CONFLUENT

Confluent Thesis

Recommendation: I have taken a position in Confluent at $16.2 a share (~20% discount to today's price) and believe that at that price I am being compensated (20% 5-year IRR) for the competitive risk in the business. However, we top out at $16.2 and would only recommend adding to our position slightly below our cost basis (>25% discount to today's price). Confluent is at a cross-roads in their competitive evolution, having changed their price / performance selection & fully launching their streaming platform. We believe that at our cost basis, market prices do not reflect Confluent's upside ability to strengthen their competitive advantage, enabling them to become a highly thematic data infrastructure *platform*.

Thesis: Below we outline a simplified version of the key contentions

- Kafka will be on the right side of history:
  - Infrastructure companies must stay relevant and open-source communities are the lifeblood for keeping a project modern & active. These communities create entry barriers, given the community network effects + developer mindshare, and they contribute to source code innovations. Kafka continues to be relevant, with monthly project contributors reaccelerating over the past few years (+38% L2Y)
  - Data streaming should benefit from the increase in AI workloads given the type of customers who will be most-likely to build these products (e.g., digital native / tech-forward clients) and the applications that are tightly coupled with real-time data, like customer service + search
- Confluent's cross-sell has strong industrial logic & a compelling fan:
  - Confluent has the opportunity to improve their unit economics / profitability by selling additional products to their base. Customers need to get data out of systems before they worry about governance and processing, and Confluent has shown that they can cross-sell data integration in their on-prem base (*80% attach*). Cloud has unique dynamics given the wider range of data sources & targets, but Confluent has been able to daisy chain "Confluent Connect" with other products (80% of cloud connect users leverage a second product). We model platform attach for one product reaching ~40% (28% today), two products 30% (18% today), and three products at ~15% (6% today) - driving a ~113% DBNR over the hold / (*~16% upsell ARR CAGR '24-'30*)
  - There is competitive risk in the connector / product adjacencies, with several partners, like MongoDB – the most popular connector by customer count, launching competitive products. These competitive inroads are offset by recent partnerships with share gainers, like Databricks & Postgres databases, which position Confluent as a "Data Switzerland"
- Confluent's competitive positioning against CSPs / Ankle Biters has significantly improved
  - Confluent's new logo adds have fluctuated significantly over the past few years with gross new logos peaking in 2021 at ~1.5k and dropping to ~714 logos in 2023. Over the past year, Confluent has changed their price / performance curve by introducing freight clusters + acquiring WarpStream, which offer ~50%+ infrastructure cost savings and open up Confluent to a whole new set of latency tolerant / high throughput Kafka users (*network centric*)
  - This has been a historical battleground where players like AWS + Red Panda have been able to find success, but with an expanded selection tailored for these log + telemetry use cases in tandem with the complete platform, Confluent has the right to increase their win rate 750-1000bps and drive new logo ARR acceleration (*3% CAGR '24-30*)
- Operating Margins should continue to scale: Confluent has successfully driven to non-GAAP operating margins due to leverage in shared functions & some cost-reduction in S&M. As they continue to mix into adjacent products and drive SoW with existing customers (ARR & Revenue '24-'30 CAGR at ~19% & 18% respectively), this should unlock ~20%+ non-GAAP EBIT margins
- ARR is tightly coiled: Expansion dollars & DSP attach depend on the magnitude and composition of go-forward cohorts, that we model re-accelerating due to improved pricing & packaging. This dynamic is compounded by the upsell dynamics, where Connect is a pre-requisite to other modules
- Management has shown deep technical expertise, but needs to execute on GTM for a home-run outcome

Intro: Confluent is the commercial parent of the Kafka project, a decades long open-source collaboration that enables real-time distributed systems via the decoupling of senders & receivers. The product is highly technical & monetization depends on compute, storage, and network requirements for distinct workloads. This is a name that we have been tracking for the past year given our interest in real-time digital experiences that require real-time data & the modernization of legacy systems where important institutional information is at rest.

David Morales Lam

Opportunity Cost: Confluent is one of several open-source data infrastructure point solutions, and the whole basket seems to be experiencing the same re-acceleration question (*although Mongo's Q2 earnings seems to have temporarily ameliorated investor concerns*). Ultimately, we believe all the companies in this basket are battling with a similar question: will AI workload / technology re-accelerate & strengthen our positioning against CSPs & ankle-biters. Here is how we have thought about the opportunity cost:

- Vs Elastic: We lean into Confluent's stronger competitive positioning against the CSPs
- Vs Mongo: We think that these are actually tightly coupled bets, but Confluent is nearly half the price (*see more in the Confluent "MongoDB" connector section*). Moreover, Mongo has to compete with the rise of Postgres, which Confluent is insulated from.

Jugular Questions:

- What does terminal adoption look like for their adjacent products?
- Will new logos re-accelerate due to pricing / performance changes and expansion impact?
- How will recent pricing and packaging changes impact Confluent's SoW with existing customers?
- What will end-state margins be?

We believe these are the appropriate jugular questions for two reasons:

Kafka's Dominance: This is one of many messaging / data transfer approaches, with IBM + Tibco providing popular message buses. While there are alternative methods to send real-time data, like message queues or brokers – **Kafka by far is the dominant approach with 80% of F500 companies using the platform & over 50k[1]+ companies listing Kafka practitioners**. Moreover, Kafka has 2x the number of open-source contributors / month vs. RedPanda & 4x RabbitMQ – with a decade more of history vs. the former[2]. Given the existing base, conversion & endurance of different OSS / CSP Kafka segments is more important to the underwrite vs. a totally new Kafka user. Kafka's relevance is driven by both the inherent technological advantages it has against legacy systems and enduring community it has built.

Operating Leverage: Confluent currently spends $726M to generate ~$1B of revenue, with roughly 42% of that spend being allocated towards S&M – note that this is on a non-GAAP basis (on a GAAP basis they spend roughly ~$1.2B, with a significant portion tied up in S&M headcount). The nature of software means that these FTEs function like operating assets over time – especially QBSRs that have a semi-fixed outlay where incremental dollars deliver high CMs. Thus, we think margin & growth, and in particular the type of growth are highly reflexive.

Product 101: Kafka is a real-time broker, that decouples services from interacting simultaneously and instead enables **asynchronous system interaction**. Put simply, Kafka helps you drop off data, such that the producer of this data is not 'kept' sending the message until it's received, and multiple receivers can automatically consume this data. To send 'low-latency' data you need to know a) where that change occurred, b) what constitutes a relevant data change, and c) who is going to receive that data. The kafka broker 'listens' for relevant changes, as defined by the user, and neatly organizes the data into topics – which are ordered logs. From there, any consumer who subscribes to this channel will receive that data update idempotently (*message is written once into the topic even if it sent multiple times*). This creates two benefits for development teams: 1) real-time data is handled by the broker, abstracting the IDing of system changes and data organization; 2) scalability is inherent to the system because consumers of data simply need to subscribe to a centralized broker as opposed to dealing with independent upstream services

On the infrastructure side, Kafka takes advantage of distributed systems to allocate processing to the best positioned clusters – set of computers working together as one unit. Kafka brokers write events to a partition, within a topic, within a cluster (e.g., "North_America" partition; "Online_Orders" topic, "Cluster_A"). As you can imagine there's a lot of complexity around handling these logs & properly allocating them to clusters – especially when clusters go down. This is where Confluent started to monetize. Pricing for streaming is based on consumption driven by number of clusters and their latency, throughput, data retention policies, etc.

If the typical request-response paradigm is a phone call where both people need to pick up to transfer data, and a message queue / broker is like an email where you still need to identify the receiver prior to sending, then Kafka is a newsletter engine that allows users to easily sign up for updates without approval from producers

[1] LI lists 500k+ Kafka engineers non self employed, we assume ~10 Kafka engineer per org for ~55k Kafka enterprises

[2] *Selecting the right metric for activity in open-source projects is tricky, for instance – more lines of code may mean less efficiency. We index on the number of contributors per month (avg. of L3M) to gauge community activity – whereas metrics like commits may be driven by power users / contributors*

- Phone Call = Request / Response
- Queue = Email
- Kafka = Newsletter

In addition to core streaming, Confluent has focused on widening the breadth of their platform offerings for customers. They have focused their efforts on four different products

Connectors: Pre-built data transfer tools, that enable users to easily link data from & to Kafka brokers. Over the past 5 years, Confluent has focused on simplifying the usability of these connectors (no-code) + expanding the range of offerings for ecosystem tools, with a recent emphasis on AI data tools (e.g., Anthropic, Langchain, etc.). They currently offer ~120 pre-built connectors for tools like Kinesis, PostgresSQL, Salesforce, and MongoDB. This is by far Confluent's most widely adopted add-on product with ~50% of Confluent customers using connect[3], and this traction is unsurprising given the industrial logic of driving efficiency in deployments via data capture abstraction. Users with the labor bandwidth to create these connectors or relatively consistent data sources / targets see less value in a solution like this. For instance, if you already have a polling service or a pre-built method for hitting the salesforce API and the data model is simple, there is limited need for a managed service from Confluent. Priced based on throughput, latency, and number of connectors

Flink: This is the popular open-source stream processing framework that was started ~3 years after Kafka and has emerged as the default runtime engine for handling real-time compute. In 2023, Confluent acquired Immerok – the leading Flink managed service with several of the project's leading contributors. This helps organizations run queries / actions on the real time data stored in Kafka brokers, for low latency processes (e.g., fraud detection). Priced based on # of functions (e.g., think excel formulas)

Govern: Allows users to implement data policies that standardize & ensure data pipelines via data contracts, data lineage, and data catalog. Similar to data governance tools like Informatica's MDM solution, but in a streaming context. Priced based on # of schemas created and applied

Tableflow: Newly released product for Apache Iceberg tables, that automatically transforms events into tables. Can be thought of as a productized connector for any stream, that targets various end-analytical warehouses, like Databricks. Priced similar to connectors

Why Does This Matter For AI?

While Confluent should benefit from increased AI investment, we think the benefit is a bit more nuanced than clear AI winners like Databricks, Nvidia, and TSMC. AI's impact on Kafka usage largely depends – on *the % of new services that have real-time SLAs / communication patterns*. We further contextualize the drivers below:

AI Investment: Surveys indicate an overwhelming excitement for agent adoption with 70-80%[4] of enterprises developing and launching their own agents over the next few years. At their core, agents are simply autonomous systems that use LLMs to help prioritize & execute tasks, leveraging tools like APIs to access different data stores & the internet. From an infrastructure perspective, these agents aren't too distinct from a microservice, they represent another workload that may need real-time data to take actions.

Architectural Rationale: There are two technical considerations for Confluent's relevance in AI workloads

- RAG: Short for retrieval augmented generation, which is just a method for enriching prompts with enterprise context via embeddings in a vector database to help prevent model drift. The bearish side of the RAG debate points out that as context windows scale, i.e., LLM's ability to handle more tokens, there is less of a need to externally enrich prompts. Imagine you are studying for a test, the RAG bears would say that the smarter student with more memory capacity doesn't need their notes to help them navigate through the test. However, the reality is that context windows do not scale linearly and typically model performance degrades with longer context windows as the model struggles to parse through the vast # of tokens & potential distractors. Moreover, longer context windows, inherently raises the # of tokens & costs + compliance risk that enterprises are likely to be sensitive towards.

[3] Confluent Investor Day 2025
[4] https://www.langchain.com/stateofaiagents?utm_source=chatgpt.com + https://www.allganize.ai/en/blog/ai-agents-are-rising-ai-agent-integration-companies-to-grow-82-within-3-years#:~:text=Agents%20Are%20Rising!

- Agent Communication: In this context, agents can be likened to micro-services, with individual nodes that are running their own task and need to communicate with other agents, for instance to upload, process, and push a PO order to an ERP. Confluent is the connective tissue that helps store information created by an upstream agent, say continuous PO data, and then distributes it to downstream agents to ensure resilience & real-time execution.

Crucially, if a system was not real-time before-hand it's not as if Gen AI automatically makes it latency-sensitive, rather these architectural patterns help illustrate the incremental pipelines / tasks that Confluent can be expected to run in existing real-time systems.

Applications: We hypothesize AI should help disproportionately drive real-time services

- (+) Consumer facing SLAs: AI has helped introduce a new wave of workloads[5] targeting customer service, voice agents, and real-time search. While some of these explicitly serve consumers, AI applications blur the lines in the workforce with several companies like Cursor, OpenAI, and Anthropic existing in both personal & professional workflows. Real-time responses for these new workflows play a crucial role in driving consumer experience. Across each of these domains, consumer satisfaction is almost directly correlated with latency + accuracy
    - Search (Google = slowing down search by 100 to 400 m/s leads to 20-60bps fewer searches – from 2009[6])
    - CX (ZenDesk = 72% of customers expect immediate service)
    - Voice (latency above 4 seconds significantly degrades quality of experience / with leading players targeting 300 m/s p50)

In a LangChain survey of 1k+ professionals, latency was ranked as the fourth biggest blocker to enterprise-wide agent deployments, by 15% of organizations. When looking only at MM / ENT, latency climbed to the third most cited blocker with 17-18% of respondents, behind performance & cost.

Confluent should be a beneficiary of AI workloads given breakthrough applications will require real-time data for decision making. Consumers care about latency, and to fulfil search, voice, and cx SLAs Kafka will be required. While longer context windows & MCP (model context protocol) may reduce the need for basic RAG Kafka pipelines, the underlying latency SLA, which is determined by the type of application, still requires an event broker to create resilient & fault-tolerant systems. Moreover, Kafka is mixing towards customers that we think are more likely to launch agentic workflows, with digital natives that are more addressable thanks to warpstream + freight clusters (*more on this below*)

Competitive Section: The alternatives to Confluent largely boil down to the open-source version of Kafka (OSS), Hyperscaler offerings, & anklebiters (e.g., Red Panda). The open-source alternative gets a deeper dive in the following section given the popularity of the OSS version.

Hyperscalers: Based on market commentary the most formidable hyperscaler is AWS, given their set of infrastructure focused services, early release of Kinesis (in early 2010s), and dual offering of Kinesis (AWS proprietary streaming) + MSK (AWS managed Kafka).

