The Role of Data Centers for Ancillary Businesses and Northwest Indiana's Technology Stack

Co-Authors:
Saurav Roychoudhury, Capital University - School of Management
Justin M. Ross, Indiana University Bloomington - School of Public & Environmental Affairs (SPEA)

Embargoed until release: September 10, 2026

Executive Summary

This report examines the types of ancillary business activity that may accompany hyperscale data-center development in Northwest Indiana. Its primary question is not whether data centers generate economic activity in the aggregate, but which portions of their construction, operating, infrastructure, and service requirements could plausibly be supplied by firms and workers within the region. The economic-impact estimates in Section 2 provide context for the scale and composition of the relevant expenditures; they are not forecasts of locally realized outcomes.

The analysis combines a regional input-output scenario, a six-layer framework for organizing the data-center support economy, qualitative evidence from established data-center markets, and an assessment of when proximity to a data center affects business-location decisions. The results identify several activities for which regional participation appears comparatively plausible, including critical-facilities maintenance, equipment repair, commissioning, fiber and network services, warehousing and specialized logistics, remote hands, security, systems integration, engineering, and related professional services. Conversely, large expenditures on servers, GPUs, switchgear, and other specialized equipment do not, by themselves, imply that the associated manufacturing will locate near a data-center campus.

Evidence from Quincy, Washington, central Ohio, and Loudoun County, Virginia, indicates several pathways through which ancillary activity may develop. Existing firms may adapt their services to meet data-center requirements; specialized vendors may enter after recurring demand becomes sufficiently visible; and some businesses may locate nearby when proximity provides a specific operating advantage. The evidence does not indicate that these outcomes follow automatically from data-center investment. Their realization depends on such factors as procurement practices, vendor qualification, workforce availability, infrastructure conditions, the number of potential customers, and the capacity of regional firms to compete for contracts.

The relevance of proximity also varies by business function. Location near a data center may matter because of latency and interconnection requirements, access to customers or data, vendor-response times and repeated contracts, or risk-diversification objectives. These mechanisms imply different geographic scales and should not be combined into a general assumption that data-center proximity benefits all technology-related businesses.

Applied to Northwest Indiana, the analysis identifies differentiated potential roles across the region. Lake County and Hammond have characteristics relevant to interconnection and operating services; Porter

County and Burns Harbor offer logistics and industrial assets; LaPorte County combines an announced campus with workforce and controls-related institutions; Jasper County presents host-site infrastructure considerations; and Newton, Pulaski, and Starke counties may have more limited or conditional roles related to workforce, fiber, maintenance, and customer-specific digital applications. These classifications are screening hypotheses rather than county-level forecasts.

The report concludes by identifying information and institutional mechanisms that would permit these hypotheses to be evaluated more directly, including supplier registries, procurement data, employer-validated training, vendor-qualification records, and reporting on awarded contracts and resident employment. Such evidence would allow subsequent research to distinguish announced investment and potential demand from realized regional business activity. The analysis does not estimate the net social benefits of data-center development or provide a benefit-cost assessment of particular projects or public policies.

1. Introduction

When hyperscale data centers are built in the Northwest Indiana region, what kinds of business activity appear along with them? This is a challenging question for any region, because hyperscale data center constructions are a relatively new phenomena at the scale and prevalence they are being observed. Precise estimates of the trends vary, but the volume of data centers has approximately doubled from 2020 to 2025 and forecasts regularly predict another doubling within the subsequent five years (e.g. see JLL's 2025 Global Data Center Market Outlook). Economic data, particularly at local and regional levels, lags several years: at the time of this writing the most recent County Business Patterns data available from the U.S. Census is just 2023. Furthermore, when the nature of economic activity changes, it doesn't necessarily fit into pre-existing categories or with a level of resolution on its connection to specific activities like data centers. These newer, hyperscale, data centers are the exact kind of stress test relative to their older counterparts. Particularly where they are AI-serving, the physical construction differs in part by their sheer scale, the technology of system-architecture inside the warehouse differs, and the opportunities they offer change.

This study seeks to investigate the character of prospective business growth accompanying hyperscale data centers in the Northwest Indiana region (the combined counties of Lake, Porter, Jasper, LaPorte, Pulaski, Starke, and Newton). It considers this both through the direct construction and operations of such data centers, as well as through its prospects as building the technology stack in the region. To address the challenges described above, it proceeds in two complementary strategies. First, it investigates supply chain relationships from a hypothesized new build of a 1-gigawatt campus following the specifications in Roychoudhury and Ross (2026). This part is based on the most recently available data on supply chain relationships within the region estimated by the US Bureau of Economic Analysis and released in their 2026 Regional Input-Output Modeling System (RIMS II). For instance, this allows us to trace which subsectors of the regional economy are predicted to be impacted in terms of jobs and output based on those historically estimated relationships. This plays a joint role of helping us establish an inventory of existing regional assets.

The second part complements this with qualitative case-study research of existing prominent areas with significant data center presence. We employ Artificial Intelligence as an automated research assistant to substantially scale up major reviews of individual businesses. A set of industry classification codes are identified from the RIMS II data from the supply chain, literature review, and other complementary research where businesses might be affected by the location of a data center. We identify relevant industry codes where a data center could be a customer to a business, or for which the location of a data center is not a customer but conveys local benefits that are desirable. These latter types of co-location are the most likely where there are complementary relationships in the technology stack to which data centers belong (e.g. advantages in network speeds, energy supply, particular types of human labor, etc.). We then orchestrate AI agents to conduct structured internet searches of these data center regions to identify local businesses using search terms associated with those industry classification codes. We then send additional AI agents to conduct individual research projects on each of those businesses to identify where there is evidence of a data center connection or motivation to their presence in the region. This allows us to gain a sense of businesses that exist in a region which adapt to provide services to data centers, those that move into a region to supply data centers, and finally those which find data center presence to be advantageous without having them as customers. These provide illustrative examples of how businesses change with more real-time information that is not possible at local levels in national data sets that take years to formulate.

The rest of the paper proceeds as follows. Section 2 investigates the supply chain impacts in the region as modeled by the BEA's RIMS II model data. Section 3 draws on case studies of mature data center clusters to better understand adaptations and evolutions in the supporting data center ecosystem. Section 4 investigates the data center "tech stack" that might be drawn to the ecosystem surrounding a data center. Section 5 recaps the findings.

2. Supply Chain Mapping from BEA Data into Regional Sector Impacts

This section begins with a hypothetical $10 billion capital investment into producing a 1 gigawatt, hyperscale data center in the Northwest Indiana region (Lake, Porter, LaPorte, Jasper, Newton, Pulaski, and Starke Counties) from Roychoudhury and Ross (2026). It then uses RIMS II data from the Bureau of Economic Analysis (BEA) to map this outside investment onto the local sector demands within the regional economy. This begins with the initial construction over a three year period, followed by 17 years of operations that include hardware refreshes on a 5 year cycle. The 2026 RIMS II data for these relationships is derived from national supply chain relationships in 2017, and anchors them to the NWI regional data in 2024 (BEA, 2026).

The notion of the supply chain impacts from the original capital outlay can be considered in a sequence like the following:

1. A outside investor outlays $10 billion to construct a hyperscale data center, which is apportioned across land acquisition, the facility, equipment, and other purchases. The size of this investment similarly determines the future annual operational expenditures. In both cases, only some of those expenditures are served by the region's own existing supply chain.
2. The expenditures by the capital investor on local construction, operations, and vendor purchases. These first round pulls from the local suppliers are **Direct effects** that result in local earnings and employment.
3. Locally, suppliers to the data center have their own suppliers, some of which are also local to the region. **Indirect effects** are those business-to-business exchanges that are measured as their own output, earnings, and employment.

4. Lastly, these transactions result in household level incomes within the region that is recirculated on local activities. The are **induced effects** from workers spending their earnings on housing, food, retail, etc.

Because the regional presence of businesses differs across different supply chain components, as well as their own suppliers, this results in different multipliers stemming from the initial investment. We use the RIMS II analysis from BEA data in the region to categorize and explain the ancillary impacts from the initial investment. This tracks the economic sectors impacted according to their North American Industry Classification System (NAICS) code. For instance, annual operations will require some amount of maintenance and equipment refurbishment, which is tracked in the NAICS sector 811310 for "Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance." The BEA data in RIMS II indicates a regional purchase coefficient of 0.70, indicating that 70-cents over every dollar spent by the data center would go to in-region suppliers.

Impacts accruing in future years are discounted by 3 percent following guidance from the Office of Management and Budget Circular A-4.

2.1 The Data Center Build and Operations

A hyperscale campus is a portfolio of purchases, not a single industry. The table below provides the assumed allocation of the initial $10 billion of capital spending into construction, imported equipment, contingency, and land-related services. The apportionments of the assumptions are explained and defended in Roychoudhury and Ross (2026).

Capital category	Share	Scenario value	Regional treatment
Warm-shell construction	25.0%	$2,500M	Gross $2,500M; component-level multipliers, RPC-adjusted
IT equipment: servers, GPUs, networking	55.0%	$5,500M	Margin-adjusted; only $682M (12.4%) treated as in-region
Contingency / reserves	17.5%	$1,750M	Allocated across construction sectors
Land acquisition	2.5%	$250M	Transfer of an existing asset; $18M of transaction services modeled
Total	100.0%	$10,000M	Scenario total; not all regional final demand

Note: Adopted from Northwest Indiana hyperscale baseline model in Roychoudhury and Ross (2026).

Similarly, after the build is completed there is continued annual expenditures on annual operations. The central case assumes $860 million in annual gross operating spending, of which $550 million is treated as in-region final demand. The table below provides the initial apportionment of annual operation expenditures across NAICS codes.

Operating component	NAICS / BEA mapping	Annual spend	RPC	In-region	Out-of-region
Electricity	2211	$570M	0.60	$342M	$228M
Infrastructure maintenance	811310	$150M	0.70	$105M	$45M
On-site staffing / systems services	541512	$90M	1.00	$90M	$0M
Network and bandwidth	517110/517111	$30M	0.40	$12M	$18M
Software and security	511210/513210	$20M	0.05	$1M	$19M
Total	-	$860M	-	$550M	$310M

Source: Roychoudhury and Ross (2026), Table III-2. RPC = regional purchase coefficient.

As can be seen above, for instance, very little of the annual operating expenditures made by the data center on software and security are likely to go to the types of businesses found in the NWI region. On-site staffing for system services, however, do have essentially a 1 to 1 ratio from annual expenditures according to the BEA RIMS II data.

Under these assumptions, the initial build and operations of the hypothesized data center have the following aggregated effects on the regional economy after working through different stages of the supply chain:

Phase	Layer	Output	Earnings	Employment
Construction (3 years)	Direct	$3,815.2M	$792.8M	10,787 job-years
Construction (3 years)	Indirect	$1,637.8M	$283.6M	3,867 job-years
Construction (3 years)	Type I subtotal	$5,453.0M	$1,076.4M	14,651 job-years
Construction (3 years)	Induced	$1,263.2M	$265.4M	5,481 job-years
Construction (3 years)	Type II total	$6,716.6M	$1,341.9M	20,133 job-years
Operations (annual)	Direct	$550.0M	Not separately additive	926 jobs
Operations (annual)	Indirect	$141.0M	Not separately additive	1,164 jobs
Operations (annual)	Type I subtotal	$691.0M	$175.9M	2,090 jobs
Operations (annual)	Induced	$154.8M	$43.4M	896 jobs
Operations (annual)	Type II total	$845.8M	$219.3M	2,986 jobs

Source: Roychoudhury and Ross (2026). Values are reproduced as published; displayed source rows and totals can differ slightly because of underlying aggregation and rounding. Construction employment is job-years; operations employment is annual jobs supported.

Quantitative guardrails: how to read the employment figures
Construction. The 14,651 Type I and 20,133 Type II figures are job-years accumulated across a three-year build. Type II adds induced household spending. Neither figure represents permanent jobs or necessarily distinct workers.
Annual operations. The 300 on-site FTE assumption is facility headcount. The 926 first-round jobs include labor supported by the broader operating bill of goods. The 2,090 Type I regional jobs add indirect supplier effects; the 2,986 Type II regional jobs add induced household-spending effects.

The construction result shows a large but temporary supplier market. Directly, the construction outlays commission a little over $3.8 billion in local output that produces $793 million in earnings and support 10,787 job-years. The strict on-site staffing assumption is 300 permanent FTE in the 1 GW case. Annual operations directly support 300 on-site employees plus an additional 626 contract and vendor workforce for total direct employment of 926 jobs, annually.

However, the indirect effects follow these as knock-on output, earnings, and employment as many of these direct expenditures have additional subsequent rounds of their own supplier services. Combining direct and indirect (Type I) impacts, construction generates about $5.4 billion in output, $1.1 billion in earnings, and 14,651 job-years. Indirect output is approximately 30% of the Type I construction total, and induced output adds another amount equal to roughly 23% of the Type I total. The operations result is smaller but persistent: supplier employment (1,164 jobs) exceeds the model's broader first-round direct employment (926 jobs), underscoring the importance of utilities, maintenance, carrier operations, field technicians, and other vendors. The RIMS II first-round direct employment concept is broader because it includes labor supported by the entire operating bill of goods. The 2,090 Type I jobs are jobs supported across the region, not workers inside the building. Accounting for estimated impacts from additional rounds from local household spending grows the estimated employment effect to almost 3,000 annually, and a little over 20,000 during the 3-year construction build.

2.2 The B2B supply-chain sectors

The previous section went from initial investment to total local impacts after aggregating across the supply chain. We now look at the supply chain to gain a clearer picture of what industries are impacted, according to the BEA RIMS II data.

The following abridged table lists the 14 highest Type II output multipliers among the target industries that appear along the data center's various supply chains, followed by two leakage examples. To illustrate, each $1 million of final demand for architectural and engineering services generates approximately $1.68 million in total regional economic output, and supports 8.451 local jobs. Again, these are numbers specific to the Northwest Indiana region. A data center build out is going to have different levels of final demand across each of these sectors, but these multipliers provide a sense of which sectors generate the largest total regional response as measured by output and employment.

Rank	Industry function	NAICS / BEA mapping	Type II output multiplier	Jobs per $1M	NWI interpretation
1	Office/commercial facility construction	236220 / BEA 2332OC	1.8242	7.3429	Immediate construction capture
2	Warehousing and storage	493110	1.7963	9.9561	Secure staging, spares, logistics
3	Industrial machinery repair	811310	1.7200	10.1845	Generators, pumps, mechanical systems
4	Architectural and engineering services	541330	1.6747	8.4510	Design, commissioning, compliance
5	Electronic and precision repair	811210	1.6656	11.9255	Controls, sensors, electronic service
6	Computer systems design	541512	1.6615	8.2861	Integration, cyber, managed services
7	Power and communication structures	237130 / BEA 2332PC	1.6323	6.6660	Substations, line and fiber construction
8	Other electronic components	334419	1.6237	8.5769	Assembly/service opportunity
9	Other computer-related / facilities services	541519	1.6143	8.7061	Operations support and remote hands
10	Communication and energy wire	335929	1.5760	4.0811	Cable, connectivity, field supply
11	Wholesale trade	42	1.5539	5.1135	Electrical/HVAC/IT distribution
12	Data processing and hosting	518210	1.4659	4.5163	Core digital infrastructure
13	Electric power generation/transmission/distribution	2211	1.4555	3.2630	Large final-demand base; lower job intensity
14	Wired telecommunications	517111	1.3729	4.2555	Fiber/carrier operations
Leakage	Electronic computer manufacturing	334111	1.0000	0.0000	Do not assume local server production
Leakage	Switchgear/switchboard manufacturing	335313	1.0000	0.0000	Target service, testing, distribution

Source: Extracted from BEA RIMS II Type II table for Northwest Indiana Region, May 2026 data vintage. Final-demand output is total regional output per $1; employment is jobs per $1M of regional final demand. Rank in the first column 1 is according to the Type II final demand output multiplier.

The next table interprets the data center annual operating expenditures in which final demand is modeled to occur in the Northwest Indiana region. The largest component of this demand comes from $342 million in expenditures on local electricity. Although electric power (NAICS 2211) ranks as only 13th among the sectors by the type II multiplier (1.4555), its much larger expenditure base produces $497.8 million in regional output and supports 1,116 jobs. This makes electricity the largest stimulating component of the data center investment in terms of total employment in the region. These 1,116 jobs include those employed in the electric power industry itself as well as those stimulated indirect along the supply chain and induced by household spending across the Northwest Indiana economy.

