talk?

What happened in 2006 that is relevant to this talk?

Press release

Amazon Web Services Launches

Share

March 14, 2006 at 3:00 AM EST

S3 Provides Application Programming Interface for Highly Scalable Reliable, Low-Latency Storage at Very Low Costs

SEATTLE--(BUSINESS WIRE)--March 14, 2006-- Amazon Web Services today announced "Amazon S3(TM)," a simple storage service that offers software developers a highly scalable, reliable, and low-latency data storage infrastructure at very low costs. Amazon S3 is available today at http://aws.amazon.com/s3.

Amazon S3 is storage for the Internet. It's designed to make web-scale computing easier for developers. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Amazon S3 Functionality

Amazon S3 is intentionally built with a minimal feature set. The focus is on simplicity and robustness.

AWS News Blog

Amazon EC2 Beta

by Jeff Barr | on 25 AUG 2006 | Permalink | Share

Innovation never takes a break, and neither do I. From the steaming hot beaches of Cabo San Lucas I would like to tell you about the Amazon Elastic Compute Cloud, or Amazon EC2, now open for limited beta testing, with more beta slots to open soon.

Amazon EC2 gives you access to a virtual computing environment. Your applications run on a "virtual CPU", the equivalent of a 1.7 GHz Xeon processor, 1.75 GB of RAM, 160 GB of local disk and 250 Mb/second of network bandwidth. You pay just 10 cents per clock hour (billed to your Amazon Web Services account), and you can get as many virtual CPUs as you need. You can learn more on the EC2 Detail Page. We built Amazon EC2 using a virtual machine monitor by the name of Xen.

Amazon EC2 works in terms of AMIs, or Amazon Machine Images. Each AMI is a pre-configured boot disk — just a packaged-up operating system stored as an Amazon S3 object. There are web service calls to create images, and to assign them to virtual CPUs to run your application. If your application consists of the usual web server, business logic, and database tiers, you can built distinct AMIs for each tier, and then spawn one or more instances of each type based on the load.

In a previous post, Sometimes You Need Just a Little..., I alluded to the new world of scalable, on-demand web services. In that post I talked about the fact that sometimes a little bit of storage is all you need.

WESTERN SYDNEY UNIVERSITY

Fast forwarding to 2021

WESTERN SYDNEY UNIVERSITY

Fast forwarding to 2021

Netflix Streaming:
February 2007

Fast forwarding to 2021

NETFLIX

Netflix Streaming:
February 2007

1st iPhone:
July 2007

Fast forwarding to 2021

Netflix Streaming:
February 2007

1st iPhone:
July 2007

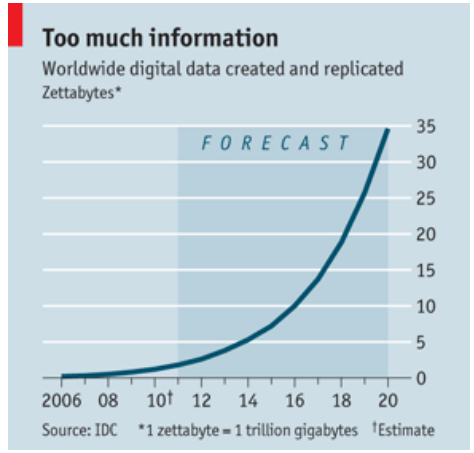
4G:
Early 2010's

Fast forwarding to 2021

Netflix Streaming:
February 2007

1st iPhone:
July 2007

4G:
Early 2010's



IoT & Big Data expansion
Since mid 2010's

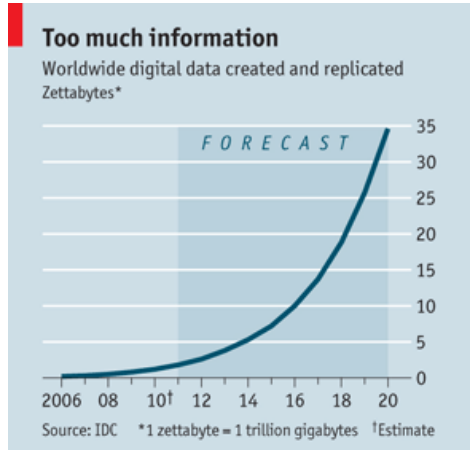
http://www.economist.com/node/18226961?story_id=18226961

