System Design Interview Questions - Comprehensive Collection

This document contains all system design interview questions organized by category and complexity level.

Table of Contents

Distributed Systems & Infrastructure

Large-Scale Web Services

1. **Design a URL Shortener (like bit.ly)**
   - Handle 100M URLs shortened per day
   - Support custom aliases
   - Provide analytics and click tracking
   - Ensure high availability and low latency

2. **Design a Content Delivery Network (CDN)**
   - Global distribution of static content
   - Cache invalidation strategies
   - Load balancing and failover
   - Performance monitoring and optimization

3. **Design a Load Balancer**
   - Support multiple load balancing algorithms
   - Health checking and failover
   - SSL termination
   - Rate limiting and DDoS protection

4. **Design a Distributed Cache System**
   - Support for Redis/Memcached protocols
   - Consistent hashing for data distribution
   - Replication and failover
   - Memory management and eviction policies

Monitoring & Observability

1. **Design a Metrics and Monitoring System**
   - Collect metrics from thousands of services
   - Real-time alerting and dashboards
   - Time-series data storage and querying
   - Anomaly detection and forecasting

2. **Design a Distributed Logging System**
   - Collect logs from microservices
   - Real-time log processing and search
   - Log aggregation and retention policies
   - Integration with monitoring and alerting

3. **Design a Service Discovery System**
   - Dynamic service registration and discovery
   - Health checking and load balancing
   - Configuration management
   - Multi-datacenter support

Infrastructure & Platform

1. **Design a Container Orchestration Platform**
   - Container scheduling and resource management
   - Service mesh and networking
   - Auto-scaling and load balancing
   - Security and multi-tenancy

2. **Design a Distributed File System**
   - Petabyte-scale storage
   - Replication and consistency
   - Metadata management
   - Performance optimization

3. **Design a Database Sharding System**

   - Automatic data partitioning
   - Cross-shard queries and transactions
   - Rebalancing and migration
   - Consistency and availability guarantees

Cloud & Serverless

1. **Design a Serverless Computing Platform**

   - Function execution and scaling
   - Event-driven architecture
   - Resource isolation and security
   - Cold start optimization

2. **Design a Multi-Cloud Management Platform**

   - Resource provisioning across clouds
   - Cost optimization and governance
   - Security and compliance
   - Disaster recovery and migration

3. **Design an API Gateway**

   - Request routing and transformation
   - Authentication and authorization
   - Rate limiting and throttling

- Analytics and monitoring

Storage & Databases

1. **Design a Distributed Database**

   - ACID transactions across nodes
   - Consensus and replication
   - Query optimization and indexing
   - Backup and recovery

2. **Design a Time-Series Database**

   - High-throughput ingestion
   - Efficient compression and storage
   - Real-time querying and aggregation
   - Retention policies and downsampling

Security & Compliance

1. **Design a Distributed Authentication System**

   - Single sign-on (SSO) across services
   - Multi-factor authentication
   - Session management and security
   - Integration with identity providers

2. **Design a Secrets Management System**

   - Secure storage and rotation
   - Access control and auditing
   - Integration with applications
   - Compliance and governance

Performance & Optimization

1. **Design a Global DNS System**

   - Anycast routing and load balancing
   - DNS caching and TTL management
   - DDoS protection and security
   - Performance monitoring

2. **Design a Distributed Rate Limiting System**

   - Global rate limiting across services
   - Different limiting algorithms
   - Real-time enforcement
   - Analytics and monitoring

3. **Design a Circuit Breaker System**

   - Failure detection and recovery
   - Configurable thresholds and policies
   - Monitoring and alerting
   - Integration with service mesh

Advanced Infrastructure

1. **Design a Multi-Region Disaster Recovery System**

   - Cross-region replication
   - Automated failover and recovery
   - Data consistency and integrity
   - RTO/RPO optimization

2. **Design a Distributed Configuration Management System**

   - Real-time configuration updates
   - Version control and rollback
   - Environment-specific configs
   - Security and access control