Operating component	Local final demand	Type II output multiplier	Modeled Type II output	Jobs per $1M	Modeled jobs
Electricity	$342.0M	1.4555	$497.8M	3.2630	1,116
Infrastructure maintenance	$105.0M	1.7200	$180.6M	10.1845	1,069
On-site staffing / systems services	$90.0M	1.6615	$149.5M	8.2861	746
Network and bandwidth	$12.0M	1.3729	$16.5M	4.2555	51
Software/security and rounding residual	$1.0M	Not separately reported	$1.4M	Not separately reported	4
Published annual total	$550.0M	-	$845.8M	-	2,986
Note: Values are annual steady-state operations, not a cumulative 20-year distribution. Modeled Type II output equals local final demand multiplied by the output multiplier; modeled jobs equal local final demand in millions multiplied by the jobs-per-$1M coefficient. The software/security line is the residual to the published total after the four separately reported coefficients and rounding. Construction values above cover the three-year build, and construction employment is reported as job-years.					

In reviewing the results above, we can see that annual infrastructure maintenance is the next largest contributor. Network and bandwidth fiber for wired communications is ultimately a much smaller total impact despite a meaningful multiplier.

2.3 The Data Center Support Economy as a Six Layer Ecosystem

Having now overviewed the major logic of the data being tracked, the industrial codes with the heaviest multiplier bite, we now turn towards examining Northwest Indiana regional supply information with more granularity to try and understand them as complementary economic ecosystems.

In this section we have categorized them as six layers. We continue them as a sequence from temporary construction demand to permanent operations, but fill in with associated sectors where infrastructure, skills, professional capacity, and community spending occur. The layers overlap, but each has a different time horizon, business model, workforce, and public-policy lever. The table below summarizes the six layers to be investigated in turn.

Layer	Ecosystem function	Time profile	Indicative NAICS	NWI readiness
1	Construction & Skilled Trades	Temporary / high intensity	236220, 237130, 238210, 238220, 238290	Strong base; capacity and scheduling constraints
2	Operations & Maintenance	Permanent / recurring	518210, 811210, 811310, 561210, 561612	Buildable through vendor qualification

Layer	Ecosystem function	Time profile	Indicative NAICS	NWI readiness
3	Energy Infrastructure	Permanent system + project build	2211, 237130, 335311, 335313, 335910	Strategic scale; cost and reliability governance essential
4	Workforce Development	Growing enabling system	611210, 611519; training intermediaries	Strong institutional base; employer conversion must be formalized
5	Professional Services	Permanent / project and recurring	541310, 541330, 541620, 541110, 531, 524	Good metro access; expand NWI firm participation
6	Support Economy	Permanent / broad but dispersed	44-45, 531, 621-624, 721, 722, 81	Capture depends on worker residence and service capacity

In addition, as we proceed through the layers, we discuss instances where the RIMS II model and BEA data built from historical data might under or overstate relationships through the supply chain. For instance, the relationships between sectors were built on national data in 2017 when hyperscale data centers are in their infancy and the data center sector better resembled office buildings. To the extent that hyperscale data centers induce local agglomeration effects, those historical relationships might be underestimated and local multipliers would be larger. Similarly, the 2024 data on industry compositions in the region might have idiosyncrasies that make it better or worse suited.

2.3.1 Layer 1: Construction & Skilled Trades

The first layer of the data-center ecosystem is the construction and skilled-trades activity required to bring a facility online. This begins with site preparation and civil work and extends through utility connections, the building shell, electrical and mechanical systems, cooling, fire and life-safety systems, controls, fiber, security systems, equipment installation, testing, and final commissioning. The principal industries involved include commercial construction (NAICS 236220), power and communication line construction (237130), electrical contractors (238210), plumbing and HVAC contractors (238220), other building equipment contractors (238290), and engineering services (541330). Although this layer is temporary by definition, it represents the most concentrated period of local labor and business demand associated with data-center development.

The corresponding workforce is broad but heavily weighted toward skilled construction trades and technical occupations. Important job families include electricians (SOC 47-2111), plumbers, pipefitters, and steamfitters (47-2152), HVAC mechanics and installers (49-9021), operating engineers and other construction-equipment operators (47-2073), construction laborers (47-2061), and first-line construction supervisors (47-1011). Data-center projects also create demand for electrical and mechanical technicians (17-3023 and 17-3027), commissioning specialists, safety professionals, riggers, truck drivers, and project-controls staff. The relevant business ecosystem therefore extends well beyond general contractors. It includes union subcontractors, engineering and testing firms, controls integrators, surveyors, geotechnical firms, equipment-setting contractors, and secure logistics providers.

The scale of this activity can be substantial. As discussed in section 2.1, the report's central 1 GW scenario, construction supports an estimated 14,651 Type I construction job-years over a three-year build period, increasing to 20,133 Type II job-years once the additional employment supported by household spending is included. These figures should not be interpreted as 20,133 permanent jobs or even necessarily 20,133 distinct workers.

This concentration of activity creates both opportunities and pressures for surrounding communities. A large construction pipeline can increase wages, expand apprenticeship opportunities, and provide local contractors with experience on technically demanding projects that may position them for future work that in turn makes the area desirable for new entities that require those skills. At the same time, a surge in construction can tighten the supply of skilled crafts, place additional demands on roads and freight infrastructure, increase demand for lodging and rental housing, and add pressure to emergency-response systems. The economic significance of this layer therefore depends not only on the number of jobs generated, but on the region's ability to supply qualified workers and businesses.

Northwest Indiana has an established industrial and skilled-trades base, but that capacity is a starting condition rather than proof of available workers or local contract awards. Documented assets and announced projects support different, though nonexclusive, county roles: Porter County's Burns Harbor port, rail, and heavy-industrial platform supports materials handling and logistics (Ports of Indiana, 2026); the larger labor and industrial base and operating/interconnection assets in Lake County support regional service delivery; the planned Microsoft campus in LaPorte County creates a current project and workforce anchor (IEDC, 2024); and the announced Amazon development in Jasper County concentrates host-site civil, utility, and community-infrastructure needs (Amazon, 2026). These are evidence-backed hypotheses, not county-specific impact estimates or proof that contracts have been awarded locally. Accordingly, the central capture question is whether workers and firms can obtain the qualifications, certifications, scale, and project experience required, and whether those resources can be mobilized in the sequence and on the timetable required by hyperscale construction.

2.3.2 Layer 2: Operations & Maintenance

Once construction is complete, the data-center ecosystem shifts from a temporary surge in project-based activity to a smaller but permanent base of operations and maintenance employment. This layer encompasses data-processing infrastructure (NAICS 518210), computer systems and facilities services (541512 and 541519), industrial and electronic equipment repair (811310 and 811210), facilities support services (561210), physical security (561612 and 561621), network services (517111), and field service provided by equipment manufacturers and specialized vendors. Unlike the construction phase, data-center operations are continuous. Facilities operate around the clock, reliability standards are exceptionally high, and responsibility for keeping the site functioning is often divided among the data-center operator's own employees, specialized contractors, original-equipment-manufacturer technicians, and security and facilities vendors.

The resulting occupational mix combines information technology with electrical, mechanical, and facilities expertise. Priority roles include critical-environment technicians and engineers, network support specialists (SOC 15-1231), systems administrators (15-1244), information security analysts (15-1212), electrical and electronic repairers (49-2094), industrial machinery mechanics (49-9041), HVAC mechanics, general maintenance and repair workers (49-9071), inventory and asset-management specialists, security officers (33-9032), and site managers. Many of the most promising business opportunities similarly lie at the intersection of industrial maintenance and information infrastructure. These include generator, uninterruptible-power-supply, and battery-energy-storage-system service; chiller and cooling-system maintenance; controls and sensor calibration; remote-hands services; regional parts depots; electronic repair; secure IT asset disposition; and specialized cleaning and security services.

The employment scale of this layer is substantially smaller than that of construction, but its economic character is different because the activity persists for the life of the facility. Under the report's central 1 GW scenario, a strict assumption based on workers physically located at the data-center sites yields approximately 300 permanent full-time-equivalent positions. A broader RIMS II first-round direct concept produces an estimate of 926 jobs, while the Type I regional total reaches 2,090 jobs and the supplementary Type II estimate, including induced household-spending effects, reaches 2,986 jobs. These estimates represent progressively broader concepts of the employment supported by operations and should therefore not be interpreted as alternative estimates of the number of people working inside the facilities themselves.

This distinction is important for understanding the regional opportunity. Data centers are not especially labor-intensive facilities once operational, particularly when compared with the scale of the capital investment required to construct them. Their permanent economic contribution instead comes from a combination of relatively specialized on-site employment, recurring payroll, and an ongoing stream of contracts for maintenance, security, equipment service, parts, and other supporting functions. A region that captures those vendor relationships can therefore obtain more lasting economic value than is suggested by the on-site headcount alone. Moreover, repeated contracts can create career ladders and support firms whose customer base extends across multiple data centers rather than depending on a single facility.

Access to these jobs, however, differs from access to ordinary commercial or industrial employment. Twenty-four-hour operations create substantial demand for shift work, while reliability and security requirements can require specialized training, certifications, background screening, and familiarity with mission-critical environments. For Northwest Indiana, the opportunity is therefore not simply to supply workers to individual facilities. It is to translate the region's existing industrial maintenance, electrical, mechanical, and technical capabilities into a workforce and supplier base that can meet the unusually stringent reliability standards of hyperscale operations. If that transition occurs, this layer provides the first durable bridge between the temporary construction boom described in Layer 1 and a permanent regional data-center economy.

2.3.3 Layer 3: Energy Infrastructure

The third layer is the energy infrastructure required to supply data centers with electricity at hyperscale. This includes generation, transmission and distribution, substations and interconnection facilities, protection and control systems, high-voltage testing, battery storage, renewable generation, demand management, and broader grid modernization. Relevant industries include electric power generation, transmission, and distribution (NAICS 2211), power and communication line construction (237130), transformer manufacturing (335311), switchgear manufacturing (335313), battery manufacturing (335910), engineering services (541330), and environmental consulting services (541620). Unlike the construction activity in Layer 1, much of this infrastructure remains in place for decades and becomes part of the region's permanent utility and industrial capital stock.

The workforce required for this layer combines utility trades with engineering and specialized technical expertise. Important occupations include electrical power-line installers and repairers (SOC 49-9051), electricians, electrical engineers (17-2071), electrical and electronic engineering technologists and technicians (17-3023), relay and protection technicians, civil engineers, environmental technicians, heavy-equipment operators, commissioning engineers, vegetation and land-management contractors, and utility operations personnel. Northwest Indiana is unlikely to manufacture every major component required for a large data-center buildout. Large turbines, power transformers, switchgear, and other highly specialized equipment will often be sourced from national or global manufacturers. The more plausible regional opportunities lie in construction, field testing, controls, engineering support, commissioning, maintenance, and the logistics involved in delivering, installing, and servicing this equipment.

Energy is also the largest recurring input in the report's economic model. Under the central 1 gigawatt scenario, data-center electricity consumption represents approximately $570 million in annual gross spending, of which an estimated $342 million is retained within the region. This scale distinguishes energy infrastructure from nearly every other part of the data-center supply chain. A sustained increase in electricity demand can support utility investment, accelerate modernization of transmission and distribution systems, and create recurring demand for engineering, testing, maintenance, and other field services. Some of the infrastructure built to serve large-load customers may also strengthen the wider regional grid or facilitate other industrial development.

The economic value of data-center electricity demand cannot be evaluated simply by treating utility purchases as additional regional spending. Though beyond the scope of the current analysis, the distribution of costs and risks among the data-center operator, utility shareholders, existing ratepayers, and future customers is itself an important component of the regional economic impact.

NIPSCO's proposed GenCo framework is intended to address this problem by separating costs associated with very large loads and structuring the arrangement so that credits generated by those customers can flow back to existing ratepayers (NIPSCO, 2026). If successful, this model could allow Northwest Indiana to pursue the economic benefits associated with large new electricity loads without shifting their infrastructure costs onto existing households and businesses.

A possibility that is not transparent to this study is that legacy manufacturers in the region have behind the meter energy generating capacity that might help to supplement the local grid or data centers. It is well known that manufacturing has become far more employment efficient, but it has also become energy efficient as well. Retired process load facilities with onsite generation or industrial fuel streams might have capacities that make the area in a stronger position than the REMIS II model is capable of catching. Gary Works, Cleveland-Cliffs Indiana Harbor, U.S. Steel Midwest Plant, Unilever in Hammond, and others might find data center energy demands a stabilizing source of revenue under the right set of strategic investments. The authors' of this study have no inside knowledge as to whether this is the case, but there are analogs around the country from aluminum smelters, steel mills, paper mills, and other forms of industrial campuses where legacy power position from abandoned manufacturing was picked up by data center development. Century Aluminum in Hawesville-Kentucky was idled in 2022, and sold the property in February 2026 to TeraWulf for redevelopment as a high-performance computing and AI campus; In July 2026, Anthropic agreed to occupy about 401 megawatts of the critical IT capacity there (Hart, Wall Street Journal, 2026).

Workforce preparation provides another opportunity for local capture. Training programs that build competency in medium-voltage systems, relay and protection systems, battery-energy-storage-system servicing, power quality, electrical one-line diagrams, test instrumentation, and commissioning can help translate large utility investments into employment and contracting opportunities for Northwest Indiana firms and workers. These capabilities are useful beyond data centers themselves, giving them potential value across advanced manufacturing, utilities, renewable generation, and other electricity-intensive industries.

2.3.4 Layer 4: Workforce Development

Workforce development differs from the first three layers because it is not itself a large data-center industry, but an adjacent-industry supporting sector. Local workforce development determines how much of the employment and business activity generated elsewhere in the ecosystem can be captured locally. This layer includes community colleges, high-school career and technical education programs, union apprenticeship and Joint Apprenticeship and Training Committee programs, WorkOne, universities, customized employer training, mobile laboratories, employer academies, and services that help workers overcome barriers to participation. The objective should not be to maximize enrollment in programs labeled as data-center training. It should be to translate verified employer work packages and actual job openings into local hires, apprenticeships, placements, and supplier capacity.

Northwest Indiana already possesses much of the institutional infrastructure needed to perform this function. WorkOne Northwest Indiana and the Northwest Indiana Workforce Board serve all seven counties, providing a natural regional framework rather than requiring each county to develop a separate workforce strategy. Within that system, Lake, Porter, and LaPorte Counties are well positioned to serve as larger laboratory, classroom, and apprenticeship hubs. Jasper County can function as a rural host-site location and as a satellite for construction, utilities, and water-related training. Newton, Pulaski, and Starke Counties can participate through mobile and hybrid instruction, high-school CTE programs, SCILL, and access to regional laboratories and apprenticeship programs. This hub-and-spoke structure matters because the employment opportunity generated by data-center investment need not be confined to the county in which a particular facility is located.

A potentially important addition to this platform is the 2026 memorandum of understanding between Ivy Tech Community College and Microsoft establishing a framework for a Data Center Academy (Ivy Tech, 2026). The agreement may support curriculum alignment, workforce training, internships, experiential learning, industry engagement, and career exploration. It is best understood, however, as a platform from which a workforce pipeline can be built rather than evidence that one already exists. An institutional agreement does not by itself establish the number of Northwest Indiana residents who will be trained, hired, or retained in data-center employment. Those outcomes will depend on the programs subsequently created, employer participation, paid work-based learning opportunities, and actual placement results.

For this reason, workforce programming should be organized around occupational clusters tied to identifiable demand rather than around a generic "data center" credential. The principal clusters identified in the regional training crosswalk are electrical power and controls; HVAC, mechanical, and water systems; construction and commissioning; fiber and networking; information technology and critical facilities; cybersecurity and managed-service providers; industrial maintenance and logistics; and facilities and security. Many of these skills are transferable across the wider Northwest Indiana economy, which reduces the risk associated with training for a single employer or technology. Electrical controls, industrial maintenance, HVAC, networking, cybersecurity, and utility operations all have applications in manufacturing, health care, logistics, utilities, and other infrastructure-intensive sectors.

The most important design principle is that training should begin with the conversion mechanism between instruction and employment. Before a cohort is recruited, the participating institutions should be able to identify the employer tasks students are being trained to perform, current job descriptions, required credentials, the paid work-based learning opportunity available to participants, and the outcomes that will subsequently be reported. This shifts workforce policy away from measuring success primarily by enrollment or certificates awarded and toward measuring whether residents actually move into the jobs and contractor opportunities created by data-center investment.

