ChatGPT

Ceneca – Pre-Seed Investor Research Briefing

Product Description

Ceneca is a self-hosted **AI data analysis agent** designed for on-premises deployment. It connects directly to a company's databases (e.g. PostgreSQL, MongoDB, vector stores like Qdrant) and allows users to query data **in natural language** [1]. Essentially, Ceneca translates plain English questions into SQL (or other database queries) and returns results along with analysis or visualizations, functioning like a "ChatGPT for your internal data." By operating on-prem, Ceneca keeps sensitive data within the organization's own environment, addressing data privacy and compliance concerns. The agent can interface with multiple data sources (structured and unstructured) to provide unified insights. According to its open-source repo, it "allows users to query databases using natural language" and converts those questions into SQL, providing analysis of the results [1]. This enables non-technical business users to get insights from enterprise data without writing code or moving data to an outside cloud service. In summary, **Ceneca's value proposition** is an AI-powered data assistant that brings conversational analytics **behind the company firewall**, making data exploration easier while preserving security.

Market Overview: Enterprise AI Analytics (On-Prem Focus)

Enterprise AI analytics is a large and rapidly growing market. Businesses worldwide are seeking to leverage artificial intelligence for data-driven decision making, leading to what Gartner calls *"augmented analytics"* – tools that use AI/ML to automate data preparation, insight generation, and natural language querying. The global augmented analytics market is projected to grow from about **$13.6 B in 2024 to $16.5 B in 2025** (21% YoY growth) [2], on track to reach **$41 B by 2034** [3]. Broader definitions that include all business intelligence (BI) and analytics software peg the market at even higher values (e.g. **$84.6 B by 2026** for BI/analytics combined) [4]. This growth is fueled by the convergence of big data, cloud computing, and more accessible AI (e.g. GPT-4) driving demand for smarter, easier analytics.

Within this space, **on-premises AI analytics** is a significant segment. Many enterprises – especially in regulated industries like finance, healthcare, and government – *require solutions that can be deployed in private clouds or on-prem data centers* due to data security and compliance. Despite the cloud analytics boom, **55% of organizations still rely on traditionally managed on-prem systems** alongside cloud infrastructure [5]. In other words, over half of enterprises maintain data on-premises that they cannot freely send to cloud SaaS tools. This opens a clear need for AI analytics solutions that **work behind corporate firewalls**. The ability to deploy an AI agent *on-prem* appeals to companies worried about sensitive data leakage or those operating in regions with strict data sovereignty laws.

Additionally, the "**agentic analytics**" trend is emerging: companies are exploring autonomous or conversational agents that sit on top of their data to provide insights proactively [6]. The popularity of ChatGPT in 2023 has validated the power of natural language interfaces, and enterprises now want that capability on their *own* data. However, many cloud-based AI analytics startups don't meet the needs of data-sensitive enterprises. Ceneca's focus on on-prem deployment directly targets this gap by offering **the benefits of generative AI in analytics without forcing data into a third-party cloud**. Given that nearly 80% of IT leaders believe generative AI will help them use data better [7], but over 40% struggle with data being too complex or siloed [8], a solution like Ceneca (which uses plain English to access data silos securely) is timely.

Key market drivers include: the explosion of data volumes, shortage of data scientists (making self-service AI tools attractive), cloud cost concerns (on-prem can be cheaper at scale), and regulatory pressure on data handling. In summary, the **global enterprise analytics market is huge and growing (TAM in tens of billions)**, and within it there is a strong niche for **on-prem or private-deployable AI analytics** for enterprises that require security and control. Ceneca sits at the intersection of these trends: AI + data analytics + on-prem deployment.

TAM, SAM, and SOM Estimates (with Assumptions)

To size the opportunity for Ceneca, we break down the **Total Addressable Market (TAM)**, **Serviceable Available Market (SAM)**, and **Serviceable Obtainable Market (SOM)**:

- **TAM (Total Addressable Market):** We use the global enterprise analytics software market as a proxy for TAM. This includes all organizations that could potentially use AI-driven data analysis tools. According to industry research, the *augmented analytics* segment (AI-powered BI) will be **~$16.5 B in 2025** and growing ~25% CAGR [3]. Including traditional BI and analytics, TAM is even larger – roughly on the order of **$30–$85 B** in the mid-2020s by various estimates [4]. For a conservative number, we can say **TAM ≈ $16–20 B in 2025**, representing all global spend on software that enables data insight generation (AI or BI tools). *Assumptions:* This assumes all industries worldwide, any deployment model (cloud or on-prem), and includes both large enterprises and mid-market firms that have significant internal data to analyze.

- **SAM (Serviceable Available Market):** This is the portion of TAM that **Ceneca can realistically serve in the medium term**, focusing on customers who need on-prem or private cloud solutions. We assume that roughly **30–50% of the TAM prefers or requires on-prem deployments** (given the data that 55% of orgs still run on-prem systems [5]). Additionally, Ceneca will initially target English-speaking markets and industries with acute data privacy needs (finance, healthcare, government, etc.). If we take 50% of the augmented analytics TAM for on-prem-focused buyers, **SAM in 2025 would be on the order of ~$8 B** (e.g. half of $16B). To be more conservative, we might take 30% of TAM, giving **SAM ≈ $5 B**. *Assumptions:* Ceneca's reachable market excludes customers that are cloud-only or small businesses that lack databases. It includes primarily mid-to-large enterprises in regions where Ceneca can sell (initially North America, Europe, and key Asia-Pacific markets including India). This ~$5–8B SAM encapsulates organizations that have the budget and need for an AI data agent deployed in their own environment.