3. **Design a Service Mesh Architecture**

   - Traffic management and routing
   - Security and encryption
   - Observability and monitoring
   - Policy enforcement

4. **Design a Distributed Tracing System**

   - Request tracing across microservices
   - Performance analysis and debugging
   - Sampling and data retention
   - Integration with monitoring

5. **Design a Global Load Balancing System**

   - Traffic distribution across regions
   - Health checking and failover
   - Performance optimization
   - DDoS protection

Edge Computing

1. **Design an Edge Computing Platform**

   - Distributed compute at edge locations
   - Data synchronization and consistency
   - Low-latency processing
   - Resource management

2. **Design a Global Content Optimization System**

   - Dynamic content optimization
   - Image and video processing
   - Caching strategies
   - Performance monitoring

3. **Design a Distributed IoT Platform**

   - Device management and provisioning

- Data ingestion and processing
- Real-time analytics
- Security and compliance

4. **Design a Multi-Tenant SaaS Platform**

- Tenant isolation and security
- Resource sharing and optimization
- Billing and metering
- Customization and configuration

5. **Design a Distributed Backup and Recovery System**

- Incremental and differential backups
- Cross-region replication
- Point-in-time recovery
- Compliance and retention

Data & AI/ML Systems

Data Processing & Analytics

1. **Design a Real-time Analytics Platform**
   - Stream processing for millions of events/second
   - Real-time dashboards and alerting
   - Complex event processing
   - Integration with batch processing

2. **Design a Data Lake Architecture**
   - Petabyte-scale data storage
   - Schema evolution and governance
   - Data quality and lineage
   - Multi-format support (structured, semi-structured, unstructured)

3. **Design a Data Warehouse System**
   - OLAP queries and aggregations
   - ETL/ELT pipelines
   - Data modeling and optimization
   - Business intelligence integration

4. **Design a Distributed ETL System**
   - Data extraction from multiple sources
   - Transformation and validation
   - Error handling and recovery
   - Scheduling and monitoring

Machine Learning Infrastructure

1. **Design an ML Training Platform**
   - Distributed training across GPUs/TPUs
   - Experiment tracking and versioning

- Hyperparameter optimization
- Resource scheduling and management

2. **Design an ML Model Serving System**
   - Real-time and batch inference
   - Model versioning and A/B testing
   - Auto-scaling and load balancing
   - Monitoring and performance optimization

3. **Design a Feature Store**
   - Feature computation and storage
   - Real-time and batch feature serving
   - Feature versioning and lineage
   - Data quality and monitoring

4. **Design an ML Pipeline Orchestration System**
   - Workflow definition and execution
   - Dependency management
   - Error handling and retries
   - Monitoring and alerting

Search & Recommendation

1. **Design a Search Engine**
   - Web crawling and indexing
   - Query processing and ranking
   - Real-time updates and freshness
   - Personalization and relevance

2. **Design a Recommendation System**

   - Collaborative and content-based filtering
   - Real-time recommendations
   - Cold start problem handling
   - A/B testing and evaluation

3. **Design an Elasticsearch-like System**

   - Distributed indexing and search
   - Real-time data ingestion
   - Query optimization and caching
   - Cluster management and scaling

Data Streaming & Processing

1. **Design a Stream Processing System**

   - Real-time data processing
   - Windowing and aggregations
   - Fault tolerance and exactly-once processing
   - Backpressure handling

2. **Design a Message Queue System**

   - High-throughput message delivery

- Ordering and durability guarantees
- Consumer groups and scaling
- Dead letter queues and error handling

3. **Design a Data Pipeline Monitoring System**

- Data quality monitoring
- Pipeline performance tracking
- Anomaly detection and alerting
- Lineage and impact analysis

AI/ML Applications

1. **Design a Computer Vision Platform**

- Image and video processing
- Model training and inference
- Real-time object detection
- Edge deployment and optimization

