Design a News Feed ML Ranking System - Meta

9 billion monthly active users, 1. 8 billion daily active users Volume : 4. 75 billion posts per day, 100+ billion content interactions daily Latency : Sub-200ms p95 response time for feed generation Throughput : 10,000+ requests per second during peak hours Key ML Challenges and Technical Requirements Real-time ranking : Process millions of posts in milliseconds Personalization : Individual user preferences and behavior modeling Content diversity : Balance relevance with content variety Cold start : Handle new users and new content Multi-objective optimization : Engagement, diversity, and safety Interview Discussion Flow and Technical Deep-Dives 1. Problem formulation (10 minutes) 2. Data strategy and feature engineering (15 minutes) 3. Model architecture and training (20 minutes) 4. Serving and optimization (10 minutes) 5. Advanced topics and edge cases (5 minutes) 2. Mathematical Formulation : Score(u , p) = f(θ, x_u, x_p, x_c, x_t) Where: u : User features p : Post features c : Context features t : Temporal features θ : Model parameters ML-Specific Success Metrics and Evaluation Criteria Primary : Time spent on platform, daily active users Secondary : Click-through rate, like rate, share rate, comment rate Diversity : Content type distribution, source diversity Safety : Harmful content detection rate Scale Requirements and Technical Constraints Users : 2. 9B monthly, 1. 8B daily Content : 4. Feature retrieval : Real-time feature lookup 2. Preprocessing : Feature normalization and transformation 3. Inference : Model prediction 4. Performance : Simple models : Fast inference, easy to interpret Complex models : Better accuracy, slower inference Trade-off : Balance between accuracy and efficiency Online vs. Batch Learning : Online learning : Real-time updates, adaptive to changes Batch learning : Stable training, better convergence Trade-off : Adaptability vs. stability Accuracy vs. Diversity : High accuracy : Relevant content, user satisfaction High diversity : Content variety, user exploration Trade-off : Relevance vs. exploration Advanced Follow-up Questions and Technical Challenges Algorithm Questions : "How would you handle the cold start problem for new users. " "What's the computational complexity of your ranking algorithm. " "How would you implement online learning for your model. " System Questions : "How would you handle model serving at scale. " "What's your strategy for A/B testing ML models. " "How would you detect and handle model drift. " Research Questions : "How would you incorporate recent research in transformers. " "What's your approach to multi-modal learning. " "How would you implement causal inference in ranking. "Design a multi-armed bandit algorithm for content recommendation" 2. "How would you implement Thompson sampling for exploration-exploitation. " 3. "What's the difference between collaborative filtering and content-based filtering. " 4. "How would you handle the curse of dimensionality in feature engineering. " Mathematical Questions : 1. "Derive the gradient of a neural network with respect to its parameters" 2. "Explain the mathematical foundation of attention mechanisms" 3. "How would you prove the convergence of stochastic gradient descent. " 4. "What's the relationship between regularization and bias-variance trade-off. " System Design Questions : 1. "Design a distributed training system for large-scale neural networks" 2. "How would you implement model versioning and rollback. " 3. "Design a real-time feature store for ML systems" 4. "How would you handle model serving with zero downtime. " Complex System Design Follow-ups Scalability Questions : 1. "How would you scale your system to handle 10x more users. " 2. "What's your strategy for handling peak traffic. " 3. "How would you optimize for different geographic regions. " Performance Questions : 1. "How would you reduce inference latency by 50%. " 2. "What's your strategy for memory optimization. " 3. "How would you handle model serving on mobile devices. " Reliability Questions : 1. "How would you ensure 99. 9% uptime for your ML system. " 2.

"What's your disaster recovery strategy. " 3. "How would you handle data corruption in production. Technical Depth : Demonstrate deep understanding of ML algorithms, mathematical foundations, and system design 2. Problem Solving : Show ability to break down complex problems and design elegant solutions 3. Trade-offs : Understand and articulate various trade-offs in ML system design 4. Research Integration : Stay current with latest research and apply it to real-world problems 5. Scalability : Design systems that can handle massive scale and real-time requirements 6. Evaluation : Implement rigorous evaluation and monitoring strategies 7. Innovation : Propose novel approaches and improvements to existing systems