- **SOM (Serviceable Obtainable Market):** This is the **near-term, practical market share** Ceneca could capture, say in the next 2–3 years. Being a pre-seed startup, Ceneca will start with a narrow subset – perhaps a specific vertical or a set of design partner customers. For example, if focusing first on, say, financial services firms in a region: suppose there are ~1,000 such target companies globally that *urgently* need on-prem NL analytics. If Ceneca could acquire ~1% of those in the near term (10 customers) at an average annual contract value of $50k, that's ~$500k ARR – a starting point. Stretching to a longer-term horizon (5 years), one could aim for on the order of **100 customers** with $50k–$100k ACV. That would be **$5–$10 M in annual revenue**, which is still a small fraction of the multi-billion SAM (well under 0.5% of SAM). So, SOM might be expressed as *tens of millions of dollars* in revenue potential within a few years of launch, given successful execution. In terms of market size, this corresponds to **<0.1% of TAM** (which is realistic for an early-stage company). *Assumptions:* Ceneca's initial go-to-market might target a specific niche such as mid-size enterprises that use PostgreSQL and are in a data-regulated sector. We assume moderate pricing (in line with enterprise software norms) and a gradual customer acquisition

ramp. The SOM basically represents achieving a foothold: e.g. perhaps **50–100 enterprise installations** in the next 3 years. This would validate the concept and position the company to pursue a larger chunk of the $5B+ SAM subsequently.

In summary: **TAM is in the tens of billions ($16B+)**, reflecting the global appetite for AI analytics. **SAM is several billion ($5–8B)**, focusing on on-premise deployments and relevant industries. **SOM is in the tens of millions**, reflecting an achievable slice (initial cohort of customers) that Ceneca could capture in the near term given its stage. These estimates rely on assumptions about adoption rates (on-prem preference) and target customer size; as Ceneca grows, it can expand beyond the initial niche to gradually address more of the SAM and eventually even the full TAM.

Competitive Landscape

The landscape for data analytics and BI is crowded, and AI is the new frontier for many of these competitors. Below is a summary of key players and how they compare:

- **ThoughtSpot:** A pioneer in search-driven analytics. ThoughtSpot's platform allows users to type natural language queries and get dynamic charts/answers. It's a mature company (unicorn valued ~$4.2B) with ~1,000+ enterprise customers. ThoughtSpot has recently doubled down on AI: it launched *ThoughtSpot Sage*, integrating GPT-3/4 into its search experience [9]. This enables users to ask even more complex questions in natural language and get AI-generated insights. ThoughtSpot's strength is a slick user experience for business users and the ability to handle large-scale cloud data (it's now offered as a cloud SaaS, though in the past they had on-prem appliances for large deployments). **Competitive note:** For Ceneca, ThoughtSpot is a direct competitor in *natural language analytics*. However, ThoughtSpot is now mostly cloud-based (though it can connect to on-prem data sources, many deployments are in cloud environments). Ceneca's differentiator would be being lighter-weight and truly on-prem for all components. Also, ThoughtSpot's costs are high, targeting Fortune 500 firms – leaving an opportunity to serve mid-market or cost-sensitive clients with an on-prem agent.

- **Tableau (Salesforce):** Tableau is the market leader in BI visualization, known for its drag-and-drop dashboards. It traditionally wasn't focused on natural language, but that is changing. Salesforce (which acquired Tableau) announced **Tableau GPT**, an integration of generative AI to let users converse with their data and automatically generate visualizations/insights [10]. Tableau also has an "Ask Data" feature (natural language query interface) and is building *Einstein GPT for Analytics* across its platform. With a huge user base, Tableau is incorporating AI to maintain its lead. Tableau can be deployed on-prem (Tableau Server) or in private cloud, which appeals to companies with data control needs – though its new AI features may depend on Salesforce's cloud AI services. **Competitive note:** Tableau is an incumbent and likely partner in some cases (Ceneca could integrate with Tableau dashboards), but as Salesforce infuses AI, it encroaches on Ceneca's conversational analytics value. Ceneca must offer deeper or more tailored analysis of raw data via language, whereas Tableau is still primarily a visualization tool adding AI assistance. Tableau's strength is enterprise trust and completeness; Ceneca's opportunity is agility (providing answers directly from databases without heavy BI development) and privacy (keeping AI logic in-house).

- **Microsoft Power BI:** Power BI is another dominant BI platform, especially among Microsoft enterprise customers. It offers a Q&A feature where users can ask questions in natural language to create charts. In 2023, Microsoft introduced **Copilot in Power BI** – a generative AI assistant (currently in preview) that helps users analyze data with natural language and even generate

entire reports automatically [11]. Given Microsoft's investment in OpenAI, Power BI's Copilot is expected to be very powerful. Power BI can be run in the cloud (Power BI Service) or via an on-premises report server for certain cases, but many advanced AI features may require Azure services. **Competitive note:** Power BI, being deeply integrated into Office 365 and Azure, is a formidable competitor. Its AI features (like Copilot) aim to do what Ceneca does – *make data querying conversational*. However, not all organizations use Power BI, and some might prefer an independent, database-agnostic agent. Ceneca's cross-database ability (Postgres, MongoDB, etc.) in one interface is a differentiator; Power BI tends to work source-by-source or after data is ETL'd into their models. Additionally, Microsoft's solution may send data or prompts to cloud AI models, whereas Ceneca could be positioned as **completely self-hosted (private)**. Organizations that are wary of sending data to Microsoft's cloud could favor a tool like Ceneca.