2. **Design a Natural Language Processing System**

- Text processing and analysis
- Language model training and serving
- Real-time inference and scaling
- Multi-language support

3. **Design a Fraud Detection System**

- Real-time transaction analysis
- Machine learning model deployment
- Rule engine and decision making
- Feedback loop and model updates

4. **Design a Personalization Engine**

- User behavior tracking
- Real-time personalization
- Content optimization
- Privacy and compliance

Big Data Storage

1. **Design a Distributed Graph Database**

- Graph storage and querying
- Distributed graph algorithms
- Real-time updates and consistency
- Visualization and analytics

2. **Design a Multi-Model Database**

- Support for multiple data models
- Query optimization across models
- Consistency and transactions

  - Performance and scaling

3. **Design a Data Catalog System**

  - Metadata management and discovery
  - Data lineage and governance
  - Search and recommendation
  - Integration with data tools

Advanced Analytics

1. **Design a Real-time Anomaly Detection System**

  - Streaming data analysis
  - Machine learning model deployment
  - Alert generation and routing
  - False positive reduction

2. **Design a Predictive Analytics Platform**

  - Time series forecasting
  - Model training and validation
  - Real-time predictions
  - Business intelligence integration

3. **Design a Data Science Collaboration Platform**

  - Notebook sharing and collaboration
  - Experiment tracking and versioning
  - Resource management and scheduling
  - Model deployment and monitoring

4. **Design a Customer Data Platform (CDP)**

  - Customer data unification
  - Real-time profile updates
  - Segmentation and targeting
  - Privacy and compliance

Specialized Systems

1. **Design a Geospatial Data System**

  - Location data storage and indexing
  - Spatial queries and analysis
  - Real-time location tracking
  - Map rendering and visualization

2. **Design a Time Series Forecasting System**

  - Historical data analysis
  - Model training and validation
  - Real-time forecasting
  - Accuracy monitoring and improvement

3. **Design a Data Governance Platform**

   - Data classification and tagging
   - Access control and auditing
   - Compliance monitoring
   - Policy enforcement

4. **Design a Multi-Cloud Data Platform**

   - Cross-cloud data integration
   - Cost optimization and governance
   - Security and compliance
   - Performance optimization

5. **Design a Real-time Decision Engine**

   - Rule-based decision making
   - Machine learning integration
   - A/B testing and optimization
   - Performance and latency optimization

Real-time & Communication Systems

Messaging & Communication

1. **Design WhatsApp/Telegram**
   - Real-time messaging for billions of users
   - End-to-end encryption
   - Media sharing and storage
   - Group chats and broadcast lists
   - Offline message delivery

2. **Design Slack/Discord**
   - Team communication and collaboration
   - Channels and direct messaging
   - File sharing and integrations
   - Voice and video calling
   - Search and message history

3. **Design a Video Conferencing System (Zoom)**
   - Real-time video and audio streaming
   - Screen sharing and recording
   - Scalable infrastructure for millions of concurrent users
   - Quality adaptation and optimization
   - Security and privacy features

4. **Design a Live Streaming Platform (Twitch)**
   - Real-time video streaming
   - Chat and interaction features
   - Content delivery and optimization

- Monetization and analytics
- Global distribution

Social Networks & Content

1. **Design Twitter/X**
   - Real-time tweet delivery
   - Timeline generation and ranking
   - Trending topics and hashtags
   - Media handling and storage
   - Global distribution and caching

2. **Design Instagram**
   - Photo and video sharing
   - Real-time feed generation
   - Stories and live streaming
   - Discovery and recommendation
   - Global content delivery

3. **Design TikTok**
   - Short-form video platform
   - AI-powered recommendation engine
   - Real-time video processing
   - Global content distribution
   - Creator tools and analytics

4. **Design LinkedIn**
   - Professional networking platform
   - News feed and content sharing
   - Job recommendations and search
   - Messaging and connections
   - Professional insights and analytics