Training track	SOC families	Required conversion mechanism	Delivery logic
Construction and MEP	47-2111, 47-2152, 49-9021, 47-2061	Apprenticeship / paid rotation / contractor interview	Lake, Porter, LaPorte, Jasper; portable access regionwide
Critical facilities and controls	49-2094, 49-9041, 17-3023, 17-3024	Vendor lab + operator task validation + paid placement	LaPorte/Lake hubs; Porter/Jasper support
Fiber and networking	49-9052, 15-1231, 15-1244	Supervised field hours, splicing/test sign-off	Lake/Porter/LaPorte plus Newton/Pulaski satellites

Training track	SOC families	Required conversion mechanism	Delivery logic
Cybersecurity and MSP	15-1212, 15-1244, 15-1231	Client project, NOC/SOC/help-desk internship	Lake/Porter/LaPorte lead; hybrid regionwide
Facilities, water, and security	51-8031, 19-4042, 49-9071, 33-9032	Utility/facility rotation and licensing progress	Host counties plus mobile delivery

This framework also provides a way to evaluate whether the regional workforce system is successfully capturing the opportunity. For construction and mechanical-electrical-plumbing trades, the relevant measure is not simply course completion but entry into an apprenticeship, paid rotation, or contractor interview. Critical-facilities and controls programs should culminate in validated tasks and paid placements. Fiber programs should include supervised field hours and demonstrated competency in splicing and testing. Cybersecurity and managed-service training should connect students to real client projects or internships in network- or security-operations environments. Facilities, water, and security pathways should similarly incorporate rotations, licensing progress, and employer-verified competencies.

Seen this way, workforce development is the principal mechanism through which the temporary and permanent labor demand described in the preceding layers becomes a regional economic-development strategy. Data-center investment can create substantial demand for labor without automatically creating substantial opportunity for Northwest Indiana residents. The difference between those outcomes will depend on whether training providers, unions, employers, contractors, and workforce agencies can coordinate around actual hiring requirements early enough to prepare workers before demand peaks and labor is imported. The most useful workforce investment is therefore not necessarily the program with the largest enrollment.

2.3.5 Layer 5: Professional Services

The fifth layer consists of the professional and technical services that support data-center development before, during, and long after construction. These include architecture (NAICS 541310), engineering and commissioning (541330), environmental consulting (541620), legal services (541110), surveying and geotechnical work, planning and construction management, finance, insurance, commercial real estate, tax advisory, cybersecurity, and other specialized technical consulting. Unlike the construction trades, many of these services are not concentrated solely in the initial build. They enter during site selection, due diligence, design, permitting, and construction, then recur as facilities expand, contracts are renegotiated, insurance is renewed, environmental requirements are monitored, tax obligations are reviewed, cybersecurity systems are tested, and assets are managed over time.

The occupational base is correspondingly diverse and comparatively high-skilled. Priority occupations include electrical, mechanical, and civil engineers; engineering technicians; architects; construction managers; environmental scientists and technicians; surveyors; land-use and utility attorneys; commercial real-estate professionals; insurance brokers; tax professionals and accountants; project-controls analysts; cybersecurity consultants; and commissioning authorities. These occupations matter not only because of their wages, but because professional-services spending can generate substantial secondary economic activity. In the detailed Northwest Indiana input-output table, architectural and engineering services have a Type II output multiplier of 1.6747 and support approximately 8.451 jobs per $1 million of final demand. Capturing a greater share of these contracts therefore provides one avenue for retaining more of the value associated with data-center investment inside the regional economy.

For Northwest Indiana, however, this layer also illustrates the importance of defining the regional economy carefully. The Chicago metropolitan area provides an exceptionally deep pool of legal, engineering, architectural, financial, and technical expertise. That proximity is an advantage to data-center developers because sophisticated services are readily available, but it does not necessarily translate into economic activity captured within Northwest Indiana. Under the seven-county RIMS II geography used in this report, a contract awarded to a firm across the state line in Illinois represents leakage from the regional economy even if the provider is geographically nearby. The relevant economic-development question is therefore not whether data-center projects can obtain the professional services they require, but how much of that work can be performed by firms and workers located within Northwest Indiana.

This creates a different local-capture challenge than the one facing the skilled trades. Northwest Indiana firms may not initially have the scale, specialized portfolio, or national relationships necessary to serve as prime consultants on the largest hyperscale projects. They can nevertheless participate as subcontractors, joint-venture partners, local specialists, and recurring service providers. A deliberate strategy would make regional firms more visible to national architecture, engineering, environmental, and construction-management companies while encouraging local firms to develop specialized capabilities that are difficult to substitute from a distance. Particularly promising niches include large-load utility and interconnection work, industrial-site reuse, cooling and water analysis, environmental monitoring, acoustical and sound modeling, grid and campus cybersecurity, and testing and commissioning.

These specializations have an additional economic-development advantage: they are potentially exportable. A Northwest Indiana engineering or consulting firm that develops expertise serving one hyperscale project can subsequently sell that knowledge to projects elsewhere in Indiana, the Midwest, or nationally. The same is true for commissioning, environmental monitoring, utility consulting, cybersecurity, and specialized legal or tax services. In this respect, professional services offer a pathway by which a one-time regional investment can contribute to the formation of durable intellectual and organizational capital. The goal is not simply to capture a larger share of a developer's initial spending, but to help regional firms acquire credentials, relationships, and expertise that expand their markets beyond the original project.

At the same time, local professional capacity serves a second purpose that should remain distinct from supplier development. Communities, utilities, and public agencies need access to independent technical expertise when evaluating projects whose consequences extend beyond the developer's property line. Maintaining regional capacity in utility analysis, environmental review, fiscal analysis, land use, and infrastructure planning allows public officials to evaluate claims and negotiate agreements without relying exclusively on studies financed by the project developer. Thus, professional-services development produces two complementary benefits: it enables Northwest Indiana firms to capture more of the high-value work generated by data-center investment, while also strengthening the region's own capacity to evaluate and govern that investment.

2.3.6 Layer 6: Support Economy

The sixth layer is the broader support economy surrounding data-center construction and operations. It includes housing, retail, restaurants and food services, lodging, health care, child care, transportation, personal services, recreation, and the civic infrastructure used by workers and their families. Unlike the earlier layers, most businesses in this category are not suppliers to data centers themselves. They benefit because wages earned through construction, operations, and supporting industries are subsequently spent in the regional economy. Communities may also receive additional fiscal resources through property taxes or negotiated payments associated with data-center development. These are two economically important but analytically distinct channels.

That distinction is essential when interpreting the economic-impact estimates in this report. The induced effects reported by RIMS II measure economic activity generated when workers spend their earnings on goods and services such as housing, groceries, restaurants, health care, and recreation. They do not measure local government revenue. Property taxes, payments in lieu of taxes, school allocations, tax abatements, infrastructure obligations, and the cost of providing public services must instead be evaluated through a separate fiscal analysis. Combining the two would risk treating the same development as producing benefits through fundamentally different mechanisms without accounting for the corresponding public costs.

The induced effects are nevertheless substantial. In the central 1 GW scenario, construction generates approximately $1.263 billion in induced output and 5,481 induced job-years during the construction period. Once the facilities become operational, household spending associated with the permanent employment base supports approximately $154.8 million in annual induced output and 896 jobs. As with the earlier construction estimates, the 5,481 figure represents job-years accumulated during the build rather than an equivalent number of permanent positions. The operational effects are smaller but recurring, making them part of the continuing economic footprint of the facilities.

The extent to which Northwest Indiana captures these supporting effects depends heavily on geography and worker behavior. Permanent workers contributing to the supply chain may reside or shop closer to Chicago or in other parts of Indiana. The magnitude of local induced effects therefore depends on where workers live, where temporary workers lodge, where households shop, and whether local businesses have sufficient capacity to meet the resulting demand. This is particularly important for rural host communities, where a major construction project can be physically located within the county while much of the accompanying household spending leaks into larger neighboring markets.

The regional opportunity consequently differs by county. Lake and Porter Counties already possess relatively large housing, retail, health-care, restaurant, and personal-service sectors and are well positioned to capture spending generated throughout the regional data-center ecosystem. LaPorte and Jasper Counties may have a particularly important opportunity during construction waves, but capturing that activity will depend more deliberately on lodging capacity, local food procurement, worker transportation, and the extent to which residents are hired into project employment. Newton, Pulaski, Starke, and other rural portions of the region should be especially cautious about assuming that payroll generated at a nearby project will automatically circulate through their local economies. Strategies that increase resident employment and make local establishments accessible to project workers can materially change how much of the induced effect remains nearby.

This distributional issue makes the fiscal side of the support economy particularly important. For example, the City of La Porte's 2026 agreement with Microsoft states that the company will pay full local property taxes and that 15 percent of the resulting revenue will be allocated to the school corporation for 20 years. Such arrangements can create a direct channel through which a large capital investment supports public services in the host community. The ultimate fiscal benefit, however, cannot be inferred from the announced investment value alone. Actual receipts will depend on assessed value, the applicable tax structure, implementation of the agreement, and how revenues and service responsibilities are distributed among local governments.

The appropriate regional strategy is therefore not simply to maximize the size of the induced-spending multiplier. It is to increase the probability that workers reside and spend within Northwest Indiana while ensuring that the communities absorbing growth have the fiscal and physical capacity to accommodate it.

That means connecting resident hiring to the workforce strategies described in Layer 4, anticipating temporary housing and transportation needs during major construction waves, helping local businesses prepare for large but potentially short-lived increases in demand, and separately evaluating the revenues and public-service costs generated by each project. In this final layer, the central economic-development question is ultimately one of retention: how much of the income created by the preceding layers circulates again through Northwest Indiana households, businesses, and public institutions before leaving the regional economy?

2.4 The Data Center Economic Ecosystem

The supply chain analysis and six-layers demonstrate that data centers are not limited to a small operating payroll and initial construction jobs as their economic footprint. Instead, the supply chain analysis illustrates how data centers integrate themselves through business-to-business transactions in the economic sectors that have a strong presence in Northwest Indiana. The nature of the data center build is that it results in complementary impacts in adjacent sectors, particularly related to construction, but also including energy infrastructure and professional services.

This supply chain analysis being rooted in data, was necessarily retrospective. The next section looks to case studies of mature data center sectors to get a clearer impact of other developments that might be induced outside of the region’s historical strengths.

3. Ancillary Impact Evidence from Case Study Investigations

The previous sections described business and industry in terms of what it was already doing, but the aggregates do not paint a vivid picture of what data center adjacent businesses look like. Furthermore, it is necessarily tied to the historical existence of what the Northwest Indiana region already has and does.

This section looks at three mature areas with substantive data center activity to gain a clearer picture of what this might look like. We employ Artificial Intelligence in a systematic, structured, internet search of these areas today for active business activity in pre-selected industry code designations, then task sub-agents to collect evidence files on their historical operations in the area to build evidence files. These evidence files are in turn reviewed and coded by agents and passed on to the authors for human review. The details of this process, from industry code selection to coding rubrics, are fully described in Technical Appendix A.

The purpose of this exercise is proof of concept illustrations, not census level statistics to provide forecasts or representative statistics. In fact, as we are identifying from sources like businesses own webpages, press releases, and other statements, there is selection bias in coverage that prevents drawing representative sample conclusions. However, we do gain insights on industries that are not available from government data. For example, even in recent historical data, useful statistics on business establishments at the city or county level are often suppressed at the six-digit NAICS code level to preserve privacy. This makes it very hard to observe with any granularity what establishment or employment trends will look like in a given area for many years.

3.1 Cases and Activity Definitions

The geographic areas of the case studies are:

- Quincy and the Columbia Basin (Washington)
- Central Ohio (Columbus and New Albany)
- Loudoun County (Virginia)

For coherency, we organize illustrations from these case studies according to three types of activities. **Type A** is the adaptation of an existing business. These are businesses that already operated in the area, but adapted its services or products for data center customers. We are most interested in the types of industry and businesses that are commonplace and would appear in Northwest Indiana, and these case studies offer examples of how they find new opportunities in serving data center operations. **Type B** is a business attracted as a data center supplier. These are businesses establishing where they can be close to their customers. A electrical contractor who opens a regional office to serve data center construction would be an illustrative example. **Type C** is a businesses that is attracted by proximity benefits to data centers. The business is not primarily selling to data centers, but is located in the area because being near them is valuable: dense fiber, low network delay, direct connections into cloud providers, a technical labor pool, etc.

3.2 Adaptations of Existing Businesses (Type A)

For Northwest Indiana, Type A is arguably the most consequential category, because the businesses in it are the kinds of businesses the region already has. These are firms and institutions that operated in their areas before the data centers arrived, often by decades, and that found ways to sell their existing capabilities into the new demand. Two patterns recur across the case studies. First, the earliest and largest contracts flow disproportionately to whoever already owns local capacity in the earliest trades, particularly sitework, electrical, and heavy equipment. Second, what changes for these incumbents is often the shape of demand as much as its volume: longer deployments, new service lines for existing trades, and an added marketing identity.

A clear example of incumbency rewarded is George J. Igel & Co., a Columbus sitework and earthmoving contractor founded in 1911, when its crews hauled excavation spoil by horse-drawn wagon. The firm now claims on its own marketing page to have "performed sitework on more data centers than any other contractor in central Ohio" [1]. From January to November 2022, Igel worked on Meta's $1.5 billion New Albany campus alongside soil-stabilization specialists, among the first contractors on site [3], and its portfolio records half a million cubic yards of earth moved for a single million-square-foot data center building [1]. Notably, the firm did not reinvent itself to win this work. Its own history page attributes its data center growth to "the company's variety of construction services" [2]: demand found the incumbent, not the reverse.

The same dynamic appears at the very beginning of the Quincy, Washington cluster. When Microsoft purchased 70 acres outside Quincy in 2006, Valley Electric (a large regional electrical contractor) already reached north-central Washington to serve the area's agricultural and industrial economy and was positioned to capture the first wave of construction. The firm's portfolio dates its "Columbia Data Center" project in Quincy to 2006, an installation of thirteen additional diesel backup generators that brought the facility to 37 generators in total, for an owner listed only as "Confidential Developer" [4]. Today "mission critical/data centers" is one of eight named sectors in the firm's business [5]. The pre-positioned regional branch won the year-one work, and the data-center specialty became a durable line within a diversified contractor.

Data center construction also changes the rhythm of business for the supporting equipment economy. ALL Crane Rental's Columbus branch has rented cranes in central Ohio since 1982. Its own 2020 case study describes the New Albany building boom (Facebook's 1.5 million-square-foot campus and Google's $600 million data center) and draws the same lesson: "Data center construction projects are done in multiple phases and can take years. That means consistent work for our branch" [6]. By 2025 the branch general manager described having "dozens of cranes working on one site," typically on "nine- to 12-month rentals, mostly crawlers, doing everything from foundation steel and footers to precast, structural steel, and setting major utility components on the roof" [7]. The branch's traditional hospital and university work continues alongside; what the data centers added was a radically longer and denser utilization pattern for the same fleet.

Incumbent adaptation is not limited to private firms. Loudoun Water, the public water and sewer authority for Loudoun County, Virginia, turned its treated wastewater into a product line. By its own account, in 2010 the authority "constructed the first pipelines of the reclaimed water distribution system to supply irrigation and industrial cooling water to Loudoun County's rapidly growing data center industry" [8]. The system has grown to 20 miles of pipeline, and in 2025 delivered more than 750 million gallons of reclaimed water, conserving an equal volume of potable supply [8]. The first documented data center customer, RagingWire, timed its hookup to a campus expansion and cited cost alongside sustainability in the decision [9]. For Northwest Indiana's municipal utilities, the case shows that data centers buy water at industrial volumes as well as power, and that the cooling-water contract can go to an incumbent public authority monetizing a waste stream.

The operations phase leaves its own trace on incumbents, visible even in job titles. In the Quincy region, N C Power Systems has advertised for "Data Center Generator Technicians" in the Wenatchee/Quincy area at $36.85 to $42.25 an hour, responsible for "diagnosing and maintaining all Data Center generators" in the field [10], and its electric-power page now lists data centers among the applications it supports [11].