- **Apache Superset:** An open-source BI platform created by Airbnb, now an Apache project. Superset is a **modern data exploration and visualization tool** that can be deployed on-prem for free [12]. It provides an SQL IDE, dashboards, and can connect to many databases. Superset does not natively support natural language queries – it's more for analysts who can write SQL or for developers embedding it. **Competitive note:** Superset competes as a low-cost, on-prem BI solution. For a technical team, Superset can be customized, whereas Ceneca would add value by offering *natural language accessibility* on top of data. In theory, one could extend Superset with an NL interface or use it in conjunction with Ceneca (Ceneca to query and get results, Superset to visualize if needed). The existence of a strong open-source BI tool means Ceneca has to offer something significantly *beyond visualization* – namely, the AI-driven querying and explanation. Also, Superset lacks the AI/ML insight generation that Ceneca aims to provide. Ceneca can differentiate by focusing on the **AI agent aspect (conversational, smart insights)** rather than being just another dashboard tool.

- **MindsDB:** A venture-backed startup (~$55M raised) that offers an open-source "AI layer" for databases [13]. MindsDB's platform acts as an AI **query engine**; it allows users (or applications) to query data in natural language or via SQL to get AI-assisted answers [14]. They integrate with many databases and even enable machine learning models to be deployed at the data source. MindsDB's tagline is about democratizing ML and enabling "AI tables" – for example, one can ask questions and get predictions within the database. Notably, MindsDB is also deployable anywhere (including on-prem) and positions itself as a developer-friendly tool to build AI into data pipelines. **Competitive note:** MindsDB is quite aligned with Ceneca's space, emphasizing **natural language to query data across sources with accurate answers** [14]. The difference may be in target user: MindsDB often pitches to developers (to embed AI in apps via SQL), whereas Ceneca might target business analysts/end-users with a chat interface. That said, as an "AI-as-a-database" concept, MindsDB is a strong competitor. Ceneca might differentiate by a more out-of-the-box *agent* experience (MindsDB still might require some setup/engineering to use effectively). Also, Ceneca could focus specifically on enterprise analyst workflows and provide a nicer UI, whereas MindsDB is more like an underlying engine. Both being open-source friendly and on-prem capable, the competition could come down to feature maturity and community adoption.

- **Other Competitors / Alternatives:** Beyond the above, there are other players in related areas:

- Traditional BI like **Qlik** or **SAP BusinessObjects** have large install bases (primarily on-prem historically) and some AI features (Qlik Insight Bot, etc.), though they lag in the generative AI race.
- **Oracle, IBM, SAS** offer advanced analytics platforms that certain enterprises use on-prem, but these tend to be complex and not focused on natural language interfaces.

- Emerging startups and open-source projects are tackling “ChatGPT for data” in various forms. For example, open-source frameworks like **LangChain** allow developers to build custom AI agents connected to databases (there are community demos of “ChatGPT SQL agents”). A Reddit community example is *Wren AI* (a toolkit with UI + LLM for data analysis) [15]. Also, projects like **LlamaIndex/GPT Index** can connect LLMs to internal data. These DIY solutions are indirect competitors – a resourceful engineering team might attempt to build an internal tool instead of buying one. However, a startup product like Ceneca can outshine DIY by offering a polished, supported solution with enterprise features (auth, UI, etc.).
- Vertical-specific analytics AI: e.g. ThoughtSpot’s success has inspired niche players focusing on certain domains (customer support analytics, marketing analytics with AI, etc.). Ceneca may encounter point solutions in different departments as competition for budget.

In summary, **Ceneca faces competition from both established BI giants (Tableau, Power BI) integrating AI and from newer AI-native platforms (ThoughtSpot, MindsDB)**. Ceneca’s **key differentiators** should be: **on-prem deployment** as a first-class design (unlike most new SaaS tools), **agnostic connectivity** to many data sources, and an **easy conversational interface** requiring minimal setup. The competitive landscape is dynamic – many incumbents are adding generative AI capabilities quickly (e.g. Salesforce’s Tableau GPT, Microsoft’s Copilot) to defend their turf [10] [11]. This validates the market need for AI-driven analytics, but also means Ceneca must move fast and build a technological edge (perhaps a superior NL-to-SQL engine fine-tuned for enterprise schemas, or a unique approach to keeping LLMs accurate with on-prem data). Given that **no single competitor perfectly fulfills all needs (on-prem, multi-db, NL querying, and simple setup)**, Ceneca has an opportunity to carve out a strong position if it can deliver a solution that is *secure, accurate,* and *easy to use*.

Pre-Seed Fundraising Benchmarks

At the pre-seed stage (the very first institutional funding round), there are established benchmarks for round size, valuation, and founder dilution. Recent data and industry benchmarks are as follows:

- **Round Size (Amount Raised):** Pre-seed rounds can vary widely in size, but most are relatively small checks intended to prove out a concept. According to Carta data, **about 75% of pre-seed rounds are under $900K** [16], and roughly 40% are under $250K (often friends & family or angel rounds) [16]. The median pre-seed raise in 2024 was around **$275K** (on a SAFE, per Carta) [17]. That said, the **average** is higher because a minority of companies raise larger pre-seeds: about 4–5% of pre-seed rounds exceeded **$5M** in recent quarters, particularly *hyped AI startups* requiring capital for compute [18]. In general, a “typical” venture-backed pre-seed in tech hubs might fall in the **$500K to $1.5M** range. Especially in 2023–2025, many AI-focused startups have commanded around $1M+ in pre-seed due to investor enthusiasm (some even $2–3M, though those are top-tier cases). The key is that pre-seed is usually enough money for 12–18 months of runway, which for a lean software team often means around a million dollars or less.