Real-time Systems

1. **Design a Real-time Notification System**
   - Push notifications across platforms
   - Email and SMS delivery
   - Personalization and targeting
   - Delivery tracking and analytics
   - Rate limiting and optimization

2. **Design a Real-time Collaboration Tool (Google Docs)**

   - Collaborative document editing
   - Operational transformation
   - Conflict resolution
   - Version history and recovery
   - Real-time synchronization

3. **Design a Live Chat Support System**

   - Real-time customer support
   - Agent routing and load balancing

- Chat history and analytics
- Integration with CRM systems
- Automated responses and bots

4. **Design a Real-time Gaming Platform**

- Low-latency multiplayer gaming
- Matchmaking and lobbies
- Anti-cheat and security
- Leaderboards and statistics
- Global server distribution

Streaming & Media

1. **Design a Music Streaming Service (Spotify)**

- Audio streaming and caching
- Playlist and recommendation engine
- Offline playback and sync
- Social features and sharing
- Artist analytics and royalties

2. **Design a Video Streaming Service (Netflix)**

- Video encoding and adaptive streaming
- Content recommendation engine
- Global content delivery
- Offline downloads and sync
- Analytics and personalization

3. **Design a Podcast Platform**

- Audio content distribution
- Subscription and discovery
- Analytics and monetization
- Offline playback and sync
- Creator tools and insights

Real-time Analytics

1. **Design a Real-time Dashboard System**

- Live data visualization
- Real-time metrics and KPIs
- Interactive charts and graphs
- Alert generation and notifications
- Performance optimization

2. **Design a Real-time Bidding System (Ad Tech)**

- Millisecond auction processing
- Bid optimization and targeting
- Fraud detection and prevention
- Analytics and reporting

- Global distribution

3. **Design a Real-time Fraud Detection System**

   - Transaction monitoring and analysis
   - Machine learning model deployment
   - Real-time decision making
   - Alert generation and response
   - Feedback loop and learning

Communication Infrastructure

1. **Design a WebRTC Signaling Server**

   - Peer-to-peer connection establishment
   - NAT traversal and STUN/TURN
   - Signaling and media relay
   - Quality monitoring and optimization
   - Security and encryption

2. **Design a Push Notification Service**

   - Cross-platform notification delivery
   - Device registration and management
   - Message queuing and delivery
   - Analytics and tracking
   - Rate limiting and optimization

3. **Design a Real-time Location Tracking System**

   - GPS data collection and processing
   - Real-time location updates
   - Geofencing and alerts
   - Privacy and security
   - Battery optimization

Event-Driven Systems

1. **Design an Event Streaming Platform**

   - High-throughput event ingestion
   - Real-time event processing
   - Event sourcing and replay
   - Schema evolution and compatibility
   - Monitoring and observability

2. **Design a Real-time Auction System**

   - Bid processing and validation
   - Real-time price updates
   - Fraud detection and prevention
   - Settlement and payment
   - Analytics and reporting

3. **Design a Real-time Sports Scoring System**

   - Live score updates and statistics
   - Real-time data ingestion
   - Fan engagement features
   - Global distribution and caching
   - Analytics and insights

Collaborative Platforms

1. **Design a Real-time Code Collaboration Platform**

   - Collaborative code editing
   - Version control integration
   - Real-time synchronization
   - Conflict resolution
   - Code review and comments

2. **Design a Virtual Whiteboard System**

   - Real-time collaborative drawing
   - Multi-user synchronization
   - Vector graphics and optimization
   - Version history and recovery
   - Integration with video conferencing

3. **Design a Real-time Project Management Tool**

   - Task tracking and updates
   - Team collaboration features
   - Real-time notifications
   - Progress tracking and reporting
   - Integration with other tools

Advanced Real-time Systems

1. **Design a Real-time IoT Data Processing System**

   - Sensor data ingestion
   - Real-time analytics and alerting
   - Device management and control
   - Edge computing integration
   - Security and compliance