Fast forwarding to 2021

Netflix Streaming:
February 2007

1st iPhone:
July 2007

4G:
Early 2010's



IoT & Big Data expansion
Since mid 2010's

http://www.economist.com/node/18226961?story_id=18226961

Coming Soon:
5G

Making the long story short

In this 15 years, the way web applications are designed and used changed a lot

- Users are on the move
- Multimedia content is much larger, mostly streaming
- Social networks based on videos emerged
- IoT is practical and widely adopted
- Too much data is generated and transferred
- 5G will increase bandwidth availability and consequently demand

Making the long story short

In this 15 years, the way web applications are designed and used changed a lot

- Users are on the move
- Multimedia content is much larger, mostly streaming
- Social networks based on videos emerged
- IoT is practical and widely adopted
- Too much data is generated and transferred
- 5G will increase bandwidth availability and consequently demand

Consequence: the model of centralized Cloud DCs is becoming obsolete and expensive

Enter Edge/Fog/Mobile computing

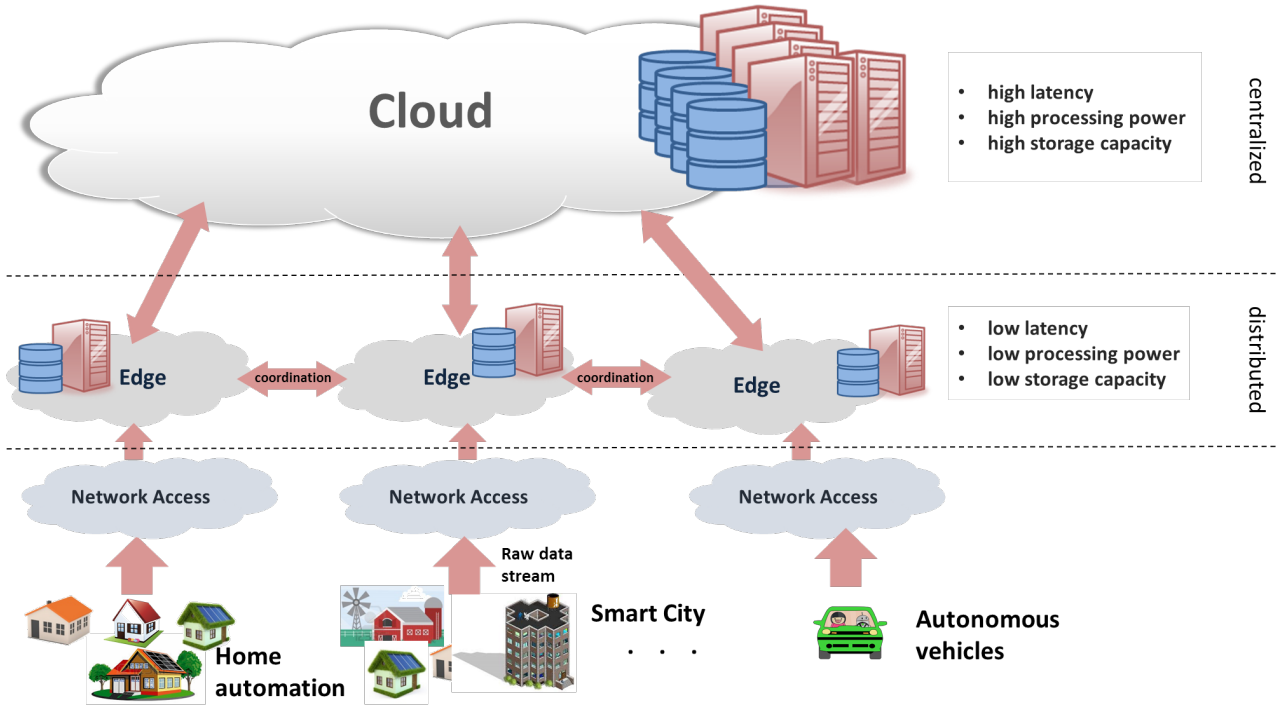
What:

Decentralized compute and storage, in a location closer to end users or sensors than cloud DCs are

Where?

User servers, cloudlets, networking equipment…

In the literature, the *where* matters. Some works define nomenclature according to how the decentralization happens



From Edge to Fog and other (confusing) "paradigms"

Satyanarayanan (2017, but has been using the term in earlier publications): "*Edge computing is a new paradigm in which substantial computing and storage resources—variously referred to as cloudlets, micro datacenters, or fog nodes—are placed at the Internet's edge in close proximity to mobile devices or sensors.*"

From Edge to Fog and other (confusing) "paradigms"

Satyanarayanan (2017, but has been using the term in earlier publications): "*Edge computing is a new paradigm in which substantial computing and storage resources—variously referred to as cloudlets, micro datacenters, or fog nodes—are placed at the Internet's edge in close proximity to mobile devices or sensors.* "

Shi and Dustdar (2016): "*[in Edge Computing] data processing occurs in part at the network edge, rather than completely in the cloud.*"

From Edge to Fog and other (confusing) "paradigms"

Satyanarayanan (2017, but has been using the term in earlier publications): "*Edge computing is a new paradigm in which substantial computing and storage resources—variously referred to as cloudlets, micro datacenters, or fog nodes—are placed at the Internet's edge in close proximity to mobile devices or sensors.* "

Shi and Dustdar (2016): "*[in Edge Computing] data processing occurs in part at the network edge, rather than completely in the cloud.*"

Dastjerdi et al. (2016): "[Fog computing is] *a distributed computing paradigm that fundamentally extends the services provided by the cloud to the edge of the network*"

From Edge to Fog and other (confusing) "paradigms"

Satyanarayanan (2017, but has been using the term in earlier publications): "*Edge computing is a new paradigm in which substantial computing and storage resources—variously referred to as cloudlets, micro datacenters, or fog nodes—are placed at the Internet's edge in close proximity to mobile devices or sensors.* "

Shi and Dustdar (2016): "*[in Edge Computing] data processing occurs in part at the network edge, rather than completely in the cloud.*"

Dastjerdi et al. (2016): "[Fog computing is] *a distributed computing paradigm that fundamentally extends the services provided by the cloud to the edge of the network*"

Cisco: *"Fog computing is a standard that defines how edge computing should work, and it facilitates the operation of compute, storage and networking services between end devices and cloud computing data centers."*

From Edge to Fog and other (confusing) "paradigms"

Satyanarayanan (2017, but has been using the term in earlier publications): "*Edge computing is a new paradigm in which substantial computing and storage resources—variously referred to as cloudlets, micro datacenters, or fog nodes—are placed at the Internet's edge in close proximity to mobile devices or sensors.* "

Shi and Dustdar (2016): "*[in Edge Computing] data processing occurs in part at the network edge, rather than completely in the cloud.*"

Dastjerdi et al. (2016): "[Fog computing is] *a distributed computing paradigm that fundamentally extends the services provided by the cloud to the edge of the network*"

Cisco: *"Fog computing is a standard that defines how edge computing should work, and it facilitates the operation of compute, storage and networking services between end devices and cloud computing data centers."*

Dew computing…

What really matters?