AWS's ICP are AWS only shops (*no need for multi-cloud*), with moderate-high understandings of Kafka, and latency-tolerant workloads where the customer can tailor their deployment. This segmentation exists for both logos & workloads, meaning that AWS can win logos or a share of deployments.

A former AWS streaming engineer characterizes their ICP vs. Confluent

> ***For other folks that have already invested pretty heavily in Kafka, that Kafka is not new to them, they don't need all the bells and whistles as people will say. MSK can be a compelling offering because they can leverage more of what they know***. *They don't need all the easy-to-use features. In fact, it gets in the way of what they're used to using sometimes. I think MSK can be a good offering for folks there, too. And then just finally****, the obvious in that it can be an easier lift for folks when they're already working with Amazon to simply add another service. And I'm talking the bureaucracy of approved vendor and all the things that you have to do when you want to buy software at large organizations***. *These are some of the thoughts that I think about when approaching that question of MSK versus Confluent.*"
>
> Former Data Streaming Lead at AWS

AWS does not report Kafka revenue but we can estimate it with a few data points

Quantity: Industry reports estimate that AWS has >4M customers as of 2025. Being conservative, we assume that this underestimates true customer count, so we center around ~4.5M logos.

[5] LangChain survey: 58% use agents for search, 53% for productivity / personal assistants, and 46% for CX.
[6] We think that is has gotten more acute over time as users have become used to quicker internet / results

Addressability: We apply a few addressability cuts to our logo count

Single Cloud: The majority of published Kinesis / MSK customers on AWS's site are single cloud which tracks with market commentary (single-cloud; Zillow, Nextdoor, Poshmark). While the AWS market report places 95% of orgs as single cloud, there are a few dynamics at play that change the direction of the number. (-) The vast majority of the 95% are small logos, (-) large organizations are more likely to be multi-cloud given they have the resources and data volume to justify multi-homing, (+) AWS tends to have more customers be single cloud given the wide breadth of their infrastructure offerings

Thus, while market surveys place single-cloud adoption at ~20%, we think that AWS likely performs slightly better than their competitors so we center around 25% of non-SMB customers as single cloud

Spend: Huge spenders (1M+ of cloud costs) are most likely to adopt MSK / Kinesis, given this signals operational complexity & the resources needed to manage a barebones Kafka implementation. Based on market reports for spend by vertical we estimate that 1-2% of AWS customers spend greater than $1M a year on AWS. The sense check here is that at 4-5M total AWS logos, 1-2% of huge spend, and $1.5M ACVs this would mean that huge spenders make up ~$100B of AWS spend – which does ~$120B.

Vertical: Lastly, we apply % likelihood to buy based on the end-market with media and internet as fully addressable given their orientation towards real-time digital experiences, 75% retail given their mix of legacy + modern businesses, and 50% of manufacturing given their varied tech stacks.

Q: This leaves you with ~10-20k customers – where we then apply a slightly higher adoption vs. Confluent's (40%), given the easy "press-a-button" dynamic AWS services have

P: We take a 30% discount to Confluent's core streaming ACV (141k), due to several interviews that explain MSK is usually cheaper – justifiably so given the weaker AWS adjacent offerings (e.g., connect & schema registry).

Our final estimate lands us at around ~$525-625M revenue business for Kinesis (around ~50-60% the size of Confluent). One last pressure test is that AWS cites "tens of thousands of customers" on Kinesis. The resolution with our ~5k of huge spenders is that typically the top ~30-40% of customers drive the bulk of revenue for these infrastructure businesses – for instance, the top 40% of Confluent customers drive 95% ARR. When this dynamic is applied on our AWS analysis – the total inferred customer count is (5.5 / 0.4) -> ~12-15k customers, which fits neatly with cited numbers of adoption

An important nuance that Confluent's management would add is that the clouds can both be a top of funnel partner and competitor. They can help introduce clients to managed versions of Kafka, and when the customer decides they need more functionality like connectors or schema registry, they will then speak to Confluent.

- Connect: Both Kinesis and MSK offer limited pre-built connectors with MSK choosing to route folks to the open-source connector ecosystem, which naturally has fewer enterprise grade ready tools. While this may be worthwhile for mature Kafka organizations with the teams ready to build and maintain connectors, this stands in stark contrast to the pre-built enterprise ready connectors that you can easily plug in with Kafka
- Schema: AWS offers a basic data governance service via AWS glue that lacks a method for easily querying your data and tracking the movement of your data through pipelines. Moreover, you can only use the service in AWS, meaning you can't point it at non-AWS kafka pipelines
- Flink: For the most part, AWS managed Flink isn't too distant from Confluent's.

This co-opetition dynamic will be increasingly important going forward given some of the pricing & packaging changes, namely Warpstream + freight cluster, that Confluent has rolled out. One former Cloud employee explains that roughly 25% of Confluent cloud logos came from MSK / Kinesis conversion & in fact they were easier to sell to given their familiarity with a cloud managed version of Kafka vs. the DIY culture of open-source. In their most recent earnings call, Confluent pointed out they won ~25 new customers that quarter from one CSP (*we think Amazon given their size*). Given the massive book of business that AWS has and the organizational priorities that are likely pointed towards more mission critical parts of the AI stack – we view this as potential upside on new logo growth.

Red Panda vs. Confluent: RedPanda is the most formidable of the ankle biters (e.g., StreamNative, Hazelcast, Pulsar) having raised ~$265M, with their latest $100M round valuing them at $1B, 1/6th the value of Confluent. Red Panda's competitive risk is most acute in the digital native / MM segment, given their nuanced price / performance curve. However,

we believe the risk can be well-bounded thanks to the limited market that Red Panda serves and Confluent's acquisition of Warpstream.

The business was founded in 2019 by Alex Gallego and has several top VCs like Lightspeed and Sapphire ventures on their cap table. Based on channel checks we believe the business ended FY'24 around~$25-30 with 100% growth but a 2x burn rate. Their value proposition a few years ago was quite clear: we are a faster version of Confluent with less overhead and wider deployment possibilities – that can afford to be aggressive on price to win business often entering at 25-35% discounts. Below we outline the product innovations and competitive responses from Confluent

Performance: Estimates from SIs suggest that RedPanda is 15-20% more performant over Confluent in the cloud. The difficulty is that incremental performance is extremely niche – these need to be extremely low latency use cases like high frequency trading.

"**The latency could be better, yes, but also, I haven't seen like a ton of folks asking for better latency than what I get from Kafka right now. So, there's a concept of like what's good enough**, and then there's like if you're a low latency game, you wouldn't use any of this."

Financial Services Confluent + RedPanda Customer 2023

Zookeeper: This ensures that stream data is eventually consistent, and nodes are properly prioritized / pruned. RedPanda eliminated this service. However, this advantage has been eroded away since Confluent sunset ZooKeeper in 2024 for the cloud.

VPC*:* We discuss this in the Warpstream section but VPC basically means that the customer gets to retain elements of the deployment in their own cloud managed account (e.g., storage stays in customer AWS account). RedPanda enabled users to be more flexible with the CSP resources they routed through their cloud account vs. RedPanda's – a significant consideration for organizations born in the cloud or compliance focused.

While these are all genuine product innovations – the platform significantly lags Confluent from a connector (pre-built data integration tools) and data governance perspective, which limits usability for enterprise customers. Moreover, RedPanda started 7 years later than Confluent, who really innovated the Kafka protocol and therefore experience a stronger brand halo with larger organizations. One account executive explains what this means for Red Panda and their ICP

"You're selling to these back-end system teams and at an enterprise level, like **$500 million in revenue or above, most companies have, from what I've seen, already adopted some sort of streaming solution. That means they're either already using Confluent or they've already built their own open-source platform** that they're self-maintaining…**That's why really, if you look at Redpanda and how much we've sold over the past few quarters, the vast majority has been coming from commercial**. I foresee that continuing for the foreseeable future."

RedPanda Account Executive 2025

The difficulty for RedPanda is that they offered a few architectural advantages: performance, overhead, and deployment. The overhead has been eliminated with the latest releases of Confluent cloud. The performance advantage is genuine but the number of organizations willing to incur the downgrade in adjacent functionality and switching costs for performance upgrades are quite niche given how performant Kafka is already. Thus, our contention is that RedPanda's competitive positioning in their core MM / Digital native segment has worsened over the past year and their burn rate actually may get worse going forward, given their 2x burn rate happened prior to Confluent eliminating Zookeeper and acquiring Warpstream. The final contributor to this is the lack of stickiness that Red Panda experiences given their more nascent platform offering

Connectors: While RedPanda claims to have 300+ connectors, this includes "processors", which are really in-process functions (e.g., append or delete) that do not constitute a separate connector. Removing processors, we get ~75 connectors for RedPanda (>40% smaller than Confluent). Moreover, Confluent has a huge advantage in key ecosystems (CSPs, Databricks, Postgres, SF + ServiceNow)

Schema Registry: RedPanda does not offer this as a separate product instead the schema linking lives inside the broker reducing the overhead but limiting the governance operations to localized simple checks vs. cross-cluster / region governance

Processing: RedPanda offers their own stateful processing but it's much more nascent, whereas Confluent capitalizes on the most popular open source real-time processing engine with Flink as a managed service

This more nascent platform results in customers characterizing Redpanda as an easy lift & shift to Confluent

"**Like today, I could theoretically lift Redpanda out and replaced it with Confluent or my own Apache Kafka instance because all I'm using is the Kafka API and then the Confluent, SerDes and Kafka Connect stuff**."

Redpanda Customer 2023

They could continue burning and discounting significantly (~30-40% off Confluent prices) but that will inevitably cap their burn rate improvements. Assuming the balance sheet gives them ~$125M of cash (100M raise + 25M remaining from series C) – if RedPanda wants to continue doubling, that would mean they have ~18 months of runway assuming stable burn rates at higher dollars, a lean in assumption given the competitive moves Confluent has made.

So why would I use Confluent vs. Kafka

In addition to the typical enterprise certainty & code auditing that commercial vendors provide on top of open-source projects, Confluent offers adjacent modules (e.g., schema registry, connectors, and processing) in addition to simplified deployment options. Confluent cloud abstracts away three key technical problems for customers

- Virtual / logical resource partitions, organizations have to decide how they want to allocate virtual machines into physical machines and then how to track these partitions
- Tiered Storage, orgs have to decide which storage their clusters write to, in what time frame, and how long their storage resources retain that data
- Load Balancing, when a cluster goes down or is at capacity organizations have to formulate a method to replicate / move data to new clusters

All of these cloud complexities in addition to the security and enterprise considerations, means that in order to self-manage a solution, you either need to have an extremely simple deployment (<3 clusters) or staff a team of 15-20 engineers. The typical role required is a data engineer / platform engineer whose salary is 125k+. Thus, the ROI at <$3-5M is quite clear, save a whole engineering team's time by abstracting away the management and maintenance of these Kafka clusters.

However, this dynamic creates interesting examples where customers at the high end of this spend range end up deciding to in-source their development – there have been several customers that have felt that past $3-5M ARR is too expensive and subsequently in-sourced their deployment (Roblox, Lyft, Open-AI).

The nuance is that this price tag is only for core Kafka management, which does not include incremental time / engineering labor dollars required for additional services like Flink, Schema registry, and connect. While these are certainly value adds, we think the incremental time required to build out these tools once you have 15-20 engineers, at least for schema registry + connect, is likely an additional 3-6 months or ~$60k.

There are two different drivers that could change the ROI on in-housing

- Data engineers in demand: Data engineers help build, maintain, and monitor the data pipelines that feed into analytical databases or any hybrid services in operational applications. Thus, AI is poised to be a massive driver for the rise of hiring given the importance of preparing data for downstream agents[7]. Market surveys show that data engineers were the third most popular job posting related to AI in both 2023 and 2024. Their importance is reflected in their salaries with U.S. data engineers typically being paid around ~$150k, with a higher salary ceiling vs. other roles (*less than AI / ML engineers but more than front-end developers)*. Customers have added that right now the bottleneck is finding qualified / seasoned data engineers who can manage complex data pipelines – a trend that makes the tradeoff of time spent on Kafka vs. AI projects
- Increased productivity: The tradeoff here is the rise of data engineering productivity tools that should help smaller teams complete more tasks, while it's still early – we know that tools like Copilot and cursor significantly reduce the time to shift, and there are several data engineering specific features in the major analytical databases that help users build queries / checks easier. However, Kafka is likely more sheltered than other parts of the tech stack given it runs in production and deals with infrastructure heavy tasks, like provisioning servers / dealing with data retention.

Ultimately, we believe that the calculus should remain largely unchanged with >$3-5M ARR customers being at risk. Productivity tools will certainly reduce the number of engineers needed to manage a Kafka cluster, but the tradeoff is that this engineering time will be increasingly directed towards preparing data for enterprise wide AI deployments.

Network Centric Workloads

[7] The BLS groups data engineers & data scientists and estimates that demand for data engineers is set to grow ~30% overall + 78% of customers not ready to capitalize on AI due to data foundational issues

In addition to launching the cloud product, the team has been hyper focused on driving Kafka conversions from the 50k+ open-source enterprise users (management figure closer to 100k but we round down) by expanding support for different deployment structures

- Real-time event brokers often do not 'store' data – like how other API brokers may – because they want to maximize system processing & memory must be balanced with compute. Therefore, the customer has a key decision to make about where that data is stored – often influenced by latency requirements, cost constraints, and privacy concerns. Typically, confluent cloud would write log data to their managed disks in the cloud, SSDs, which have lower latency but tend to be more expensive & public. However, start-ups like Warpstream, which Confluent acquired last year, have created translation agents that ensure log data is written to customer managed S3 buckets (VPCs). Confluent has effectively extended their deployment architecture across cloud & on-prem, with a hybrid option (storage being in a private cloud) that targets more cost conscious / privacy focused customer with relaxed latency requirements but high throughput (IOPS).

