SAMPLE

Building the Edge Computing of Tomorrow

Designing Where the Work Runs When the Cloud Is Too Far

Aiman JoeHanz

Free sample: Introduction & Chapter 1

putrabooks.github.io

Buy the full book:

Carousell carousell.sg/p/building-the-edge-computing-of-tomorrow-ebook-pdf-1471312610

Ko-fi

Payhip

Contents

Introduction: The Placement Discipline

Chapter 1: The Millisecond Mandate

Chapter 2: The Nearby Computer

Chapter 3: The Fog's Constitution

Chapter 4: The Offloading Decision

Chapter 5: The Carrier's Edge

Chapter 6: Edge Intelligence

Chapter 7: The Estate's Edge

Chapter 8: The Working Edge

Chapter 9: The Edge's Failures

Chapter 10: The Honest Decade

Introduction: The Placement Discipline

Every technology eventually gets the technology it deserves, and the technology of the decade's networks got a word that means, in its original geography, the border where the land ends and the water begins. Edge computing is the discipline that grew up at that border, and the word's honesty depends on which of its two founders a speaker has in mind, because the edge was founded twice, once by engineers who measured milliseconds and once by marketers who measured enthusiasm, and the two foundations share a word and almost nothing else. This book is written from the first foundation, and its first job is to separate the two so cleanly that the reader can perform the separation on any sentence the decade will produce.

The engineering foundation is easy to state and hard to dispute. Computation that must answer in milliseconds cannot afford to ask its questions of a server an ocean away, because light itself, the fastest courier the universe permits, takes longer to make the round trip than the answer is allowed to take. The physics does not negotiate. A round trip across a continent consumes a budget that a control loop at a crane, or a safety stop at a robot, or a frame at a screen simply does not have, and no improvement in the cloud's speed changes the distance. The only remedy physics allows is to move the computation closer, and the discipline of deciding what moves, where it moves, how it is kept trustworthy at the new location, and what happens when the location fails is edge computing done as engineering rather than as decoration.

The book's definition, offered here in the series' tradition of defining before building, is a contract in three clauses. Edge computing is the placement of computation, storage, and decision within one radio hop, or one otherwise-trustworthy nearby hop, of the data and the decision it serves, chosen against a written latency budget, and operated with a plan for the failures of proximity, which are the failures this book's ninth chapter catalogs and no marketing sentence ever mentions. A deployment that cannot show its latency budget is not an edge, however near its hardware sits. A deployment that cannot show its failure plan is not an edge either, however fast its answers arrive. The contract is the discipline, and the word is only the word.

The reader will notice that the definition says nothing about hardware size, brand of server, or ownership, and the restraint is deliberate, because the discipline's honest taxonomy, worked in Chapter 3 from the standards body that gave the field its constitution, sorts the paradigm family by role rather

than by acreage. A gateway on a factory wall is an edge. A rack in a shipping container beside a port crane is an edge. A platform in a telephone operator's site is an edge, and the mist, the lightest tier of the family, lives on the devices themselves. The family has names for its members, the cloudlet, the carrier's edge, the fog that spans cloud to things, and the book teaches the names not as vocabulary but as load-bearing distinctions, because the industry's worst procurement documents are written by people who believe the names are synonyms.

The series' two working rules hold in this volume as in the others. First, the reader will meet no brand names and no market figures, and every quantity in the arithmetic is an illustrative assumption, declared where it appears and reconciled before the chapter ends. Second, the organizations are invented: a container port automating its cranes, a regional hospital network triaging its imaging, a restaurant chain wiring its kitchens, an offshore wind farm, and a city's traffic authority, each built to demonstrate one structure of the discipline, because structure outlives the corporate gossip that dates a book. The register behind the volume, the standards body's fog model, the founding papers of the nearby computer, the surveys that mapped the paradigm family and its engineering, and the carrier's white paper on the mobile network's integration, are credited by name as their claims surface.

The volume's claim, stated once and defended in chapters: the edge is not a place but a placement decision, and the decision is worth making only when it is made as a contract, with a budget, a trust clause, and a failure plan, because the decade will build more edges than it will operate honestly, and the difference between the two groups will not be hardware. It will be the presence of the contract, and the willingness to fund it, and this book exists to make that difference visible before the reader signs anything.

