



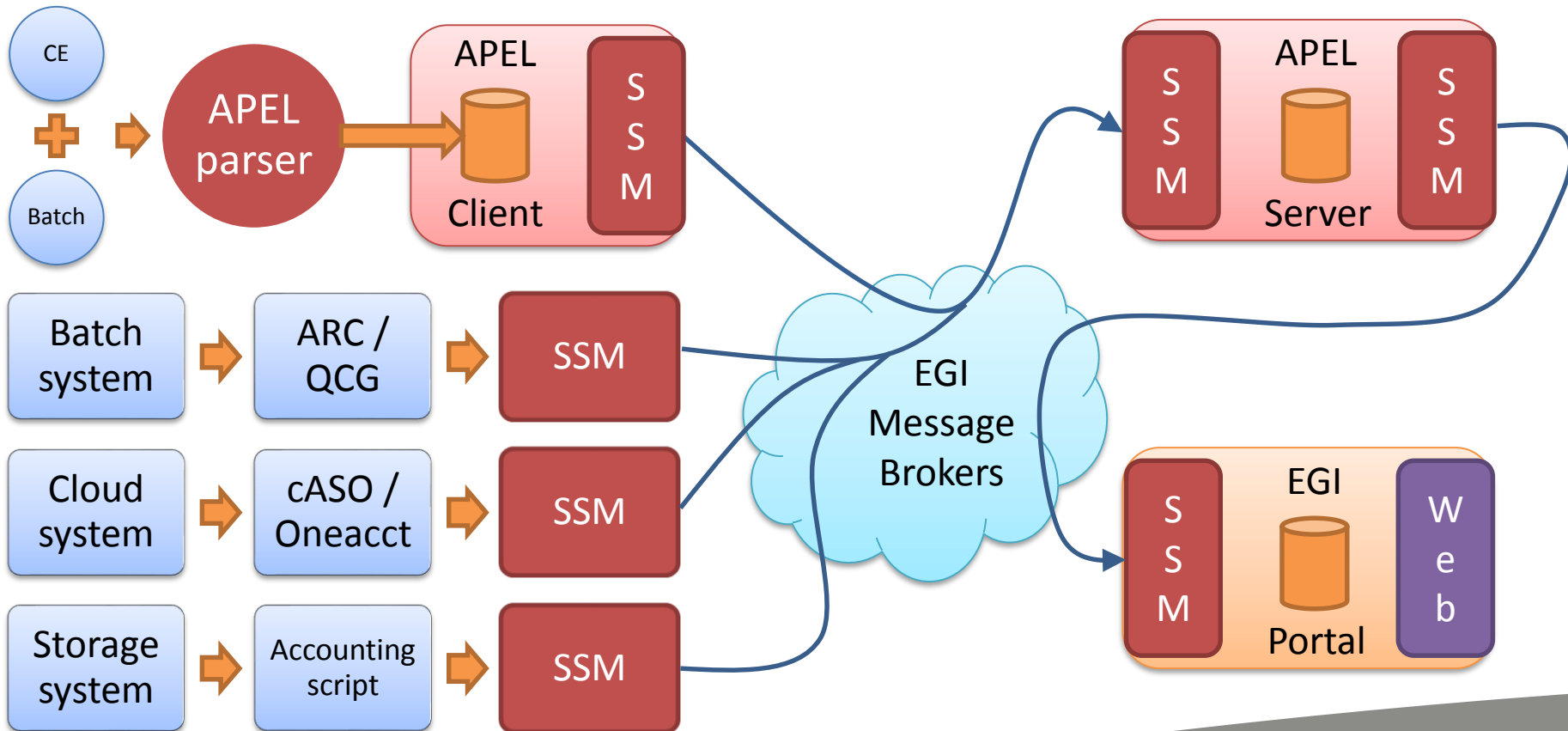
# APEL Accounting: Data Flow and Work Plan