- **Valuations (Pre-money/Post-money):** Pre-seed valuations have normalized somewhat after the 2021 bubble. According to recent PitchBook data, the **average pre-seed pre-money valuation is ~$5.7M** (median ~$5.3M) [19]. This typically yields post-money valuations in the ~$6–8M range depending on round size. In Silicon Valley, it’s common to see pre-seed companies raising ~$1M at a ~$6–7M pre-money (leading to ~$7–8M post). There is regional variation: U.S. and Bay Area startups tend to get higher valuations, while in other ecosystems (e.g. India domestically), pre-seed rounds might be smaller with pre-money valuations in the $2–4M range. However, strong global founders often incorporate in Delaware and raise at US benchmarks. For instance, many recent pre-seeds for AI startups (with solid teams) have been in the $5–10M post-money

valuation range. In summary, **valuations around $5M (pre-money)** are a rough median, with higher outliers for competitive deals and lower outliers for very early or local-market deals.

- **Equity/Dilution:** Investors at pre-seed typically seek **10–20% ownership**. Founders often try to minimize dilution at this stage (since it's high risk capital), but giving up around 15% equity for a significant pre-seed round is common. For example, raising $1M on a $5M pre-money gives 16.7% post-money to investors. Many structured as SAFEs with valuation caps: e.g., a SAFE with a $6M cap and $1M investment effectively is similar (~14% dilution when it converts). If the raise is small from angels, sometimes cumulative dilution can be under 10%, but venture-led rounds usually end up in that 10-20% range. It's worth noting that Y Combinator's standard deal (if applicable) is $500K on an uncapped SAFE with MFN (for $125K) plus a $375K SAFE at a later of next round – this often results in ~7-10% dilution at seed, which is a special case. For non-accelerator cases, **founders targeting ~15% dilution at pre-seed is a good rule of thumb**. The data supports this: a median $5.3M pre-money and ~$300K raise [19] implies ~5% dilution (because many raises are tiny). But in more meaningful raises around $1M, you'd see closer to 15%.

- **Global/India-to-SF context:** For a team from India relocating or reaching to SF investors, it's notable that global investors now frequently back strong India-based teams at global valuations, provided the market is global. The diaspora of "India-to-SF" founders has grown, and many incorporate in the US to access larger pools of capital. Pre-seed deals for such founders in 2023-2024 have ranged widely, but an anecdotal benchmark: if raising from Silicon Valley VCs, the numbers won't differ much from any other SF startup (so ~$500K–$1M on $5–8M pre). If raising from India-based funds or angels only, the round might skew smaller (e.g. ₹1–3 Cr, which is ~$125K–$375K, at valuations of ₹20–30 Cr which is ~$2.5–4M). However, given the interest in AI, even Indian seed funds are paying more – Inc42 notes Indian GenAI startups (140+ of them) have raised $1.5B since 2020 [20], indicating a healthy funding environment. A *global approach* (India team + US market) can combine the best of both: cost-efficient R&D in India and higher willingness to pay (and invest) in the US. Some recent pre-seeds for India/SF hybrid startups in AI have closed >$1M. The key for such founders is often to show they can access the US market (or global market) despite being based in India – doing so can justify valuations equal to US peers.

In summary, **pre-seed funding rounds typically are ~$1M or less, with post-money valuations around $6–8M** in today's market, implying investors take ~10–15% ownership. Of course, there's a wide range: a scrappy pre-seed might just be $200K on a $2M valuation (especially outside major hubs), whereas a hot AI startup with credentials might do $2M on $10M+. Overall, though, the averages suggest modest rounds. It's also worth noting that the trend in late 2023 into 2024 was a slight pullback – more startups raising smaller pre-seeds (sub-$500K) as investors became more cautious [21]. But *for a deep-tech AI startup with global ambitions*, one can still target the higher end of the typical range because the upside is attractive to investors.

Suggested Fundraising Amount for Ceneca

Given Ceneca's profile (deep tech product, early stage, likely pre-revenue), a **sensible pre-seed raise** would be in the range of **$1.0 to $1.5 million**. This recommendation is based on the team's needs and the benchmarks above:

- **Runway and Team Needs:** Assuming the Ceneca team currently consists of the core founders (likely with strong technical backgrounds) and maybe a couple of early engineers, a ~$1M raise should comfortably fund 18+ months of development. This would cover expanding the

engineering team (hiring 2–3 additional developers or ML engineers), product refinement, and early pilot deployments at a handful of client sites. If the team plans to split between India (engineering) and a presence in SF (for business development), $1M can stretch far – India's cost base is lower, and only minimal US travel or salary would be needed at pre-seed. If the founders anticipate needing more staff (e.g., a dedicated data scientist for model fine-tuning, or additional full-stack developers to build a polished UI), or if they want a cushion to experiment with custom LLMs, **raising toward $1.5M** gives extra flexibility. Essentially, $1M is likely the minimum to hit meaningful milestones (MVP to some initial traction), and anything beyond that would accelerate development or extend runway to ~24 months.

- **Investor Appetite and Story:** Ceneca sits at an attractive intersection of trends (enterprise AI, data, on-prem), which should garner interest from pre-seed investors, especially those focused on B2B AI or dev tools. If there is strong interest (multiple funds/angels wanting in), aiming for ~$1.5M is reasonable and still within typical pre-seed norms. However, if the team can execute with lean resources, staying closer to $1M avoids unnecessary dilution early. We recommend preparing a plan for ~$1.2M as a middle-ground, which could be adjusted up or down based on interest. Notably, raising much above $1.5M at pre-seed could push valuation expectations higher and make a follow-on seed round trickier unless growth is spectacular – so it's usually wise not to over-raise at pre-seed. Since **most pre-seeds are under $900K** [16], a $1M+ raise for Ceneca would signal strong conviction and likely require a lead investor with deep pockets in AI. This amount aligns with what we see for enterprise AI infrastructure startups (many raise ~$1M to build an MVP then go for a larger seed when they have pilot customer validation).