The book's map, so the reader can hold the whole while reading the parts. Chapter 1 establishes the physics: the latency budget, the five tolls, and the staircase of tiers, with the port's crane as the standing arithmetic. Chapter 2 works the founding lineage, the trusted nearby computer, its discovery problem, and the placement contract that turns proximity into a legal instrument. Chapter 3 reads the standards body's constitution for the fog, and sorts the paradigm family, mist, cloudlet, carrier's edge, and fog, into load-bearing names, closing with the definition charter. Chapter 4 is the discipline's core arithmetic, the offloading decision, with its contract and its partitioning method. Chapter 5 crosses into the carrier's territory, the mobile

network's own edge, and works the machinery of sessions, user planes, and the relocation that mobility forces. Chapter 6 places intelligence at the edge, the compressed models, their grading, their drift, and their rehearsed fallbacks. Chapter 7 scales from the placement to the estate, the register, the seams, the federation, and the shared services that hold many edges coherent. Chapter 8 is the operating shift: the observability thread, the remote hands, the update calendar, and the quiet screen the estate shows when the link is down. Chapter 9 catalogs the failures by their signatures, seven of them, each with its remedy, and closes with the review that grades systems rather than staff. Chapter 10 delivers the verdict decade: what the edge is for, what it will not do, the renewal economics, and the annual governance cycle that keeps the whole honest. Ten chapters, ten stealable pages, and a reader who finishes the book with a pocketful of instruments and no illusions.

A note on how to read the running organizations. The port, the hospital, the chain, the farm, and the city are not five separate case studies but one composite estate, seen from five altitudes, and the same decisions recur in each at different scales: the port's cloudlet is the city's cabinet is the farm's platform, and the restaurant's gateway is the smallest honest edition of the same instrument. The reader whose organization resembles one of the five should read that organization's passages as worked examples of her own contracts, and the reader whose organization resembles none should invent the sixth, because the book's method, define the loop, budget the road, name the custody, calendar the renewal, works on any organization the reader can describe in a sentence, which is the test of a method and the book's quiet promise.

A word on the pages themselves, because the series' stealable pages are an instrument with a use's manual, and the manual is three sentences. Each chapter ends with a page, a compact restatement of the chapter's practice, named after the instrument it is: the latency budget, the placement contract, the definition charter, the offloading contract, the continuity contract, the inference placement, the estate seam, the shift contract, the postmortem discipline, and the governance cycle. The pages are written to be lifted from the book without explanation, and the reader's honest theft is the series' intended use: take the page, adapt its lines, put it in the estate's documents, and cite nothing, because an instrument carried in the pocket is worth ten quoted in the library. The pages' order is the discipline's order, budget, placement, definition, decision, continuity, intelligence, estate, operation, failure, and governance, and the order is not accidental but architectural: each

page assumes the ones before it, and the reader who follows the order will find the tenth page turning the first nine into a routine, which is the whole of engineering governance and the quiet point of the series' tenth books.

And a final word on honesty's price, because the reader deserves the book's terms of trade before the first chapter begins. The honest edge costs more than the decorated one, in measurement campaigns, in rehearsals, in reviews, in the keeper's weekly hour, and in the budget's explicit lines, and the honest edge returns the investment in the failures that do not happen, which no ledger records and every annual review confirms, in the tides caught, the Saturdays survived, the quarters without surprises. The trade is the reader's to make, and the book will not pretend the price is zero; it will claim, chapter by chapter, that the price is the cheaper side of the trade, and the claim's evidence is the ninth chapter's catalog, the estate's failure history priced honestly, and the reader who finishes the catalog and still prefers the decorated edge has read a different book than the one written, and the series wishes her estate the luck that discipline would have bought.

Chapter 1: The Millisecond Mandate

The chapter begins where the discipline begins, with a number the marketing copy never quotes, because the number is not flattering to any product and is instead flattering to physics. The round trip is the unit of the edge's thought. A computation that must decide whether a crane's load has swung too far does not ask a question and wait for an answer in the abstract; it spends a budget, and the budget is measured in milliseconds, and the budget is spent in five tollbooths that no architect can abolish. The first toll is the sensor, which converts a physical event to a number and takes its own time doing so. The second is the wire or the radio between the sensor and the computation, which is where distance imposes its tax. The third is the computation itself, which is the only toll the hardware industry's progress has actually been reducing. The fourth is the return path, the answer traveling back to the actuator, which pays the distance tax again. The fifth is the actuator, which converts the answer to a physical motion and takes its own sweet time about it. The engineer who cannot name the five tolls for her deployment cannot know why her loop is slow, and the industry is full of deployments whose engineers cannot.