- In addition to Warpstream, Confluent launched freight clusters, which serve a similar purpose for high throughput & relaxed latency requirements that are more cost conscious. While this is still deployed in the cloud, the changed SLAs open a new paradigm of price / performance customers, mainly focused on "observability, analytics, security, and IoT."

Confluent made their product significantly cheaper & more privacy centric for compliance focused organizations to easily manage their own storage

How Much Cheaper: Freight clusters + Warpstream are high throughput clusters that Confluent announced in 2024 and then fully released in 2025. Cost savings are achieved via autoscaling, which automatically allocate resources depending on cluster performance, and networking cost reduction. This cluster type saves on networking by writing data directly to cheaper object storage (in a VPC or cloud format) circumventing costly data replication charges (i.e., data being written across nodes to confluent managed cloud) & reduces AZ costs by pulling data clusters in their availability zones.
Simplification: Think of your laptop as a server, with an excel sheet (Kafka) gathering share price data. If that laptop crashes you want data to exist so you write it to another laptop with another excel sheet, Warpstream / Freight clusters let you write that data to a cheaper store but trades off latency to get that data

*"It's worth repeating: **the most costly part of running Apache Kafka at scale is the inter-availability zone (inter-AZ) networking and bandwidth costs. It's up to 88% of infrastructure costs when self-managing Kafka**."*

Product Manager, Confluent

*"**You could have 80% of your cost on the networking side, which is data transfer charges, which is your egress charges**. And in those cases, we had to do more innovation to bring that cost down. We had to build several features that had not been built."*

Former Sr. Director of Product & Cloud, Confluent

The ideal workloads for Freight Clusters + Warpstream are typically higher throughput, with relaxed latency requirements, which is a result of architectural decisions like writing to object storage vs. storing in-memory or writing to disk. Common examples are machine logs (IoT) and security logs (observability + security), where there is more data that can be sent slowly. We can sense check mgmt's cost savings estimate via hardware costs. For reference, writing to disc, which given the typical performance SLAs likely means writing to file or block storage with SSDs may mean paying ~$1000 a tb / year vs. AWS's object storage which tends to be ~$300 a tb a year. While this is a simple storage comparison that does not capture the nuance of retention policies or the speed of read / writes in both scenarios, the 50%+ cost savings seems reasonable.

How Much More Compliant: In tandem with the introduction of Freight Clusters, Warpstream enables log data to be written to customer managed S3 buckets (VPCs). Confluent has effectively extended their deployment architecture across cloud & on-prem, with a hybrid option (storage being in a private cloud). Customers can now "own" the storage of their Kafka cloud deployment, which provides an easier way for heavily regulated industries to abide with data retention policies & digital-native companies with more VPC services to match their Kafka deployment to the rest of their services.

JQ1A: Adoption:

At the latest investor day, management stated that ~50% of customers were using connect (*we assume this references logos*).

This makes sense given the fundamental access pattern of connectors, which act as a template for receiving & sending messages from Kafka brokers to producers & consumers respectively. The right way to think about these integrations are like excel model templates, which are pre-built for commonly accessed applications, that the developer can easily configure with their own set of parameters (i.e., assumptions). The main reason why customers prefer to use Confluent's version vs. coding their own is the initial fixed cost of developing the integration for a specific application & then the ongoing cost of maintaining the integration when there are updates for those data models, which takes 3-4 months to build and configure[8].

The reason this is a critical part of the portfolio is because it almost acts like a pre-requisite functionality prior to attaching other parts of the DSP. Management explains

"*Let's start with connectors. Connectors may seem mundane, but they are in fact a key capability. Indeed, many ETL and integration products differentiate in large part on their pool of connectors.* ***They are central to our vision as well. To build a central nervous system for your business you have to be able to connect all of your systems to capture the real time streams of data***."

Q2 2023 Earnings Call

Notably, management goes on to explain that connectors enable the expansion of data flows, which magnifies the need for both security & governance and stateful processing.

- Schema Registry: Independent services that developers configure with rulesets that producers & consumers ping to dedupe / transform data into pre-set data format. Put simply, schema registry acts like a lawyer that enforces a contract of communication between two parties. Connectors act like a gateway by expanding the ranges of apps / services communicating, thereby increasing the need for data contracts that negotiate across heterogenous producers / receivers where the data format is inherently more likely to change
- Flink: Enables customers to make stateful changes to messages in transit. What this practically means is that you can make more complicated changes to events. Connectors once again play a crucial role as a gateway here, because often-times the stateful processing that Flink provides stems from data in-transit. Connectors offer a lightweight version of processing via SMT (single message transforms), but these changes are stateless and only enable basic queries. For instance, if I was piping order data from Oracle to Mongo via Kafka, a flink query would help me transform that by appending the user_name from another database to a Mongo document.

While it's entirely possible to use Schema Registry & Flink without managed Connectors, and we know some customers do, we think the type of organization that can write their own integrations is more likely to have developer time to build out incremental functionality over open source. For instance, a popular video game developer that churned off Confluent cloud (reportedly was spending up to $2M a year) after building out their own Kafka team of 15-20 people decided to eventually churn off schema registry after realizing they could leverage their team to build on top of an open-source registry.

JQ1A: Platform vs. Cloud Adoption:

On-Prem connector adoption is significantly ahead of cloud adoption – a fact that was pointed out by management in the Q2 2023 earnings call

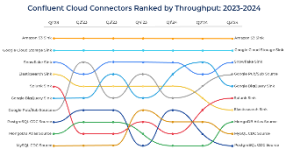
"***We are still early in monetizing this area in Confluent Cloud as fully unlocking it requires ease of use across cloud networking layers and disparate data and SAS systems***. *We took a major step towards this in Q2 with the release of our custom connectors offering, which allows running any open source connector inside Confluent Cloud, expanding our reach beyond the set of connectors we*

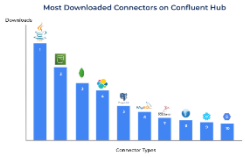
[8] Low-end of management's estimate and corroborated by customer interview

*ship with out-of-the-box. We believe this is still in the early phases of full unlock. **On premise in our Confluent platform product Connect has approximately an 80% adoption rate**. But we are still in the early days of ramping that level of usage in our fully managed offering Confluent Cloud."*

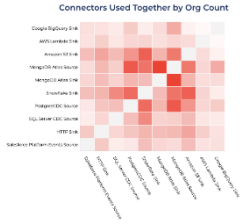
Management has mentioned that 28% of Cloud customers (48% of logos) are currently using one product in addition to the Confluent streaming platform (*We assume that the chart says that 48% of all logos use the cloud and of these 28% use +1 product).*

Confluent publishes good data on their cloud connector usage with the most common destinations being S3, GCP, and snowflake followed by a mix of cloud messaging services, logging tools, and transactional databases (e.g., Mongo, MySQL, Postgres). While these are the most popular connectors for cloud based on throughput, on-prem deployments have a notable addition: JDBC, a legacy standard java API, that streams data from relational databases to Kafka over TCP. On the right-hand side is a chart of the total downloads of self-managed connectors across both on-prem and the cloud, with JDBC being 1.2x S3 and 2x larger than MongoDB. Recall, our estimate is that on-prem self-managed connector adoption outpaces cloud by 3.75x, and we believe around ~30% of this gap can be attributed to the lack of need for JDBC in the cloud. JDBC is less relevant in the cloud because the database sits behind a firewall & therefore attempting to connect over TCP complicates the integration





Luckily, Confluent provides data on the types of workloads that create a need for cloud connectors with MongoDB at the top of the list for both the most popular connectors based on overall usage & first-time use, with some ETL driven workloads in the second tier (e.g., Snowflake Sink) – the latter is relevant given the release of a dedicated connector for ETL workloads using Apache iceberg ("Tableflow").



Top connectors used together, ranked by number of Confluent orgs:

1. MongoDB Atlas Source + MongoDB Atlas Sink
2. MongoDB Atlas Source + PostgreSQL CDC Source
3. Amazon S3 Sink + Snowflake Sink
4. PostgreSQL CDC Source + Snowflake Sink
5. MongoDB Atlas Source + Snowflake Sink
6. MongoDB Atlas Source + Amazon S3 Sink
7. SQL Server CDC Source + Snowflake Sink
8. PostgreSQL CDC Source + Amazon S3 Sink
9. Amazon S3 Sink + HTTP Sink
10. MongoDB Atlas Source + Google BigQuery Sink

JQ1B: The Theaters of Cloud Connectors: Armed with the most popular connectors (both overall and for first-time customers) we can dive into the specific ecosystem partners to determine the direction of travel for connectors in each of these theaters

OLTP / Transactional: From the 2025 investor day we know that most of Confluent's deployments (>80%) tend to be focused on operational applications (think mobile banking app user feature). Thus, it's no surprise that the top connectors used & the most popular introductory cloud connectors focus on prominent cloud OLTP databases: MongoDB & Postgres. At a business level, organizations will take data from their OLTP databases, like clickstream data, and then either send it to the cloud for cheaper storage & processing or enrich it with other data before using it to drive an application feature. The latter is an extremely encouraging access pattern, that shows up as the #1 connector used together, as it means that Kafka is being used to enrich operational data – a key workflow for AI applications. While we think the uplift from AI workloads is quite early as corroborated by MDB in their Q1 FY26 earnings call

"But what we see is that **enterprises are still early in the adoption of AI. The barriers include this, you know, limited set of skills, and experience with AI. Trust with AI systems that are probabilistic**, which is another way of saying the risk hallucinations."

MDB CEO, June 4 2025

We are confident that Kafka will be more relevant for Mongo workloads that are increasingly focused on feeding real-time info for agents in the form of relevant enterprise data. While we believe this was already happening to a certain extent,

given the popularity of Mongo as both a source & a sink, AI workloads have a much larger need for this data to reduce the hallucinations that Mongo's CEO calls out.

Mongo Relationship

On the whole it seems like MongoDB and Confluent have a bit of a co-opetition dynamic, with Confluent & Mongo having several marketing materials that emphasize real-time data systems & Mongo even being named Confluent partner of the year in 2024. Moreover, Confluent joined MDBs AI initiative in December of last year – which makes sense given the high usage of MongoDB as a source and destination in the cloud. However, MongoDB also offers a lightweight version of stream processing (released in Feb 2024) that with at least one customer has replaced Kafka connectors & eliminates Flink as a cross-sell. This lightweight stream processing allows Atlas users to easily query and process events in Mongo vs. having to leverage a Kafka connector & Flink to process & ship to Mongo. While the broader relation with Mongo seems healthy, this is a material risk to the connector & flink adoption story – Confluent's most popular cloud connector partner has a tool that is used & marketed as a lightweight replacement to Confluent's DSP. We think the two major mitigants to this risk are:

- Separate Kingdoms: Mongo's stream processing solution offers an improved developer experience for Mongo only users, with a native method of collecting event data in Mongo. However, the solution is limited in its ability to handle high throughput (hundreds of mb/s), several partitions, and multiple endpoints. As the MongoDB customer example illustrates, stream processing from Mongo will be great for customers who do not have tight SLAs for their streaming data and are keeping the streaming data contained in Mongo instances. Thus, we think that for much of the base that Confluent currently monetizes the performance tradeoff is too risky
- The Rise of Postgres: Postgres is a relational structured database that retains SQL querying & happens to be in 3 of the top 10 combinations for Confluent's connector. This is a true open-source database, a reaction to the commercial dominance that Oracle has over the first version of SQL databases and has several but no dominant commercial backers – Postgres recognizes 15+ managers, with several of these being $100M ARR+. The upshot is that Postgres has been able to take share from Mongo in relational workloads (*see Stack Overflow Survey[9] Mongo developer mindshare has decreased from ~28% in 2020 to 25% in 2025 and from an enterprise perspective usage has stalled while Postgres usage has monotonically increased*) due to improved document / unstructured data handling (JSON features). As Postgres continues to threaten MongoDB's flow share of Oracle migrations & workloads, Confluent has the benefit of being neutral to vendors for moving data between them (the second most popular access pattern).

Ultimately, organizations that are going to keep all of their data in Mongo & run vector, search, and stream workloads in their Atlas instances – should use Atlas stream processing, which will impact Confluent's base of existing & new users. However, we think the right way to size the impact is on a flow share basis for new logos that are likely to have latency tolerant / more limited downstream consumers. From that perspective, we feel quite good that Postgres has risen as a legitimate challenger to Mongo due to architectural improvements in unstructured data handling + serverless deployments coupled with the limited risk of commercial concentration from the project.

OLAP / Analytical: The smaller but fast-growing segment of Confluent & Connect are analytical databases like the CSPs + Snowflake / Databricks. In this access pattern, event data is processed by confluent from devices, OLTP databases, or software applications and then dropped into warehouses. The typical paradigm streamed that data into a data prep process called "the medallion staging", which would clean data via ETL. However, this pattern meant that customers incurred costs for cleaning data in their DW / DLs, and with the popularity of Iceberg as a true vendor-agnostic data table format, organizations can now pre-process their data to avoid paying cleaning fees from their OLAP databases. Databricks has significantly closed the gap in terms of market share to nearly 50-50, winning ~1000bps over the past 18 months between the two major independent cloud vendors (AWS Redshift, Azure Fabric, and Google BigQuery all boast their own data warehouses / lakehouses) Two major subpoints here:

Snowflake Relationship: While Snowflake has historically been an important partner / driver of Confluent connect, listed in 4/10 combinations, Confluent recently introduced Tableflow in partnership with Databricks. Relations between Snowflake & Confluent have likely frayed following rumors of Snowflake's offer to acquire Redpanda, a Confluent competitor, for

[9] We combine the Stack Overflow survey & the DB Engine ranking to help capture developer mindshare & production enterprise usage. For instance, Oracle scores quite low in the stack overflow survey because it's seem as a bit more legacy with lock-in but it's a massive / the most-popular production database & which the DB-Engine ranking properly captures.

~50x ARR. At Snowflake investor's day there was an allusion to the deal breaking down based on a massive price difference, likely fueled by the latest $1B valuation of the company in early 2025, as the CEO commented

"We continue to be very active in M&A but on our terms. **I don't want to pay a $1B or $2B for a company that is making single digit millions, it's just like the math doesn't work, we don't live in that world**."