- Data from sensors IoT sensors (or other data sources) go to the cloud or to the near device
- Near device aggregates and preprocesses before uploading to the cloud
    - Or computes locally
    - Or offloads directly
- Pricing in the edge layer is likely to be different from the cloud layer
    - Nature of the variation is contentious, and depends on the definition used
- Computing power in the edge is likely to be heterogeneous and smaller than in the cloud

02.
Challenges

Practical issues

- Existing frameworks are not intuitive to use
    - Sometimes harder than just deploying the application in the cloud and in a Raspberry Pi
- There are still many proprietary devices out there
    - SDKs may be restricted to one type of system, forcing the use of extra connector devices
- Offloading is a difficult task to execute
- Conditions for offloading do not match with practical scenarios
    - Too broad generalizations
    - Expect too much from end users/Dev and Ops teams
    - Cost of implementation of solution may outweigh its benefits
    - "Magical" solution for deployment

Image credits: https://medium.datadriveninvestor.com/a-broken-system-pt-i-birth-c9c215895fba

Other challenges

1. Heterogeneity in the edge layer
2. Deployment
3. Programming models for the edge
4. Handling failures
5. Security and trust in the edge layer

Image credits: https://www.searchenginejournal.com/seo-challenges/212614/

Other challenges

1. Heterogeneity in the edge layer
2. **Deployment**
3. **Programming models for the edge**
4. Handling failures
5. Security and trust in the edge layer

Image credits: https://www.searchenginejournal.com/seo-challenges/212614/

03.
Emerging trends

Serverless computing (FaaS)

- Pioneered with AWS Lambda (2014)
- One of the most misleading terms in our area
- It actually means that the client only needs to provide the application code and set triggers for the code to run (events)
- No concerns with hardware, platform, scalability, etc.
- Strict programming model
- Stateless, asynchronous computation with short time out
- Open Source options available (Apache OpenWhisk, OpenStack Qinling)

Serverless computing (FaaS)

- Pioneered with AWS Lambda (2014)
- One of the most misleading terms in our area
- It actually means that the client only needs to provide the application code and set triggers for the code to run (events)
- No concerns with hardware, platform, scalability, etc.
- Strict programming model
- Stateless, asynchronous computation with short time out
- Open Source options available (Apache OpenWhisk, OpenStack Qinling)

Main issues for adoption by developers:

- Applications may require large refactoring to fit the required model
- Applications may not fit the model at all
- Applications may need to be designed as a workflow of functions

Main issues for adoption by Operations:

- Edge scenarios are much more complex than clusters to which Open Source platforms are designed to work (access via WAN, heterogeneity, volatility)
- Data placement

Serverless computing (FaaS) opportunities

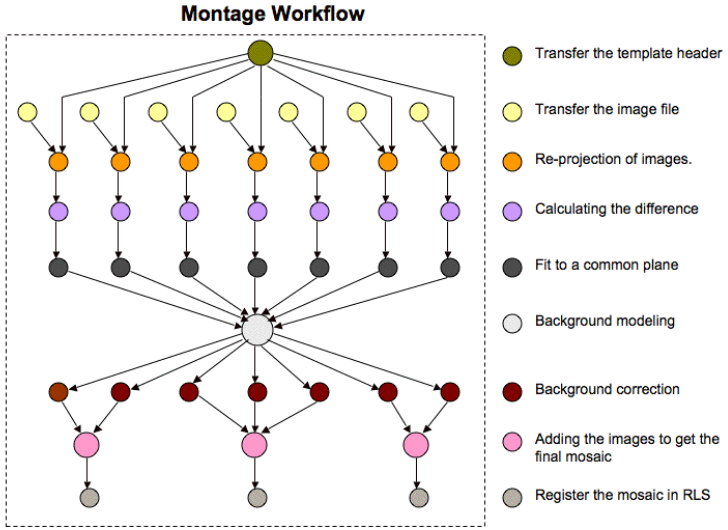
- Supporting statefullness
- Automating workflow description
- Service discovery
- Coordination
- Handling heterogeneity
- Handling volatility
- Best approaches to hold application data