Greg Corbett, Adrian Coveney

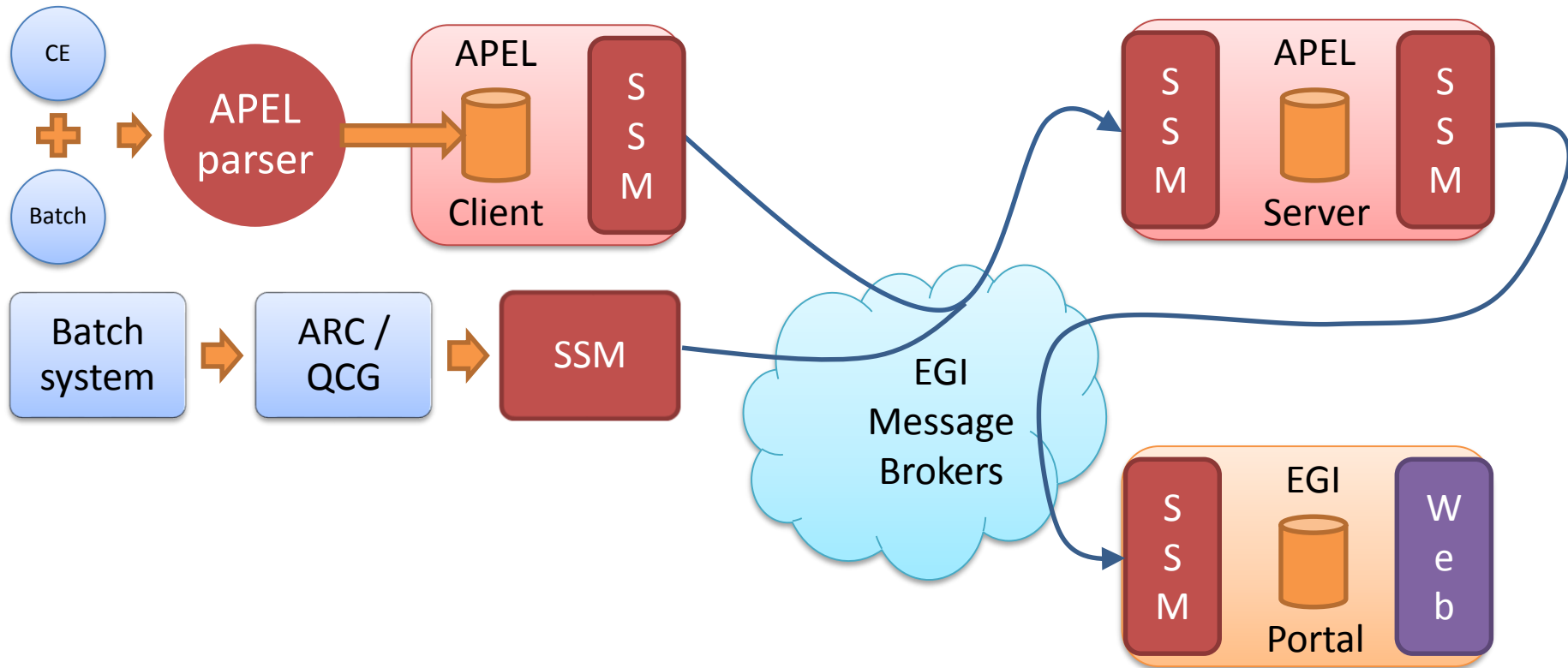
[apel-admins@stfc.ac.uk](mailto:apel-admins@stfc.ac.uk)