The port's crane loop, illustrative throughout the book, makes the tolls concrete. The crane reads its load position twenty times a second, which is every fifty milliseconds, and the control decision must land inside that window with room to spare, so the loop's budget is written as thirty milliseconds end to end. The sensor takes three. The actuator takes five. The computation, on hardware the port can afford, takes eight. The remaining fourteen milliseconds are for the road, both directions, and here physics presents its bill without apology. A signal in fiber or copper travels at roughly two thirds the speed of light, which is about two hundred kilometers per millisecond, and a radio hop is faster to start but pays its own scheduling delays. If the port's computation sits in a regional cloud four hundred kilometers away, the road alone costs eight milliseconds round trip in the best case, before a single switch, queue, or protocol retransmission adds its share, and the budget survives, barely, until the day the traffic on the shared path is heavy, at which point it does not. If the computation sits beside the crane, the road costs less than one millisecond, and the budget has thirteen milliseconds of slack where it had none. The difference is not the cloud's speed, which is exemplary, or the edge's cleverness, which is irrelevant. The difference is distance, and distance is the one term no vendor can improve.

The founding measurement, quoted from the discipline's first paper and

verified in the book's register, lands the same lesson at human scale. The researchers who made the case for the nearby computer measured interactive applications run as thin clients, with all significant computation far away and only pixels traveling, and found that even a modest network latency of thirty-three milliseconds caused the frame rate of the interaction to drop sharply, because the thin-client protocol, striving to keep up, dropped frames, and the smoothness of the interaction degraded in a way the user felt immediately. The finding is half a century's worth of lesson in one number. Perceived responsiveness is governed by the round trip, the round trip is governed by distance and queueing, and queueing is governed by sharing, and sharing is what the wide network is made of. The engineer who moves the computation one hop away removes the sharing from the critical path, which is the whole of the trick, stated plainly.

The chapter's second movement is the taxonomy of the present, because the reader cannot budget a distance she has not mapped. The industry's deployment tiers, named here in the discipline's own vocabulary and worked harder in Chapter 3, form a staircase from the device to the cloud, and each stair trades latency against capability. On the device itself runs the mist, the lightest tier, computation embedded in the sensor or the phone or the camera, with microsecond access to its own data and no capacity to spare. One hop up sits the edge proper, the gateway or the micro data center or the cloudlet, a computer near enough to serve a site or a street corner, with millisecond access and real capacity. Above it sits the regional aggregation point, the carrier's central office or the operator's site, tens of milliseconds from many users, where the network itself keeps computation. Above that the regional cloud, and above that the distant cloud, whose hundred milliseconds and more are a comfortable home for anything that answers in seconds and a catastrophe for anything that answers in frames. The staircase's engineering content is the placement decision, and the placement decision is the act of choosing the lowest stair that still has the capacity, the trust, and the survivability the workload requires, which sounds simple and is the discipline's hardest recurring judgment.

The staircase's honest complication is that workloads are not monoliths. The port's crane loop lives on the bottom stair, the port's berth scheduling lives on the middle stair, and the port's annual maintenance forecasting lives in the cloud without shame, and the three are one system. The naive reading of the edge's decade, the reading that says the edge will replace the cloud, inverts the truth: the edge exists because the cloud succeeded, and the pair is a

working division of labor, not a contest. The hospital's imaging triage will pre-screen scans at the edge to decide which need a radiologist's eye within the hour, and send the flagged ones, with their edge-computed summaries, to the cloud's heavier models and the region's specialists, and both halves are necessary and neither is redundant. The discipline's placement question is never cloud versus edge; it is which loop, at which budget, on which stair, funded by which line, and the engineer who frames it as a war has already mispriced her own project.

The chapter's practice, in the book's tradition of one stealable page, is the latency budget, and the page is a table of five tolls with four columns, written before any hardware is purchased. The first column names the toll: sensor, road out, computation, road back, actuator. The second states the budget in milliseconds, per loop, with the worst case rather than the average in mind. The third states the measured or contractual reality, and is left blank on purpose until measurement fills it, because the budget's discipline is the refusal to guess. The fourth names the evidence, the log line, the contract clause, or the test that will stand behind the number when the budget is audited, because a budget without evidence is a wish with a spreadsheet's formatting. The port's budget for its crane loop, filled in from the illustrative numbers above, reads three, four out and four back written as the road's allowance, eight, five, and thirty total, and the road's allowance is the line that decides the architecture, because four milliseconds each way is a local hop and nothing else.