Hybrid computing models

- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community

Hybrid computing models

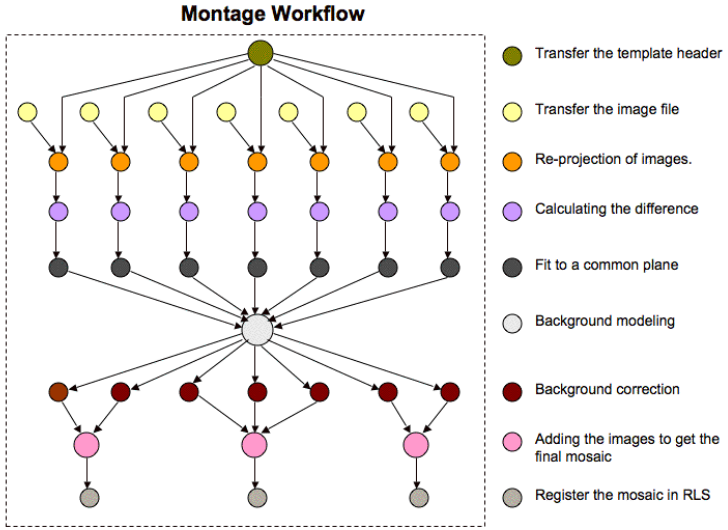
- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community



https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

Hybrid computing models

- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community

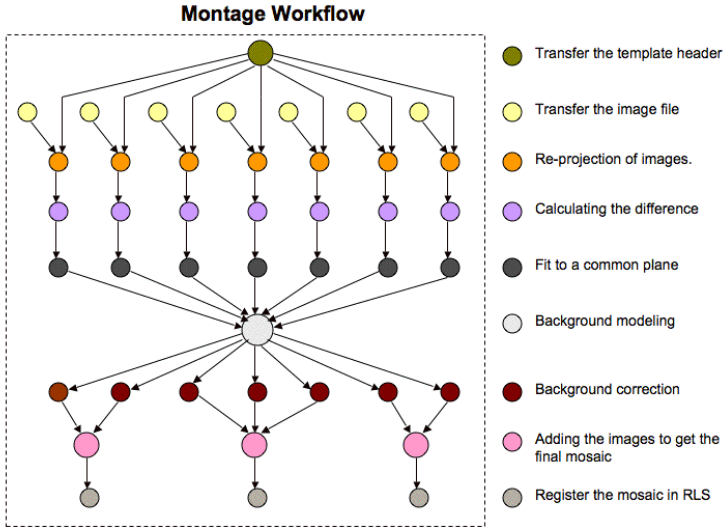


https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

- It can be seen as a type of "batch job" that requires HPC or distributed resources

Hybrid computing models

- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community



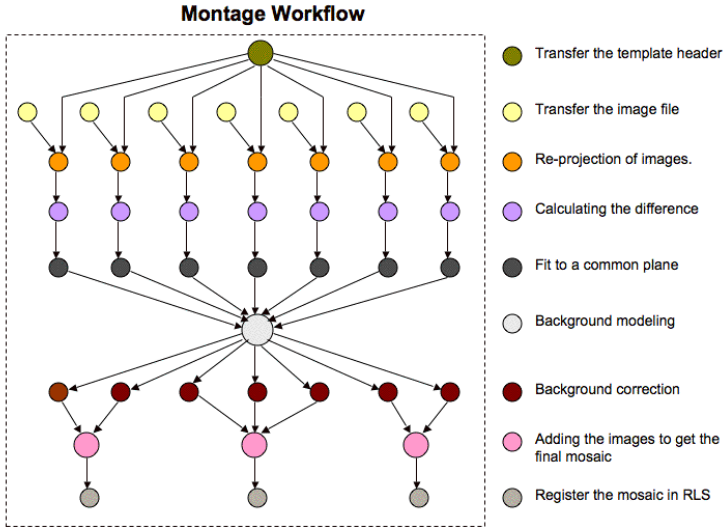
https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