Finally, the clearest cross-region rhyme in the entire study involves community colleges, and it bears directly on Northwest Indiana's own Ivy Tech–Microsoft Data Center Academy memorandum discussed in Section 2.4.4. In Quincy, Microsoft has provided scholarship funding since 2018 to Big Bend Community College students in its data center technician certificate program, and by fall 2022 local high-school students could graduate into the program through a career-launch pathway [12, 13]. In Columbus, Amazon Web Services approached Columbus State Community College and by November 2018 the college had launched a Data Center Technician certificate with $1,000-per-semester AWS scholarships and, later, registered apprenticeships paying $18 to $26 an hour during study [14, 15]. Northern Virginia Community College runs the same play at a larger scale in Loudoun. In all three regions, the hyperscalers subsidized incumbent public institutions to manufacture the labor they need.

Below these headline cases sits a broader repositioning visible mainly in marketing. A Columbus fuel distributor founded in 2011 to collect used oil now names the "New Albany & Data Center Corridor" as a primary service area and markets "mission-critical" diesel delivery and fuel polishing [16]; a Columbus e-waste recycler founded in 2000 sells certified data center decommissioning [17]. Some caution on those self-descriptions is warranted since neither names a specific data center client, but the pattern itself is informative: ordinary local businesses, of kinds Northwest Indiana has in abundance, observe the demand and build an offering around it.

3.3 Businesses Attracted as Data Center Suppliers (Type B)

Type B businesses established a presence in a case study region because the data centers are their customers. The record divides into two variants. In the first, established firms from outside the region follow the work in: Initially traveling to the region, then opening offices, and in the most mature version acquiring local firms outright. In the second, entirely new entities are created around a single piece of data center demand, most dramatically in energy and water infrastructure. Both variants matter for Northwest Indiana: the first describes the competition local incumbents should expect, and the second describes categories of investment that will not appear in any historical baseline of the regional economy.

The plainest statement of the first variant belongs to VECA, a Seattle electrical contractor, which explains on its own website that it "established our Wenatchee office to serve our central and eastern Washington data center and mission critical projects" [18]. The office's portfolio now lists roughly 250 megawatts of mission-critical electrical projects, including three 45-megawatt hyperscale jobs [19]. Holmes Electric, another Seattle contractor, shows what the end state of this progression looks like: after years of traveling to Quincy for work it attributes to Microsoft, Sabey, and Intergate campuses, the firm marked its 80th anniversary in 2025 by purchasing Tobin Electric, the town's own electrical shop to expand "through the purchase of Tobin Electric in Quincy, Washington" while it "continues to grow our data center and industrial electrical work" [20, 21]. Data center demand became a permanent local address, acquiring a local incumbent in the process.

The scale of investment flowing through these supplier relationships can be substantial. In a single Microsoft expansion project in Quincy, total contract to mechanics were claimed to total $70.9 million for the regional firms involved in diesel generators and cooling tower installations. Apollo Mechanical Contractors, a Kennewick-based mechanical firm, claimed $6.79 million of that total [22].

As per the indirect effects described in section 2, the supplier ecosystem becomes a customer base of its own. In January 2019, City Electric Supply opened a wholesale branch in Sterling, Virginia (Loudoun County) and its branch manager described the target market candidly: "There are several HVAC and mechanical contractors we are targeting in the area as well as the local data centers and businesses that will need both electrical and HVAC/mechanical supplies" [23]. At the time, trade coverage put the county at roughly 60 data center facilities and 10 million square feet [23]. This vendor supplying to data center suppliers, and is clear evidence that a cluster generates its own commercial center of gravity beyond the campuses themselves.

Fiber construction shows the same mechanism operating across a decade. SummitIG completed its original 170-mile dark-fiber network in Loudoun County in 2014, its CEO explaining that "we built our original network in Loudoun County as 70% of the world's daily Internet traffic runs through this area" and has since grown past 1,200 route miles [24]. Nine years later the cluster was still summoning new entrants: Harbor Peak, a venture announced in January 2023, was created expressly to build a seven-mile, 864-count fiber conduit ring connecting data centers through Ashburn, breaking ground that May [25, 26]. A company that did not previously exist was called into being by a single piece of cluster infrastructure.

The second variant of purpose-built entities reaches its fullest expression in energy. In New Albany, Ohio, a subsidiary of the Williams Companies is constructing the Socrates North and South facilities: two 200-megawatt gas-fired power plants that will never connect to the grid, built to serve a single data-center customer with the Ohio Power Siting Board approving the southern facility in June 2025 [27, 28]. Feeding the plants required creating yet another company: Will-Power Pipeline OH, LLC, formed to build an approximately 17.9-mile, 24-inch gas pipeline whose only role is supplying the generation that supplies the campus [29]. One hyperscale siting decision thus instantiated a vertical chain of separately registered local businesses: data hall, dedicated power plant, dedicated pipeline each in a different industry code. This behind-the-meter pattern is directly relevant to the NIPSCO GenCo framework discussed in Section 2.4.3, as it illustrates the industrial structure such arrangements are likely to take.

Water shows the same purpose-built logic, and provides the study's cleanest cross-regional pairing. In Quincy, Microsoft financed roughly $31 million of construction so that the city could open Washington's first industrial water reuse utility. This a municipally owned system of ten treatment facilities connected by 35 miles of pipe, operational in June 2021, treating data center cooling water and saving an estimated 138 to 390 million gallons of potable water per year, depending on the source of the estimate [30, 31]. The mechanism is the same one Loudoun Water arrived at as an incumbent (Section 3.2), executed instead as a new entity a continent away: where the cluster's water demand appears, either an existing utility adds the product or the product summons a new utility. The examples of utility engineering capacity following the same pull appear: when Quincy-area load growth outran the county utility's own bench, Grant PUD handed national engineering firms parallel contracts within a program projected at $80 million to build or rebuild substations, with Google's and Microsoft's demand named as the driver [32, 33].

3.4 Businesses Attracted by Proximity Benefits to Data Centers (Type C)

Type C businesses are not primarily selling to the data centers; they locate in a cluster because being near it is valuable. This is the most heterogeneous of the three categories, and interpreting it requires two cautions established in our coding rubric. First, some Type C entrants are themselves later-wave data center operators; their arrival is evidence that a cluster attracts more of its own industry rather than evidence of ancillary business formation, and we present them as such. Second, in the Columbia Basin cases, inexpensive hydropower is a confounding attractor as some firms drawn "to the cluster" were substantially drawn to the same cheap power that attracted the cluster in the first place.

The simplest proximity story is relocation into the cluster's name. In September 2015, DSI Tech, an IT services firm, moved its headquarters, around-the-clock network, and security operations center to Ashburn while announcing the move as an arrival in "Data Center Alley" (by that name), its CEO calling it "a new era for our company within a booming technology based region" [34]. The motive statement is the company's own, but it captures something real: by 2015 the cluster's designation had itself become a business address worth advertising.

The later-wave operators say most clearly why clusters compound. CloudHQ, a data-center developer founded in 2016 assembled a 97-acre campus in Ashburn, and its vice president explained the logic to a visiting U.S. senator in 2020: "There's an ecosystem that's been built, taxes, power, people, location—it's working" [35, 44]. Sabey Data Centers, a Seattle operator that helped build Quincy's cluster, crossed the country to open in Ashburn because, in its general manager's words, Loudoun is the "Super Bowl" of data-center markets — "anybody who's any type of competitor wants to be in Ashburn" [36, 45]. These are operator location decisions, not ancillary businesses; we include them because they articulate, in first-party language, the agglomeration force that every other Type C case rides. It also implies additional options can compound.

Proximity value can also arrive without a single local hire. In June 2020, Megaport, an Australian network-as-a-service provider, established a presence in Sabey's Quincy campus in the form of a cross-connect giving tenants there one-click access to what the companies described as 171 cloud on-ramps worldwide, including Amazon Web Services, Google Cloud, and Microsoft Azure [37]. No local jobs, payroll, or premises are documented, so this illustrates that some of what a cluster attracts is capability for the place rather than employment in it, and a farm town's racks gaining Ashburn-grade cloud connectivity is real economic infrastructure even though it never appears in payroll-level employment data.

The most consequential recent Type C pattern is the AI-compute wave arriving as a second-generation tenant of cloud-era infrastructure. In June 2025, Lambda, an AI cloud provider, launched NVIDIA B200 GPU clusters. This is an installation Cologix's announcement valued at "hundreds of millions" of dollars inside their COL4 facility in Columbus [38, 39]. The stated reason was not land or power alone but accumulated interconnection: a Columbus platform with more than 50 network and cloud providers, two cloud on-ramps, and an internet exchange, with Cologix's chief revenue officer calling Columbus "one of the fastest-growing digital corridors in the country" and the launch a delivery of "coastal-level AI infrastructure into the region" [38]. The claims all trace to the companies' joint announcement, but the structural point stands: each wave of infrastructure investment became the attractor for the next.

Two Quincy cases round out the range of what proximity can mean. Washington State's own technology agency, WaTech, has operated a disaster-recovery colocation site in Quincy since 2015, making the state government a tenant of the cluster its inexpensive public power helped attract [40]; a secondary site required an eastern-Washington location in any event, so the cluster explains the choice of Quincy specifically rather than the existence of the footprint, a distinction relevant to Northwest Indiana's own disaster-recovery positioning discussed in Section 4.4. And when Intuit exited its Quincy data center for the public cloud, H5 Data Centers bought the building and converted it into multi-tenant colocation marketed on sub-three-cent hydropower [41, 42]. A mature cluster's cast-off buildings had become an asset class: attraction to the cluster's physical remains rather than to its customers.

3.5 Summary Findings

Many unanticipated and hard to categorize cases manifested that is befitting a new and growing industry. One example is the introduction of new pricing tiers, particularly for utilities, directed at data centers. These do not per se distinguish expansions, but are business opportunities. In addition, a common concern is what will become of these campuses at the end of their lifecycle. We document many instances of refurbishments, and data center operators often describe how to build for upgrades as a means for improving the ability to recover the large initial investments. We also see nascent evidence of conversion, such occurred in the Columbia Basin's with crypto-to-AI. In November 2025, Bitfarms announced a fully funded $128 million agreement to gut a Moses Lake bitcoin mine and rebuild it as an 18-megawatt AI data center, its CEO projecting that the single converted site "could potentially produce more net operating income than we have ever generated with Bitcoin mining" [43]. The wider lesson of the category, taken up in Section 4, is that "near a data center" means different things to different businesses.

The results of this section are drawn from what has grown into a modern, mature data center cluster and often at the result of deliberate efforts by key agencies. For instance, Loudoun County demonstrates the cumulative result of decades of interconnection, power, permitting expertise, workforce, repeat construction, and vendor density. Loudoun County Economic Development reports more than 25 million square feet of operational data centers, more than 3,500 technology companies housed within them, more than 12,000 data-center ecosystem jobs, and over 13 years without a day lacking data-center construction. These case studies demonstrate that the economic opportunities associated with a data center site do not end with its initial configuration or operating cycle. NWI does not need to necessarily imitate these paths, and can potentially leverage its own unique strengths and advantages. The next section investigates the data center's role in the tech stack of other related industries.

4. Proximity-Driven Business: Data Centers in the Tech Stack

Beyond the businesses involved in constructing, operating, maintaining, and ultimately upgrading data center campuses, a separate question is which businesses actually benefit from locating near them. Businesses may value proximity for at least four distinct reasons: network latency and interconnection, access to customers and data, rapid vendor response and recurring contracts, or risk diversification. Each motive implies different proximity requirements and site-selection criteria. This lends itself to development in technologically inclined industries. This section reviews key industries that contribute to the tech stack, how they relate to data center proximity, and identifies pertinent Northwest Indiana regional assets. Sections 4.6 and 4.7 translate these proximity rationales into site-selection tests and potential strategic positions for Northwest Indiana's submarkets.

Proximity logic	What actually matters	Typical users	Site test
Latency proximity	Milliseconds or less matter to user experience or machine control	Edge inference, real-time manufacturing, market connectivity, interactive media	Measured route latency, jitter, packet loss, private interconnect
Vendor proximity	Travel and response time matter more than fiber distance	Remote hands, HVAC/electrical repair, security, spares, commissioning	24/7 access, service radius, workforce, parts inventory
Customer/data proximity	Compute should sit near the people, machines, or regulated data	Industrial IoT, health imaging, logistics, connected assets	Industrial customer density, data rights, fiber/5G, power
Resilience proximity	Connected enough to replicate; separated enough to avoid shared failure	Disaster recovery, backup, business continuity	Hazard separation, independent grid/fiber routes, RPO/RTO tests

4.1 Cloud-dependent startups

A hyperscale data center does not, by itself, create a startup ecosystem. Most software-as-a-service firms can reach public cloud regions from virtually anywhere and therefore have little reason to locate next to the physical building that houses the servers. The stronger proximity case applies to startups whose products depend on very low latency, direct interconnection, high-volume data movement, specialized test environments, or repeated access to nearby customers. Examples include firms developing real-time AI inference, industrial analytics, digital twins, robotics, streaming applications, cybersecurity tools, quantum-safe networking, and other data-intensive systems. For these businesses, however, the relevant question is not simply whether a data center is nearby, but whether the location provides the interconnection, testbed access, customers, and technical infrastructure their products actually require.

Northwest Indiana's strongest startup opportunity is therefore likely to be vertical and customer-led rather than a generic effort to attract software companies. The region's industrial base creates potential markets for industrial AI, equipment reliability and predictive maintenance, digital twins, and automation for steel and advanced manufacturing. Its ports, railroads, trucking activity, and intermodal facilities create applications in logistics optimization and connected freight. Large energy users and data-center campuses themselves can create demand for grid, cooling, water, and energy-management analytics, while health systems and local governments can provide additional applications for edge computing, cybersecurity, and data-intensive services. Cybersecurity and post-quantum migration may be an especially distinctive niche. Quantum Corridor's fiber connection between Chicago's 350 Cermak interconnection hub and Digital Crossroad in Hammond, together with demonstrations involving quantum key distribution and precision timing, provides NWI with an unusual platform for applied security, networking, and communications testbeds (Quantum Corridor, 2025–2026).

The assets that matter to these firms are correspondingly different from those of a conventional office-based startup program. Cloud-dependent startups may need flexible office and laboratory space, high-capacity carrier-neutral connectivity, affordable power for test equipment, direct cloud on-ramps and credits, access to university researchers and specialized mentors, and—most importantly—customers willing to test and purchase their products. This suggests that public and philanthropic resources should initially emphasize pilot procurement, testbed access, industry partnerships, and introductions to potential customers rather than speculative incubator real estate. A program built around steel producers, logistics firms, utilities, health systems, municipalities, and data-center operators could help startups move from demonstration to commercial adoption. The appropriate performance measures are therefore paid pilots, follow-on contracts, recurring revenue, and outside investment tied to demonstrated customers—not registrations, accelerator participation, or event attendance.

The economic-development implication is that data-center proximity is valuable only when it lowers a real technical or commercial barrier. NWI should therefore treat the data-center buildout as one component of a broader industrial technology ecosystem, using connectivity and compute infrastructure to deepen relationships between startups and the region's existing concentrations of manufacturing, logistics, energy, and institutional customers.

4.2 Managed Service Providers (MSPs)

Managed service providers (MSPs), managed security service providers, network integrators, backup and disaster-recovery firms, and remote-hands providers can benefit from locating near concentrations of enterprise customers and digital infrastructure. Their proximity requirement, however, is usually operational rather than latency-driven. What matters is the ability to reach customers and facilities quickly, maintain 24/7 service obligations, recruit technicians, access multiple carriers and colocation platforms, and convert those advantages into recurring contracts. For that reason, the relevant geography is generally a service radius rather than the hyperscale campus fence line. A firm may gain more from being thirty minutes from several industrial customers and a colocation facility than from occupying a site immediately adjacent to a hyperscale data center that purchases few local IT services.

Northwest Indiana has a plausible MSP opportunity because it combines digital infrastructure with a substantial base of industrial, health-care, logistics, and public-sector customers. Digital Crossroad and the broader Chicago-Hammond connectivity network provide an interconnection anchor, while the region's established enterprises create demand for hybrid-cloud management, network segmentation, backup and recovery, managed networking, cybersecurity, and operational-technology security. These services are particularly relevant for manufacturers whose production systems increasingly connect conventional information technology with plant-floor controls and sensors. Data-center growth can deepen that market by adding facilities requiring remote hands, network support, cybersecurity, equipment staging, and other recurring technical services, but the broader enterprise customer base is what makes the business model durable.