2. **Design a Real-time Financial Trading System**

   - Low-latency order processing
   - Market data distribution
   - Risk management and compliance
   - Settlement and clearing
   - Analytics and reporting

3. **Design a Real-time Emergency Response System**

   - Incident detection and alerting
   - Resource coordination and dispatch
   - Real-time communication
   - Location tracking and mapping
   - Analytics and optimization

Product & Platform Systems

E-commerce & Marketplace

1. **Design Amazon/E-commerce Platform**
   - Product catalog and search
   - Shopping cart and checkout
   - Payment processing and fraud detection
   - Order management and fulfillment
   - Recommendation engine and personalization

2. **Design Uber/Lyft**
   - Real-time ride matching
   - Dynamic pricing and surge
   - GPS tracking and navigation
   - Payment processing
   - Driver and rider management

3. **Design Airbnb**
   - Property listing and search
   - Booking and reservation system
   - Payment processing and escrow
   - Review and rating system
   - Host and guest communication

4. **Design eBay/Auction Platform**
   - Auction mechanics and bidding
   - Product listing and categorization
   - Payment and escrow services
   - Fraud detection and prevention
   - Seller and buyer protection

Food & Delivery

1. **Design DoorDash/Food Delivery**
   - Restaurant and menu management
   - Real-time order tracking
   - Delivery optimization and routing
   - Payment processing
   - Driver dispatch and management

2. **Design Instacart/Grocery Delivery**
   - Inventory management and availability

- Personal shopper assignment
- Real-time order updates
- Substitution and communication
- Delivery scheduling and optimization

Financial Services

1. **Design Venmo/PayPal**
   - Peer-to-peer payments
   - Transaction processing and settlement
   - Fraud detection and security
   - Social features and feed
   - Integration with banks and cards

2. **Design a Digital Banking Platform**
   - Account management and transactions
   - Mobile banking and security
   - Loan and credit services
   - Investment and wealth management
   - Compliance and regulatory requirements

3. **Design a Cryptocurrency Exchange**
   - Order book and matching engine
   - Wallet management and security
   - Trading pairs and liquidity
   - KYC/AML compliance
   - Market data and analytics

Content & Media

1. **Design YouTube**

   - Video upload and processing
   - Content delivery and streaming
   - Recommendation and discovery
   - Monetization and analytics
   - Creator tools and community

2. **Design Medium/Blogging Platform**

   - Content creation and publishing
   - Reader engagement and discovery
   - Subscription and monetization
   - Social features and following
   - Analytics and insights

3. **Design Pinterest**

   - Visual content discovery
   - Board and pin management
   - Search and recommendation
   - Shopping and commerce integration
   - Creator tools and analytics

Travel & Hospitality

1. **Design Booking.com/Travel Platform**

   - Hotel and accommodation search
   - Availability and pricing
   - Booking and reservation management
   - Review and rating system
   - Payment processing and cancellation

2. **Design Expedia/Flight Booking**

   - Flight search and comparison
   - Booking and ticketing
   - Price tracking and alerts
   - Itinerary management
   - Integration with airlines and GDS

Professional Services

1. **Design GitHub**

   - Git repository hosting
   - Collaboration and code review
   - Issue tracking and project management
   - CI/CD integration
   - Package and release management

2. **Design Jira/Project Management**

   - Issue and task tracking
   - Workflow and automation
   - Team collaboration and reporting
   - Integration with development tools
   - Analytics and insights

3. **Design Salesforce/CRM Platform**

   - Customer relationship management
   - Sales pipeline and forecasting
   - Marketing automation
   - Custom applications and workflows
   - Integration and API platform

Education & Learning

1. **Design Coursera/Online Learning**

   - Course creation and delivery
   - Video streaming and progress tracking
   - Assessment and certification
   - Student engagement and community
   - Payment and subscription management