# Data Flow

# Data Flow

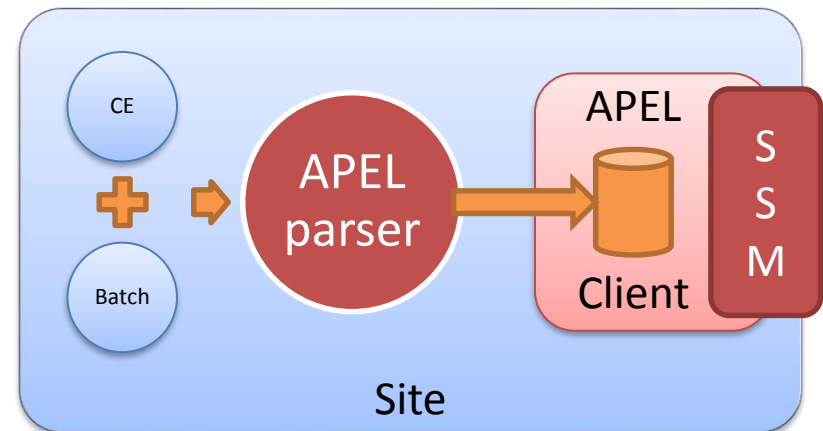


# Data Flow - Grid



# Data Flow - Site

- Sites install the APEL software
- Client parses the submit host (CE) logs and the worker node logs to create JobRecords
- JobRecords stored in client database
- Client can then either send individual records or summaries



# Data Flow - Site

- Configuration

/etc/apel/parser.cfg

/etc/apel/client.cfg

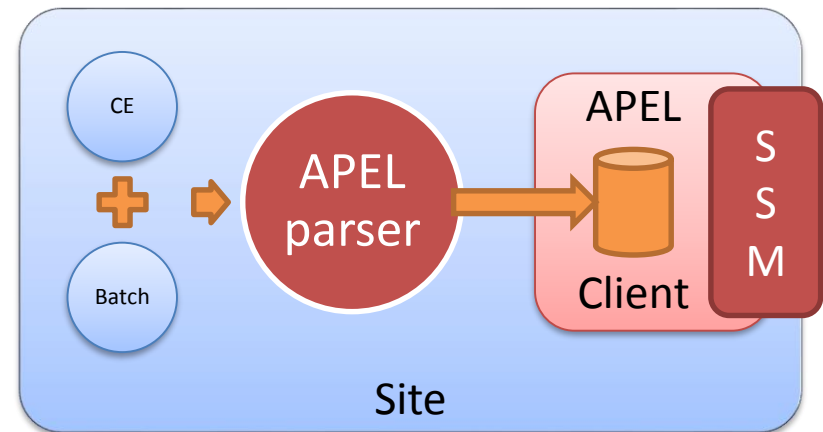
/etc/apel/sender.cfg

- Logging

/var/log/apelparser.log

/var/log/apel/client.log

/var/log/apel/ssmsend.log



# Data Flow - Site

- APEL Client Supported CEs
  - Cream CE
- APEL Client Supported Batch Systems
  - LSF
  - PBS
  - SGE/OGE
  - SLURM
  - HTCondor

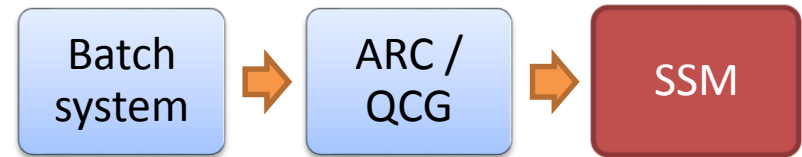
# Data Flow - Site

- For some batch systems, scripts parse raw batch logs into a '|' delimited format
  - [https://github.com/apel/apel/blob/dev/scripts/slurm\\_acc.sh](https://github.com/apel/apel/blob/dev/scripts/slurm_acc.sh)
  - [https://github.com/apel/apel/blob/dev/scripts/htcondor\\_acc.sh](https://github.com/apel/apel/blob/dev/scripts/htcondor_acc.sh)
- Parsers then read that format, mapping items to fields based on position
  - <https://github.com/apel/apel/blob/dev/apel/parsers/slurm.py#L48>
  - <https://github.com/apel/apel/blob/dev/apel/parsers/htcondor.py#L37>

