PDF 2 of 3 — English + Hindi Bilingual

# Scalability

Every English paragraph followed immediately by its Hindi translation — so you can learn System Design and build your professional English + technical vocabulary at the same time. Includes a glossary at the end.

English पूरी तरह सुरक्षित

पैराग्राफ-दर-पैराग्राफ Hindi

शब्दावली / Glossary

ENGLISH

Scalability is the ability of a system to handle an increasing workload, either by adding more resources or by making the existing ones bigger.

हिंदी

Scalability किसी सिस्टम की वह क्षमता है जिससे वह बढ़ते हुए workload को संभाल सके — या तो अधिक resources जोड़कर, या मौजूदा resources को ही बड़ा (bigger) बनाकर।

ENGLISH

Those are the two directions, and they have names.

हिंदी

यही दो दिशाएँ (directions) हैं, और इनके नाम भी हैं।

## Horizontal Scaling

हिंदी — शीर्षक का अर्थ

Horizontal Scaling यानी अधिक machines जोड़कर सिस्टम को बड़ा करना।

ENGLISH

Horizontal scaling, also called scaling out, means adding more machines or nodes to a system so the workload is distributed evenly across them.

हिंदी

Horizontal Scaling, जिसे scaling out भी कहा जाता है, इसका मतलब है सिस्टम में अधिक machines या nodes जोड़ना ताकि workload उन सभी में evenly (समान रूप से) बँट जाए।

ENGLISH

No single node has to absorb the growth. More requests arrive, more machines share them.

हिंदी

किसी एक node को अकेले ही पूरा growth झेलना नहीं पड़ता। जैसे-जैसे अधिक requests आती हैं, उतनी ही अधिक machines मिलकर उन्हें share कर लेती हैं।

ENGLISH

This suits distributed systems well. Capacity can be added while the system runs, which makes it a cost-effective way to handle workloads that rise and fall, and having many machines is what makes high availability possible in the first place.

हिंदी

यह तरीका distributed systems के लिए बहुत उपयुक्त है। Capacity को सिस्टम के चलते (running) हुए भी जोड़ा जा सकता है, जिससे यह घटते-बढ़ते workloads को संभालने का एक cost-effective तरीका बन जाता है। साथ ही, कई machines का होना ही सबसे पहले high availability को संभव बनाता है।

ENGLISH

Cassandra and MongoDB are good examples. Both are built so you add machines to meet growing demand.

हिंदी

Cassandra और MongoDB इसके अच्छे उदाहरण हैं। दोनों इस तरह बनाए गए हैं कि बढ़ती हुई demand पूरी करने के लिए आप आसानी से machines जोड़ सकें।

## Vertical Scaling

हिंदी — शीर्षक का अर्थ

Vertical Scaling यानी एक ही machine को अधिक शक्तिशाली (bigger) बनाना।

ENGLISH

Vertical scaling, also called scaling up, means increasing the capacity of an individual node by upgrading its hardware: more CPU, more memory, more storage.

हिंदी

Vertical Scaling, जिसे scaling up भी कहा जाता है, इसका मतलब है किसी एक individual node की capacity को उसके hardware को upgrade करके बढ़ाना: अधिक CPU, अधिक memory, अधिक storage।

ENGLISH

The same single machine now handles more work.

हिंदी

अब वही एक machine पहले से अधिक काम संभालती है।

ENGLISH

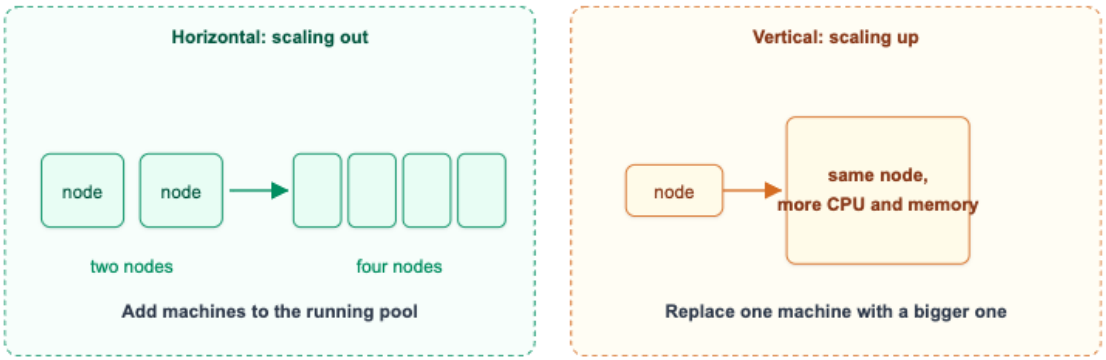
MySQL is a good example. Moving from a smaller machine to a bigger one is a straightforward way to get more capacity, and the process usually involves downtime.