2. **Design Khan Academy/Educational Platform**

   - Interactive learning content
   - Progress tracking and analytics
   - Personalized learning paths
   - Teacher and parent dashboards
   - Accessibility and localization

Health & Fitness

1. **Design Teladoc/Telehealth Platform**

   - Doctor-patient video consultations
   - Appointment scheduling and management
   - Medical records and history
   - Prescription and pharmacy integration
   - Insurance and billing

2. **Design MyFitnessPal/Health Tracking**

   - Food and exercise logging
   - Calorie and nutrition tracking
   - Social features and challenges
   - Integration with wearables
   - Goal setting and progress tracking

Gaming & Entertainment

1. **Design Steam/Gaming Platform**

   - Game distribution and downloads
   - User library and achievements
   - Social features and community
   - Workshop and user-generated content
   - Payment and DRM

2. **Design Twitch (Platform Features)**

   - Streamer tools and analytics
   - Monetization and subscriptions
   - Community and moderation
   - Clips and highlights
   - Integration with games

Productivity & Tools

1. **Design Notion/Productivity Platform**

   - Document creation and collaboration
   - Database and knowledge management
   - Templates and customization
   - Team workspaces and permissions
   - Integration with other tools

2. **Design Zoom (Platform Features)**

   - Meeting scheduling and management
   - Recording and transcription
   - Webinar and event hosting
   - Integration with calendar systems
   - Analytics and reporting

Specialized Platforms

1. **Design Reddit**

   - Community creation and management
   - Content submission and voting
   - Comment threading and discussion
   - Moderation tools and policies
   - Advertising and monetization

2. **Design Stack Overflow**

   - Question and answer platform
   - Reputation and gamification
   - Search and discovery
   - Community moderation
   - Knowledge base and documentation

3. **Design Yelp/Review Platform**

   - Business listing and information
   - Review and rating system
   - Search and discovery
   - Business owner tools
   - Local recommendations

4. **Design Dropbox/File Storage**

   - File synchronization and sharing
   - Version control and history
   - Collaboration and permissions
   - Mobile and desktop clients
   - Business and enterprise features

5. **Design Shopify/E-commerce Platform**

   - Store creation and customization
   - Product and inventory management
   - Payment processing and checkout
   - App ecosystem and integrations
   - Analytics and reporting

System Design Interview Tips

Approach and Methodology

1. **Clarify Requirements**
   - Ask clarifying questions about scale, features, and constraints
   - Define functional and non-functional requirements
   - Understand the scope and priorities

2. **Estimate Scale**
   - Calculate read/write ratios
   - Estimate storage requirements
   - Determine bandwidth and QPS needs
   - Consider peak load scenarios

3. **High-Level Design**
   - Start with a simple architecture
   - Identify major components and services
   - Show data flow and interactions
   - Consider API design

4. **Detailed Design**
   - Deep dive into critical components
   - Discuss data models and schemas
   - Address scalability and performance
   - Consider edge cases and failure scenarios

5. **Scale and Optimize**
   - Identify bottlenecks and solutions
   - Discuss caching strategies
   - Consider database scaling (sharding, replication)
   - Address monitoring and observability

Key Concepts to Master

- **Scalability**: Horizontal vs vertical scaling, load balancing
- **Reliability**: Fault tolerance, redundancy, disaster recovery
- **Consistency**: ACID properties, CAP theorem, eventual consistency
- **Performance**: Latency, throughput, caching, CDNs
- **Security**: Authentication, authorization, encryption, compliance

Common Patterns

- **Microservices Architecture**: Service decomposition, API gateways
- **Event-Driven Architecture**: Message queues, event sourcing
- **CQRS**: Command Query Responsibility Segregation
- **Circuit Breaker**: Fault tolerance and resilience
- **Bulkhead**: Isolation and resource management

This document contains comprehensive system design questions covering distributed systems, data platforms, real-time systems, and product platforms. Practice these questions by working through the complete design process from requirements to implementation details.