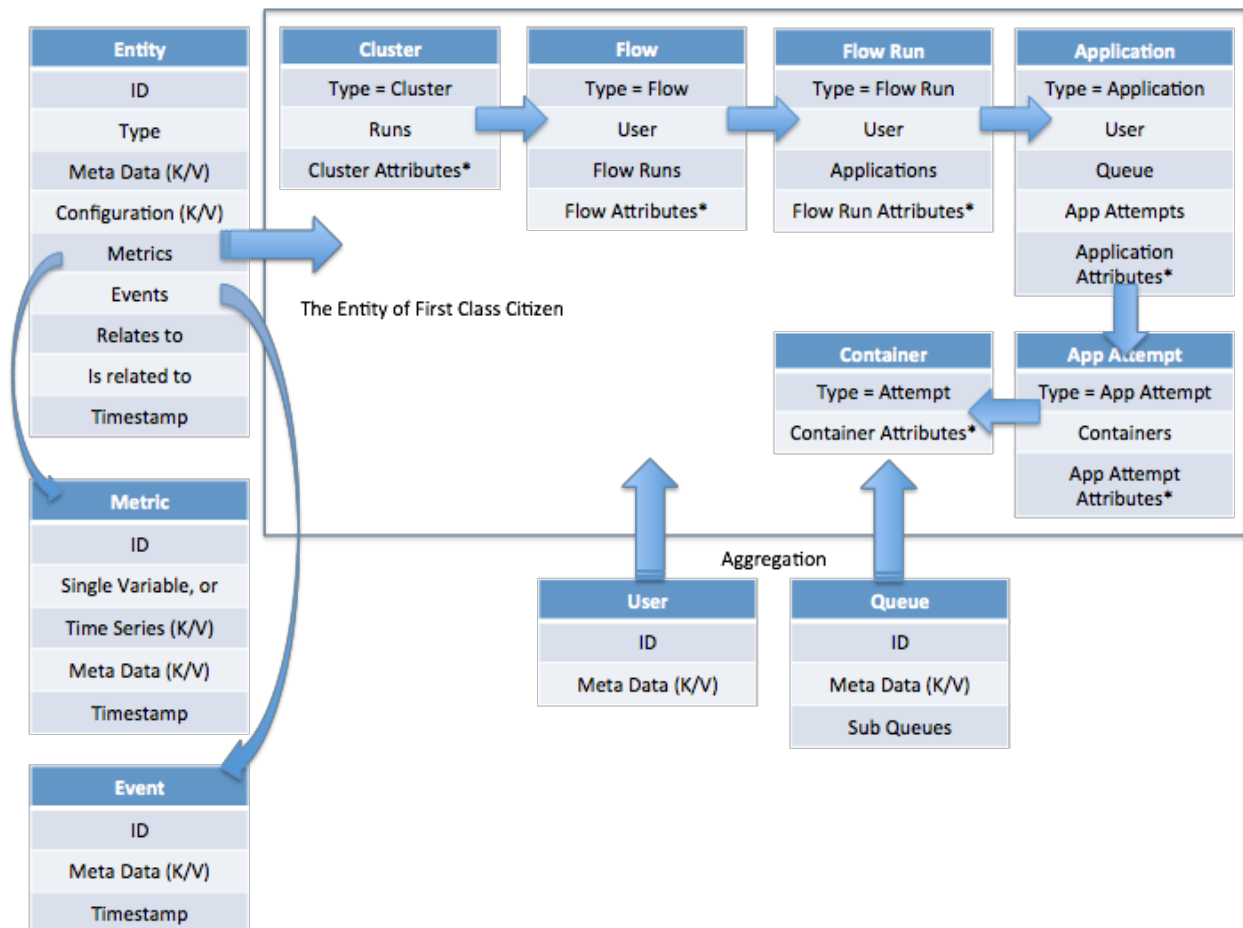


Timeline service next gen data model



Above is diagram of the data model:

- It inherits from the data model of the timeline service of current gen, and is based the entity/event schema.
- In addition to the events, an entity may have a couple of metrics, each of which contains either single metric data or time series. Typically the metric can be associated to an application entity, but we can still aggregate the metrics of all applications in a flow.
- Learning from the experience of previous data model, the relationship between entities could be directional.
- Configuration could be a key/value pair section of an entity.
- To accommodate the data model proposed in the timeline service next gen design doc, we predefine TS first class citizen entities - Cluster, Flow, FlowRun, Application and ApplicationAttempt and Container, and maintains the parent-child relationship among them.
- The hierarchy among them [Cluster -> [Flow -> [FlowRun -> [FlowRun -> ...]]]] -> Application -> ApplicationAttempt -> Container. FlowRun can be nested. In the simple use case we have orphan Application.

- We add the support of the lineage from workflow to application while we open the option to further define fine-granularity entity in the scope of an application.
- In addition, we have two other concepts, User and Queue, which allows we run aggregation from these two aspects too.

Some more comments:

- The predefined entities can have special attributes other than the meta-data, we can define them later on demand.
- The data model are somehow the protocol between user and TS, it may not be equivalent to the storage schema of TS backend.