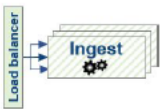
4.4. Ingest process flow

The *Ingest* service is comprised of multiple nodes that process files identically, from any reporter. **Order event** data files as well as **reference** data files are delivered to the SFTP service where they are stored until pulled and processed by the *Load Balancer*, which distributes them to the Ingest nodes for processing, as described below.



① The ***Load Balancer*** distributes incoming order event and reference data files equitably across a pool of available *Ingest* nodes. In release 2.0, *any* node can process files from *any* reporter.

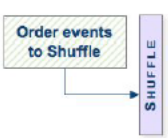
② The data file is received by the *Ingest* node and a reception acknowledged (ACK) message is sent (viat the *Reporter Portal*) to the reporter who submitted it.

③ **Order event** data files, in non-processed, "as-reported" condition, are submitted for archival storage on the *Object Store* S3 repository. This data becomes viewable to authorized users of the *Query* system.

④ The data file is *validation* tested for formatting, content and "process-ability". When issues are uncovered, the submitting reported is notified via the *Reporter Portal*.

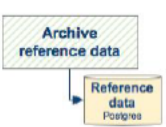
⑤ The validated order event data file is *translated* into proprietary binary format - i.e., the format in which it is stored on and accessible from the CAT.

⑥ After conversion to binary (but *before* submission to *Shuffle* processing) the non-lifecycled order event data is submitted for storage on the *Object Store* S3 repository, where it can be viewed using the *Query* service.



⑦ At this point, ingested order event data is submitted to the *Shuffle* layer for distribution to the *Lifecycle* process for linkage processing (described in subsection below).

Note: Reference data is *not* sent to the *Shuffle*/*Lifecycle* process pipeline.



⑧ The ingested **reference data** is stored in a Postgres database, which resides in the *Core* service architecture.