Lake County is the strongest initial node because it offers the greatest combination of customer density, digital infrastructure, workforce, and access to the Chicago market. Porter and LaPorte counties can support industrial clients, field-service teams, and technicians serving facilities farther east. In Newton and Pulaski counties, the opportunity is less likely to involve large standalone MSP headquarters and more likely to involve rural fiber support, field technicians, small managed-service operations, and extensions of firms headquartered elsewhere in the region. A regional model can therefore combine centralized technical capacity with distributed field service rather than attempting to reproduce the same technology cluster in every county.

MSPs generally need secure office or flex space, redundant carrier connectivity, reliable 24/7 power, direct or partner access to cloud and colocation platforms, laboratory and equipment-staging rooms, and a workforce pipeline in networking, systems administration, cybersecurity, and field service. The strategic objective is not simply to place MSPs near data centers, but to use the expanding digital-infrastructure market to strengthen a locally rooted technical-service sector capable of serving data centers, manufacturers, health systems, governments, and other enterprises throughout Northwest Indiana.

4.3 Disaster recovery and business continuity

Disaster recovery reverses the usual proximity argument. A secondary processing or storage site must remain connected well enough to replicate data and resume operations, but it should also be sufficiently separated from the primary environment that a single disruption does not disable both. The National Institute of Standards and Technology’s (NIST) current SP 800-53 Rev. 5 expresses this directly: its alternate-processing-site control calls for a site sufficiently separated from the primary site to reduce susceptibility to the same threats. NIST's contingency-planning guidance similarly emphasizes separation from common hazards, accessibility during area-wide disruptions, and recovery within established recovery-time objectives. The relevant threats can therefore include not only flood or severe weather but also common electric infrastructure, telecommunications failures, transportation disruptions, and other shared points of failure. For resilience applications, adjacency to the same data center campus may consequently be a disadvantage rather than an advantage, but they seek out areas with data centers in service of customers in separate areas.

Northwest Indiana's potential role is as a secondary location for Chicago-area enterprises that remains connected to the Chicago digital ecosystem while providing meaningful separation in site, jurisdiction, and infrastructure. The strongest value proposition is therefore not the closest possible backup site, but demonstrably independent and tested recovery capacity. A competitive site would combine diverse fiber paths, electric service that avoids unnecessary common points of failure with the primary location, suitable flood and hazard characteristics, secure 24/7 access, backup generation and fuel arrangements, and the ability to meet customers' specified recovery time and recovery point objectives. NIST likewise recommends redundant communications paths, elimination of single points of failure, backup power, and data replication as elements of robust contingency design. NWI's economic-development proposition should therefore be framed around measurable resilience rather than simple mileage from downtown Chicago.

Site evaluations for disaster recovery services begin with a shared-fate assessment that asks which failures could simultaneously affect the primary and proposed recovery locations. That assessment should trace electric supply and substations, backup generation and fuel, carrier routes and physical conduits, flood and other natural hazards, water dependencies, emergency access, cybersecurity controls, and staffing. Fiber diversity in particular should be evaluated by physical route rather than carrier count alone: purchasing service from two providers does not provide meaningful resilience if the supposedly redundant paths ultimately depend on the same physical bottleneck. This is consistent with NIST's emphasis on redundant communications paths and eliminating single points of failure. Data centers typically exist where these conditions have similar standards, and therefore are useful proxies of viable locations.

4.4 Edge computing operations

Edge computing places selected computing and storage resources closer to the people, machines, or data sources that use them. The economic value of proximity can come from lower latency, reduced movement of large volumes of data, faster local processing, or requirements to keep particular workloads close to an on-premises system. AWS Local Zones illustrate this model: AWS describes them as extensions of its cloud infrastructure into metropolitan areas for latency-sensitive applications, data-residency needs, AI and machine-learning inference, and hybrid-cloud applications. AWS currently operates two Local Zones in the Chicago market. The important geographic relationship is therefore between the edge node and the user or data-generating system, not necessarily between the edge node and the hyperscale data center. An industrial edge application, for example, may be most valuable inside or near a manufacturing plant even if its parent cloud infrastructure is hundreds of miles away.

This distinction favors a customer-led edge strategy for Northwest Indiana. The region's strongest potential applications are those in which local processing solves a specific operational problem: computer vision, equipment monitoring, predictive maintenance, and digital twins in steel and advanced manufacturing; port, rail, and fleet optimization in logistics; connected-vehicle systems; energy and grid optimization; health imaging and other data-intensive health applications; public-safety systems; and agricultural sensors and analytics. AWS identifies real-time inference and applications such as patient diagnostics among the types of workloads for which placing computing closer to where data are generated and consumed can be valuable. Lake, Porter, and LaPorte counties offer the strongest initial markets because of their concentrations of industrial and logistics users. In Newton and Pulaski counties, a more plausible opportunity is targeted deployment for agriculture, health care, rural connectivity, or other applications where a specific customer need and adequate network infrastructure can first be demonstrated.

Edge facilities are generally much smaller than hyperscale campuses, but their location can be more sensitive to infrastructure. Attractive sites combine proximity to the relevant customers with network junctions, diverse fiber or wireless connectivity, low-latency paths to the parent cloud, reliable but comparatively modest power capacity, physical security, remote monitoring, and access to technicians for field service. AWS, for example, supports private Direct Connect connections into Local Zones, illustrating how edge infrastructure can combine geographically local processing with private connectivity to a broader cloud environment. The appropriate configuration will vary by application: a manufacturing workload may require a node on or very near the plant, while an edge facility serving several customers may be better located at a regional network junction.

4.5 Tech-enabled manufacturing and logistics companies

Manufacturers and logistics firms can benefit from a growing data center cluster, but generally for different reasons than software, edge-computing, or interconnection-dependent businesses. Data center investment can improve regional fiber and cloud connectivity, deepen the pool of electrical, mechanical, cooling, and cybersecurity expertise, and create a large recurring market for equipment, components, replacement parts, and specialized logistics. These benefits can make a region more attractive to technology-intensive manufacturers and distributors. For most firms in this category, proximity to a data center is that the opportunity comes from being part of the surrounding industrial market, not from occupying property immediately beside the hyperscale campus.

Northwest Indiana is well positioned to compete for this activity because many of the assets required by data center suppliers overlap with the region's existing industrial strengths. Burns Harbor is a particularly strong platform. Ports of Indiana reports that the port hosts more than 30 companies, including 15 steel-related companies and three steel mills, and provides heavy-industrial zoning, Foreign-Trade Zone status, 24-hour security, utility access, and development space for warehousing, distribution, and manufacturing. Its transportation network includes direct highway access, connections to all Class I railroads, and Great Lakes and inland-waterway access. Those capabilities create potential locations for equipment staging, specialty packaging and crating, heavy-haul logistics, spare-parts inventories, component distribution, electrical-panel and controls integration, thermal-management equipment and services, and industrial-edge applications.

Recent industrial leasing patterns suggest that this supplier opportunity can be economically meaningful. Cushman & Wakefield (2026) examined more than 388 million square feet of leasing across six major U.S. data-center markets from 2022 through 2025 and classified 10.4% of new industrial leasing as related to the data center ecosystem. The share increased to 14.4% in 2025, when data center-related leasing grew 44% from the previous year. The firms occupying this space are not limited to highly specialized technology companies. Construction firms, electrical contractors, and electrical and HVAC equipment suppliers are among the most active firms with direct relationships to data center development, and related industrial transactions are concentrated substantially in buildings of roughly 30,000 to 400,000 square feet.

The Chicago results are particularly relevant to Northwest Indiana. In the Cushman & Wakefield (2026) study, 47% of Chicago's data-center-related tenants were classified as "third-degree" users (e.g. firms supplying materials, infrastructure, or services that have applications well beyond data centers). This is an important distinction for regional strategy in Northwest Indiana, because it need not develop an entirely new industrial base specialized exclusively in data centers. Its existing capabilities in steel, fabrication, electrical equipment, construction materials, transportation, warehousing, and industrial services may be adaptable to an expanding data-center supply chain while continuing to serve traditional customers.

4.6 Site-selection criteria by business type

This screen translates tech sectors of Sections 4.1-4.5 into selected site-selection determinants.

Business type	Site selection determinants
Cloud/industrial startup	Talent, customers, direct connectivity, flexible lab/office, pilot procurement, cloud access
MSP / cyber / remote hands	Carrier options, client density, 24/7 access, service radius, secure staging, technician pipeline
Disaster recovery	Independent hazards and utilities, diverse physical routes, RPO/RTO, backup power/fuel, secure access

Business type	Site selection determinants
Edge node	Measured latency/jitter, user/data proximity, fiber/5G, modest scalable power, security, remote operations
Manufacturing/logistics	Port/rail/highway, industrial land, workforce, staging, power, heavy haul, customer and supplier access
OEM service depot	Installed equipment base, multi-campus demand, SLA radius, trained technicians, parts inventory, airport/highway access

4.7 Strategic positioning by NWI submarket

The submarket framework applies those tests to documented NWI assets and constraints. It proposes complementary roles rather than county forecasts; every proposition still requires project-specific utility, fiber, customer, workforce, and supplier-demand verification.

Submarket	Recommended position	Priority business proposition
Lake / Hammond	Interconnection and operating-services district	Carrier-neutral connectivity, MSP/cyber, colocation, quantum-safe services, edge, remote hands, DR gateway
Porter / Burns Harbor	Mission-critical logistics and industrial-edge district	Secure warehouses, spares, heavy haul, packaging/crating, equipment service, industrial AI/logistics technology
LaPorte / Radius corridor	Cloud-campus workforce and controls district	Critical facilities, controls/HVAC, inventory/asset services, training, supplier qualification, professional services
Jasper / Wheatfield area	Power-infrastructure and accountable rural-host district	Grid/civil work, maintenance, water/site support, rural construction access; verify recurring digital-service case
Newton / Pulaski	Fiber-enabled rural services and edge corridor	Fiber, rural MSP, ag-tech/health/logistics edge, maintenance; use utility/site feasibility gates
Starke	Zoning-first workforce and supplier node	SCILL-supported trades/automation, regional project access, smaller digital uses after standards and feasibility

The framework omits claims that proximity alone will create a broad technology or equipment-manufacturing cluster and retains only propositions tied to modeled demand, a documented asset, or a testable customer use case.

5. Conclusions

This report examines the ancillary-business implications of data-center development in Northwest Indiana. The findings suggest that the relevant regional opportunity is best understood as a portfolio of activities rather than as a single data-center industry. This portfolio includes temporary construction demand, recurring operations and maintenance, infrastructure and logistics services, professional and technical services, workforce-development functions, and a narrower set of businesses for which proximity may affect location decisions.

The analysis also distinguishes potential demand from realized regional capture. A data center may purchase substantial amounts of construction, equipment, electricity, maintenance, and technical services without obtaining those inputs from firms located in the surrounding region. The presence of expenditure therefore does not establish that a corresponding local industry will develop. Regional participation depends on existing capabilities, procurement requirements, vendor qualification, workforce availability, competitive conditions, and the geographic organization of suppliers. Local labor agreements have also become quite standard.

Among the activities considered, maintenance, commissioning, controls, equipment repair, fiber, carrier services, specialized logistics, remote hands, security, engineering, and systems integration appear to have comparatively direct connections to data-center construction or operation. Construction may provide a substantial but temporary market, whereas operating and maintenance activities offer the possibility of recurring demand. In contrast, the purchase of servers, GPUs, switchgear, and other specialized equipment provides little basis, in the absence of additional evidence, for expecting the associated manufacturing to locate near a campus.

The qualitative cases indicate that ancillary-business development may occur through the adaptation of existing firms, the entry of specialized suppliers after demand is demonstrated, or the location of businesses for which proximity provides an identifiable operating benefit. They do not support a general conclusion that data centers independently produce broad technology clusters. The relevance of proximity is conditional on particular mechanisms, including latency, interconnection, customer access, response-time requirements, repeat contracting, and risk diversification.

The Northwest Indiana submarket analysis similarly indicates variation within the region. Lake County and Hammond have assets relevant to interconnection and operating services; Porter County and Burns Harbor have logistics and industrial capabilities; LaPorte County combines an announced project with workforce and controls-related institutions; Jasper County presents a distinct set of host-site infrastructure issues; and the potential roles of Newton, Pulaski, and Starke counties appear more conditional on fiber, workforce, utility feasibility, and identifiable customer applications. These classifications are intended as testable propositions, not predictions of county-level effects.

Several limitations qualify the findings. The input-output estimates describe modeled regional relationships rather than observed project outcomes. The case studies rely primarily on documentary and business-level evidence and do not provide causal estimates of the effect of data-center entry. Announced projects, planned expenditures, and stated employment figures may also differ from realized activity. In addition, this report does not measure environmental, fiscal, utility-rate, land-use, or other social costs and therefore should not be interpreted as a benefit-cost analysis.

Further evaluation would benefit from project-level evidence on procurement, supplier locations, awarded contracts, worker residence, wages, establishment entry and expansion, and the persistence of vendor relationships after construction. These measures would permit researchers to assess whether data-center investment is associated with durable changes in the regional business base and to distinguish those changes from activity that would have occurred in the absence of the projects.

Sources and Analytical Notes

Primary quantitative sources

1. JLL. “2026 Global Data Center Market Outlook: Navigating AI Demand, Power Constraints and Global Opportunities in 2026.” January 5, 2026. https://www.jll.com/en-us/insights/market-outlook/data-center-outlook
2. Roychoudhury, Saurav and Justin M. Ross. 2026. “Modeling Economic Impacts of a Hyperscale Data Center Campus.” SSRN Working Paper 7087540. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7087540
3. U.S. Bureau of Economic Analysis. RIMS II May 2026 release for the seven-county Northwest Indiana region; 2017 benchmark I-O and 2024 regional data. https://apps.bea.gov/regional/rims/rimsii/

3. Cushman & Wakefield. From Megawatts to Multipliers: The Data Center Effect on Industrial Demand, Jobs and Local Revenue. June 2026. Client-provided PDF.

4. NWI_NAICS_SOC_Vocational_IvyTech_Training_Crosswalk.xlsx and related seven-county strategy/evidence files. August 2026. Client-provided.

Current project, infrastructure, workforce, and benchmark sources

5. Indiana Economic Development Corporation (IEDC). 2024. Planned $1 billion Microsoft data center in La Porte; 245,000 square feet, 489 acres, up to 200 jobs by 2032. https://iedc.in.gov/events/news/details/2024/06/04/gov-holcomb-announces-plans-for-new-1b-microsoft-data-center-in-northwest-indiana

6. City of La Porte. Updated Microsoft tax agreement; full property taxes and 15% school allocation for 20 years. March 3, 2026. https://www.cityoflaporte.com/news/microsoft-dermody-eupdate-tax-agreement-to-include-historic-funding-for-la-porte-schools

7. Amazon. 2026. "Amazon to Invest $15 Billion in Indiana for New Data Centers." Corporate announcement; planned investment, capacity, and job values are not independently audited. https://www.aboutamazon.com/news/company-news/amazon-15-billion-indiana-data-centers

8. NIPSCO. GenCo framework for large-load service and existing-customer protection. Utility statements and forecasts remain implementation-dependent. https://www.nipsco.com/our-company/news-room/news-article/powering-indiana-s-future--how-nipsco--genco--and-amazon-keep-energy-reliable-and-affordable

9. Ivy Tech Community College and Microsoft. Data Center Academy MOU. June 17, 2026. https://www.ivytech.edu/about-ivy-tech/newsroom/news/all-locations/2026/microsoft-mou/

10. WorkOne Northwest Indiana / NWI Workforce Board. Seven-county service geography and governance. https://www.in.gov/dwd/workonenwi/nwi-workforce-board/

11. NYI. Digital Crossroad-350 Cermak-Oak Brook connectivity ring. February 2021; facility specifications are dated. https://www.nyi.net/resources/news/digital-crossroad-partners-with-nyi/

12. Quantum Corridor. Chicago-Hammond live commercial network and public milestones. https://www.quantumcorridor.com/news-and-events-posts/quantum-corridor-toshiba-demonstrate-first-cross-state-quantum-key-distribution-over-live-commercial-metro-fiber-network

13. Ports of Indiana. Burns Harbor industrial, rail, road, water, security, and FTZ capabilities. Accessed August 14, 2026. https://www.portsofindiana.com/locations/burns-harbor/

14. Loudoun County Economic Development. Data Centers at a Glance. Mature-cluster benchmark. https://loudounpossible.com/business-sector/data-centers

15. NIST SP 800-34 Rev. 1 and SP 800-53 Rev. 5. Alternate-site separation and contingency planning. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-34r1.pdf

16. AWS. Chicago Local Zone and low-latency edge use cases. https://aws.amazon.com/about-aws/whats-new/2021/09/general-availability-aws-local-zones/

17. U.S. Census Bureau. 2022 NAICS classification standard. https://www.census.gov/naics/

18. U.S. Bureau of Labor Statistics. SOC occupation structure. https://www.bls.gov/oes/current/oes_stru.htm

Sources for Qualitative Research Sections 3.2–3.4

1. Igel Construction. “Data Centers / Mission Critical.” https://www.buildwithigel.com/markets/data-centers-mission-critical/ (accessed Aug. 12, 2026).