Snowflake CEO, 2025 Investor Day

The massive overlay is that the introduction of Apache Iceberg has meant that now workloads / data that was previously proprietary or locked into vendor-specific formats, can now easily be moved between analytical warehouses. A fact that is highly relevant given the fierce competition for analytical data storage & compute in the enterprise. Ultimately, we believe the Confluent & Snowflake connection has been damaged by the RedPanda acquisition rumors in addition to the rise of Databricks – and if Snowflake ever acquired RedPanda or a smaller competitor that would significantly impact connector attach for ETL workloads (likely less than 20% of attach).

Databricks

Contrasts Snowflake's more closed ecosystem / storage-centric focus with an open ecosystem targeting AI / ML researchers + data scientists. While Databricks is absolutely taking share, they are usually winning a disproportionate share of dollars vs. replacing logos. In fact, a survey of 1,800 customers showed that ~60% of Databricks accounts also run Snowflake and 40% of Snowflake accounts also have Databricks. While there is certainly some migration, the sheer data gravity that these platforms have in the cloud with years of data stored there (likely >5-10PB) make it extremely difficult to switch.

Under this backdrop, Confluent launched Tableflow, a pre-built connector that automatically transforms stream data to Apache Iceberg tables, enabling transformations in real-time with Flink, before dumping data into Databricks or Snowflake. The reason we highlight this under the Databricks section is the major partnership this has led to between the two organizations

"Additionally, **our partnership with Databricks includes comprehensive go-to-market efforts, encompassing field and partner enablement, solution architectures, co-marketing and co-selling initiatives**."

Q4 Earnings Presentation

"And **so 11 SIs have agreed to combine their Confluent and Databricks practices to capitalize on this announcement** and our plans to deliver products together. So we're super excited about that."

2025 Investor Day

While we were a bit surprised not to see Databricks as a major target for Confluent connectors, Tableflow represents a huge opportunity for Confluent to drive similar adoption in the Databricks ecosystem as they have with Snowflake. Just to provide some datapoints – roughly 70% of joint Snowflake & Databricks customers plan to use or evaluate Open Table formats over the next 6 months. Within this group ~60% are planning on using / already using Iceberg (the # 1 table format outside of proprietary AWS, Hive, and Parquet).

From a modeling perspective we model attach as a % of the total customer base, and more specifically we bifurcate pre-2025 logos and 2025+ logos, to isolate our go-forward new logo drivers from our attach drivers. Three main assumptions we make in the pre-2025 build

- Retention: We model logo retention on the whole customer base (assumed around 95%) slightly higher than the dollar retention, which hovers around the low 90's, given there is no down sell included, which we know plays a big role in the Confluent ARR story. Our pre-2025 base benefits from no new logos that are at higher risk of churning & therefore the 5.8k total logos eventually end up at ~4.9k logos
- Cloud vs. On-Prem: We conservatively assume this split remains even post churn (52% on-prem vs. 48% cloud), which means that we do not account for any on-prem customers who add a cloud deployment or switch to a cloud deployment – which should represent a higher LTV opportunity given the wider set of products they can use on the cloud side
- Connector Attach: The bulk of the DSP ARR is modeled as a function of connector attach, given the importance of connectors as a gateway to other products. As of now we believe that 75% of +1 DSP customers are leveraging

connectors, 80% of customers who use connectors leverage a second solution and 22% leverage a third solution. At exit we drive connector attach to 55%, hold secondary attach at 83%, and drive +2 attach to 60% as Flink matures. While the 55% may seem like a big jump – the actual logo count grows from a base of 900 to 1.3k over 5 years

- ASP: We model a 5% increase each year, mostly driven by workload increases (i.e. more data) being added to these products. Likely a bit conservative given how much data volumes are growing

The post 2025 base is highly variable depending on the # of new logos we model, a key assumption we dive into in the next section. However, the basic build follows the same logic with different numbers

- Retention: An output given we already calculate global retention and pre-2025 retention, the number starts at around 90% and then scales as the cohort ages
- Cloud: Assume an 80% flow share that maxes out at 85% share at exit, consistent with the qualitative commentary that management has given about the flow share of net new logos being cloud (read: "vast majority") – the on-prem cohort resembles pre-2025 in terms of % attach (80%) and ASP
- Connector Attach: Given the nascency of this base we model attach reaching the same levels as the pre-2025 cohort, with the main difference being that it takes 5 years for the 2025+ cohort to reach the same level whereas the pre-2025 cohort reached these attach levels in less time
- ASP: Same treatment as above

Across both cohorts we get to a combined cloud attach rate of

- ~37% attach for +1 DSP product vs. 28% today
- ~27% attach for +2 DSP product vs. 18% today
- ~17% attach for +3 DSP product vs. 6% today

Resulting in ~$350M ARR of Cloud DSP ARR at exit and ~$200M ARR of Confluent Platform DSP ARR at exit, **~$325-350M ARR of incremental net expansion uplift at exit**. However, while this build isolates key assumptions (pre-2025 vs. 2025+ cohorts) – we also check what this build means for the main connectors / workloads we win

Upshot: The 10,000 ft view of Cloud DSP from a workload perspective is that defined narrowly, you need to believe Tableflow / Databricks partnerships drives 2-3x more connector attach in the Databrick's customer base as we have seen with Snowflake & you have to believe despite Mongo inroads, that the rise of OLAP + OLTP workloads + Postgres are enough to drive attach from 3-5% to 7-10% of the OLTP customers. Both of which we feel is quite conservative, given the excitement around agents and alignment with powerful share gainers (Databricks + Postgres) with solid bases (Snowflake + Mongo). Additionally, this narrow definition ignores the rapid expansion of connectors to new applications (ServiceNow, Salesforce, etc.) – which we view as incremental to the underwrite

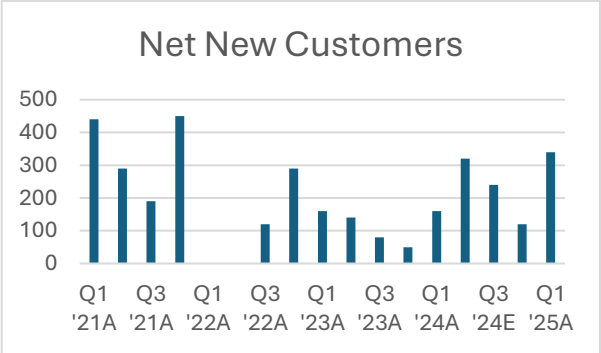
JQ2: How have recent pricing & packaging changes impacted Confluent's new logos and how should we think about that going forward

We start with a top-down view of how many new logos Kafka adds every year. In order to do this we simply take a retention assumption and apply it to the number of net new logos added every year. We assume this is around 93 to 95%, given there is no impact of down-sell and gross-dollar tends to hover around the low 90s. The imputed new logo counts are therefore

- New Logos: 2021: ~1.5k / 2022: 1.3k / 2023: 715 / 2024: 1.1k
- Net New Logos: 2021: ~1.4k / 2022: 1.1k / 2023: 430 / 2024: 840

In our base scenario we model new logos climbing to slightly above peaks in 2028 and then descending to nearly the 5 year average of ~1.2k. While new logos climb past historical levels, net new logos never surpass 2021 momentum and are fairly consistent around 1.1-1.2k logos until falling to 750. While this seems counterintuitive in the face of weak demand signals and a soft Q2, we believe that the selection expansion that Confluent has undertaken over the past year should help them drive new logo count to pre-2023 levels. Moreover, we think it's worth unpacking what we think is happening in the net new logo momentum

Confluent has struggled to build consistency in their new logo motion, which maps to the technical complexity of the product & required familiarity with streaming that creates a typical awareness funnel. The latter point is worth pausing on because while some customers may immediately enter in a higher spend band, we believe the <$20k segment, typically filled with pay as you go customers, are an important top of funnel metric (*Confluent Investor Day 2023 50% of new cloud logos were from self-serve)*. Below we show how QoQ net new customers have trended for all size bands, with the punchline being that they add around 100-350 net new customers with relatively little seasonality but had an encouraging Q1 (*modified reporting this metric from quarterly to yearly cadence)*.



While Q2 was absolutely difficult from a net new logo perspective (only 46 added >20k, 50% QoQ dec.) and sentiment seems to be quite weak, we think this is driven by the rise of the <20k segment as a key driver for new logos, which Confluent stopped reporting on a quarterly cadence (Q2 first quarter without it) and is unlikely to be picked up in surveys.

JQ2A: What do our new logo assumption mean for our win rate and jump-ball visibility?

For our back-up work we model this out based on the number of logos we believe transact every year, which we solve for via the historical win rate x % contracts we see. We estimate these numbers using the F500 Confluent market share, which is around 53% as of FY24 and the number of new logos that we win every year.

Win Rate: Confluent claims a 90%-win rate today and calls out that this has improved over time (we take a discount for modeling purposes given the historical improvement = ~85%). However, the win rate for F500, closer to Confluent's historical ICP, are higher than down-market customers, so we discount the non-F500 win rate to 70%

Jump Ball Visibility: The reported win rate only uses RFPs that Confluent is involved in. Given the 50% market share in the F500 and ~85% historical win rate, the F500 jump ball visibility is around 60-70%. Naturally, the jump ball visibility should be lower for non F-500, we place it at 55%

> For our analysis we assume our historical flow share & market share in the non-F500 is around ~38%. Applying these percentages to our historical new logo results in roughly ~2-3.5k logos that transact every year.

As a quick sense check, we use the ~50-70k orgs on LI[10] that have at least one Kafka engineer, which means that 4-5% of logos transact each year, which feels fair given the maturity of the project. Going forward we model the win rate increase 770bps from 70% to 77.5% and hold the RFP participation at 55%, given the improved price / performance offerings from the bolstered adjacent offerings + wider selection of deployment options.This is corroborated by the material increases over the past year in both the win rate and the types of customers we are winning, with mgmt.. provided win rates jumping to 90%+ across both CSPs and start-ups.

"Every WarpStream deal closed since acquisition was from our digital native cohort including companies like Elastic, the search AI company used by over half of the Fortune 500, and Cursor, the AI code editor that's become one of the hottest names in AI. **And the vast majority are net-new customers to Confluent**."

Confluent Q4 FY24 Earnings Call

"The first is replacing **CSP streaming offerings with Confluent. We've had success displacing these CSP offerings with win rates well above 90%. This is an area where we feel our product capabilities and TCO story have improved enormously over the last year with differentiated offerings like Freight Clusters, Enterprise Clusters and WarpStream**.

Already in Q2, **we saw more than two dozen displacements against a single CSP offering**. We plan to amplify this success by intentionally targeting these offerings and increasing our number of at bats against these competitor."

[10] *Confluent boasts that there are 150k+ orgs using Kafka, we do not think that all of these are addressable, and this number may be inflated*

Using our improved win rate in this segment and stable RFP visibility, our new logo estimate means that 3-4k logos transact each year – a similar range from the one we imputed in historical counts. We feel comfortable modeling a slight increase in total at bats given the improved IT spending forecast from AI budgets

		2020A	2021A	2022A	2023A	2024A	2025E	2026E	2027E	2028E	2029E	2030E
# of New Logos (Est.)		1,335	1,512	1,273	714	1,111	1,265	1,420	1,574	1,728	1,568	1,408
(-) # of New F500 Logos (Est)		(6)	(6)	(6)	(13)	(13)						
# of non-F500 Logos		**1,329**	**1,506**	**1,267**	**701**	**1,099**	**1,265**	**1,420**	**1,574**	**1,728**	**1,568**	**1,408**
							14%	28%	42%	56%		
Historical & Impued Flow												
Win Rate @ 70%		1899	2151	1810	1002	1569						
JB Visibility @55%		3452	3911	3291	1822	2853						
Total Kafka Customer at bats		**3452.3**	**3910.5**	**3291.2**	**1821.8**	**2853.4**	**2968.4**	**3330.2**	**3692.1**	**4053.9**	**3678.6**	**3303.2**
Avg '20-'24	3065.8											
Premium / (Discount)		13%	28%	7%	-41%	-7%	(3%)	8%	17%	24%	17%	7%
New Win Rate	78%						1632.62	1831.64	2030.65	2229.67	2023.22	1816.77
JB Visibility	55%						2968.40	3330.25	3692.09	4053.94	3678.57	3303.21

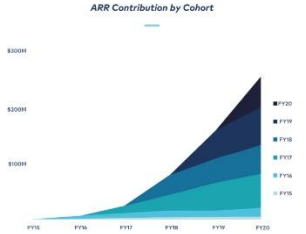
Ultimately, we view Q1 as validation of this hypothesis with 340 net new logos being added, the highest number since Q4 ’21. While Q2 only had 46 customers added in the >20k segment, Confluent did not report in the <20k segment, which made up 86% of their Q1 net adds (LTM is around 80%). It’s only one real quarter of data, but the qualitative voice-over from management and the industrial logic seems to support in part a re-acceleration of logos. However, it’s still early and therefore we limit net new logo growth to peak at historical levels when Confluent was much smaller

JQ2B: Will new customer segments look the same as previous customer segments given the new deployments and use cases?

We hypothesize that new logos should re-accelerate driven by this new segment of customers that we are better able to serve, however a new composition of workloads / SLAs also means a potentially different expansion journey. An extremely crucial part of understanding Confluent, and really any open-source infrastructure business, is that they have baked in expansion. Whenever Confluent flips a customer, they are likely not new to Kafka and already have production deployments running. Thus, while they may sign up for an initial cluster they have moved over to the commercial version with a plan to convert some share of their deployments – there is built in expansion as Confluent wins these dollars. While the dynamic depends on the pool of Kafka the customer already has deployed, Confluent also has the crucial opportunity to ideate use cases to drive expansion in that first 1–2-year period. Notably, Confluent lands at ~$40k-60k of ARR but the average customer spends ~$150-175k – a key data point that showcases the built-in expansion dynamics.