- **Use of Funds:** With ~$1–1.5M, Ceneca can allocate budget roughly as: engineering salaries (the bulk, to refine the NL-to-SQL models, expand database connectors, and harden the platform for on-prem installs), some spending on cloud or on-prem infra for development/testing (perhaps setting up a lab with various databases, and possibly GPU resources if fine-tuning LLMs or vector DB work is needed), and business development efforts (travel to meet design partners, maybe a small presence at an industry event or two, marketing website content). At pre-seed, heavy marketing/sales spend is not expected; the focus is building a product that wows initial users. So the ask is primarily to fund **product development and pilot execution**. $1M could fund a team of ~5-6 people for 18 months (assuming a mix of SF-level and India-level salaries). $1.5M might allow hiring a few more key people (e.g., a UX designer or a solutions engineer) or simply provide 6 more months of burn, which can be valuable if enterprise sales cycles are long.

- **Valuation and Equity in this raise:** If Ceneca raises ~$1.2M, we would expect the pre-money valuation to be in the ballpark of **$6–8M** (giving ~15% to the new investors). For example, $1.2M on $6.8M pre ($8M post) would mean investors get 15% post-money. This fits within the average pre-seed valuation range [19]. Given the deep tech nature, some investors might accept a slightly higher valuation if they believe in the team (pushing pre-money to maybe $8–10M). But as first-time founders (if that is the case), staying grounded around ~$6M pre might make it easier to close the round. The *global/India-to-SF* dynamic could allow a bit of premium if pitched as "we already built a robust prototype with minimal capital" – essentially showing high efficiency could justify valuing the company more. Nonetheless, targeting a realistic valuation that leaves room for a healthy step-up at Seed is prudent. So, **suggested ask: $1.2M on ~$7M pre-money** (adjust as needed in negotiation). This communicates confidence but not excess.

In conclusion, **raising on the order of one million dollars is appropriate for Ceneca's pre-seed**. It provides enough capital to reach the next value inflection (product-market fit signals, some paying beta customers) without undue dilution. The benchmarks indicate this is slightly above median but well within normal for a venture-backed pre-seed, especially in the AI sector where 10-15% of rounds do

exceed $1M [21]. The exact figure can be fine-tuned (e.g. $1.0M vs $1.5M) based on interest levels, but anything in that range sets Ceneca up with ~1.5 years of runway. Importantly, the pitch to investors should clarify that this amount gets the company to specific milestones (e.g. "with $1.2M we will deploy at 2 Fortune 500 design partners and build an open-source community edition, positioning us for a $ seed round in 18 months"). That will assure investors that the ask is grounded in a roadmap.

Key Slide Recommendations for the Pitch Deck

For a pre-seed investor pitch, **Ceneca's deck** should be concise (10–12 slides) and hit all the key points that VCs expect. Here are the essential slides and what they should cover:

- **Team:** Introduce the founding team and key advisors. Highlight relevant experience (e.g. expertise in AI/NLP, enterprise software, prior startups or domain knowledge in data analytics). Investors at pre-seed bet heavily on people, so showcase why *this* team can execute this vision (technical skills, industry connections, complementary co-founder roles). If the founders have notable accomplishments (publications, open-source projects, etc.), include those briefly.

- **Problem:** Clearly articulate the problem Ceneca solves. For example: *"Data-rich companies struggle to extract insights quickly because business users can't query databases directly and data teams are a bottleneck. Existing BI tools are too complex or require moving data to cloud, creating friction and compliance issues."* Frame the pain in terms of lost time, lost opportunities, or costs. Use a simple story if possible (e.g. a data analyst spending days writing queries for executives, or a regulated bank unable to use cloud BI for customer data). This slide should make investors *feel* the pain in the current status quo.

- **Solution:** Present Ceneca as the solution to that problem. Explain *what the product is* and *how it works* at a high level (an AI agent that lets you talk to your data, deployed on-prem). Emphasize the key benefits: *natural language interface (no SQL needed), direct connection to internal databases (no data migration), fast insights, and privacy/security by design.* If you have a product screenshot or diagram, this is a good place for it. Show a quick example: e.g. a user asks "What were our top 5 products by revenue last quarter?" and how Ceneca returns the answer from the database. The goal is to make the solution tangible and differentiated. Keep it focused (don't dive into technical architecture deeply on the deck; that can go in an appendix or demo).

- **Market:** Define the market opportunity. Use some of the TAM/SAM/SOM points – e.g. "We're targeting the $X billion enterprise analytics market" [2]. Highlight trends: "Data is growing exponentially, and 80% of enterprises want generative AI in analytics [7], but 50% can't use cloud solutions [5]." Essentially, convey that the **market is huge and growing** and that **now is the right time** (because of AI momentum and unmet on-prem needs). If you have a specific initial target segment (say, pharma companies or mid-market banks), mention how many such companies exist and their typical spend – to make the opportunity concrete. Investors want to see both a big vision (huge market) and an entry point (how you start slicing into it).

- **Product (Tech) Details:** Depending on the audience, you may include a slide on "How it works" or the tech stack. This could be a brief architecture diagram showing components: e.g. Ceneca's interface, the NL->SQL engine, the connectors to DBs, etc. You might also highlight any early traction or prototype status here ("Working alpha connecting to Postgres and Mongo, with 100+ sample queries successfully handled"). If you have a demo or pilot, mention results (e.g. "in a pilot, we enabled 5x faster data querying for analysts"). Keep text minimal; use visuals to indicate

the workflow. This slide essentially reinforces that *Ceneca is not just an idea; it's a real product in the making* and that the team has a handle on the technical implementation.