2. Igel Construction. “History.” https://www.buildwithigel.com/about-us/history/ (accessed Aug. 12, 2026).

3. Mt. Carmel Stabilization Group. “Meta Data Centers New Albany.” https://mtcsg.com/project/meta-data-centers-new-albany/ (accessed Aug. 12, 2026).

4. Valley Electric. “Columbia Data Center” (project page). https://www.velectric.com/project/columbia-data-center/ (accessed Aug. 12, 2026). Project completion year per the company’s own portfolio.

5. Valley Electric. “About.” https://www.velectric.com/about/ (accessed Aug. 12, 2026).

6. ALL Crane Rental Corp. “Buckeye Bytes: Tech and Warehouse Building Boom in Greater Columbus.” March 6, 2020. https://www.allcrane.com/resources/case-study/article/buckeye-bytes-tech-and-warehouse-building-boom-in-greater-columbus-a-strong-national-economy-encourages-construction-opportunities (accessed Aug. 12, 2026).

7. ALL Crane Rental Corp. “Region Report.” Feb. 17, 2025. https://www.allcrane.com/resources/news/article/blog/2025/02/17/region-report (accessed Aug. 12, 2026).

8. Loudoun Water. “Reclaimed Water Program.” https://www.loudounwater.org/commercial-customers/reclaimed-water-program (accessed Aug. 12, 2026).

9. Data Center Knowledge. “In Ashburn, Recycled Water to Keep Data Centers Cool.” Sept. 9, 2013. https://www.datacenterknowledge.com/business/in-ashburn-recycled-water-to-keep-data-centers-cool (accessed Aug. 12, 2026).

10. N C Power Systems. “Data Center Generator Technician” (job posting, Caterpillar Dealer Tech Jobs, East Wenatchee, posted ca. Feb. 2026). Original URL (https://dealertechjobs.caterpillar.com/job/1467/data_center_generator_technician) now offline; text reconstructed from search-index captures.

11. N C Power Systems. “Cat Electric Power.” https://cat.ncpowersystems.com/electric-power (accessed Aug. 12, 2026).

12. Microsoft Datacenter Academy blog. “Big Bend Community College cultivates a hometown tech workforce.” https://local.microsoft.com/blog/big-bend-community-college-cultivates-a-hometown-tech-workforce/ (accessed Aug. 12, 2026).

13. North Central Educational Service District. “Quincy Data Center Tech Career Launch.” https://www.ncesd.org/news/quincy-data-center-tech-career-launch/ (accessed Aug. 12, 2026).

14. Columbus State Community College. “Columbus State and AWS Data Center Technician Program.” Nov. 19, 2018. https://www.cscc.edu/about/news/2018/AWSDataCenterTechnician.shtml (accessed Aug. 12, 2026).

15. Columbus State Community College. “Data Center Technician” (program page). https://www.cscc.edu/academics/departments/computer-it/data-center-technician.shtml (accessed Aug. 12, 2026).

16. OilWorks LLC. “Industries We Serve in Columbus, OH.” https://oilworksllc.com/industries-we-serve/ (accessed Aug. 12, 2026).

17. Cinco Technologies. “Data Center Decommissioning in Columbus, Ohio.” https://www.cincoer.com/recycling-resources/data-center-decommissioning-columbus-ohio (accessed Aug. 12, 2026).

18. VECA. “Wenatchee” (location page). https://www.veca.com/location/wenatchee (accessed Aug. 12, 2026).

19. VECA. “Projects.” https://www.veca.com/projects (accessed Aug. 12, 2026).

20. Holmes Electric. “Our History.” https://www.hecowa.com/home/our-history/ (accessed Aug. 12, 2026).

21. Holmes Electric. “Microsoft – Quincy” (portfolio page). https://www.hecowa.com/portfolio/microsoft-quincy (accessed Aug. 12, 2026).

22. Levelset. “Quincy, Wash. Microsoft Data Center Construction Claims Top $70 Million.” Sept. 18, 2020 (updated Dec. 10, 2021). https://www.levelset.com/news/quincy-wash-microsoft-data-center-construction-claims-70-million/ (accessed Aug. 12, 2026).

23. tEDmag. “City Electric Supply Opens New Branch in Dulles.” Jan. 9, 2019. https://tedmag.com/city-electric-supply-opens-new-branch-in-dulles/ (accessed Aug. 12, 2026).

24. GlobeNewswire. “SummitIG Completes 170-Mile Dark Fiber Network Build in Virginia.” Nov. 18, 2014. https://www.globenewswire.com/news-release/2014/11/18/950913/0/en/SummitIG-Completes-170-Mile-Dark-

Fiber-Network-Build-in-Virginia.html (accessed Aug. 12, 2026). Network-mileage growth per summitig.com (captured Aug. 5, 2026).

25. Harbor Network Solutions. “Harbor Peak Breaks Ground on a New Fiber Conduit System Connecting Data Centers Throughout Ashburn, Virginia.” May 2, 2023. https://harbornetworksolutions.com/harbor-peak-breaks-ground-on-a-new-fiber-conduit-system-connecting-data-centers-throughout-ashburn-virginia/ (accessed Aug. 12, 2026).
26. The Daily Record. “Harbor Link joins ‘Harbor Peak’ digital infrastructure venture.” Jan. 17, 2023. https://thedailyrecord.com/2023/01/17/harbor-link-joins-harbor-peak-digital-infrastructure-venture/ (accessed Aug. 12, 2026).
27. Williams Companies. “Socrates Power Solution Facilities.” https://www.williams.com/expansion-project/socrates-power-solution-facilities/ (accessed Aug. 12, 2026).
28. Pipeline and Gas Journal. “OPSB Greenlights Williams-Backed Power Plant for Meta Data Center in Ohio.” June 2025. https://pgjonline.com/news/2025/june/opsb-greenlights-williams-backed-power-plant-for-meta-data-center-in-ohio (accessed Aug. 12, 2026).
29. POWER Magazine. “New 200-MW Gas-Fired Plant in Ohio Will Power Meta Data Center.” June 10, 2025. https://www.powermag.com/new-200-mw-gas-fired-plant-in-ohio-will-power-meta-data-center/ (accessed Aug. 12, 2026). Pipeline scope per Will-Power Pipeline OH, LLC, OPSB Letter of Notification (filed June 23, 2025).
30. U.S. Environmental Protection Agency. “Water Reuse Case Study: Quincy, Washington.” https://www.epa.gov/waterreuse/water-reuse-case-study-quincy-washington (accessed Aug. 12, 2026).
31. Microsoft Local blog. “Partnering with the City of Quincy to open Washington’s first industrial water reuse center.” https://local.microsoft.com/blog/partnering-with-the-city-of-quincy-to-open-washingtons-first-industrial-water-reuse-center/ (accessed Aug. 12, 2026).
32. Stanley Consultants. “Grant County 10 Sub-Stations and Transmission Line.” https://www.stanleyconsultants.com/solutions/engineering-design/project/grant-county-10-sub-stations-and-transmission-line (accessed Aug. 12, 2026).
33. Grant County PUD. FYI Packet, Contract Completion Report, Nov. 28, 2023. https://www.grantpud.org/block/documents/655cfda67f49d-2023-11-28-fyi-packet.pdf (accessed Aug. 12, 2026).
34. DSI Tech. “DSI Moves Headquarters to Larger Facility in ‘Data Center Alley’.” Sept. 2015. https://dsitech.com/dsi-moves-headquarters-data-center-alley-ashburn-virginia/ (accessed Aug. 12, 2026).
35. Loudoun Now. “Warner Hears of Data Centers’ Local Impact.” Feb. 28, 2020. https://loudounnow.com/2020/02/28/warner-hears-of-data-centers-local-impact (accessed Aug. 12, 2026).
36. Loudoun Now. “Amid Pandemic, Data Center Alley Going Strong.” Sept. 18, 2020. https://loudounnow.com/2020/09/18/amid-pandemic-data-center-alley-going-strong (accessed Aug. 12, 2026).
37. Megaport. “Sabey Data Centers Partners with Megaport” (media release). June 30, 2020. https://www.megaport.com/media-releases/2020-07-01.pdf (accessed Aug. 12, 2026).
38. Lambda. “Cologix and Lambda Launch First NVIDIA HGX B200-Accelerated AI Clusters in Columbus at COL4.” June 3, 2025. https://lambda.ai/blog/cologix-and-lambda-b200-col4 (accessed Aug. 12, 2026).
39. Cologix. “Columbus Cologix data center to house Lambda’s AI computing installation worth ‘hundreds of millions’.” https://cologix.com/news/columbus-cologix-data-center-to-house-lambdas-ai-computing-installation-worth-hundreds-of-millions/ (accessed Aug. 12, 2026).
40. Washington Technology Solutions (WaTech). “Colocation Quincy.” https://watech.wa.gov/services/Colocation-Quincy (accessed Aug. 12, 2026; page has since been consolidated into WaTech’s general services pages).

Program dates corroborated by WaTech, Data Center Onboarding Guide v2.0, Nov. 6, 2015, https://watech.wa.gov/sites/default/files/2023-11/colocation-data-ctr-onboarding-guide.pdf.

41. Data Center Knowledge. "Intuit Sells Quincy Data Center to H5 at a Loss as It Moves to AWS." July 2018. https://www.datacenterknowledge.com/h5-data-centers/intuit-sells-quincy-data-center-h5-loss-it-moves-aws (accessed Aug. 12, 2026).
42. H5 Data Centers. "Quincy Colocation." https://h5datacenters.com/quincy-colocation.html (accessed Aug. 12, 2026).
43. Bitfarms Ltd. "Bitfarms Announces Plans for Conversion of Washington Site to HPC/AI Workloads" (press release). Nov. 13, 2025. https://investor.bitfarms.com/news-releases/news-release-details/bitfarms-announces-plans-conversion-washington-site-hpcai (accessed Aug. 12, 2026).
44. CloudHQ. "About." https://cloudhq.com/about/ (accessed Aug. 12, 2026).
45. Sabey Data Centers. "SDC Ashburn." https://sabeydatacenters.com/locations/ashburn-data-center (accessed Aug. 12, 2026).

Technical Appendix A: Methodology for AI-Powered Qualitative Research Assistant

This appendix describes the process used to identify and classify the businesses that appear around large data-center campuses discussed in section 3. When very large computing campuses are built in a region, what new kinds of business activity appear alongside them? Answering that question with public records alone is difficult, because government statistics report employment and payroll by industry and county but never say *why* a firm is in a place. A regional total can rise for many reasons at once. So instead of inferring cause from totals, this study identifies individual businesses one at a time, gathers the written record for each one, and classifies each business according to the relationship it has to the data centers, using the firm's own words wherever possible.

A1. Overview of the Process

This study approach is ordinarily called qualitative case-study research, and it is normally limited by how many businesses a single research team can investigate. This study is designed to run that same qualitative process at a much larger scale by combining a written protocol with automated research assistants.

The process has seven stages.

Stage one: build the industry frame. Before any searching begins, the study fixes a list of industries in which data-center-related activity could plausibly appear. Each industry is identified by its official government industry NAICS code, and each is placed in one of four categories according to the kind of relationship it would have with a data center. The list is fixed in advance and is not expanded or narrowed once searching starts, so that a region cannot appear more connected simply because more industries were examined there. Section A2 provides the full details of the industry frame.

Stage two: define the regions and the search geography. The study covers three regions with very different histories and sizes. For each region, a core search county area is identified for most activities, and the wider metropolitan area used for a smaller set of industries. In addition, the starting year of the region's data center era is selected. These choices are locked before discovery begins and are documented with a government source. Section A3 describes the time and region scope.

Stage three: discovery. For each industry code in each region, a standard set of search queries is run: the industry's own terms combined with local place names, a directory-style query, and a query that pairs the industry with the phrase "data center." In addition, each region gets a one-time sweep of local institutional sources — economic development organizations, financing authorities, chambers of commerce, local newspapers, industry trade press, and government regulatory filings — which is harvested for every industry code at once. Every business that surfaces is entered in a register with its name, its candidate industry code, its locality, the query that found it, and the web address where it was found. Every query is written to its region's search log whether or not it produced anything.

Stage four: evidence gathering. Each registered business then gets its own evidence file. The research assistant works through a fixed sequence of sources: the company's own website; a search pairing the company name with the phrase "data center"; a search pairing the company name with the names of the local data-center operators; local news outlets; press releases; and finally the Internet Archive's snapshot history of the company's own website (via the Wayback Machine) which is used to bracket the date when data-center language first appeared on the site by recording the last archived version of the page without that language and the first version with it. Everything relevant is copied into the evidence file as a direct quotation, with the web address, the date the page was read, and the date the page was published or archived. If nothing is found after the full sequence, the file records that fact explicitly. The evidence file contains only source material and citations; it does not contain the researcher's conclusion in the section that a second coder is later shown.

Stage five: coding. Each business is then assigned one classification from a fixed set, together with a rating of how strong the underlying evidence is and the earliest date at which the connection can be documented. The classification scheme, the evidence ratings, and the twelve written decision rules that resolve recurring hard cases are given in Section 4. Where the evidence does not fit the rubric, the business is marked for review rather than forced into a category, and only the principal researcher may add a new rule.

Stage six: review. A first pass proposes a classification for each business from its evidence file. A second, more capable model then re-reads every evidence file, applies the written rules, and produces the final code; this pass systematically corrects whole groups of records at once when it finds that a rule was being applied inconsistently. Cases that remain genuinely ambiguous are listed with both readings and go to the principal researcher, whose rulings are then written back into the rubric as numbered decision rules so that the same question is answered the same way across regions. Separately, blinded packets are prepared for a second coder: they contain the evidence files with every conclusion removed, along with the rubric and a blank coding form, so that agreement between coders can be measured on the evidence alone.

Stage seven: assembly and reporting. All coded records are loaded into a single workbook with one row per business. Summary tables count businesses by industry and by classification, and a second sheet arranges first-evidence dates by year so that the appearance of connected businesses can be compared with the construction history of the campuses. Percentages are always computed on what the study calls the found frame: the set of businesses actually discovered by this protocol, after removing the data-center operators themselves and after removing businesses whose claimed local presence could not be verified. Section 5 describes how the coding is checked by independent second coders, and Section 6 reports the counts that this process produced.

Three habits run through all seven stages and are what make the work auditable. First, every query is logged, so the search history can be replayed. Second, every factual claim is a quotation with a web address and an access date, so no assertion rests on the memory of the person or system that made it. Third, every classification cites the specific lines of the evidence file that support it, so a reader can move from a number in a summary table to the sentence on a company website that produced it.

A2. The Industry Frame

The frame is a list of industries that data centers could plausibly impact to make a deep internet search worthwhile. The study uses the North American Industry Classification System, the official system that the United States, Canada, and Mexico use to sort businesses into industries. Each industry has a numeric code, and longer codes describe narrower industries. The industry frame for this study contains 88 such identifiers, drawn from three sources: data from the RIMS II BEA study described in sections 2 and 3, a broad search of literature on the possible channels through which a data center could affect a local economy, and a review of published case studies of data-center regions.

The four categories

Every identifier is assigned to one of four categories. The categories describe *how* a business would come to be connected to a data center, and that distinction matters because the four kinds of connection have different geographic reach, different timing, and different strength of evidence.

Category 1 — Recurring operations and infrastructure. These are industries that a data center needs continuously, for as long as it runs. Electric transmission and grid control (221121) is the clearest example: a campus that draws hundreds of megawatts requires substations, transformers, and ongoing grid management, and that demand does not end when construction ends. The same logic covers the utility that supplies water (221310), the utility that treats the water afterward (221320), the fiber-optic carriers that connect the buildings to the internet backbone (517111), the firms that guard the site (561612), the firms that clean it (561720), and the firms that repair the servers and network equipment inside it (811210). Because the demand recurs, these industries can support a permanent local establishment, and because the work is physical and time-sensitive, the establishment usually has to be nearby.

