

CLAIMS

1. A computer-implemented method, comprising operations for:
 - receiving a description of a multi-tiered edge topology with a plurality of nodes, a multimedia stream analytics composition that is executed on the plurality of nodes, and performance objectives, wherein the multimedia stream analytics composition includes tasks that use analytic models;
 - optimizing the analytic models;
 - clustering the optimized analytic models to form clusters of optimized analytic models;
 - selecting a representative optimized analytic model from each of the clusters of optimized analytic models;
 - determining a configuration recommendation that indicates deployment of the tasks and of each selected representative optimized analytic model on the plurality of nodes to meet the performance objectives;
 - generating one or more workflows from the configuration recommendation to execute the tasks of the multimedia stream analytics composition on nodes of different tiers of the multi-tiered edge topology; and
 - executing the one or more workflows on the plurality of nodes to generate output for the multimedia stream analytics composition.
2. The computer-implemented method of claim 1, wherein the performance objectives comprise any combination of accuracy, precision, recall, and throughput.
3. The computer-implemented method of claim 1 or claim 2, wherein optimizing the analytic models comprises compressing the analytic models.
4. The computer-implemented method of any preceding claim, wherein the clustering is based on benchmark results, and wherein the benchmark results are generated by benchmark testing of the optimized analytic models using test data and the plurality of nodes in the multi-tiered edge topology.
5. The computer-implemented method of any preceding claim, wherein one or more of images and metadata are transmitted between the nodes of the plurality of nodes at different transmission speeds.
6. The computer-implemented method of any preceding claim, further comprising operations for:
 - for each node in the multi-tiered edge topology, defining an objective function based on performance objectives for performance, accuracy, precision, and recall;
 - for each node, defining a cumulative objective function based on a cumulative performance, a cumulative accuracy, a cumulative precision, and a cumulative recall; and

creating an overall objective function for the multi-tiered edge topology based on the cumulative objective function of each node.

7. The computer-implemented method of any preceding claim wherein the plurality of nodes includes a central server in a cloud infrastructure, wherein the central server stores one of images and metadata for the multimedia stream analytics composition in a data store.

8. The computer-implemented method of any preceding claim, wherein a Software as a Service (SaaS) is configured to perform the operations of the method.

9. A computer program product comprising a computer readable storage medium having program code embodied therewith, the program code executable by at least one processor to perform operations for:

receiving a description of a multi-tiered edge topology with a plurality of nodes, a multimedia stream analytics composition that is executed on the plurality of nodes, and performance objectives, wherein the multimedia stream analytics composition includes tasks that use analytic models;

optimizing the analytic models;

clustering the optimized analytic models to form clusters of optimized analytic models;

selecting a representative optimized analytic model from each of the clusters of optimized analytic models;

determining a configuration recommendation that indicates deployment of the tasks and of each selected representative optimized analytic model on the plurality of nodes to meet the performance objectives;

generating one or more workflows from the configuration recommendation to execute the tasks of the multimedia stream analytics composition on nodes of different tiers of the multi-tiered edge topology; and

executing the one or more workflows on the plurality of nodes to generate output for the multimedia stream analytics composition.

10. The computer program product of claim 9, wherein the performance objectives comprise any combination of accuracy, precision, recall, and throughput.

11. The computer program product of claim 9 or claim 10, wherein optimizing the analytic models comprises compressing the analytic models.

12. The computer program product of any of claims 9 to 11, wherein the clustering is based on benchmark results, and wherein the benchmark results are generated by benchmark testing of the optimized analytic models using test data and the plurality of nodes in the multi-tiered edge topology.

13. The computer program product of any of claims 9 to 12, wherein one or more of images and metadata are transmitted between the nodes of the plurality of nodes at different transmission speeds.
14. The computer program product of any of claims 9 to 13, wherein the program code is executable by the at least one processor to perform further operations for:
- for each node in the multi-tiered edge topology, defining an objective function based on performance objectives for performance, accuracy, precision, and recall;
 - for each node, defining a cumulative objective function based on a cumulative performance, a cumulative accuracy, a cumulative precision, and a cumulative recall; and
 - creating an overall objective function for the multi-tiered edge topology based on the cumulative objective function of each node.
15. The computer program product of any of claims 9 to 14, wherein the plurality of nodes includes a central server in a cloud infrastructure, wherein the central server stores one of images and metadata for the multimedia stream analytics composition in a data store.
16. The computer program product of any of claims 9 to 15, wherein a Software as a Service (SaaS) is configured to perform the operations of the computer program product.
17. A computer system, comprising:
- one or more processors, one or more computer-readable memories and one or more computer-readable, tangible storage devices; and
 - program instructions, stored on at least one of the one or more computer-readable, tangible storage devices for execution by at least one of the one or more processors via at least one of the one or more memories, to perform operations comprising:
 - receiving a description of a multi-tiered edge topology with a plurality of nodes, a multimedia stream analytics composition that is executed on the plurality of nodes, and performance objectives, wherein the multimedia stream analytics composition includes tasks that use analytic models;
 - optimizing the analytic models;
 - clustering the optimized analytic models to form clusters of optimized analytic models;
 - selecting a representative optimized analytic model from each of the clusters of optimized analytic models;
 - determining a configuration recommendation that indicates deployment of the tasks and of each selected representative optimized analytic model on the plurality of nodes to meet the performance objectives;
 - generating one or more workflows from the configuration recommendation to execute the tasks of the multimedia stream analytics composition on nodes of different tiers of the multi-tiered edge topology; and

executing the one or more workflows on the plurality of nodes to generate output for the multimedia stream analytics composition.

18. The computer system of claim 17, wherein the performance objectives comprise any combination of accuracy, precision, recall, and throughput.

19. The computer system of claim 17 or claim 18, wherein optimizing the analytic models comprises compressing the analytic models.

20. The computer system of any of claims 17 to 19, wherein the clustering is based on benchmark results, and wherein the benchmark results are generated by benchmark testing of the optimized analytic models using test data and the plurality of nodes in the multi-tiered edge topology.

21. The computer system of any of claims 17 to 20, wherein one or more of images and metadata are transmitted between the nodes of the plurality of nodes at different transmission speeds.

22. The computer system of any of claims 17 to 21, wherein the operations further comprise:
for each node in the multi-tiered edge topology, defining an objective function based on performance objectives for performance, accuracy, precision, and recall;
for each node, defining a cumulative objective function based on a cumulative performance, a cumulative accuracy, a cumulative precision, and a cumulative recall; and
creating an overall objective function for the multi-tiered edge topology based on the cumulative objective function of each node.

23. The computer system of any of claims 17 to 22, wherein the plurality of nodes includes a central server in a cloud infrastructure, wherein the central server stores one of images and metadata for the multimedia stream analytics composition in a data store.

24. The computer system of any of claims 17 to 23, wherein a Software as a Service (SaaS) is configured to perform the operations of the computer system.