- **Go-to-Market Strategy:** Explain how you plan to acquire and grow customers. Early on, this might be via **direct sales or pilot deployments** in a few friendly organizations (maybe leveraging personal networks or industry contacts). For instance, "We will start with 2-3 design partners in banking and refine the product, then expand sales to Fortune 1000 companies in regulated sectors." If you plan an open-source or community edition to drive adoption (common for developer tools), mention that strategy (e.g. "Offer a free open-source core to build developer community, then upsell enterprise features/support"). Also address who the buyer/user is – e.g. will the Head of Data or CIO buy Ceneca? Or will analysts adopt it bottom-up? A credible go-to-market might be: *Pilot in one department -> expand usage internally -> enterprise license.* If relevant, mention partnerships (maybe partnering with cloud providers for marketplace listing, or consulting firms for integration). This slide assures investors you have a plan to reach the market, not just build tech in a vacuum.

- **Business Model:** Outline how Ceneca will make money. Likely this is a **B2B SaaS or license model**. Even if on-prem, you can sell subscriptions for software licenses or support. For example, "Annual enterprise license per server or per user" or "subscription based on number of data sources and users." You might say, "We plan to charge **$50K–$100K annually per enterprise deployment**, which is in line with what companies pay for BI tools." If you have tiers (community vs enterprise) or a usage-based component, note it. The business model should tie to the market: e.g., "With 100 target customers at ~$50K ARR each, that's a $5M potential – showing room to scale revenue." At pre-seed, this can be tentative, but showing you've thought about pricing and revenue generation is important. Also, if you're leveraging open-source, clarify the monetization (e.g. open-core model, paid premium features or services).

- **Competition:** A slide to acknowledge competitors and articulate **Ceneca's unique edge**. List key competitors (ThoughtSpot, Tableau, PowerBI, Superset, MindsDB, etc.) in a table or quadrant, and highlight how Ceneca differentiates [22] [10]. Often startups use a quadrant or feature comparison chart (with checkmarks) to show where all others fall short on certain criteria. For example, a table of features: "Natural Language Query, On-Prem Deployment, Multi-DB Support, Requires Data Move?, etc." and show Ceneca as the only one that ticks all the boxes needed by a certain customer segment. The point is to assure investors that, while there are big players, **there is a gap** that Ceneca fills. You can also frame competition as validation: "Big companies are investing in AI analytics (Microsoft, Salesforce) – validating the space – but their solutions force cloud usage or lack cross-platform ability. We target those who are left out." Keep the tone respectful but confident – acknowledge others but emphasize Ceneca's *differentiators* (e.g., privacy, simplicity, flexibility, etc.).

- **Roadmap:** Lay out the next 12–24 months of milestones. This slide shows what you will accomplish with the pre-seed funding. Break it into Q3 2025, Q4 2025, etc., for instance:

- *Q3 2025:* Complete MVP with Postgres, MongoDB support; deploy at 1 design partner.
- *Q4 2025:* Add support for additional databases (MySQL, Snowflake), implement user management and on-prem installer; 2 pilot customers using product.
- *Q1 2026:* Beta launch, early paid trials; integrate user feedback, improve NL query accuracy with domain-specific tuning.
- *Q2 2026:* General Availability of v1.0; first $ revenue from 3 customers; begin seed fundraising.

Having a roadmap instills confidence that the team has a focused execution plan. It also helps investors see when the next funding round might be and what progress will be made by then. Emphasize critical milestones like product releases, user metrics, or any POCs converting to paying customers.

- **Fundraising Ask:** Conclude with what you are asking for and what it will achieve. For example: "**Raising $1.2M** seed capital to reach 12-month milestones" – list use of funds: "hiring 3 engineers, launching 5 pilots, achieving $100k ARR" (whatever the goals are). Specify the round type (SAFE or equity) if known, and any notable current commitments (e.g., "$500k committed from XYZ Ventures and angels" – if applicable – as that can trigger urgency). Also, if you have room, you can mention projected runway ("gives 18 months runway"). Essentially this slide is the **terms and purpose** of the round. It should match what you discuss verbally. Keep it simple: amount, broad allocation (product, go-to-market, etc.), and the key outcomes you plan to deliver with that capital (which tie back to the roadmap). Make sure the **ask is consistent** with earlier narrative (if you said you want to hire or do X, ensure the money requested covers that).

These slides cover the basics. Additionally, some founders include **"Vision"** (what's the big picture 5 years from now), or **"Why Now?"** (market timing) explicitly, though those points can be weaved into Problem/Market. For a deep tech startup, sometimes a slide on **Technology/IP** (if there's proprietary AI innovation or patents) can be included to show defensibility. Also, **Customer quotes or LOIs** (if you have any endorsements from pilots) can add credibility. However, given a pre-seed deck should be tight, the above list is likely sufficient. Remember to keep slides uncluttered – minimal text, more visuals – and use the deck to support your storytelling. Clarity and focus on these key aspects will make the pitch more compelling.

What VCs Look For at Pre-Seed (Especially in Deep Tech AI Startups)

Investors evaluating a pre-seed company like Ceneca will primarily be assessing the **potential** – since early-stage startups won't have much revenue or traction yet, VCs focus on qualitative factors that indicate a high chance of future success. Here's what they typically look for, with an emphasis on deep tech/applied AI startups:

- **Team & Founder "Market Fit":** The team is often the #1 factor. VCs will ask: *Does this founding team have the capability to build this ambitious product and adapt as needed?* For Ceneca, having founders with strong AI/ML backgrounds or significant experience in data analytics is a big plus. If the founder is a ex-Googler who built internal data tools, or an AI PhD, that instills confidence. Deep tech investors also value teams that can cover both research and engineering – i.e., they can not only prototype an NLP model but also ship enterprise-ready software. Since pre-seed is before meaningful traction, VCs are essentially **"betting on the jockey."** They look for passion, domain expertise, coachability, and a clear vision from the founders [23]. A track record (previous startup or relevant project) helps but is not mandatory if the vision and expertise shine through.