Category 2 — Construction, site development, and professional services. These industries serve the campus while it is being built, and again each time it expands. A data-center general contractor (236220) manages the building; excavation and grading firms (238910) prepare the site; electrical contractors (238210) install switchgear, distribution, and internal cabling; mechanical contractors (238220) install chillers, cooling loops, and fire suppression; power-line and substation builders (237130) connect the campus to the grid. The category also holds the professional services that surround construction: surveying (541370), engineering design (541330), testing and commissioning before the building is energized (541380), environmental review (541620), and private building inspection (541350). The distinguishing feature is that the work comes in waves tied to construction phases rather than continuously, and that a large share of it can be performed by firms based elsewhere that send crews in.

Category 3 — Equipment manufacturing and regional distribution. These industries supply the physical goods a campus consumes. The category spans a wide range of how local the supply can realistically be. Ready-mix concrete (327320) is at one extreme: concrete has a short delivery radius, so it is almost necessarily supplied locally, which makes it one of the most inherently local links in the chain. Structural-steel fabrication (332312), sheet-metal and duct fabrication (332322), and generator service centers (333611 and 335312) are intermediate — the manufacturing may happen far away, but service depots and fabrication shops can localize. Server and networking-equipment manufacturing (334111, 334112, 334210) is at the other extreme: a single campus buys this equipment on a global market, and only a very large cluster could plausibly attract assembly or configuration work. The category also holds the wholesale branches that keep contractors supplied with wire, controls, cooling parts, and replacement electronics (423610, 423690, 423730, 423830).

Category 4 — Potential proximity and digital-agglomeration beneficiaries. These industries are not suppliers to data centers. They are businesses that might choose to locate near data centers because being nearby is itself valuable — because of dense fiber routes, low network delay, direct connections into cloud providers, cheap and abundant power, or a concentrated pool of technical workers. Additional colocation, edge-computing, and content-delivery facilities (518210) are the central case: the fiber, the power infrastructure, the experienced contractors, and the familiar permitting environment make the next facility easier to build there than elsewhere. Software publishers (513210), custom software and analytics developers (541511), streaming and online-media firms (516210), payment processors (522320), and telecommunications resellers that resell capacity and connect other businesses to cloud services (517121) are all in this category. Two features set Category 4 apart from the first three. First, the relevant geography is wider during the search in stage 2: the benefit of a cloud region or a fiber hub extends across a metropolitan area, not just a county, so these codes are searched at the metropolitan level.

Two codes deserve special mention because they sit across a boundary. Code 541519, disaster-recovery and specialized computer services, is a recurring operational service and also a form of clustering, and it carries both category assignments in the master list; it is counted under Category 1 in the tables below. Code 518210 is the code under which the data-center operators themselves would fall; the coding rubric handles them separately, as Section 4 explains.

The full frame

Categories are shown as 1, 2, 3, and 4 in the order described above. "Current" in the status column means the identifier is a current six-digit industry in the 2022 revision of the classification system. Identifiers that are not current are searched through the current equivalent shown in the second table that follows.

Code	Industry	Category	Status
22111	Electric-power generation (five-digit group)	1	Current, five-digit group
221112	Dedicated fossil-fuel electric generation	1	Current
221114	Solar electric-power generation	1	Current
221115	Wind electric-power generation	1	Current
221121	Electric transmission and grid-control operators	1	Current
221122	Electric-distribution utilities	1	Current
221210	Natural-gas distribution utilities	1	Current
221310	Water-system operators	1	Current
221320	Sewage and wastewater facilities	1	Current

221330	District-energy and waste-heat network operators	1	Current
423930	Recyclable-material wholesalers (equipment disposition)	1	Current
424720	Backup-generator fuel distributors	1	Current
486210	Natural-gas pipeline operators	1	Current
517110	Wired telecommunications carriers (older version of the code)	1	Superseded by 517111
517111	Fiber-optic and wired telecommunications carriers	1	Current
532490	Temporary generator and industrial cooling rental firms	1	Current
541512	Computer-system integrators	1	Current
541513	Outsourced computer-facility operators	1	Current
541519	Disaster-recovery and specialized computer services	1 (also 4)	Current
561210	Integrated facilities-support contractors	1	Current
561612	Security-guard firms	1	Current
561621	Alarm, access-control, and surveillance firms	1	Current
561720	Janitorial and specialized cleaning firms	1	Current
562111	Solid-waste collection	1	Current
562920	Materials recovery and equipment disposition	1	Current
611210	Community colleges (data-center training programs)	1	Current
811210	Server, network, and electronic-equipment repair	1	Current
811300	Commercial and industrial equipment repair (aggregate)	1	Bureau of Economic Analysis aggregate
236220	Data-center general contractors	2	Current
237110	Water- and sewer-line contractors	2	Current
237120	Oil and gas pipeline construction contractors	2	Current
237130	Power-line, fiber, and substation construction firms	2	Current
237990	Other heavy and civil engineering construction	2	Current
238110	Concrete-foundation contractors	2	Current

238120	Structural-steel and precast installers	2	Current
238210	Electrical and data-wiring contractors	2	Current
238220	Heating, cooling, plumbing, and fire-suppression contractors	2	Current
238290	Other building-equipment contractors	2	Current
238390	Other building-finishing contractors	2	Current
238910	Excavation, grading, and site-preparation firms	2	Current
484000	Truck transportation (aggregate)	2	Bureau of Economic Analysis aggregate
484110	Local trucking and construction logistics	2	Current
493110	Warehouses and project-staging facilities	2	Current
532412	Heavy-equipment and crane rental firms	2	Current
541300	Architectural and engineering services (aggregate)	2	Bureau of Economic Analysis aggregate
541310	Architectural services	2	Current
541330	Engineering services	2	Current
541350	Private building-inspection services	2	Current
541370	Surveying and mapping firms	2	Current
541380	Testing, calibration, and commissioning firms	2	Current
541620	Environmental consultants	2	Current
561320	Temporary-help firms	2	Current
562991	Portable-sanitation providers	2	Current
327320	Ready-mix concrete suppliers	3	Current
327390	Other concrete-product manufacturers	3	Current
332312	Structural-steel fabricators	3	Current
332322	Sheet-metal and ductwork fabricators	3	Current
333415	Industrial cooling-equipment manufacturers	3	Current
333611	Turbine and generator-set manufacturers and service centers	3	Current
334111	Computer and server manufacturers	3	Current
334112	Storage-device manufacturers	3	Current
334210	Telephone and networking-equipment manufacturers	3	Current
335311	Transformer manufacturers	3	Current
335312	Motor and generator manufacturers and service centers	3	Current

335313	Switchgear and switchboard manufacturers	3	Current
335910	Battery manufacturers	3	Current
335921	Fiber-optic cable manufacturers	3	Current
335999	Other electrical-equipment manufacturers (power conversion)	3	Current
420000	Wholesale trade (aggregate)	3	Bureau of Economic Analysis aggregate
423430	Computer and equipment wholesalers	3	Current
423610	Electrical-equipment wholesalers	3	Current
423690	Electronic-parts wholesalers	3	Current
423730	Heating and cooling equipment wholesalers	3	Current
423830	Industrial-machinery wholesalers	3	Current
424710	Petroleum or renewable-fuel bulk terminals	3	Current
449210	Electronics and appliance retailers	3	Current; retained for accounting only
511200	Software publishers (aggregate)	4	Bureau of Economic Analysis aggregate
512191	Video postproduction, animation, and rendering firms	4	Current
513210	Software publishers and hosted-software firms	4	Current
516210	Streaming, social-network, and online-media firms	4	Current
517121	Telecommunications resellers and network-service providers	4	Current
517410	Satellite telecommunications and network-operations hubs	4	Current
518210	Colocation, edge-computing, and content-delivery facilities	4	Current
519290	Web-search and online-information firms	4	Current
522320	Payment processors and financial-technology infrastructure	4	Current
523150	Securities and electronic-trading firms	4	Current
541511	Custom software, artificial-intelligence, and analytics developers	4	Current
541715	Computer-science and engineering research establishments	4	Current

Seven identifiers in the table are not searched directly, because they are either an older version of a code or a broad aggregate used for national accounting rather than an industry that a business could belong to. Each is searched through its current equivalent, so no channel is lost:

Identifier not searched	Searched instead through
517110 (older telecommunications code)	517111
811300 (equipment-repair aggregate)	811210
484000 (truck-transportation aggregate)	484110
420000 (wholesale-trade aggregate)	423430, 423610, 423690, 423730, 423830, 423930, 424710, 424720
541300 (architecture and engineering aggregate)	541310, 541330, 541370, 541380
511200 (software-publishing aggregate)	513210
449210 (electronics retail)	Not searched; the underlying source treats this retail margin as excluded

Five further identifiers appearing in the original modeling source are construction-commodity codes from the Bureau of Economic Analysis rather than industry codes at all — office and commercial structures, other nonresidential structures, power and communication structures, the electric-power aggregate, and the retail-trade aggregate. They are kept in the archive so that the frame can be matched back to that source, and they are excluded from searching.

A3. The Three Case Regions and the Geography of the Search

The study covers three regions chosen to differ in size, age, and rural or metropolitan character: a rural agricultural county in central Washington with an early hyperscale cluster; a fast-growing metropolitan area in central Ohio where the cluster is recent; and the oldest and densest data-center market in the world, in northern Virginia.

For every region, three decisions are set: the core search area, the metropolitan search area, and the reference period.

The core search area is the set of counties in which businesses are discovered for all industry codes. In Quincy, the core is Grant County, with the adjacent counties of Douglas, Chelan, and Adams admitted only when a business explicitly states that it serves the Quincy-area campuses. In central Ohio, the core is Franklin and Licking counties, which together contain the New Albany campuses. In Virginia, the core is Loudoun County, including its named communities of Ashburn, Sterling, Leesburg, and the Loudoun side of the Dulles corridor.

The metropolitan search area applies only to Category 4, the proximity industries, where the value of being near a data center extends across a labor market and a fiber network rather than stopping at a county line. In Washington this is the Wenatchee and East Wenatchee area and central Washington generally; in Ohio it is the wider Columbus region, including Fairfield County and Lancaster; in Virginia it is northern

Virginia generally, including Prince William and Fairfax counties. Businesses found in the metropolitan tier are recorded with a note stating the level of geography at which they were found.

The reference period is the year from which the region is treated as having a data-center industry. Each starting year is taken from a government body's own description of its region rather than from press coverage. New Albany's city government states that it has been home to data centers since 2010. Loudoun County's data-center presence traces back to the arrival of major internet companies in 1996 and the first facilities in 1997 and 1998, but the locked reference year is 2008, the year the county's economic development department adopted a deliberate strategy of courting the industry — the same kind of government self-description that produced Ohio's 2010 date. Quincy's buildout began around 2006 with the first hyperscale campuses; evidence from any date is recorded and dated rather than discarded.

Two further decisions differ across the regions and are noted because they shape how much was found. Quincy, as the pilot, used an exhaustive discovery standard: an industry-and-region cell was complete only when every query template had been run and the last two queries produced no business that was not already registered. Ohio and Virginia use a capped standard instead, in which each discovery session is limited to roughly thirty new businesses or twenty-five logged queries and remaining codes carry forward to a later session; at metropolitan scale, an exhaustive sweep of an industry such as janitorial services is not a meaningful stopping point.

	Quincy and the Columbia Basin, Washington	Central Ohio (Columbus and New Albany)	Loudoun County, Virginia
Core counties, all codes	Grant County; Douglas, Chelan, and Adams counties when a firm states it serves the cluster	Franklin and Licking counties	Loudoun County
Metropolitan level, Category 4 only	Wenatchee, East Wenatchee, central Washington	Columbus region, Fairfield County, Lancaster	Northern Virginia, Prince William County, Fairfax County
Reference period begins	2006 buildout; case window emphasizes 2018 onward	2010	2008
Discovery standard	Exhaustive per industry-and-region cell	Capped by session budget	Capped by session budget
Priority tiers	Four	Five	Five

Directory sweeps

Alongside the per-industry queries, each region receives a one-time sweep of institutional sources. These sweeps are run once and harvested for every industry code at once, because a single economic development announcement or a single permit docket frequently names contractors and suppliers across half a dozen industries. The sources are chosen so that each region's list plays the same functional roles as the others': a regional development agency, a public financing body, chambers of commerce, local newspapers of record, the same three industry trade publications, government and regulatory filings, and the state's business registry for confirming that a company exists under the name it advertises.

Type of source	Quincy and the Columbia Basin, Washington	Central Ohio	Loudoun County, Virginia
Regional economic development organizations	Grant County Economic Development Council	One Columbus; Franklin County Economic Development and Planning; Grow Licking County; Fairfield 33 Development Alliance (metropolitan tier only); JobsOhio	Loudoun County Department of Economic Development; Virginia Economic Development Partnership; Northern Virginia Technology Council
Public financing and port authorities	Port of Quincy tenant lists	Licking County Port Authority; Columbus-Franklin County Finance Authority; The New Albany Company, master developer of the business park	Loudoun County Economic Development Authority; Loudoun Water
Chambers of commerce	Quincy Valley Chamber; Moses Lake Chamber	Columbus, New Albany, Licking County, Pataskala, and Lancaster-Fairfield chambers	Loudoun Chamber; Dulles Regional Chamber; Leesburg Business Alliance
Local newspapers	Columbia Basin Herald; Wenatchee World	Columbus Dispatch; Columbus Business First; Columbus Underground; Newark Advocate; ThisWeek Community News; Lancaster Eagle-Gazette	Loudoun Now; Loudoun Times-Mirror; Inside Northern Virginia; Washington Business Journal; Washington Post regional bureau
Industry trade press	Data Center Frontier; Data Center Dynamics; Data Center Knowledge	Same three publications, with Ohio place names substituted	Same three publications, with Virginia place names substituted
Government and regulatory records	Washington Utilities and Transportation Commission pipeline-operator registry	Ohio Power Siting Board dockets; school-district and county-auditor tax-abatement filings; state environmental and utility filings	State Corporation Commission dockets for transmission and substation filings; county board and planning-commission minutes; county building-permit records; electric utility and cooperative infrastructure filings

State business registry	Washington Secretary of State	Ohio Secretary of State	Virginia State Corporation Commission business search

A4. The Coding Rubric

Every business in a register receives exactly one primary classification, one evidence rating, and a set of dates. All three cite specific lines in that business's evidence file.

Activity types

The three positive classifications separate three genuinely different economic events: a business that already existed and changed what it sells; a business that came to the area to sell to data centers; and a business that came to the area because being near data centers is valuable in itself.

Type A, adaptation of an existing business. The business operated in the area, or regionally, before it served data centers, and later added a specialized service, product, or division for data-center customers. Two things must both be shown: that the business predates its data center offering, and that a data-center-specific offering exists. A community college in Grant County that had operated for decades and then built a data-center technician program with the campus operators — covering networking, cooling systems, backup diesel maintenance, and site security — is a clean Type A, because both the prior existence and the new offering are documented and dated. So is the county public utility district's transmission expansion undertaken to serve hyperscale load, and so is a regional equipment dealer that added a dedicated data-center generator technician service line to a branch it had operated for years. Type A requires evidence of *change*: a business that predates the campuses and sells to them, but shows no added specialization, is not Type A.

Type B, a business attracted as a supplier. The business established its presence in the area in order to have data centers as customers. Again two things must be shown: that the local presence begins during or after the buildout, and that data-center demand motivated the entry. The purpose-built water-reuse utility created jointly by a city and a campus operator, which recycles cooling water through more than thirty miles of dedicated pipe under a thirty-year agreement, is Type B, as are the dedicated gas-fired generating plants in Ohio built specifically to power adjacent campuses. So is an electrical contractor whose own website states that it opened a regional office specifically to pursue data-center work in that part of the state. The cleanest illustration of the boundary is a matched pair from the same trade and the same small labor market: one electrical contractor opened its regional office to serve data-center construction and is Type B; another had operated a branch in the same valley for nine years before the campuses arrived and later added data-center work, and is Type A.

Type C, a business attracted by proximity benefits. The business is not primarily selling to data centers. It located in the area because being near them is valuable: dense fiber, low network delay, direct connections into cloud providers, available power, a technical labor pool, or the cluster itself. Network providers that placed equipment inside an existing campus so that their customers could reach it directly are Type C, as are the high-density computing operators that came for the same cheap, abundant power that drew the campuses, and the later colocation companies that built in a market precisely because the fiber, power, and permitting environment were already proven. The dividing line with Type B is a single question: if the data center is the firm's *customer*, the code is B; if the data center or its infrastructure is the firm's *input or valuable neighbor*, the code is C.