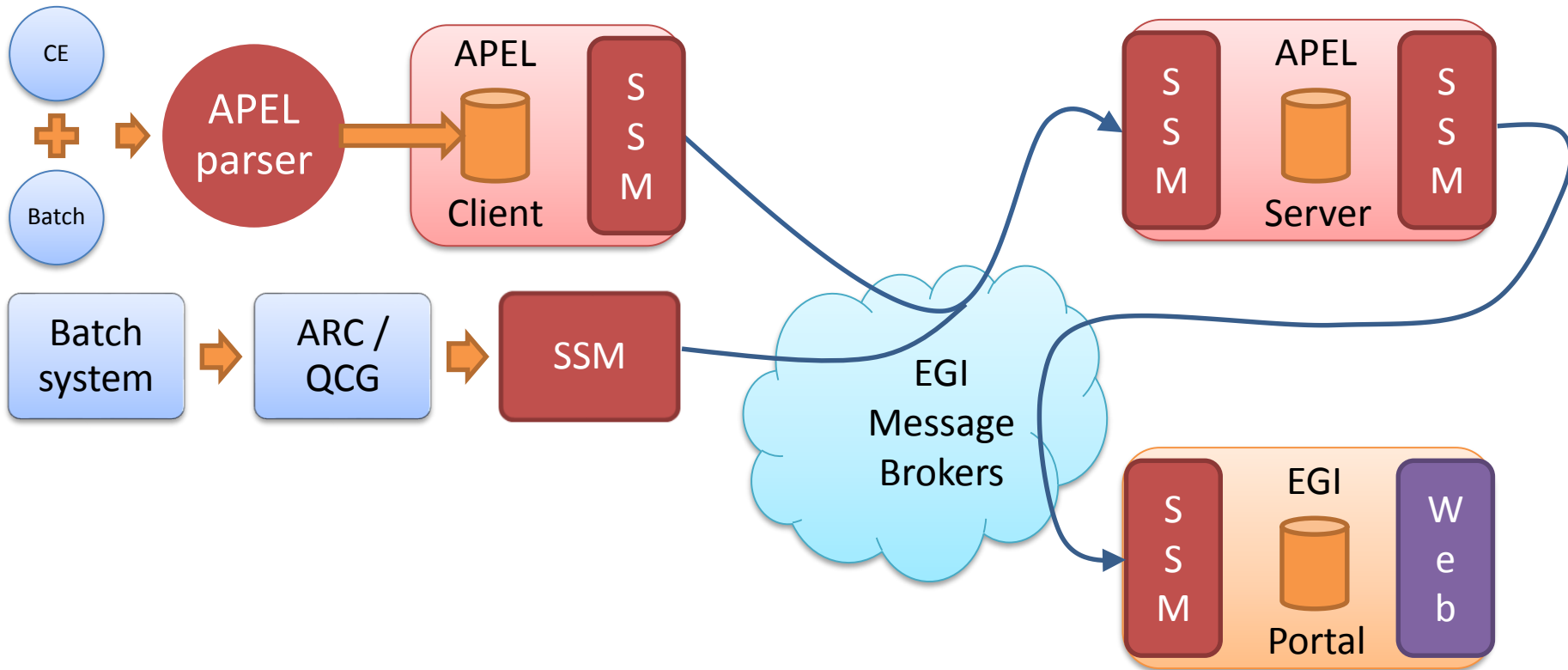


# Data Flow - Site

- For sites not using the APEL Client it's a black box as far as the APEL team is concerned.
- So long as the output is in our record format, we can ingest it regardless of the exact source.
- Sent to us via SSM and the EGI message brokers