- It can be seen as a type of "batch job" that requires HPC or distributed resources

- More recently, streaming models became popular
- Many similarities to workflows, though short-liven computations occurs in "always on" resources

Hybrid computing models

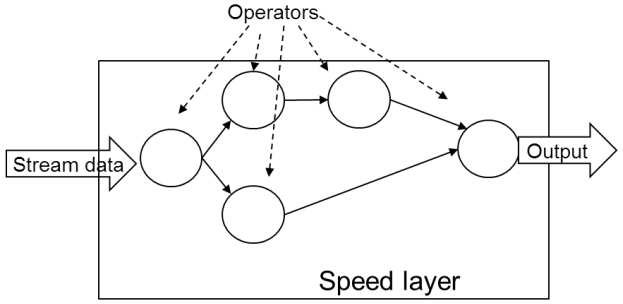
- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community



https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

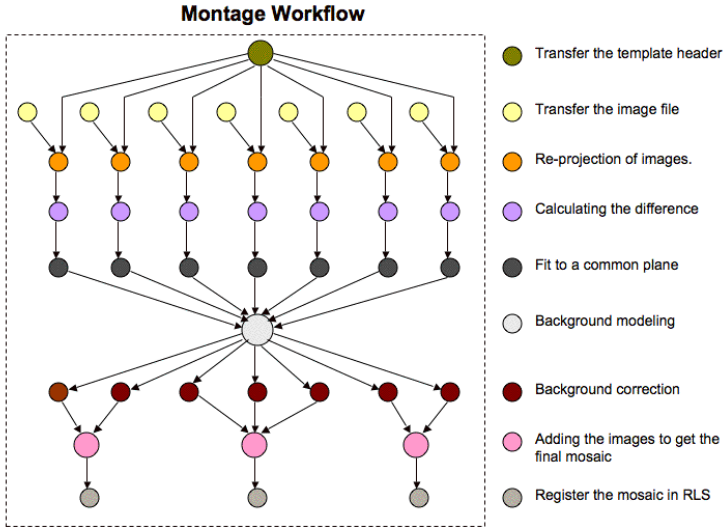
- It can be seen as a type of "batch job" that requires HPC or distributed resources

- More recently, streaming models became popular
- Many similarities to workflows, though short-liven computations occurs in "always on" resources



Hybrid computing models

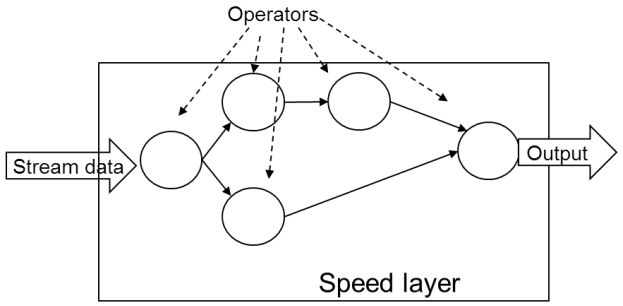
- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community



https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

- It can be seen as a type of "batch job" that requires HPC or distributed resources