- Historical expansion: We have limited data on what historical expansion has looked like but Confluent’s S1 provides a cohort expansion curve (shown below)

Leveraging this chart with the total ARR from 2021 to 2024, we can back into the estimated expansion for cohorts from T=0 to T=5



Full Cohorts (with DSP)

Cohort	FY17	FY18	FY19	FY20	FY21	FY22	FY23	FY24
FY17	$15	$22	$38	$62	$68	$74	$80	$86
FY18		$22	$43	$50	$80	$88	$95	$103
FY19			$36	$65	$91	$137	$150	$162
FY20				$55	$94	$126	$177	$194
FY21					$102	$143	$200	$220
FY 22						$52.2	$88.7	$118.0
FY 23							$37	$62.88
FY 24								$57.08
ARR	**$15**	**$44**	**$117**	**$232**	**$435**	**$619**	**$827**	**$1,003**

T=0	T=1	T=2	T=3	T=4	T=5+
	1.47	1.73	1.63	1.1	1.08
	1.95	1.16	1.6	1.1	1.08
	1.81	1.4	1.5	1.1	1.08
	1.7	1.35	1.4	1.1	
	1.4	1.4	1.1		
	1.7	1.33			
	1.7				

T=0	T=1	T=2	T=3	T=4	T=5+
	1.47	2.53	4.13	4.55	4.91
	1.95	2.27	3.64	4.00	4.32
	1.81	2.53	3.79	4.17	4.50
	1.70	2.30	3.21	3.53	
	1.40	1.96	2.16		
	1.70	2.26			
	1.70				

With the headline being that for the cohorts in the chart we know they expand by 1.5-2x in T=1 and end up cumulatively being around 2.3-2.5x by T=2. We use these benchmarks + the actual total expansion each year, with new logos booked that year taken out, to estimate the expansion curve for the rest of the cohorts & time periods.

Our curve generally follows a (t=1 / 1.4-1.7x; t=2 / 1.3-1.4x; t = 3 / 1.1-1.5x; t = 4 / 1.1x; t = 5 / 1.08x) curve. Cumulatively, we end up with an expansion of ~3.5-4x by t=4, which aligns with management’s figure in their 2023 investor day of 3.5x+ ARR expansion in 4 years (*slide 17 efficient growth*).

This curve needs to be slightly adjusted because it includes ~$180M ARR of DSP uplift and therefore mixes cross-sell with what we would call baked-in expansion. Below is the Non-DSP impacted cohort curve that takes a slight discount to the curve above. For comparison, cumulative expansion in t =5 is ~3.0-3.5x vs. >4x with DSP included

Non-DSP Cohort

Cohort	2017A	2018A	2019A	2020A	2021A	2022A	2023A	2024A
FY17	$15.0	$20.5	$33.4	$51.1	$53.6	$57.9	$62.6	$67.6
FY18		$22.0	$40.8	$43.4	$65.0	$68.3	$73.8	$79.7
FY19			$36.0	$61.4	$79.8	$111.7	$117.3	$120.9
FY20				$55.0	$88.0	$110.0	$134.5	$141.3
FY21					$101.9	$132.5	$172.3	$180.9
FY 22						$52.2	$85.1	$114.9
FY 23							$37	$60.29
FY 24								$57.08
(=) Total ARR ex. DSP	**$15**	**$43**	**$110**	**$211**	**$388**	**$533**	**$683**	**$822**

	T=0	T=1	T=2	T=3	T=4	T=5+
		1.37	1.63	1.53	1.05	1.03
		1.85	1.06	1.50	1.05	1.03
		1.71	1.30	1.40	1.05	1.03
		1.6	1.25	1.223	1.05	1.03
		1.3	1.3	1.05	1.05	1.03
		1.63	1.35	1.25	1.05	1.03
		1.63	1.35	1.25	1.05	1.03
		1.70	1.30	1.25	1.05	1.03

	T=0	T=1	T=2	T=3	T=4	T=5+
		1.37	2.22	3.41	3.58	3.86
		1.85	1.97	2.96	3.10	3.35
		1.71	2.22	3.10	3.26	3.36
		1.60	2.00	2.45	2.57	2.65
		1.30	1.69	1.77	1.83	1.88
		1.63	2.20	2.75	2.89	2.97
		1.63	2.20	2.75	2.89	2.97
		1.70	2.21	2.76	2.90	2.99

JQ2C: Who are these new customers that Confluent is unlocking with TCO / more latency sensitive workloads?

With the normal cohort curve adjusted, the question is what curve we should use going forward.

As a benchmark we take the FY22 cohort as a representative sample given the similar size cohort of $52M and 1.1k logos – to FY25 and FY26, that scales up with the business. However, prior to centering on an expansion curve it's worth double clicking on who exactly these net new logos are for Confluent & how they may differ / resemble previous cohorts

Confluent has provided detail on 9 named / 23 total net new logos over the past three quarters

- Audacy, Booking, Top-Five Video Gaming Company, World's Largest Sports Media Outlet, Fortune 100 Pharmaceuticals, Global Cruise Operator, Leading European Airliner, Elastic, Cursor, 24 CSP replacements (TCO / clusters + warpstream)

Staring at the list, and looking at the numbers, it's clear that there are two different segments that make up these new logos

- Digital Natives / TCO sensitive logos: Audacy, Booking, Elastic, Cursor, CSP replacements, Top 5 video game company
- Large-scale customer / Old World: media outlet, F100 pharma, cruise operator, airline

We focus on the first segment given management's characterization of these customers as net new to Confluent, which also means their expansion behavior is unlikely to be captured in historical cohorts.

Part of this is described in the freight cluster / WarpStream section, but as a refresher, much of Confluent's focus prior to 2024 was on compute-centric Kafka, meaning low latency workloads and low throughput. A separate problem is how do you solve data replication across multiple availability zones. This is naturally a much less important problem when latency is all that matters, but in 1ms workloads, this data replication can be 60-80% of the cost. More specifically, every time you write data to a Kafka log – that data is replicated across multiple physical / virtual machines for resilience purposes. Unless you have AZ aware clusters it's extremely difficult to allocate these replicas in a manner that avoids data egress and cross AZ data costs. Freight Cluster and Warpstream ameliorate this problem by writing to cheap object store first, straight to the cloud, which creates the resilience. The current situation is well framed by a prescient former employee

"**So the digital native segment is not currently [in 2023] served by Confluent well. I mean, honestly, there are many ways to save on cloud costs**. All the traditional stuff exists out there. And all of that has been already done. What has not been done is specifically networking charges, which is very, very unique to streaming data by definition… **I think the use cases that will come over to Confluent Cloud would be most of the use cases that is being used by the digital native segment**. So gaming, for sure. We already had four or five gaming logos, but I think there's much more to capture there. Ride sharing would be another one, 100%. Performance tracking and management, everybody that fits in that segment like Splunk and New Relic and Datadog and companies like that in application performance monitoring. That would be another use case. I'm just thinking more observability. Yes, I would say those are the main use cases."

Former Confluent Employee, 2023

While this is great detail on these incremental customers what matters is Confluent's expansion journey with these customers. The comparison is framed up nicely by the former employee who characterizes these digital native customers as established Kafka customers running deployments at scale vs. more compute-centric orgs that may be less Kafka mature and ran into a hair on fire problem while scaling. Our scorecard vs. more compute centric cohorts shows

- (+) Larger swath of Kafka deployment vs. more immature compute organizations that converted when they ran into scaling problems
- (+) Strongly levered towards greenfield real-time use cases, especially AI, given their orientation towards more consumer-centric usage (e.g., Audacy, Booking, Cursor)
- (/) Strong DIY culture that likely caps the SoW that Confluent can capture, however this DIY culture also means allocating valuable developer resources to non-core infrastructure tasks – which AI tools may nullify

- (-) Smaller budgets vs. these larger / old world organizations

Ultimately, we apply a 5% uplift over our representative cohort, which we think accurately captures the large swath of available Kafka deployments and the uplift from AI workloads for digital native customers, that also have a strong DIY culture that impairs full adoption of Confluent across all Kafka deployments.

Thus, the cohorts reach ~2.7x expansion by T=4 / 2.9x by T = 5, at a material discount to earlier / smaller sized cohorts and in-line with our base cohort. Applied to the base of new customers this means that the business ex. DSP scales from $822M ARR in FY24 to $2.2B of "embedded expansion ARR."

Non-DSP Cohort

Cohort	2017A	2018A	2019A	2020A	2021A	2022A	2023A	2024A	2025E	2026E	2027E	2028E	2029E	2030E
FY17	$15.0	$20.5	$33.4	$51.1	$53.6	$57.9	$62.6	$67.6	$73.0	$78.8	$85.1	$91.9	$99.3	$107.2
FY18		$22.0	$40.8	$43.4	$65.0	$68.3	$73.8	$79.7	$86.0	$92.9	$100.3	$108.4	$117.0	$126.4
FY19			$36.0	$61.4	$79.8	$111.7	$117.3	$120.9	$124.5	$128.2	$132.1	$136.0	$140.1	$144.3
FY20				$55.0	$88.0	$110.0	$134.5	$141.3	$145.5	$149.9	$154.4	$159.0	$163.8	$168.7
FY21					$101.9	$132.5	$172.3	$180.9	$186.3	$191.9	$197.7	$203.6	$209.7	$216.0
FY 22						$52.2	$85.1	$114.9	$143.6	$150.7	$155.3	$159.9	$164.7	$169.7
FY 23							$37	$60.29	$81.39	$101.7	$106.8	$110.0	$113.3	$116.7
FY 24								$57.08	$94.18	$117.7	$147.2	$154.5	$159.1	$163.9
FY25									$66.95	$110.5	$138.1	$172.6	$181.2	$186.7
FY26									$77.4	$127.7	$159.6	$199.5	$209.4	
FY27										$83.2	$137.3	$171.6	$214.5	
FY28											$88.6	$146.2	$182.8	
FY29												$78.0	$128.7	
FY30													$67.9	
(=) Total ARR ex. DSP	$15	$43	$110	$211	$388	$533	$683	$822	$1,001	$1,200	$1,428	$1,681	$1,944	$2,203

	T=0	T=1	T=2	T=3	T=4	T=5+
	1.37	1.63	1.53	1.05	1.03	
	1.85	1.06	1.50	1.05	1.03	
	1.71	1.30	1.40	1.05	1.03	
	1.6	1.25	1.225	1.05	1.03	
	1.3	1.3	1.05	1.05	1.03	
	1.63	1.35	1.25	1.05	1.03	
	1.63	1.35	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	
	1.65	1.25	1.25	1.05	1.03	

	T=0	T=1	T=2	T=3	T=4	T=5+
	1.37	2.22	3.41	3.58	3.86	
	1.85	1.97	2.96	3.10	3.35	
	1.71	2.22	3.10	3.26	3.36	
	1.60	2.00	2.45	2.57	2.65	
	1.30	1.69	1.77	1.83	1.88	
	1.63	2.20	2.75	2.89	2.97	
	1.63	2.20	2.75	2.89	2.97	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	
	1.65	2.06	2.58	2.71	2.79	

JQ3: How will recent pricing and packaging changes impact Confluent's SoW with existing customers?

While the prior analysis focused on what happens to future cohort expansion, this section dives into the net SoW gains that Confluent can drive in their existing base.

SoW via packaging changes: The introduction of low-cost alternatives will lead to some cannibalization, but we think that the impact will be limited past 2025. Part of the top 20 customer optimization that has spooked investors & played a material headwind to cloud growth this year is likely ***partially*** driven by these cluster / deployment optimizations. This contention is pieced together via management's voiceovers and customer anecdotes.

Culprits: Management has been quite explicit that most of these optimizations have been idiosyncratic and from a handful of top 20 customers, in addition to a large AI-native business (read: likely OpenAI)

*Equity Researcher: "I think last quarter, you mentioned **it was a handful of top 20 customers**. But just trying to get a sense of how that trended sequentially."*

Jay Kreps: "Yes. I would say it's a similar dynamic"

Q2 Earnings Call, 2025

Profile: From the latest investor day, we know that 19 of the top 20 customers have been with Confluent for over 5 years and they spend >$5M. Moreover, we know that 45% are hybrid

This last bit is particularly relevant for the new network-oriented clusters that are perfect for hybrid deployments where customer & confluent VPCs create this massive data replication charge. While management has not stated anything about this phenomenon, the possibility of customers downselling or optimizing their clusters with WarpStream / enterprise / freight clusters has been highlighted by customers

*"But then we also realized **that we are not able to leverage the full capacity of the Confluent cluster,** which we had. So **we migrated from three clusters to a single cluster**, and we were able to manage that. So I think that was one complaint we have always been with Confluent that, okay, hey, you need to come up with more SKUs or this is like too big for us. And even I think our scale is quite good. But I think that's the response… **Our plan is like if we reach the capacity of our current cluster, then we will move our data platform use cases to freight cluster and then we will see**."*

Staffing Platform, India

We think this is a unique phenomenon for the largest customers, as corroborated thus far by management where small customer consumption patterns have been stable, because they are the most likely to be operating near capacity. The simultaneous pausing of new projects until AI gets into production (per MDB CEO) and typical optimizations compounded by the additional lever of cheaper clusters helps partially explain the cloud headwind in H1 2025.

For the rest of the customer base, that is unlikely to be operating at a similar capacity as >$5M spend, we think that warpstream and the inclusion of new clusters is likely to be incremental in the long run.