- **Clear Problem & Big Vision:** Investors want to see that *Ceneca is solving a real, significant problem*. The pain point should be clearly articulated (as mentioned in the deck slides). If it's a nice-to-have, it'll be less compelling. They will gauge **market need** by thinking: *"If this works, will a lot of companies desperately want it?"* [23]. In deep tech, sometimes founders get enamored with the technology – but VCs will probe if there's a real use case and demand. At the same time, they appreciate **visionary thinking**: does this startup have the potential to create a new category or be a billion-dollar company one day? For Ceneca, the vision could be *"Every enterprise has an AI assistant for data – we want to be that default AI analyst across the Fortune 1000."* Demonstrating a

huge TAM (discussed earlier) and a pathway to capturing it, even if starting niche, is important. Essentially, VCs look for **ambition with credibility** – a grand vision backed by understanding of immediate steps.

- **Early Product/MVP Progress:** While pre-seed means early, having an MVP or prototype significantly de-risks the investment. VCs will be impressed if the team can show a working demo of Ceneca querying a live database in natural language and getting correct results. It proves technical feasibility. They will ask: *What has been built so far?* The further along, the better – e.g., a prototype with two database connectors and a basic UI is great for pre-seed. They'll also value any early user feedback: *Has the team talked to potential customers?* Letters of intent, pilot user feedback, or even a couple of beta users are gold at this stage. It signals demand and reduces market risk. In short, VCs look for evidence that **the team can build stuff fast** and that the product works at least at a basic level [24] [25]. For deep tech, this also shows that the technology isn't just theoretical. If parts of the tech are still research-y, investors will consider the risk – so demonstrating a functional NL->SQL conversion (even using an existing LLM with clever prompt engineering) can show that the concept is sound.

- **Differentiation / Moat:** Especially for AI startups in 2025, VCs are aware of the *competition and copycats*. They will question: *What is unique about your approach? What's stopping a big company or another startup from doing the same?* Since Ceneca is leveraging AI (likely large language models), a common concern is whether the startup has any **defensible IP or data advantage**. At pre-seed, you might not yet, but you should articulate a path to a moat: e.g., proprietary fine-tuning on unique query-corpus, building a library of enterprise schema knowledge, or simply first-mover in on-prem AI agents giving a distribution advantage. Deep tech investors like to see that the startup could eventually have a **technical moat** – be it algorithms, patents, or even just a head start in integrating into customer workflows. For now, a strong differentiator is the on-prem capability and the full-stack nature (others might focus just on cloud, or just on one DB, etc.). Make sure to highlight how that gives an edge (e.g., "We can land customers that others can't serve due to security constraints"). Also, being aware of competition (and ideally having a better strategy or tech) is critical – investors often ask "Why won't Microsoft or another AI model own this space?" You should have an answer that might involve agility, focus, or a unique understanding of user needs.

- **Market and Business Potential:** As discussed, the **market size** and the **business model** need to be attractive. VCs at pre-seed might not expect detailed financial forecasts, but they want to know this could be a venture-scale business (10x+ return). For deep tech, they are willing to wait longer for monetization, but they want to see a line of sight: e.g., enterprise SaaS is a proven business model if you can get enterprise adoption. They'll look for indications of **high growth potential** [26] – for example, if one bank uses Ceneca and loves it, will it virally spread to 10 other banks via word-of-mouth? Or will each customer expand usage organization-wide (land-and-expand potential)? They also consider timing – *Why now?* If AI enabling NL queries is newly viable (thanks to GPT-4 etc.), then now is the time to build this company (which is good). If on-prem is a shrinking need, that could worry them – but evidence suggests many still need on-prem, so frame that as a persistent or even growing niche due to data privacy trends.

- **Traction or Early Validation:** While revenue is rare at pre-seed, any **early traction** significantly boosts credibility. This could be pilot users, a waitlist of interested companies, partnerships, or even an open-source GitHub stars count if relevant. Even qualitative letters from potential customers saying "If you build this, we'd use it" help. VCs often ask if there are design partners or paid POCs lined up. Having one or two initial deployments (even if free trials) by the time of seed is ideal, so at pre-seed they look for your plan to get there, and any initial signs that *customers*

care. Deep tech startups sometimes focus too much on tech and not enough on customer development – showing that you've engaged real users (maybe through a pilot or user testing of the prototype) will set you apart. It demonstrates customer obsession and reduces market risk for the investor [25].

- **Specific to Deep Tech/AI:** Investors in AI will want to understand the **technical risk** and how you mitigate it. For example, they might ask: *What if GPT-4 is not reliable for SQL generation? Do you have in-house NLP research to improve accuracy?* Have a sense of your tech roadmap (maybe fine-tuning smaller models for on-prem use later, etc.). They may also discuss the compute or cost: *Can this run entirely on-prem without internet? If so, do you need to ship a large model? How will that be efficient?* Be ready to address these with either current answers or a vision ("Today we use OpenAI via API with anonymized schemas; in future we plan to develop a lightweight local model once our training data from usage grows"). Show that you understand both the **capabilities and limitations of AI** in your product. VCs also often check if an AI startup is just a "thin wrapper" over someone else's model. You want to convince them that, over time, you will develop **proprietary technology or data** that makes you more than just a UI on GPT. This could be by accumulating a unique corpus of enterprise queries and results, or integrating with systems in a clever way that's hard to replicate.