No connection (N). No evidence of any data-center connection surfaced under the protocol. This is the expected outcome for most businesses in most industries, and it is a finding rather than a failure.

Three additional labels keep the counts honest.

Operator. The data-center companies themselves are the cause under study, not the effect. They are registered so that the record is complete, but they are excluded from all percentages. The one exception is a later-wave colocation company that entered because the cluster already existed: that firm is an outcome of the cluster and is coded Type C.

Unverified. Some businesses appear only through location-targeted marketing pages — a national firm that publishes a page naming a small town it does not operate in, in order to appear in local search results. Where no address, registry entry, local news mention, or directory listing can confirm a real local establishment, the business is labeled unverified and removed from the denominator entirely rather than counted as having no connection. Detecting these pages is a standing part of the protocol.

Review. Evidence exists but the classification is genuinely ambiguous. The business is listed with both competing readings and referred to the principal researcher, who rules on it; recurring questions become new numbered decision rules.

Code	Name	What must be shown	Illustration
A	Adaptation of an existing business	The business predates its data-center offering, and a data-center-specific offering exists	A community college that added a data-center technician program with the campus operators; a utility that expanded transmission for hyperscale load
B	Attracted as a supplier	Local presence begins during or after the buildout, and data-center demand motivated the entry	A purpose-built cooling-water reuse utility; a gas-fired plant built to power an adjacent campus; a contractor that opened a regional office for data-center work
C	Attracted by proximity benefits	The firm chose the location and cited or demonstrably depends on adjacency, connectivity, power, or the cluster	A network provider placing equipment inside an existing campus; a high-density computing operator drawn by the same cheap power; a later colocation entrant
N	No connection	The full protocol produced no evidence of a data-center connection	Local trucking, fuel, and crane-rental firms whose work is agricultural
Operator	The treatment, not the outcome	The business is itself a data-center operator	The hyperscale campus owners in all three regions

Unverified	Local presence not established	Discovered only through location-targeted marketing with no verifiable local establishment	A national service brand publishing a town-named page with no local address
Review	Ambiguous under the rubric	Evidence exists but the type assignment is contested	An operator's separately incorporated in-house construction affiliate

Evidence ratings

Alongside the type, each business receives a rating describing where the evidence came from. When several ratings apply, the highest is recorded and all are cited.

Rating	Meaning	Example
E1	The business states the connection itself	A services page naming data-center clients; a press release announcing a data-center division
E2	A third party attributes the connection	A news article, an economic development agency document, a government filing, or trade press
E3	Circumstantial	Timing and specialization are consistent with a connection, but nothing states it; never used alone to support a headline finding, and reported separately
E0	Nothing found after the full protocol ran	The complete query sequence and stopping rules produced no evidence

Dates, and how they are established from archived web pages

Three dates are recorded where they can be established: the earliest dated artifact showing the connection to data centers; the date the business began operating in the area; and a statement of which specific artifacts each date rests on.

Dating is the hardest part of the evidence work, because a company website shows what a firm sells today and almost never says when it started selling it. The study handles this with the Internet Archive's Wayback Machine, a public service that has been storing copies of web pages since the late 1990s. For each business with a website, the researcher steps through the archived copies of the home page and the services page at roughly two-year intervals, then more finely once the change appears to be close, and records a pair of snapshots: the last archived version of the page that does not mention data-center work, and the first archived version that does. That pair brackets the moment the firm began presenting itself as a data-center supplier — for example, a contractor whose services page says nothing about data centers in one archived year and lists a data-center construction practice in a later one has demonstrably added that line between those two dates.

Two conventions follow from how archiving works. First, a snapshot date is an upper bound, not a date of change: it proves that the content existed by that date, not that it appeared then, and the study writes such dates with a "no later than" marker so that no later reader mistakes one for an onset. Second, archive coverage is uneven across small-business websites, so the bracket is recorded as whatever pair of snapshots actually exists, along with the list of snapshot dates inspected, rather than being reported as a single date with no indication of how tightly it is pinned.

The archive serves the evidence trail in two further ways. When a cited page has since changed or disappeared, the archived copy is what a checker reads instead, so a citation does not decay into an unverifiable claim. And when a key page has never been archived, it is saved to the archive at the time it is read, and the archive address is recorded alongside the live address.

The twelve decision rules

The rules are researcher-written for the coders to resolve the hard cases that recur. They exist so that the same situation is decided the same way in every region.

Rule	Situation	Decision
1	A national firm with a local branch	The local establishment is what is coded, on the local establishment's own evidence
2	A firm predates the campuses and sells to them, but shows no added specialization	Code N with a note, not Type A; adaptation requires evidence of change
3	Utilities and municipal entities	Eligible for coding; a purpose-built utility is Type B, an incumbent that upgraded for the load is Type A
4	Industries with a strong agricultural confounder, such as fuel, trucking, and cold storage	The evidence must name data centers specifically; regional employment concentration or timing alone is at most E3
5	Cryptocurrency and artificial-intelligence computing operators	Type C, since they consume the same power and connectivity, unless they sell services to the data centers
6	The data-center operators themselves	Coded as operators and excluded from percentages, except later-wave colocation entrants attracted by the existing cluster, which are Type C
7	One business spanning several industries	Code the primary industry, note the secondary; the business is counted once
8	Infrastructure built before the campuses arrived, even when repeatedly cited as an attraction	It is a precondition, not an outcome, and codes N unless a later, data-center-motivated addition is documented, which is then Type A on the added line only
9	A nonlocal firm that won a contract but has no local establishment	Counted if the work logically requires local activity — construction, installation, physical operations, maintenance — and not counted for purely remote deliverables such as software subscriptions, remote design, or off-site fabrication
10	A group of businesses with essentially identical profiles	May be coded as a class, requiring a written class memo defining membership, the full evidence protocol run on at least two randomly drawn members with the draw

		documented, and removal of any member whose own evidence contradicts the class
11	A claimed local presence that cannot be verified	Coded unverified and removed from the denominator; this rule takes precedence over rule 9, and readmission requires a verified local establishment, not merely documented work performed from elsewhere
12	A utility rate schedule, tariff, or customer class that names data centers	A pricing instrument is not a service or capability the business added, so it does not by itself support Type A; the incumbent codes N and the instrument is still documented

Rule 10 deserves a note, because it is what allows the qualitative process to scale without becoming shallow. When a set of businesses shares an essentially identical profile — the same role, the same relationship to the infrastructure, the same absence of a data-center service line — the group is treated as a class, and the full evidence protocol is run on at least two members drawn at random with the seed recorded. If those members come back as expected, the class closes and its members take the class code. If a drawn member comes back differently, the class hypothesis is rejected, the class is dissolved, and every member is investigated individually. For instance, in one case a proposed class of local electrical contractors was dissolved when the sampled members turned out to have real data-center ties, and each member was then evidenced on its own.

Record fields

Every coded business becomes one row with the following fields: business name; identifier slug; region; level of geography at which it was found; assigned industry code; the basis on which that code was assigned; type; secondary type if any; evidence rating; first evidence date; local presence date; the basis for those dates; the evidence file; key sources as web addresses; how the business was discovered; notes on confounding explanations; coder notes; and the outcome of the quality-control check.

A5. Checking the Coding with Independent Second Coders

A classification scheme is only as good as its repeatability. The study therefore has every classification checked by second coders who work from the same evidence but have never seen the first coder's answer. Two kinds of second coder are used: a human research assistant, and an artificial-intelligence model from a different family than the ones that did the original work. Both are given the same materials and asked to do the same job, so that their results can be compared with the original codes and with each other.

Preparing a blind packet

Second coders never receive the working files. A small program builds a separate packet for each region from that region's evidence folder. For every business, the program copies the evidence file but cuts it off at the point where the first coder's conclusion begins, so the packet contains the quotations, the web addresses, the access dates, and the archive brackets, and nothing about how any of it was interpreted. The packet also contains a blank coding form with one row per business and a manifest listing every file with a short fingerprint, so that a later reader can confirm exactly which version of which file a coder saw. The combined workbook, the register's status column, and the study's running notes are never included, and the packets are checked before release to confirm that no conclusion text survived the strip. Because the program reads the evidence folder as it stands, a packet is regenerated whenever a region's evidence base grows, so that a reliability run always works from the current material. Packets have been built for all three regions and currently hold 130, 66, and 159 blinded dossiers for Washington, Ohio, and Virginia.

Alongside the packet, both kinds of coder receive the coding rubric in full — the type definitions, the evidence ratings, and the numbered decision rules — because the point of the exercise is to test whether the written rules produce the same answer in different hands, not whether two people share the same instincts.

Instructions given to the human coder

The human instructions are written to be self-contained: a coder who knows nothing about the project can pick them up and start. They open with a short description of the study and of the seven possible codes, then set out the procedure. The coder reads one business's evidence file, keeps the rubric open beside it, and assigns a type using only what is in the file and what the rubric says — outside knowledge and fresh searching are specifically excluded from the type decision, because the exercise is testing whether the same evidence yields the same code. The coder then assigns an evidence rating and the earliest year the connection can be documented, records a confidence level from one to three, and writes two or three sentences citing the specific extracts that drove the decision. Anything that looks wrong — a broken link, an extract that claims more than its source supports, a missing date, an industry code that does not fit — goes in a flags field.

One step does send the coder online, and it is deliberately separate from coding: for each business, the coder opens two of the cited sources chosen at random and records whether the page still exists — checking the archived copy when the live page is gone — and whether it actually says what the extract claims. That step audits the evidence trail itself rather than the classification.

The conduct rules are short. Do not look at the first coder's codes or at the master workbook, and do not ask the project lead what the original answer was. When torn between two codes, apply the rubric's boundary rules; if still torn, use the review code and explain the competing readings, because a marked review is more useful than a guess. Class memos, which cover groups of near-identical businesses, are coded once as a unit, with any member whose own evidence contradicts the group flagged separately.

Instructions given to the artificial-intelligence coder

The design for the model coder mirrors the human protocol point for point, so that the two sets of results are comparable. The model is given a standing instruction describing its role as an independent second coder, the full rubric, and the same short statements of the boundary rules that the human coder works from. It receives one blinded dossier at a time, or a small batch, and returns a fixed set of fields: the business identifier, the type, any secondary type, the evidence rating, the first evidence year and whether that year is an upper bound, a confidence level, a short rationale citing the extracts, and any data-quality flags. The fields match the human coding form exactly, so both can be scored by the same procedure.

Three conditions keep the comparison fair. Web browsing is turned off for the coding pass, so the model works from precisely the evidence the human sees. The model is told explicitly that there is no expected answer and that it should not try to infer what another coder decided. And each group of industry codes is coded in a fresh session, so that dossiers read earlier cannot anchor the judgments made later. Source verification is run afterward as a separate pass with browsing enabled, checking the same kind of question the human coder answers by hand: does each cited page exist — or, if it has gone dead, does its archived copy — and does it support the claim attached to it. Keeping the two passes apart prevents anything found on the open web from contaminating the blind coding.

Comparing the results

The three sets of codes — original, human, and model — are compared on the type variable using percent agreement and Cohen's kappa for each pair, and Fleiss' kappa across all three. The review code is treated as a category of its own in a first pass and dropped in a second, so that genuine disagreement about type can be distinguished from disagreement about whether a case is decidable at all. Evidence ratings are scored both for exact agreement and for agreement within one rating, and first-evidence years are scored as agreeing when they fall within one year of each other, since many dates are archive brackets rather than exact onsets.

Disagreements are treated as findings about the rubric rather than as coder errors. Every case where all three coders differ, and every case where the original code stands alone against both second coders, goes onto an adjudication list for the principal researcher. Those rulings become new numbered decision rules, and every business affected by a new rule is recoded — the same path that produced the later rules in Section 4. This is what keeps the rubric stable as the study moves from one region to the next: a question that had to be judged once becomes a rule that is applied automatically thereafter.

A6. Counts of Evidence Items

The counts below describe the process itself: how many businesses each region's discovery stage produced, how many evidence files were written and read, how many businesses remained in the counted frame after operators and unverifiable firms were removed, and how the remaining businesses were classified.

Three conventions govern the tables. **Registered** is the number of businesses entered during discovery. **Evidence files** counts the individual research dossiers written; a smaller number of businesses are covered instead by a class memo written under rule 10, and those memos are counted separately in the regional summary below. **Source extracts** counts the individual sources read and recorded inside those dossiers — one entry for each cited page, carrying its title, its web address, the date it was read, and the quoted material drawn from it. A page consulted and found to contain nothing relevant is not an extract; the count therefore measures the source material that actually stands behind the coding, and it separates a business settled on a single page from one built from a dozen documents. **Counted frame** is the number of registered businesses remaining after removing operators and unverified local presences, and it is the denominator for every percentage. The final columns give the classification of businesses in the counted frame.

Regional summary

	Quincy, Washington	Central Ohio	Loudoun County, Virginia	All three
Queries logged	1338	812	1504	3654
— of which discovery	302	172	157	631
— of which evidence	1036	640	1347	3023
Businesses registered	176	78	163	417
Individual evidence files	130	64	155	349
Class memos	14	2	4	20
Source extracts read	445	260	612	1317
Average extracts per evidence file	3.4	4.1	3.9	3.8
Excluded as operators	7	5	9	21
Excluded as unverified	13	5	39	57
Counted frame	**156**	**68**	**115**	**339**
Coding not yet applied	0	2	0	2
Held for review	5	8	19	32
Type A	8	11	6	25
Type B	32	8	15	55
Type C	10	9	18	37
No connection	101	30	57	188
Businesses with a connection (A, B, or C)	50	28	39	117
Share of the counted frame with a connection	32.1%	42.4%	33.9%	34.7%

The share for central Ohio is computed on 66 businesses, since two registered businesses there are pre-period items that were deliberately left uncoded; the pooled share is likewise computed on 337.

By category of industry

Region	Category	Registered	Evidence files	Counted frame	A	B	C	No connection	Review
Quincy	1. Recurring operations	66	44	58	4	15	1	37	1
Quincy	2. Construction and professional services	64	53	60	2	17	0	39	2
Quincy	3. Equipment and distribution	24	19	22	1	0	0	19	2
Quincy	4. Proximity beneficiaries	22	14	16	1	0	9	6	0
Central Ohio	1. Recurring operations	32	32	29	5	5	0	16	3

Central Ohio	2. Construction and professional services	20	19	19	5	3	0	9	2
Central Ohio	3. Equipment and distribution	8	8	7	1	0	0	4	2
Central Ohio	4. Proximity beneficiaries	18	5	13	0	0	9	1	1
Loudoun	1. Recurring operations	56	52	40	4	5	3	19	9
Loudoun	2. Construction and professional services	46	46	30	2	7	0	17	4
Loudoun	3. Equipment and distribution	7	7	7	0	2	0	5	0
Loudoun	4. Proximity beneficiaries	54	50	38	0	1	15	16	6

Reading the tables

All three regions have been searched across the whole frame — Washington through all four of its priority tiers, and Ohio and Virginia through all five of theirs, including every batch of the lowest tier in each case. An industry code that appears in one region's table and not another's therefore records a search that produced no business to register, not a search that was skipped. The three tables contain 52, 43, and 53 industry codes out of the 81 searched, and the codes that appear in none of them are mostly manufacturing industries whose output reaches a campus through national markets rather than through a local establishment.

Extract counts are attributed to the individual dossiers, so a class memo contributes none of its own: the reading behind a class is recorded in the dossiers of the members who were drawn for the full protocol, and those extracts appear under the class's industry code. The same convention explains why an industry code can show more extracts than the arithmetic of its evidence files would suggest — a single contested business can carry a dozen sources, while a business with no data-center connection is often settled in two or three.

The gap between registered businesses and individual evidence files has two sources. Some businesses are covered by a class memo under rule 10 rather than by an individual dossier, which is why Quincy's fiber-carrier code shows twenty-two registered businesses and four individual files: two class memos cover the rest. Others are businesses whose claimed local presence was disproved during discovery itself and which were recorded as unverified without a full dossier.

The final archive supporting these tables consists of one register per region, one search log per region recording all 3,654 numbered queries, 349 individual evidence files and 20 class memos holding 1,317 quoted and dated source extracts, a combined workbook with one row per business and live summary formulas, and the protocol, rubric, and priority-tier documents that governed every step.