The page's corollary is the budget's review, and the review is a calendar item, not an opinion. Queues grow as sharing grows, and sharing grows as systems succeed, so the measured road that cost three milliseconds at commissioning will cost five at the second year and seven at the third unless the budget's review catches it, and the review is the difference between an edge that ages honestly and an edge that degrades silently into the thing its buyers were promised it was not. The chapter closes on its mandate: the milliseconds are the discipline's physics, the physics does not negotiate, and every placement decision this book makes from here forward will show its budget first, because an edge that cannot show its budget is a decoration, and the decade will build many.

The chapter's third movement is the queue, because distance is only the first half of the road's tax, and the second half is the tax's evil twin, the one that grows without anyone touching the hardware. A queue forms wherever a

shared resource serves arrivals that come faster than service, and the wide network is made of shared resources, so the wide network is made of queues, and the queue's cruelty is its mathematics: the average traveler may pass quickly while the unlucky traveler waits, and the unlucky traveler's wait grows with the load, not linearly but steeply, so a path that is comfortably empty at three in the morning is a lottery at half past five, and the lottery's losing ticket is a missed deadline. The control loop that shares the public internet's path with the world's video traffic answers to the queue's mathematics, and the mathematics does not care that the loop is more important than the video, because the queue's discipline is arrival order and buffering, not purpose. The edge's deepest engineering sentence follows: the placement of computation one hop away removes the sharing from the critical path, and the removal of sharing is the removal of the queue, and the removal of the queue is the flattening of the tail, and the tail is where the loop lives or dies. The engineer who can recite the sentence can explain the edge to a board in one breath: the edge is the part of the system that does not stand in line.

The perception tiers deserve their own paragraph, because the loop's budget is finally a human or a machine's tolerance, and the tolerances have known shapes. The human eye forgives a screen that answers in a hundred milliseconds, notices the answer that takes two hundred, and suffers the answer that takes five hundred; the gamer's tolerance is tighter and the reader's looser, and the designer's rule of thumb, that interaction feels instantaneous under a tenth of a second, is the whole of the consumer's latency politics in one number. The machine's tolerances are stricter and stranger: the crane's controller wants its answer inside the fifty milliseconds of its sampling window, the intersection's controller works in tens of milliseconds, the power grid's protection works in single milliseconds, and the industrial controller's decade-old standards, which name their tolerances in the language of machinery, are the edge's oldest customers and its most literal ones. The tiers' honest use is in the budget's first line: the budget is written from the tolerance backward, from the eye or the relay to the tolls, and the engineer who writes the budget from the hardware forward has written a spec for the hardware she owns rather than a contract for the user she serves.

The measurement discipline is the movement that separates the honest budgets from the decorative ones, and the discipline has three rules. The first rule is to measure the road at its worst hour, not its best, because the budget is a worst-case instrument and the average is a comfort, and the measurement campaign therefore samples the day, the week, and the season,

and the port's wharf's measurements are taken in the gusty month on purpose. The second rule is to timestamp honestly, which is a harder sentence than it looks: the two ends of the measurement must agree on the time, and the agreement is the observability thread's time seam, worked in Chapter 7, and the measurement taken across clocks that disagree by tens of milliseconds is a measurement that contains its own error and reports it as latency. The third rule is to measure the road in both directions, because the request's path and the response's path are different journeys through different queues, and the engineer who measures only the outbound trip has budgeted half the toll. The three rules, kept together, produce the number the first chapter's page demands: the road's worst-case, clock-honest, bidirectional delay, and the number is the architecture's foundation stone, and every later chapter builds on it.

The clock deserves the movement's closing paragraph, because the edge is a distributed system and distributed systems are governed by their clocks' disagreements. The estate's sites keep their own time, drift with temperature and load, and are disciplined by the time service of Chapter 7's shared services, and the discipline's tolerance is set by the estate's own needs: the observability thread that must order events across sites needs the clocks within milliseconds, the crane's loop needs its own site's clock only, and the financial records need the day's order, not the millisecond's. The reader who has followed the series will recognize the theme from the distributed ledger's and the digital twin's chapters, and the recognition is the point: the edge inherits the classic problems of distributed systems, naming, time, and agreement, and the edge's honest engineering is the old engineering performed at the network's newest frontier, which is why the discipline's grayest literature is also its youngest, and why the book keeps returning to the same three problems wearing new decade's clothes.