Below we have mapped out the deployment estates of ~8 customers, with the punch line being that even with current customers, Confluent only wins ~50% of deployment share. Importantly, customers who are not using Confluent fall into two categories:

	Confluent	OSK	MSK / Kinesis
Verizon	15%	50%	35%
Walmart	5%	90%	0%
Bloomberg	60%	40%	0%
RBC	95%	5%	0%
Fidelity	20%	80%	0%
FanDuel	80%	20%	0%
AT&T	80%	20%	0%
Block	50%	0%	50%
Total (Avg.)	51%	38%	11%
Total (Median)	55%	30%	0%

- Historical ICP (FanDuel, AT&T, Bloomberg): These are customers with latency sensitive deployments, that were early adopters of Confluent, who run mission critical workloads (e.g., customer recommendation systems). These customers have settled use cases and Confluent >60% of deployments, we do not forecast substantial increases in Confluent share of pipelines / workloads for this segment, given they usually have less than <100% for idiosyncratic reasons, for instance, Bloomberg uses OSK for DA
- Less than 20%: This segment represents 3/8 customers (~35%) & represents a huge opportunity for Confluent to win share. The most common reasons for sub-optimal SoW is:
  - Existing Data Engineering Team: Customers may have an existing set of engineers that manage their kafka deployment, which customers estimate takes 10-20 data engineers, we think as the importance of data engineers increases this is less likely to occur

*"It also depends on where you want to spend your money. Confluent is more expensive than self-hosting. **If I have a team of 10, 15, 20 engineers, do I want them to spend their time on managing the Kafka infrastructure or do I want them to spend their time providing something which is more valuable for my end customers**."*

Digital Native Customer 2024

  - Cost / Performance Reasons: The other reason why Confluent may fail to win >50% of deployments is because they simply are not a good fit for the architecture due to a captive CSP or a more latency-tolerant workload. One customer estimates that over 80% of their workloads are more latency / log based workflows

Q:"So I'm curious like roughly, I know that you cannot say it like precisely, but when you think of your use cases, how much of your use case, how many of your use cases, sorry, are, let's say, high latency, high throughput use cases. When you move a lot of data and the latency is not supercritical in terms of like 50 milliseconds. So how much -- yes, you got my question."

*A: "**I would say it's probably like 80%. The market data use cases, we certainly looked at doing with Kafka. So I would say like logs and things like that, they're not anywhere close in near real time**. The reality is like I felt that like security folks would care about I need to get like log events that somebody is trying to break into my system right away. It turns out that's not the case."*

Sizing: Taking this set of customers as directionally correct, we think there are ~35% of customers where Confluent is used for less than 1/5 of use cases. Naturally, some of these deployments won't be addressable, i.e., dev servers for testing Kafka services, and so we use the ~60% as our ceiling for addressable deployments. Driving share for more latency tolerant / TCO focused workloads leaves ~40 percentage points of SoW that Confluent can go out and win. **Taking a conservative uplift of ~25-30 percentage points, on a base of $820M of core confluent streaming (ex. DSP) – we think that means there is ~$290M of spend that at 25-30% conversion results in ~$75M of ARR SoW expansion at exit**

Below we show a sensitivity of the SoW uplift based on % of FY'25 spend addressable and % upsell amount

	15%	20%	25%	30%	35%	40%	45%	**% FY '24 Spend Addressable**
15%	$18.5	$24.7	$30.8	$37.0	$43.2	$49.3	$55.5	
20%	$24.7	$32.9	$41.1	$49.3	$57.6	$65.8	$74.0	
25%	$30.8	$41.1	$51.4	$61.7	$72.0	$82.2	$92.5	
30%	$37.0	$49.3	$61.7	$74.0	$86.4	$98.7	$111.0	
35%	$43.2	$57.6	$72.0	$86.4	$100.8	$115.1	$129.5	
40%	$49.3	$65.8	$82.2	$98.7	$115.1	$131.6	$148.0	
45%	$55.5	$74.0	$92.5	$111.0	$129.5	$148.0	$166.5	

% Upsell Amount

As an evidence point of our emerging perspective on the recent earnings call, we believe there was an indication we may be in the early innings of this phenomenon

"The large majority of our WarpStream business in Q2 is incremental. **Even in existing customers, we're seeing customers increase their spend with Confluent through WarpStream while actually lowering their overall cloud infrastructure costs. For example, two customers, a major retail investing platform and a leading prepaid mobile provider both deployed workstream for their high-volume logging and telemetry workloads in Q2. These customers increased their spend with Confluent by 30%.** while decreasing overall CSP infrastructure costs roughly 50%."

Lastly, as a sense check for our ~30% of spend addressability, roughly 25% of customers fall into the $100k-$1M and 40% are between 20k-1M – which is the segment we believe is least likely to be at capacity and therefore at higher propensity of upselling. Moreover, internal Confluent surveys shows that 37% of customers prefer a BYOC deployment, which we feel confident taking a discount on.

The simple view of the uplift is as follows

- Embedded Expansion: $2.2B ARR: Re-acceleration in new logos + Expansion curves that perform +5% better than benchmark at fully loaded LTV

- DSP + On-Prem: ~$350M of incremental ARR at exit: Attach % effectively doubles
- SoW Gains: $50-100M: Key assumption is that there are under-utilized Kafka customers where WarpStream will be incremental

This results in $2.78B ARR / 16% LTM Growth and ~$3.0B NTM Revenue Growth / 15% NTM Growth with ~113% DBNR / 92% GDR

JQ4: What will end-state margins be?

- Growth: As previously mentioned, this will be tightly coupled with the level & type of growth that Confluent drives. We have conviction that if Confluent can maintain net retention around ~115% with multi-product attach / SoW gains over the next 5 years, they should be able to achieve ~20% non-GAAP EBIT Margins
  - Mgmt. Estimate: Management forecasts that they will get to 6% non GAAP operating margin in FY25 (excluding a cash compensation adjustment of around 3-4%), 12-15% in FY '27 (*we have them hitting the low end of this range*), and 25%+ in the long term - we take the under on their estimate after looking at comps
- Comps: We select Elastic and Mongo as near-in comps due to their open-source lead gen, CSP competitive set – supplemented by adjacent substitutes, and lastly their focus on the data infrastructure ICP / market. The punchline is that both businesses have been able to drive to ~15% non-GAAP EBIT margins with HDD growth (17-19%), with healthy leverage driven across various cost buckets – however neither is yet GAAP EBIT profitable primarily due to SBC (20-25% of revs - almost 1/2x Confluent's FY24 SBC)
  - Elastic: We think that they are a lower quality business vs. Confluent mainly due to competitive dynamics & therefore this should be a floor from a rule of perspective (rule of 30) / margin target (15%). Elastic has struggled tremendously with Amazon as a competitor, in fact from 2021-2024 Elastic closed sourced their solution – causing massive community upheaval. Amazon ended up forking the open-source project – which seems to have taken developer share from the elastic project – up to 25% in some developer surveys[11]. All of this means that Elastic has a weaker new logo engine, having to compete with Amazon and other open-source tools, and they have more limited pricing power given the Amazon backed viability of the open-source project. We particularly like Elastic as a comp given, they also capitalize sales commissions (7% of revs of FY25A) like Confluent does
  - Mongo: We view as a historically higher quality business vs. Confluent, and as an aspirational rule of, – primarily due to their stronger competitive positioning vs. competitors and more mission critical product. On the former, while Mongo does have an open-source version and a self-serve option – the community is much smaller than Confluent or Elastic. The reason for this dovetails with the mission criticality of the product – you simply cannot have the database that runs your user applications miss an interaction or an entry – often this means missing revenue generating SLAs. This is exacerbated by Mongo's focus on semi-structured / more scalable applications – that require more technical complexity for ensuring write / read consistency. Lastly, Mongo does capitalize their commission expenses but it's usually half of what Confluent / Elastic expenses, providing a bit cleaner of a P&L

	Elastic					Mongo				
	0	1	2	3	4	0	1	2	3	4
			FYE Apr 31,					FYE Jan 31,		
	2021A	**2022A**	**2023A**	**2024A**	**2025A**	**2021A**	**2022A**	**2023A**	**2024A**	**2025A**
ARR	$658.2	$886.9	$1,024.5	$1,243.9	$1,447.0	$655.8	$1,032.9	$1,392.7	$1,779.8	$2,123.8
% Growth		*35%*	*16%*	*21%*	*16%*		*58%*	*35%*	*28%*	*19%*
% NRR	*130%*	*130%*	*117%*	*110%*	*112%*	*120%*	*120%*	*120%*	*120%*	*118%*
% of Growth from NRR		*86%*	*110%*	*47%*	*74%*		*35%*	*57%*	*72%*	*93%*
GAAP										
Revenue	$608.5	$862.4	$1,069.0	$1,267.3	$1,483.3	$590.4	$873.8	$1,284.0	$1,683.0	$2,006.4
% Growth		*42%*	*24%*	*19%*	*17%*		*48%*	*47%*	*31%*	*19%*
Gross Profit	$447.4	$630.2	$772.4	$937.2	$1,103.4	$413.3	$614.3	$934.7	$1,258.5	$1,471.1
% Gross Margin	*74%*	*73%*	*72%*	*74%*	*74%*	*70%*	*70%*	*73%*	*75%*	*73%*
S&M	$273.9	$406.7	$503.5	$559.6	$617.2	$325.1	$471.9	$699.2	$782.8	$871.1
% of Revenue	*45%*	*47%*	*47%*	*44%*	*42%*	*55%*	*54%*	*54%*	*47%*	*43%*
R&D	$199.2	$273.8	$313.5	$342.0	$365.8	$205.2	$308.8	$421.7	$515.9	$596.8
% of Revenue	*33%*	*32%*	*29%*	*27%*	*25%*	*35%*	*35%*	*33%*	*31%*	*30%*
G&A	$103.8	$123.4	$143.2	$160.6	$175.2	$92.3	$122.9	$160.5	$193.6	$219.2
% of Revenue	*17%*	*14%*	*13%*	*13%*	*12%*	*16%*	*14%*	*12%*	*12%*	*11%*
OpEx	$576.9	$803.9	$991.5	$1,067.1	$1,158.3	$622.6	$903.7	$1,281.4	$1,492.3	$1,687.2
% Operating Margin	*95%*	*93%*	*93%*	*84%*	*78%*	*105%*	*103%*	*100%*	*89%*	*84%*
GAAP EBIT	**(129.5)**	**(173.7)**	**(219.2)**	**(129.9)**	**(54.9)**	**(209.3)**	**(289.4)**	**(346.7)**	**(233.7)**	**(216.1)**
% EBIT Margin	*-21%*	*-20%*	*-21%*	*-10%*	*-4%*	*-35%*	*-33%*	*-27%*	*-14%*	*-11%*
D&A	$17.2	$19.7	$20.2	$18.0	$12.3	$20.1	$10.8	$13.2	$12.5	$14.5
% Revenue	*3%*	*2%*	*2%*	*1%*	*1%*	*3%*	*1%*	*1%*	*1%*	*1%*
GAAP EBITDA	(112)	(154)	(199)	(112)	(43)	(189.24)	(278.52)	(333.48)	(221.25)	(201.56)
% GAAP EBITDA Margin	*-18%*	*-18%*	*-19%*	*-9%*	*-3%*	*-32%*	*-32%*	*-26%*	*-13%*	*-10%*
SBC	$93.7	$141.2	$204.0	$250.5	$269.9	$149.3	$288.3	$399.4	$493.6	$512.2
% of Revenue	*15%*	*16%*	*19%*	*20%*	*18%*	*25%*	*33%*	*31%*	*29%*	*26%*
Non-GAAP EBITDA	**(18.6)**	**(12.8)**	**5.1**	**138.6**	**227.3**	**(39.9)**	**9.8**	**66.0**	**272.3**	**310.6**
% GAAP EBITDA Margin	*-3%*	*-1%*	*0%*	*11%*	*15%*	*-7%*	*1%*	*5%*	*16%*	*15%*
Non-GAAP EBIT	**(35.8)**	**(32.5)**	**(15.1)**	**120.6**	**215.0**	**(60.0)**	**(1.1)**	**52.8**	**259.9**	**296.1**
% Non-GAAP EBIT	*-6%*	*-4%*	*-1%*	*10%*	*14%*	*-10%*	*0%*	*4%*	*15%*	*15%*

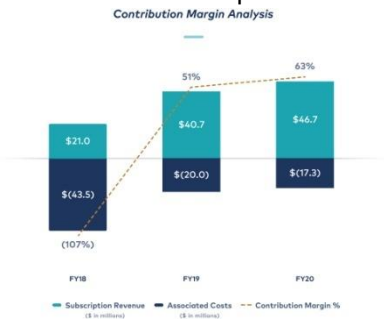
[11] https://newsletter.pragmaticengineer.com/p/the-pragmatic-engineer-2025-survey-part-2

Given management targets 25% Non-GAAP EBIT at exit, Confluent's S&M leverage via net retention endurance, and comps that bound margins at MDD and rule of at ~low 30% / ~low 40% - we feel comfortable calling for Confluent to be ~rule of high 30's / ~20% Non-GAAP NTM Margins at exit.