- **Coachability and Vision Alignment:** At pre-seed, many VCs (especially those who lead rounds and take board seats later) consider whether the founders are open to feedback and able to navigate the unknown. Deep tech founders sometimes lean academic; investors will want to see you also value business input and can iterate on go-to-market strategy. They will also gauge your **commitment**: is this a side project or are you *all in*? They generally seek founders who are **mission-driven** and resilient, because deep tech startups can take time to fine-tune product-market fit. Conveying a balance of confidence and openness can signal you are the kind of founder they can work with throughout the risky early phase.

In essence, VCs at pre-seed for a startup like Ceneca are looking for **a stellar team, attacking a big problem, with an innovative solution, and early proof that it's feasible and in demand**. They accept that many details (exact pricing, precise target customer) may evolve, but they want to see that the core ingredients are there. Particularly for applied AI, they will examine if this is a **"painkiller" solution (must-have)** rather than a novelty. Demonstrating that a few potential customers have shown strong interest can go a long way. Lastly, they will look for the **story**: a clear narrative of why Ceneca will be the one to succeed in this space.

By addressing these points in your interactions (and in the deck), you can hit the notes investors expect. For example, highlight: *Team's unique data/AI expertise (team), the inefficiency and demand for NL analytics (problem), how Ceneca works and initial success in making it work (solution + MVP), the large number of companies with this need (market), how others are missing the mark (competition), and how you'll capture this opportunity (go-to-market & business).* At pre-seed, convincing investors essentially means convincing them that **"this could be huge and we are the people to do it"** despite the early stage – everything you present should reinforce that narrative. [23] [26]

Sources:

- Ceneca Product Repo – *"Data Connector: Natural Language to SQL Query Agent… allows users to query databases using natural language… translates questions into SQL and provides analysis."* [1]

- Augmented Analytics Market – *Global market projected to grow from $13.62B in 2024 to $16.51B in 2025 (21% CAGR), reaching $41.23B by 2034* [2].

- Cloud vs On-Prem Adoption – *55% of organizations still rely on on-premises systems (2023)* [5], showing many enterprises need on-prem solutions despite cloud growth.

- BI/Analytics Market Size – *Global BI and analytics market expected ~$84.6B by 2026 (14.7% CAGR)* [4].

- ThoughtSpot Sage (Competitor) – *ThoughtSpot's new GPT-powered search experience for analytics* [9].

- Salesforce/Tableau GPT – *Salesforce introducing Tableau GPT to let users ask questions and auto-generate insights (generative AI in Tableau)* [10].

- Microsoft Power BI Copilot – *Microsoft's Copilot in Power BI uses generative AI to help uncover insights faster* [11].

- Apache Superset – *Open-source on-prem data visualization platform for exploring and visualizing data (BI alternative)* [12].

- MindsDB – *"enables … query data in natural language and SQL, and get highly accurate answers across disparate data sources"* [14] – an AI query engine for databases.

- Pre-Seed Round Stats – *75% of pre-seed rounds < $900K; ~40% < $250K* [16]*. Median SAFE raise $275K at $10M cap* [17]. Only ~10-15% of pre-seed raises exceed $2.5M [27].

- Pre-Seed Valuations – *Average pre-seed pre-money ~$5.7M, median $5.3M (Pitchbook 2023)* [19].

- What Investors Look For – *At pre-seed, investors focus on factors like quality of the idea (solving a real problem), MVP progress, and strength of founders (relevant experience)* [23], rather than financials or traction.

---

1 GitHub - hardikSrivastav/data-connector
https://github.com/hardikSrivastav/data-connector

2 3 Augmented Analytics Market Report 2025- Augmented Analytics Market Share, Industry Trends, Size 2034
https://www.thebusinessresearchcompany.com/report/augmented-analytics-global-market-report

4 Business Intelligence Statistics and Facts (2025) - Market.us Scoop
https://scoop.market.us/business-intelligence-statistics/

5 90+ Cloud Computing Statistics: A 2025 Market Snapshot
https://www.cloudzero.com/blog/cloud-computing-statistics/

6 Agentic Analytics Heart of the Autonomous Enterprise | IIA @MIT 2025
https://www.youtube.com/watch?v=jvbsCsqrRFg

7 8 10 Salesforce Introduces the Next Generation of Tableau, Bringing Generative AI for Data and Analytics to Everyone - Salesforce
https://www.salesforce.com/news/stories/tableau-einstein-gpt-user-insights/

9 Introducing ThoughtSpot Sage: AI-Powered Analytics with GPT
https://www.thoughtspot.com/blog/enhanced-ai-powered-analytics-with-gpt

11 Introducing Microsoft Fabric and Copilot in Microsoft Power BI
https://powerbi.microsoft.com/en-us/blog/introducing-microsoft-fabric-and-copilot-in-microsoft-power-bi/

12 Apache Superset - Wikipedia
https://en.wikipedia.org/wiki/Apache_Superset

13 MindsDB: AI that works at work. Turnkey. Private. | Y Combinator
https://www.ycombinator.com/companies/mindsdb

14 22 Minds[DB] - AI's Query Engine - MindsDB
https://docs.mindsdb.com/what-is-mindsdb

15 AI agent to chat with database and generate sql, charts, BI - Reddit
https://www.reddit.com/r/dataengineering/comments/1i5jng3/ai_agent_to_chat_with_database_and_generate_sql/

16 19 21 23 24 25 26 27 Pre-seed valuations in 2025: What founders need to know
https://www.zeni.ai/blog/pre-seed-valuations

17 18 State of Pre-Seed: Q3 2024
https://carta.com/data/state-of-pre-seed-q3-2024/

20 Indian AI Startup Tracker: Meet 140+ Startups Putting India On The Global AI Map
https://inc42.com/startups/indian-genai-startup-tracker/