The staircase deserves its history, because the tiers the decade deploys are the computing industry's oldest rhythm returning, and the rhythm explains why the edge's arrival surprised the industry's institutions twice. The industry began local, the mainframe's room, the minicomputer's department, the personal machine on the desk, and every era's answer was the pendulum's swing between the capital's efficiency and the user's proximity. The client-server era swung the computation halfway, the web swung it back to the center, the cloud completed the swing, and each swing's driver was a technology that made the distant end cheaper: the network made the distance survivable, the browser made the client thin, the hypervisor made the center

efficient. The edge's swing is driven by the same physics this chapter began with, the distance's tax on the answers that must be immediate, and the reader who expects the pendulum to rest at the edge has misread the mechanism: the pendulum never rests, and the next swing, whatever the next decade names it, will pull back toward the center whatever the center's new economics make cheap, and the estate that survived the last four swings knows the discipline that survives them all, which is the placement's arithmetic, performed without sentiment on each workload as its own case, because the pendulum's lesson is that no placement is permanent and every placement is a lease on physics' patience.

The data's gravity deserves the movement's second paragraph, because the offloading's decision of Chapter 4 has a shadow doctrine that deserves its name. Data has gravity, the industry's oldest aphorism says, and the gravity is the road's tax stated as an attraction: the data's mass pulls the computation toward it, and the computation that fights the gravity pays the road's tax forever, and the decade's honest architectures respect the gravity in their placement's grammar. The port's manifests are born at the wharf and are heavy with the day's operations, and the heavy data's computation belongs beside it, whatever the cloud's efficiency; the hospital's images are born in the scanner's room and are heavy with confidentiality as well as bytes, and the gravity is doubled, physics and law pulling the same direction; the chain's transactional records are small and light, born at the register and destined for the finance office, and their gravity is the center's, and the records belong to the cloud without shame. The gravity's doctrine is the placement's honesty made physical: the architect who asks where the data is born, how heavy it is, and where it must end, has asked the placement's whole question, and the tiers' staircase is the answer's landscape, and the gravity is the answer's compass.

The perception tiers' last word belongs to the machine's side, because the edge's fastest loops answer no human at all, and their budget is the physics of the process they serve. The power grid's protection relays answer in single milliseconds, and the relay's computation lives beside the current it watches, and the relay's estate is the edge's oldest citizen, deployed for a century before the word arrived; the engine's ignition and the robot's safety stop work in tens of microseconds, at the mist's deepest layer, where the computation and the sensor share a die, and the estate's staircase does not reach this depth, because the deepest mist is not infrastructure but machinery, and the machinery's maker owns its loops forever. The staircase's honest boundary is

therefore drawn by the budget: the tiers this book governs are the loops from the millisecond up, the loops that a network can serve, and the loops below the millisecond belong to the machine's own design, purchased with the machine, and the estate's architect who understands the boundary knows which budgets her contracts can promise and which budgets she must buy already-kept inside the machinery she procures.

The illustrative latency ladder deserves its worked numbers, because the staircase's tiers are best internalized as the road's price list, and the price list's figures, illustrative in the house tradition and labeled as such, give the reader the scale the book will reuse. The device's own computation, the mist: the data's arrival to the verdict's production, microseconds, the price of a bus on a die. The local hop, the mist to the cloudlet in the same building or the same cabinet: fractions of a millisecond on the wire, a few milliseconds with the radio's scheduling, the price of one honest hop. The site to the site's neighborhood, the cloudlet to the micro data center down the road: single-digit milliseconds round trip, the price of a few kilometers of owned fiber. The neighborhood to the regional aggregation, the carrier's site or the operator's office: tens of milliseconds, the price of the metro's sharing. The region to the distant cloud: the hundred milliseconds that the ocean's breadth and the internet's queues impose, the price of the world's road. The ladder's figures are the architecture's first table, and the engineer who can recite them can place any workload in one sentence, because the placement's question is the budget's ladder-matching, the loop's tolerance against the ladder's prices, and the match is the whole of the staircase's engineering, decided in one breath and audited for a decade.