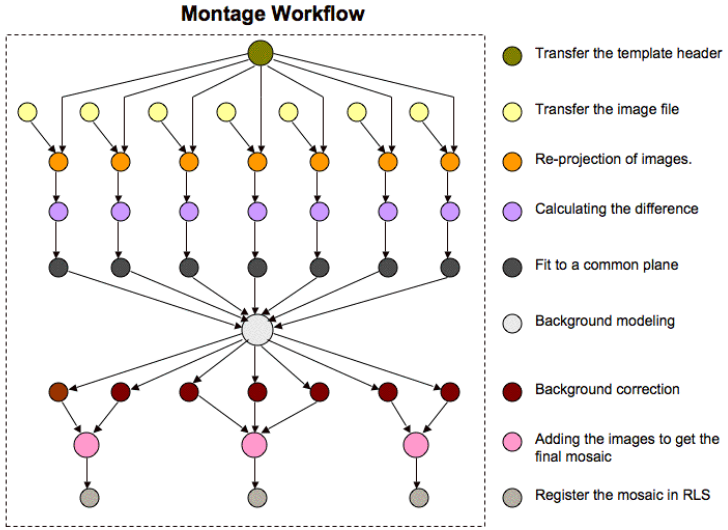
- More recently, streaming models became popular
- Many similarities to workflows, though short-liven computations occurs in "always on" resources



- Resources need to be managed differently to enable low latency of operators

Hybrid computing models

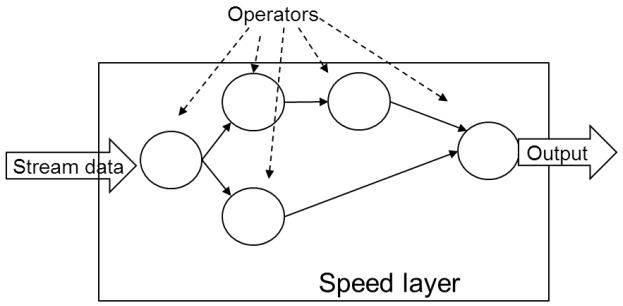
- For many years, the scientific workflow application model has been popular in the HPC/DS/Cloud community



https://vgrads.rice.edu/research/applications/images/montage-workflow/image_view_fullscreen

- It can be seen as a type of "batch job" that requires HPC or distributed resources

- More recently, streaming models became popular
- Many similarities to workflows, though short-liven computations occurs in "always on" resources



- Resources need to be managed differently to enable low latency of operators
- A hot topic today is how combine these two models on a single application

Hybrid computing models opportunities

- More accurate modelling of the problem
- Understanding applications that follow this model, and their requirements
- Fine tuning batch/streaming throughput
    - "Aggregation windows"?
    - Bigger streaming flow may increase the time of batch processing
    - Smaller flow means that data is being dropped, which may affect quality of output

Hybrid computing models opportunities

- More accurate modelling of the problem
- Understanding applications that follow this model, and their requirements
- Fine tuning batch/streaming throughput
    - "Aggregation windows"?
    - Bigger streaming flow may increase the time of batch processing
    - Smaller flow means that data is being dropped, which may affect quality of output
- This is already hard to achieve in a DC. How about when we add the Edge dimension to it?

Other interesting research questions

- How 5G can help to solve some of the problems of Edge computing
    - Overlay network over 5G-connected devices?
- Cooperative/collaborative edge computing
    - Edge nodes cooperate to perform computing tasks
    - They can also cooperate to enable storage or to distribute IoT data
    - How to protect data consistency?
- IoT-Edge-Cloud Continuum
    - Applications can seamlessly “move” across layers

03.
Concluding remarks

Take-aways

- Changes in the way application and content is consumed, occurred in the last 15 years, reduced the efficiency of centralized cloud DCs
- Edge computing addresses some of the sources of such inefficiencies by bringing computing closer to where the data is generated or needed
- Serverless computing and hybrid computing models are promising compute models for the Edge

Acknowledgements

The ideas discussed here emerged from research undertaken with, and from discussions with my colleagues A/Prof Bahman Javadi and Dr Kenan Matawie, with research undertaken by PhD students Raed Alsurdeh and Mohammad Alkhalaileh

WESTERN SYDNEY UNIVERSITY

FOG AND EDGE COMPUTING: CHALLENGES AND EMERGING TRENDS

Dr Rodrigo N. Calheiros

School of Computer, Data and Mathematical Sciences
Western Sydney University, Australia

Seminar – IEEE Dhofar University Student Branch, 04-Nov 2021