As a sense check we can see what this means for some of our major margin levers

- Illustrative Unit Economics (2021 S1) – management provides the unit economics for the 2018 cohort that we think are helpful from a cost allocation perspective but likely not representative of current margins.
  - (-) Lower GMs: Cloud was launched in late 2017 / early 2018 which meant a decent portion of the 2018 cohort likely was on prem, a higher split and therefore a higher gross margin than the vast majority of cloud customers that compose newer cohorts
  - (-) Expansion: Given the smaller size of the cohort vs. estimated recent cohorts & the cloud lever enabling upsell, we think the expansion curve for the 2018 cohort is much higher than recent customers

  We can infer two things from the above: CMs have likely stayed stable / decreased vs. the 2018 cohort, and much of the margin improvement has stemmed from operating leverage, given the overall margin improvement but stable / decreasing CMs



FY18 (*Barely Cloud*)	T=0	T=1	T=2	T=3	T=4	T=5	T=6
Revenue	$21.00	$40.70	$46.70	$80.00	$88.00	$95.04	$102.64
Cost of Revenue	$3.35	$6.48	$7.44	$13.16	$14.93	$16.62	$18.48
GMs	*84%*	*84%*	*84%*	84%	83%	83%	82%
GAAP Gross Profit	$17.65	$34.22	$39.26	$66.84	$73.07	$78.42	$84.17
% Growth		*94%*	*15%*	70%	9%	7%	7%
Cost of Sales	$40.17	$13.47	$9.89	$16.53	$17.73	$18.65	$17.96
% of Revenue	*191%*	*33%*	*21%*	21%	20%	20%	18%
% Growth				*67%*	7%	5%	*-4%*
% of Non-GAAP Total S&M Cost	*74%*	*12%*	6%				
Total Cost of Subscription Revenue	$43.52	$19.95	$17.33	$29.69	$32.66	$35.27	$36.44
CMs	($22.52)	$20.75	$29.37	$50.31	$55.34	$59.77	$66.20
CM %	-107%	51%	63%	63%	63%	63%	65%
CM growth rate			*42%*	*71%*	*10%*	*8%*	*11%*

FY20		T=0	T=1	T=2	T=3	T=4
Revenue	93%	$51.15	$86.96	$117.39	$164.34	$180.78
Cost of Revenue		$11.25	$19.13	$25.83	$36.16	$39.77
GMs		78%	78%	78%	78%	78%
GAAP Gross Profit		$39.90	$67.82	$91.56	$128.19	$141.01
% Growth			*70%*	*35%*	40%	10%
Cost of Sales		$117.65	$28.78	$24.86	$31.23	$30.73
% of Revenue		230%	33%	21%	19%	17%
% Growth					*26%*	-2%
% of Non-GAAP Total S&M Cost		*75%*	9%	5%		
Total Cost of Subscription Revenu		$128.90	$47.91	$50.69	$67.38	$70.50
CMs		($77.75)	$39.05	$66.70	$96.96	$110.28
CM %		-152%	45%	57%	59%	61%
CM growth rate				*71%*	*45%*	*14%*
NRR			*170%*	*135%*	*140%*	*110%*
CAC (k)		$88.10				

Thus, we try to depict roll forward FY18 with our imputed expansion curve and select a more representative cohort FY20 for comparison, given the larger cohort size ($52M and higher cloud mix). While the FY18 cohort reaches ~65% CMs, the lower cloud margins and higher CACs mean that the CMs stay in the low 60s for the FY20 cohort – a place we believe cohorts have stabilized given the inconsistent new logo + weaker expansion vs. earlier cohorts.

<u>BOE margin calc</u>: The main drivers of margin are

- GMs: We do not model significant operating leverage / improvement here over the hold (78% FY'24 to 80% FY'30 non-GAAP) – primarily due to CSP pricing power on infrastructure usage. AWS, GCP, and Azure all offer some form of alternative offerings & generally their hosting services are an oligopoly. GAAP gross margins have scaled +800bps from '21-'24 and continue to scale up to 75%, however this is generally the level that larger businesses like elastic / mongo tend to top out at (~75% GAAP GMs). Below we outline the directional drivers
  - <u>(/) Mix Share</u>: Cloud is a lower margin product given the hosting costs that Confluent incurs for managing the solution via their cloud accounts vs. on-prem deployments – our estimate is that On-Prem is 500-1000bps higher margin, thus as the business has scaled and continues to scale their cloud business (mgmt. target is 85% cloud share)– this should act as a downward pressure that is partially offset by purchasing economies
  - <u>(+) Multi-tenancy & Attach</u>: The offsetting consideration is the rise of multi-tenancy, which effectively allocates virtual machines / resources to physical clusters at a more efficient rate. It's extremely difficult to estimate what % of customers are deploying their solution in a multi-tenant environment. However, as we see the DSP take-off these additional services should be easier to house within the same customer account and flow through at a higher gross margin
  - <u>(+) VPC</u>: The last consideration is the rise of VPC deployments, as evidenced by the excitement and traction that WarpStream / Redpanda have been able to build – (recall our estimates are that this should be of interest for 25-30% of user base). This has a positive impact on gross margins because it offloads the storage hosting considerations to the customer, effectively acting as a partial on-prem deployment. Qualitatively, we think these deployments should continue to gain share as organizations worry about data transfer policies & therefore as the business starts to mix into VPC – this should have a positive impact on margins.
- S&M: We have non-GAAP S&M growing at a 14% CAGR over from '24-'30, with Revenue growing at an 18% CAGR over that same time-period. We model two key drivers of margin expansion

- (+) Embedded Expansion: In our cohorts above we have the FY20 contribution margin reaching ~61% after T=4, we think the combination of agentic applications targeting digital native customers + more mature Kafka deployments with these customers means that Confluent should see slightly higher expansion (+5%) over this reference cohort, driving improved contribution margins and therefore operating margins.
- (+) DSP + SoW: While embedded expansion only captures the core streaming uplift, the increased adoption / mix towards DSP & improved SoW gains from improved cluster selection should materially contribute to improvements in margin expansion.

At exit we have non-GAAP S&M reaching 34% of revenue, +800bps improvement from '24-'30 (this ratio improved +2100 bps from '21-'24). As a sense check we can see what this level of S&M spend and new logo assumptions mean for our CACs

CACs

		FYE Dec 31,				Base						CAGR	
% of S&M in period to new	75%	2021A	2022A	2023A	2024E	2025E	2026E	2027E	2028E	2029E	2030E	'21-'24	'24-'30
Revenue		$388	$586	$777	$964	$1,162	$1,394	$1,663	$1,959	$2,280	$2,631	*35%*	*18%*
# of customers		1,511.81	1,273.35	713.89	1,111.04	1,265.28	1,419.52	1,573.75	1,727.99	1,567.99	1,407.99	-10%	4%
S&M Metrics													
GAAP S&M		$319	$456	$505	$547	$587	$664	$749	$830	$917	$996		
Non-GAAP S&M		$264	$357	$380	$407	$474	$550	$634	$719	$806	$895		
Assumed Allocation to New		75%	75%	75%	75%	75%	75%	75%	75%	75%	75%		
GAAP S&M - New Logo S&M		$239.50	$342.34	$378.70	$410.53	$440.57	$498.37	$561.62	$622.34	$687.96	$746.67		
% Growth			*43%*	*11%*	*8%*	*7%*	*13%*	*13%*	*11%*	*11%*	*9%*		
Non-GAAP S&M - New Logo S&M		$197.87	$267.81	$285.27	$305.59	$355.82	$412.73	$475.39	$539.62	$604.79	$670.91		
% Growth			*35%*	*7%*	*7%*	*16%*	*16%*	*15%*	*14%*	*12%*	*11%*		
GAAP S&M - CAC		**158.42**	**268.85**	**530.47**	**369.50**	**348.20**	**351.09**	**356.86**	**360.15**	**438.75**	**530.31**	33%	6%
% Growth			*70%*	*97%*	*-30%*	*-6%*	*1%*	*2%*	*1%*	*22%*	*21%*		
Non-GAAP S&M - CAC		**130.88**	**210.32**	**399.60**	**275.05**	**281.22**	**290.75**	**302.08**	**312.28**	**385.71**	**476.50**	28%	10%
% Growth			*61%*	*90%*	*-31%*	*2%*	*3%*	*4%*	*3%*	*24%*	*24%*		
New Logo ACVs		$67,433	$40,989	$51,809	$51,372	$52,913	$54,500	$52,865	$51,280	$49,741	$48,249		

H1 2025 Estimate	
Non-GAAP S&M	$215
(x) % allocated to New	75%
(=) New Logo S&M	160.88
(/) H1 New Logos	545
(=) Est. CACs	$295
Premium / Discount to 2025 Forecast	*5%*

Two key takeaways:

- We feel good about our 2025 CACs, given our full year estimates are within 5% of H1 CACs, with an extremely difficult Q2
- While our CACs compound a lot slower from '24-'30 vs. '21-'24, we take comfort in the improved competitive positioning / platform offering for downmarket logos / CSPs, and we maintain the 700bps spread between revenue and CAC over the hold

- Tactical Levers:

  SBC: This adds another asymmetric element, that could magnify a downside scenario as slow growth could cause brain drain – especially in the R&D department where formers continually highlight great organizational talent. Confluent has issued a lot of SBC, below we compare their SBC issuance across a band of high growth and slower growth infrastructure businesses. We believe that Confluent should be closer to Elastic and Mongo in their SBC issuance – with management targeting mid-teens in the long term – for comparison we have SBC at 27% of revs in FY'27 and exit in line with management at mid-teens.

	Elastic **2025A**	Mongo **2025A**	Snowflake **2025A**	INFA **2024A**	CFLT **2024A**
Rev Growth	*17%*	*19%*	*29%*	*-4%*	*24%*
SBC % of Revenue	*18%*	*26%*	*41%*	*14%*	*43%*

Valuation & Exit

The basic building blocks of our operating scenario are

- Net New Logos & New Logo ARR accelerating until ~2028, and then decelerating given pricing and packaging changes that help them simultaneously target down-market and enterprise logos
- Net Retention that holds at ~113-115% over the hold thanks to stable expansion curves vs. benchmarks, moderate attach (~40% 1 product / ~30% 2 products / ~ 20% 3 products), and net SoW gain due to improved selection
- ~5% operating leverage gains as the business continues to scale cloud and mixes into higher margin upsell

CONFLUENT

FDSO = grows at 4% over the hold in line with '21-'24, driven by continued SBC issuance that eventually attenuates to ~12-15% of revenue

(FY $M)	Historicals				Base							CAGR	
	2021A	**2022A**	**2023A**	**2024A**	**2025E**	**2026E**	**2027E**	**2028E**	**2029E**	**2030E**	**2031E**	**'21-'24**	**'24-'30**
ARR Build													
Bookings	$207	$233	$253	$268	$293	$364	$417	$454	$509	$571		*9%*	*13%*
% Growth		*12%*	*9%*	*6%*	*9%*	*24%*	*15%*	*9%*	*12%*	*12%*			
Net New ARR		$189	$190	$192	$192	$257	$286	$315	$345	$380			*12%*
% Growth			*0%*	*1%*	*0%*	*33%*	*12%*	*10%*	*10%*	*10%*			
Net New Logos	1,370	1,060	430	840	917	1,084	1,184	1,369	1,154	948			
BoP ARR	$253	$433	$621	$811	$1,003	$1,195	$1,451	$1,738	$2,052	$2,397			
(+) New	$102	$52	$37	$57	$67	$77	$83	$89	$78	$68		*(18%)*	*3%*
(+) Upsell	$105	$180	$216	$211	$226	$287	$334	$365	$431	$503		*26%*	*16%*
(-) Downsell incl. Churn	($27)	($44)	($64)	($77)	($100)	($108)	($131)	($139)	($164)	($192)			
EoP ARR	**$433**	**$621**	**$811**	**$1,003**	**$1,195**	**$1,451**	**$1,738**	**$2,052**	**$2,397**	**$2,777**		*32%*	*19%*
% Growth		*44%*	*31%*	*24%*	*19%*	*21%*	*20%*	*18%*	*17%*	*16%*			
GDR	*89%*	*90%*	*90%*	*91%*	*90%*	*91%*	*91%*	*92%*	*92%*	*92%*			
NDR	*131%*	*132%*	*125%*	*117%*	*113%*	*115%*	*114%*	*113%*	*113%*	*113%*			
Cloud	$135	$274	$400	$552	$709	$923	$1,179	$1,481	$1,832	$2,241		*60%*	*26%*
Confluent Platform	$298	$348	$411	$451	$486	$529	$558	$572	$565	$535		*15%*	*3%*
Total ARR	$433	$621	$811	$1,003	$1,195	$1,451	$1,738	$2,052	$2,397	$2,777		*32%*	*19%*
% Growth		*44%*	*31%*	*24%*	*19%*	*21%*	*20%*	*18%*	*17%*	*16%*			
% - Cloud		*102%*	*46%*	*38%*	*28%*	*30%*	*28%*	*26%*	*24%*	*22%*			
% - Platform		*17%*	*18%*	*10%*	*8%*	*9%*	*6%*	*2%*	*(1%)*	*(5%)*			
Net New Cloud		$138	$126	$152	$157	$214	$257	$301	$351	$409			
Net New Support		$50	$63	$40	$35	$42	$30	$13	($7)	($29)			
BoE Drivers (Not Linked)													
Core Expansion (Est.)	$388	$533	$683	$822	$1,001	$1,200	$1,428	$1,681	$1,944	$2,203			
DSP ARR	$44	$89	$128	$176	$75	$108	$159	$234	$335	$461			
SoW Expansion				$0	$12	$25	$37	$49	$62	$74			
Total ARR	$433	$621	$811	$999	$1,088	$1,333	$1,624	$1,965	$2,340	$2,738			

We exit on a 8-month stubbed amount of ~$570M of NTM FCF (pre-SBC), at 28x NTM FCF (pre-SBC), which we think is fair taking a look at the comps. Specifically, we hone in on Mongo & Elastic both thematically & financially similar to Confluent at exit, and we center around Elastics 30x NTM FCF (recall Elastic is rule of low 30s on an LTM basis & is expected to be rule of 27 on an NTM basis).