The measurement campaign's instrumentation deserves its paragraph, because the honest number requires the honest instrument, and the instrument's design is the campaign's half that the budgets usually skip. The campaign's kit: the timestamp's source, the estate's time service, with the clocks' agreement verified before the campaign begins, because the measurement across disagreeing clocks contains its own error and reports it as latency; the probe's placement, one at the loop's beginning, one at its end, and one at each toll the campaign will price, because the road's honest decomposition requires the tolls' individual evidence, and the probe that measures only the whole path prices the tolls as one lump; and the probe's load, the campaign's measurements taken with the loop's real payloads rather than the network's test traffic, because the test's packets are small and the loop's are heavy, and the queues treat them differently. The campaign's

calendar: the worst hour found and sampled deliberately, the week's shape covered, the season's variation respected, and the campaign's report written beside the budget's page, the evidence and the promise on the same page, because the page's fourth column exists for exactly this meeting of number and number, and the estate that keeps the meeting keeps the truth.

The speed's arithmetic deserves its own paragraph, because the physics deserves the reader's own numbers rather than the book's recitals, and the numbers are two sentences to memorize. Light in vacuum covers three hundred thousand kilometers a second, three hundred kilometers a millisecond; the signal in fiber or copper travels slower, at roughly two thirds of light's pace, two hundred kilometers a millisecond, which is the round trip's foundational figure: a computation served four hundred kilometers away pays four milliseconds of pure propagation for the round trip, before the switches' queues and the protocols' overheads, and a computation served four thousand kilometers away pays forty, which no consumer's screen and no crane's loop can absorb. The radio's figure is different in shape: the radio hop's propagation is nearly free at the cell's scale, microseconds, but the radio's scheduling, the medium's sharing, the protocol's handshakes, impose their milliseconds, and the millisecond's honesty is the measurement's, not the datasheet's. The two figures, two hundred kilometers a millisecond on the wire and the measured radio hop, are the book's pocket physics, and the engineer who carries them can audit any vendor's latency claim in one breath, because the claim that beats the physics is a claim about a signal the universe does not stock.

The slack's policy deserves the chapter's closing sentence, because the budget's remainder is a decision rather than an accident. The slack between the tolls' sum and the loop's deadline is the budget's margin, and the margin's honest uses are three: the road's worst case, the hardware's aging, and the change that will be asked for, and the estate that spends its slack at commissioning, tuning the budget to the millisecond's elegance, has spent the year three's queue before the year three arrives, and the policy's sentence is the budget's last line: the slack is stated, owned, and defended at the annual review, in milliseconds, like every other promise in the page. The sampling arithmetic deserves its closing paragraph, because the loop's window and the sensor's rate are the same number seen from two sides, and the side the engineer forgets is the one the budget dies on. The crane reads its position every fifty milliseconds, and the loop must answer inside that window, so the budget is thirty and the sampling's cadence is the deadline's metronome; but the road's queues are not metronomic, and the answer that arrives at

millisecond forty-five has met the budget's letter and missed the loop's use, because the actuator's correction applies to the position the crane held a window ago, and the position has moved. The remedy is the budget's honest pairing with the deadline's consequence: the loop's contract states not only the round trip's worst case but the answer's staleness's cost, the position's drift over the budget's worst case, and the pairing converts the latency budget from a performance figure into a control-theory number, which is what the crane's engineers meant by it all along, and the book's pages speak the management's grammar so that the control room's grammar survives the translation, and the translation's fidelity is the difference between the budget that keeps the crane and the budget that frames it.

The Latency Budget page

The stealable page, compact: name the five tolls (sensor, road out, computation, road back, actuator); write the worst-case budget for each; measure before filling; name the evidence per line; review on a calendar, because queues grow. Total under the loop's deadline, with slack stated as a number. The road's allowance decides the architecture, and the road's allowance is a distance, and the distance is the decision.

ABOUT THE AUTHOR

Aiman JoeHanz writes the Building of Tomorrow series: structural, vendor-neutral guides to the technologies reshaping industry, labor, and daily life. Each volume takes one technology apart the way an engineer takes apart a machine: the definition and its contract, the build and its cost, the failure modes and their early warnings, the working day, and the honest decade ahead. This volume dissects the placement discipline, the edge, and the contracts that decide whether the decade's nearby computers become instruments or decorations.