

(12) PATENT APPLICATION PUBLICATION

(21) Application No.202611036681 A

(19) INDIA

(22) Date of filing of Application :26/03/2026

(43) Publication Date : 08/05/2026

(54) Title of the invention : A CLOUD-BASED RESOURCE ALLOCATION ENGINE WITH PREDICTIVE SCALING

| | | |
|---|---|---|
| (51) International classification | :G06F 9/50, H04L 12/24, H04L 12/911, H04L 29/08, G06F 9/48 | (71) Name of Applicant : 1)NOIDA INSTITUTE OF ENGINEERING & TECHNOLOGY Address of Applicant :19, Knowledge Park-II, Institutional Area, Greater Noida – 201306, Uttar Pradesh, India. Uttar Pradesh India |
| (31) Priority Document No | :NA | (72) Name of Inventor : 1)Dr. SANGEETA ARORA 2)Dr. DEEPAK UPRETY |
| (32) Priority Date | :NA | |
| (33) Name of priority country | :NA | |
| (86) International Application No | : | |
| Filing Date | :01/01/1900 | |
| (87) International Publication No | : NA | |
| (61) Patent of Addition to Application Number | :NA | |
| Filing Date | :NA | |
| (62) Divisional to Application Number | :NA | |
| Filing Date | :NA | |

(57) Abstract :

A cloud-based resource allocation engine is disclosed for predictive scaling of distributed cloud infrastructure. The engine includes a telemetry ingestion module (101), a workload profiler (102), a predictive scaling engine (103), a policy evaluator (104), an allocation planner (105), an execution controller (106), cloud resources (107), and a feedback analyzer (108). Operational signals are transformed into forecast demand estimates, policy-approved allocation envelopes, and coordinated execution plans for pre-emptive provisioning. Post-execution feedback is used to refine future scaling decisions, thereby improving service continuity, utilization efficiency, and allocation stability across dynamic cloud workloads.

No. of Pages : 24 No. of Claims : 7