**हिंदी**

MySQL इसका एक अच्छा उदाहरण है। एक छोटी machine से बड़ी machine पर जाना अधिक capacity पाने का एक straightforward (सीधा) तरीका है, और इस प्रक्रिया में आमतौर पर downtime शामिल होता है।

The same single machine now handles more work.

MySQL is a good example. Moving from a smaller machine to a bigger one is a straightforward way to get more capacity, and the process usually involves downtime.




Scaling up hits a physical ceiling, usually needs downtime, and leaves a single point of failure

*Scaling out adds more machines beside the existing ones, while scaling up replaces one machine with a bigger one*

**हिंदी — इमेज कैप्शन**

Scaling out मौजूदा machines के साथ और अधिक machines जोड़ता है, जबकि scaling up एक machine को एक बड़ी machine से बदल (replace) देता है।

## The Difference That Decides It

**हिंदी — शीर्षक का अर्थ**

वह अंतर जो असली फ़ैसला तय करता है।

ENGLISH

Horizontal scaling is easier to do dynamically. You add machines to the pool that is already running.

**हिंदी**

Horizontal scaling को dynamically (गतिशील रूप से) करना आसान होता है। आप उस pool में machines जोड़ते हैं जो पहले से ही running है।

ENGLISH

Vertical scaling is bounded by a single server's capacity. Growing past that ceiling usually means downtime, and there is an upper limit you eventually hit no matter what you spend.

**हिंदी**

Vertical scaling एक ही server की capacity तक bounded (सीमित) होता है। उस ceiling को पार करने के लिए आमतौर पर downtime लगता है, और आप चाहे कितना भी खर्च कर लें, आख़िरकार एक upper limit ज़रूर आ जाती है।

| | Horizontal (scaling out) | Vertical (scaling up) |
|---|---|---|
| **What changes** | Number of machines | Size of one machine |
| **Ceiling** | Practically far away | The biggest machine available |
| **Adding capacity** | Add nodes to the running pool | Usually needs downtime |
| **Failure of one node** | The others carry on | Everything is on that node |
| **Examples given** | Cassandra, MongoDB | MySQL |

हिंदी अनुवाद — तालिका (Table Translation)

| | Horizontal (scaling out) | Vertical (scaling up) |
|---|---|---|
| **क्या बदलता है (What changes)** | Machines की संख्या | एक machine का आकार (size) |
| **सीमा (Ceiling)** | व्यावहारिक रूप से बहुत दूर | सबसे बड़ी उपलब्ध machine |
| **Capacity जोड़ना** | चल रहे pool में nodes जोड़ना | आमतौर पर downtime चाहिए |
| **एक node का fail होना** | बाकी nodes काम करते रहते हैं | सब कुछ उसी node पर निर्भर होता है |
| **दिए गए उदाहरण** | Cassandra, MongoDB | MySQL |

ENGLISH

Read the last two rows together, because that is where the risk lives. Relying only on vertical scaling has two problems at once: there is a physical ceiling on how big one machine can get, and putting everything on that one machine creates a single point of failure. The system gets faster right up until the moment it is completely down.

हिंदी

आख़िरी दो rows को साथ में पढ़ें, क्योंकि असली risk वहीं छिपा है। सिर्फ vertical scaling पर relying (निर्भर) रहने में एक साथ दो समस्याएँ होती हैं: एक तो यह कि एक machine कितनी बड़ी हो सकती है इसकी एक physical ceiling होती है, और दूसरा यह कि सब कुछ उसी एक machine पर रखने से वह एक single point of failure बन जाती है। सिस्टम तब तक तेज़ होता रहता है, जब तक कि वह अचानक पूरी तरह से down नहीं हो जाता।

### Interview Tip

ENGLISH

In an interview, do not just say "I would scale horizontally." Say what makes it possible. Horizontal scaling only works if requests can go to any node, which means saying where the state lives. "The service is stateless, so I can put any number of instances behind a load balancer, and session state goes in a shared cache" is the actual design.

हिंदी

Interview में सिर्फ यह मत कहिए कि "मैं horizontally scale करूँगा।" यह भी बताइए कि इसे संभव किस चीज़ ने बनाया। Horizontal scaling तभी काम करता है जब requests किसी भी node पर जा सकें — यानी आपको यह बताना होगा कि state कहाँ रहता है। असली design यह है: "यह service stateless है, इसलिए मैं किसी भी संख्या में instances को एक load balancer के पीछे लगा सकता हूँ, और session state एक shared cache में रहता है।"

### Key takeaway

ENGLISH

Scalability is handling more work by adding resources. Horizontal scaling adds machines and spreads the workload across them, which is the approach distributed systems are built for. Vertical scaling makes one machine bigger, which is simple but limited by that machine's ceiling, usually requires downtime, and concentrates everything into a single point of failure.