# Data Flow

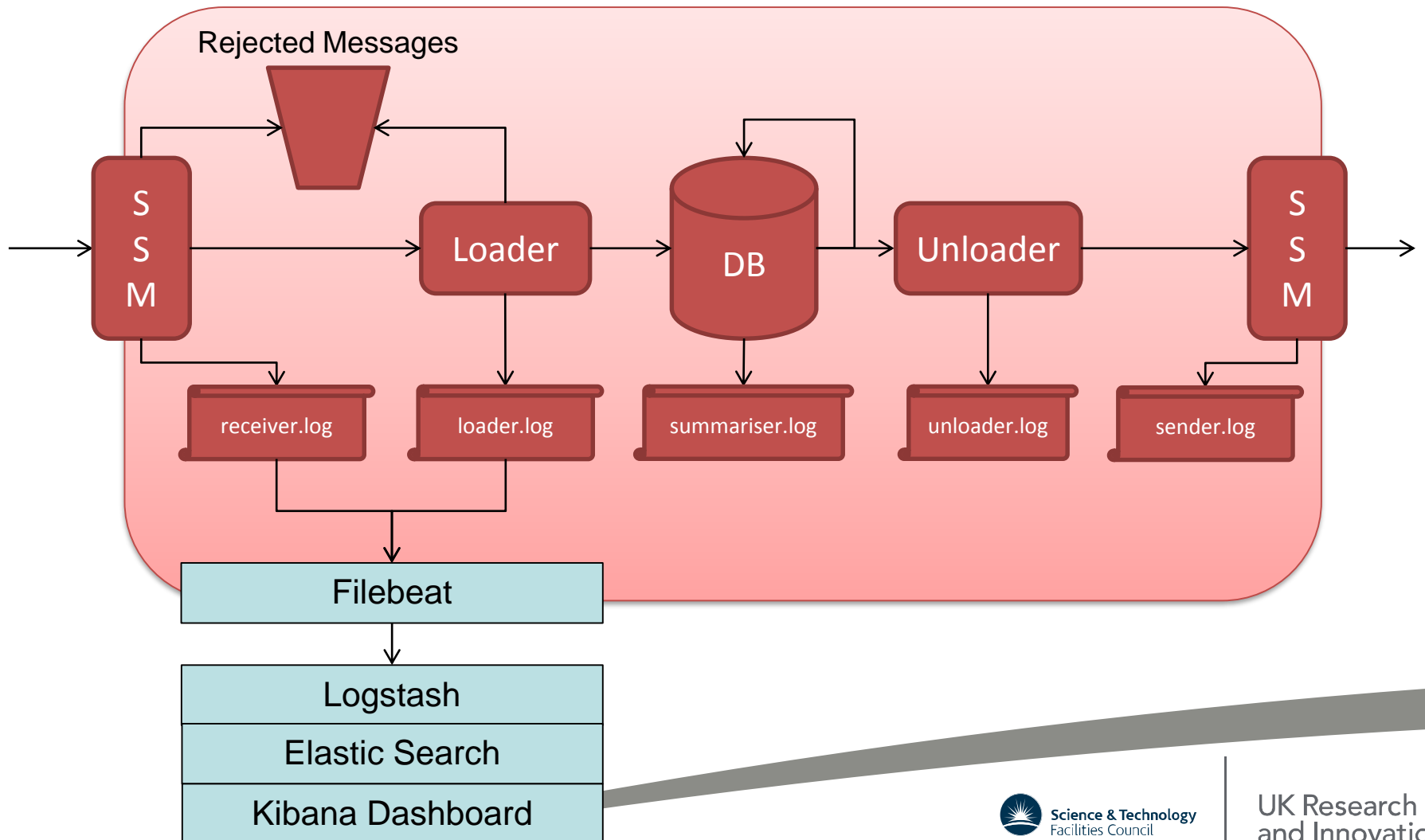


# Data Flow - SSM

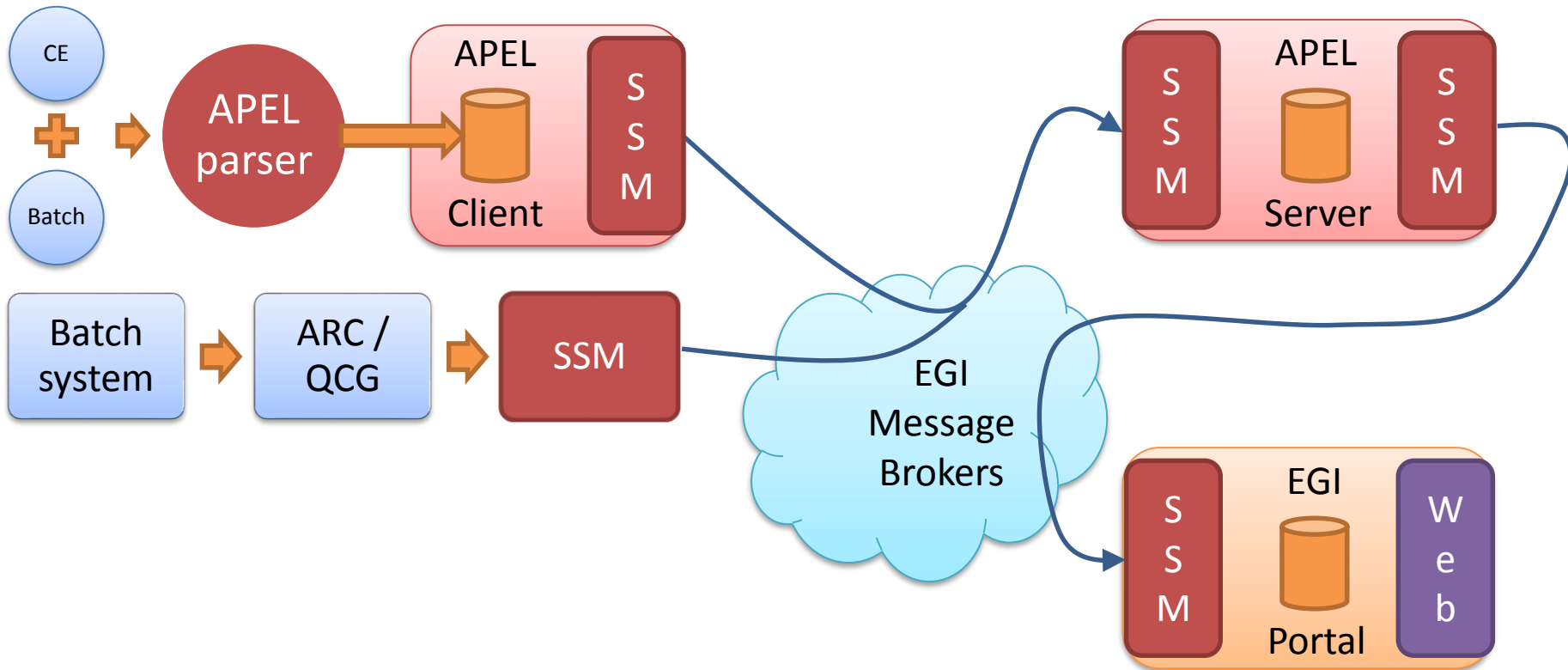


- Our messaging software uses the STOMP protocol and the EGI Message Brokers
  - Plan is to move away from the current STOMP message brokers towards HTTP based brokers, the ARGO Messaging Service.
  - <https://argoeu.github.io/guides/messaging/>

# Data Flow - Central Server



# Data Flow



# Work Plan

# Work Plan

- Raw unscaled wall clock planned to be tested as part of IRIS work.
- APEL is site-centered so collecting experiment-specific measurements may have to be a separate system. Probably doable, but possibly better as part of another system.
- Currently working with Stephen Jones on including benchmarks in local APEL client config. Future plan to enable fetching from CE JSON via GOCDDB once that's more widely used and more stable.

# Work Plan

- Having a different benchmarks ok so long as there's a conversion factor. Storing more than one benchmark is another matter. Latter would require schema change. More research required.
- Move to using ARGO Messaging Service – new version of SSM with minimal configuration changes.
- Defining a summary record for Storage Accounting to ease load on portal



# Questions?