हिंदी

Scalability का मतलब है अधिक resources जोड़कर ज़्यादा काम संभालना। Horizontal scaling machines जोड़ती है और workload को उनमें बाँट देती है, और यही वह तरीका है जिसके लिए distributed systems बनाए जाते हैं। Vertical scaling एक machine को बड़ा बनाती है, जो सरल तो है लेकिन उस machine की ceiling तक सीमित है, आमतौर पर downtime माँगती है, और सब कुछ एक ही single point of failure में समेट देती है।

ENGLISH

The next lesson, Availability, covers what it takes to keep a system reachable once it is spread across many machines.

हिंदी

अगला lesson, Availability, यह बताता है कि एक बार सिस्टम कई machines में फैल जाने के बाद उसे reachable बनाए रखने के लिए क्या ज़रूरी होता है।

## General + Technical Words Glossary

Important English words used in this lesson, with Hindi meaning and a simple explanation — to build both your professional English and System Design vocabulary.

## General English Words

| English Word / Term | Hindi Meaning | Simple Explanation |
|---|---|---|
| **increasing** | बढ़ता हुआ | Going up or growing over time. |
| **workload** | कार्यभार | The amount of work a system has to process. |
| **resources** | संसाधन | Things like machines, CPU, or memory used to do work. |
| **existing** | मौजूदा | Something that is already there. |
| **absorb** | झेलना / सोखना | To take in and handle extra load without breaking. |
| **growth** | वृद्धि | An increase in size, amount, or demand. |
| **arrive** | आना / पहुँचना | To reach or come in, e.g. requests reaching a server. |
| **distribute evenly** | समान रूप से बाँटना | To divide something fairly among all parts. |
| **rise and fall** | घटना-बढ़ना | To go up and down over time, like traffic. |
| **cost-effective** | किफ़ायती | Giving good results without spending too much money. |
| **straightforward** | सीधा / सरल | Simple and easy to understand or do. |
| **involve** | शामिल होना | To include something as a necessary part. |
| **downtime** | डाउनटाइम | Time during which a system is unavailable. |
| **dynamically** | गतिशील रूप से | While the system is running, without stopping it. |
| **bounded** | सीमित | Limited by an upper limit. |
| **ceiling** | ऊपरी सीमा | The maximum limit something can reach. |
| **upper limit** | ऊपरी सीमा | The highest point something can reach. |
| **relying (on)** | निर्भर रहना | Depending on something completely. |
| **concentrate** | केंद्रित करना | To put everything in one single place. |
| **reachable** | पहुँच योग्य | Able to be accessed or contacted. |
| **spread (across)** | फैलाना | To distribute across multiple places. |
| **meet demand** | माँग पूरी करना | To satisfy the amount of work or requests needed. |

## Technical Words

| English Word / Term | Hindi Meaning | Simple Explanation |
|---|---|---|
| **Scalability** | स्केलेबिलिटी | A system's ability to handle more work by adding resources. |
| **Horizontal Scaling** | क्षैतिज स्केलिंग | Adding more machines/nodes to share the workload. |
| **Vertical Scaling** | लंबवत स्केलिंग | Making one existing machine bigger/more powerful. |
| **Scaling out** | स्केलिंग आउट | Another name for horizontal scaling. |
| **Scaling up** | स्केलिंग अप | Another name for vertical scaling. |
| **Node** | नोड | A single machine or server in a system. |
| **Distributed System** | वितरित प्रणाली | A system made of multiple machines working together as one. |
| **High Availability** | उच्च उपलब्धता | Keeping a system usable and reachable almost all the time. |
| **Single Point of Failure** | एकल विफलता बिंदु | A part of the system whose failure brings the whole system down. |
| **Load Balancer** | लोड बैलेंसर | A component that spreads incoming requests across multiple servers. |
| **Stateless** | स्टेटलेस | A service that keeps no local session data, so any instance can serve any request. |
| **Session State** | सेशन स्टेट | Data about a specific user's ongoing interaction with a system. |
| **Shared Cache** | शेयर्ड कैश | A cache accessible by all instances of a service, e.g. Redis. |
| **Instance** | इंस्टेंस | One running copy of a service or application. |
| **CPU / Memory / Storage** | सीपीयू / मेमोरी / स्टोरेज | The core hardware resources of a machine. |
| **Cassandra** | कैसांड्रा | A NoSQL database designed to scale horizontally. |
| **MongoDB** | मोंगोडीबी | A NoSQL document database designed to scale horizontally. |
| **MySQL** | माईएसक्यूएल | A relational database traditionally scaled vertically. |