Returns are 2.2x MoM / 17% IRR at an entry price of $18 (*price as of 8/26*) & 2.5x MoM / 20% IRR at our cost basis

(=) NTM FCF	$569
(x) Multiple (NTM EBIT)	28x
TEV	**15,921**
(+/-) Net Debt	2,721
Equity Value	18,643
/ Shares	463
Share Price	$40.3
(/) Current Share Price	$18.0
MoM	**2.2x**
IRR	**17%**

($ in millions)		Equity	TEV / FCF	Rev	FCF
Company	**TEV**	**Value**	**NTM**	**NTM**	**NTM**
Data & Analytics					
MongoDB, Inc.	$17,277	$17,935	92.25	$2,360	$187
Informatica Inc.	$8,283	$7,534	18.89	$1,706	$438
Elastic N.V.	$8,233	$8,391	29.50	$1,666	$279
Teradata Corporation	$2,115	$2,016	7.76	$1,620	$273
Snowflake Inc.	$65,862	$65,831	67.03	$4,764	$983

	FYE Apr 31,				
Elastic Comp	**2021A**	**2022A**	**2023A**	**2024A**	**2025A**
FCF, Pre SBC	**18**	**11**	**57**	**169**	**286**
% Margin	*3%*	*1%*	*5%*	*13%*	*19%*
FCF, Post SBC	(75)	(130)	(147)	(81)	16
% Margin	*-12%*	*-15%*	*-14%*	*-6%*	*1%*
Share Price	114	127	65	87	98
(x) Shares Outstanding	91	94	97	102	106
Market Cap	10,357	11,977	6,340	8,835	10,380
(+/-) Net Debt	(376)	(269)	(323)	(493)	(806)
TEV	9,981	11,708	6,016	8,342	9,574
TEV / NTM FCF, Pre-SBC	932	207	35.6	29.2	29.5

CONFLUENT

Non GAAP P&L - Non-Stub	Historicals				Base							CAGR	
	2021A	2022A	2023A	2024A	2025E	2026E	2027E	2028E	2029E	2030E	2031E	'21-'24	'24-'30
Subscription Revenue	$347	$535	$729	$922	$1,116	$1,344	$1,610	$1,902	$2,221	$2,572	$2,958	*38%*	*19%*
Services Revenue	$41	$51	$48	$42	$45	$50	$54	$57	$59	$59	$67	*1%*	*6%*
Total Revenue	**$388**	**$586**	**$777**	**$963.64**	**$1,162**	**$1,394**	**$1,663**	**$1,959**	**$2,280**	**$2,631**	**$3,026**	*35%*	*18%*
% Growth		*51%*	*33%*	*24%*	*21%*	*20%*	*19%*	*18%*	*16%*	*15%*	*15%*		
Subscription Gross Profit	$265	$413	$580	$751	$911	$1,098	$1,316	$1,556	$1,820	$2,109		*41%*	*19%*
Services Gross Profit	$4	$5	$6	$2	$3	$3	$3	$3	$3	$3		*(16%)*	*3%*
Total Gross Profit	**$270**	**$417**	**$586**	**$754**	**$913**	**$1,100**	**$1,319**	**$1,559**	**$1,823**	**$2,112**		*41%*	*19%*
% Margin	*70%*	*71%*	*75%*	*78%*	*79%*	*79%*	*79%*	*80%*	*80%*	*80%*			
% Margin - Subscription	*76%*	*77%*	*80%*	*81%*	*82%*	*82%*	*82%*	*82%*	*82%*	*82%*			
% Margin - Services	*10%*	*9%*	*12%*	*6%*	*6%*	*6%*	*5%*	*5%*	*5%*	*5%*			
S&M	$260	$355	$372	$407	$474	$550	$634	$719	$806	$895		*16%*	*14%*
R&D	$111	$160	$186	$225	$263	$305	$352	$401	$450	$500		*27%*	*14%*
G&A	$60	$79	$85	$95	$112	$133	$156	$182	$208	$237		*16%*	*17%*
Operating Costs	**$430**	**$594**	**$643**	**$726**	**$850**	**$989**	**$1,143**	**$1,302**	**$1,464**	**$1,631**		*19%*	*14%*
S&M % of Revenue	*67%*	*60%*	*48%*	*42%*	*41%*	*39%*	*38%*	*37%*	*35%*	*34%*			
R&D % of Revenue	*29%*	*27%*	*24%*	*23%*	*23%*	*22%*	*21%*	*20%*	*20%*	*19%*			
G&A % of Revenue	*15%*	*14%*	*11%*	*10%*	*10%*	*10%*	*9%*	*9%*	*9%*	*9%*			
(+) D&A	$4	$8	$14	$22	$27	$32	$38	$45	$52	$60	$69	*83%*	*18%*
Non-GAAP EBITDA	**($157)**	**($169)**	**($43)**	**$50**	**$90**	**$144**	**$214**	**$302**	**$410**	**$541**	**$700**		*49%*
% Margin	*(40%)*	*(29%)*	*(6%)*	*5%*	*8%*	*10%*	*13%*	*15%*	*18%*	*21%*	*23%*		
Non-GAAP EBIT	**($161)**	**($177)**	**($57)**	**$27**	**$64**	**$112**	**$176**	**$257**	**$358**	**$481**	**$631**		*61%*
% Margin	***(41%)***	***(30%)***	***(7%)***	***3%***	***5%***	***8%***	***11%***	***13%***	***16%***	***18%***	***21%***		
(-) SBC	155.6	277.7	349.8	417.1	441.4	446.1	449.1	430.9	433.2	394.7	363.1		
% Revenue	*40%*	*47%*	*45%*	*43%*	*38%*	*32%*	*27%*	*22%*	*19%*	*15%*	*12%*		
(-) Other	23.4	8.1	71.6	29.5	38.2	1.2	1.2	1.2	1.2	1.2			
% Revenue	*6%*	*1%*	*9%*	*3%*	*3%*	*0%*	*0%*	*0%*	*0%*	*0%*			
GAAP EBIT	(339.6)	(462.5)	(478.8)	(419.1)	(377.9)	(334.4)	(273.1)	(173.4)	(75.0)	86.2	267.6		
% Revenue	*-88%*	*-79%*	*-62%*	*-43%*	*-33%*	*-24%*	*-16%*	*-9%*	*-3%*	*3%*	*9%*		
(+) D&A	3.6	7.6	13.9	22.1	26.9	33.2	39.3	46.1	53.5	61.5			
(+) Deferred Acquisition Costs	$27	$37	$46	$54	$51	$57	$66	$77	$91	$106			
(+) Other / One-Time	23.4	8.1	71.6	29.5	38.2	1.2	1.2	1.2	1.2	1.2			
(+) SBC	$156	$278	$350	417.1	$441	$446	$449	$431	$433	$395			
(-) Cash Taxes	$1	($0)	$2	$0	$0	$0	$0	$0	$0	$0			
(-) Ch. In NWC	$22	($32)	($56)	($67)	($78)	($60)	($73)	($87)	($103)	($120)			
(-) Cap Ex	($9)	($14)	($21)	($24)	($29)	($34)	($40)	($46)	($53)	($60)			
FCF, Pre SBC	**(116.0)**	**(178.5)**	**(72.1)**	**13.0**	**72.8**	**108.7**	**169.7**	**248.6**	**347.7**	**469.4**	***618.2***		***82%***
% of Revenue	*-30%*	*-30%*	*-9%*	*1%*	*6%*	*8%*	*10%*	*13%*	*15%*	*18%*	*20%*		
FCF, Post SBC	($272)	($456)	($422)	($404)	($369)	($337)	($279)	($182)	($86)	$75	$255.1		
Taxes													
GAAP EBIT				(419.1)	(377.9)	(334.4)	(273.1)	(173.4)	(75.0)	86.2	267.6		
(-) Taxes @25%				-	-	-	-	-	-	21.6	66.9		
(+) DTA Used				-	-	-	-	-	-	21.6	66.9		
Cash Taxes				-	-	-	-	-	-	-	0		
NOL BoP				$0	$419	$797	$1,131	$1,405	$1,578	1,653	1,631		
(+) New				$419	$378	$334	$273	$173	$75	$0	$0		
(-) Usage				-	-	-	-	-	-	(21.6)	(66.9)		
NOL EoP				$419	$797	$1,131	$1,405	$1,578	$1,653	$1,631	1,565		
Balance Sheet													
BoP Debt					1,092	1,092	1,092	-	-	-			
(-) Paydown					-	-	(1,092)	-	-	-			
EoP Debt					1,092	1,092	-	-	-	-			
Net Debt					*(893)*	*(1,107)*	*(1,398)*	*(1,782)*	*(2,287)*	*(2,938)*			
Metrics													
Rule of:		*22%*	*27%*	*29%*	*28%*	*30%*	*32%*	*33%*	*34%*	*36%*	*38%*		
Magic Number	*0.80x*	*0.66x*	*0.68x*	*0.66x*	*0.62x*	*0.66x*	*0.66x*	*0.63x*	*0.63x*	*0.64x*			
GM Magic Number		*0.47x*	*0.51x*	*0.52x*	*0.48x*	*0.52x*	*0.52x*	*0.50x*	*0.50x*	*0.51x*			
LTV:CAC		*2.2x*	*2.4x*	*2.5x*	*2.3x*	*2.6x*	*2.6x*	*2.7x*	*2.7x*	*2.7x*			
FDSO	334	345	358	372	390	408	425	439	452	463		*4%*	*4%*

However, this business is incredibly torqued from an ARR uplift perspective, meaning there is high co-variance with a few variables. New logo re-acceleration largely depends on the PMF of the new network-sensitive use cases. These new logo cohorts flow through to our embedded expansion and compose 50% of the logo base for product cross-sell. Moreover, this kernel of PMF for new TCO focused selection flows through our SoW assumptions, which depends on winning low-latency sensitive and high throughput workloads. All of this eventually drives contribution margins & operating margins.

FOO:

Upside scenario: DSP attach reaches ~50% for one product, 40% for two products, and ~25% for three products due to continued adoption in MongoDB + Snowflake, solid adoption in Databricks + Postgres, and moderate adoption in application connectors (e.g., SF and Service Now). Additionally, agentic workloads play a bigger role in driving expansion vs. our baseline cohort (+5% in base vs. + 15% in the upside scenario). Both growth drivers also expand margins, as our contribution margins past the point of customer acquisition reach low 70's due to an additional ~10-15 pts of gross dollar expansion per cohort

Downside Scenario: Snowflake and Mongo become increasingly competitive, either via M&A or organic product development, and Databricks fails to take off as a partner – severely limiting attach to ~35% for one product, ~25% for two products, and ~10% for three products, minimal improvements from today's attach. Moreover, newer cohorts turn out to be more difficult to win, compressing new logo, and when they do convert they have a lower ceiling of Kafka use cases due to limited AI adoption or a preference for DIY. Growth slows to LDD, and margins are similarly impacted

Fan

	2024A	Downside 2030E	CAGR '24-'30	Base 2030E	CAGR '24-'30	Upside 2030E	CAGR '24-'30
# of Logos	5,800	8,901	*7%*	12,455	*14%*	13,568	*15%*
ARR	$1,003	$2,280	*15%*	$2,777	*19%*	$3,460	*23%*
% Growth	24%	11%		16%		22%	
NRR	117%	109%		113%		119%	
NTM Revenue	$964	$2,409	*16%*	$3,026	*21%*	$3,869	*26%*
% Growth	*21%*	*10%*		*15%*		*20%*	
NTM Non-GAAP Operating Costs	$726	$1,604	*14%*	$1,798	*16%*	$2,119	*20%*
% Revenue	75%	*67%*		*59%*		*55%*	
NTM Non-GAAP EBIT	$27	$307	*49%*	$631	*69%*	$986	*82%*
% Non-GAAP EBIT Margin	3%	13%		21%		25%	
SBC + Other GAAP	$447	$290	*-7%*	$364	*-3%*	$466	*1%*
% of Revenue	*46%*	*12%*		*12%*		*12%*	
NTM GAAP EBIT	($419)	$16		$266		$521	
% EBIT Margin	*-43%*	*1%*		*9%*		*13%*	
D&A	$22	$55	*16%*	$69	*21%*	$88.7	*26%*
% Revenue	*2%*	*2%*		*2%*		*2%*	
FCF, Pre-SBC	$13.0	$368.1	*75%*	$618.2	*90%*	$864.6	*101%*
% Revenue	*1%*	*15%*		*20%*		*22%*	
% Growth							
FDSO	372.35	469.6	*4%*	462.53	*4%*	453.33	*3%*

Entry at $18			
Exit Multiple (NTM FCF)	**23.0**	**28.0**	**33.0**
Exit Multiple (NTM Rev)	**3.5**	**5.7**	**7.4**
MoM / IRR	**1.3x / 4.8%**	**2.2x / 17.5%**	**3.5x / 28.4%**

Entry at $15			
Exit Multiple (NTM FCF)	**23.0**	**28.0**	**33.0**
Exit Multiple (NTM Rev)	**3.5**	**5.7**	**7.4**
MoM / IRR	**1.5x / 8.7%**	**2.7x / 21.9%**	**4.2x / 33.2%**

Tibco: Tibco was a legacy messaging queue system that offered event-driven streaming alongside other application integration systems. In the early 2010s the rise of Red Hat, Mulesoft, and Kafka all began to steal significant share while Tibco was unable to fully transition to the cloud. They were eventually purchased by Vista in November 2014 at ~16-17x NTM PF FCF, Pre-SBC / ~20-21x NTM PF[12] FCF, Post-SBC. We leverage pro-forma because we imagine that Vista had a plan of taking out costs and raising prices for growth. Tibco has slightly higher Pre-SBC LTM CF margins ~19% vs. our 13-15%, however the business is growing half as fast from a revenue perspective. Thus, we center on a 23x NTM FCF, Pre-SBC scenario for our downside given Tibco cleared at 16-17x NTM PF FCF

	FYE Nov, 30th					CAGR	Actuals	Sponsor	No Sponso
	2009A	2010A	2011A	2012A	2013A	'09-'13	LTM	NTM	NTM
GAAP P&L									
Revenue	$621	$754	$920	$1,025	$1,070	*15%*	1,076	1,141	1,119.28
% Growth	*-4%*	*21%*	*22%*	*11%*	*4%*			6%	4%
Gross Profit	$462.34	$556.21	$672.82	$741.80	$762.51	*13%*	$762.5	$826.2	$805.9
% Gross Margins	*74%*	*74%*	*73%*	*72%*	*71%*		*71%*	72%	72%
Operating Costs	$373.73	$441.06	$518.44	$564.08	$618.09	*13%*	$632.7	$642.7	654.78
% Revenue	*60%*	*58%*	*56%*	*55%*	*58%*		*59%*	56%	59%
GAAP Operating Income	$88.60	$115.16	$154.37	$177.72	$144.42	*13%*	$129.8	$183.5	$151.1
% Margin	*14%*	*15%*	*17%*	*17%*	*13%*		*12%*	*16%*	*14%*
(+) D&A	$15.1	$13.4	$13.1	$14.9	$16.3		$19.15		
(+) Other / One-Time	14.6	19.8	29.9	21.7	32.9		24.98		
(+) SBC	23.5	32.2	48.9	61.1	53.6		46.29		
(-) Cash Taxes	(22.2)	(28.8)	(38.6)	(44.4)	(36.1)		(32.46)		
(-) Cash Interest	(2.4)	(3.0)	(3.0)	(10.5)	(14.9)		(14.9)		
(-) Ch. In NWC	40.4	74.9	40.2	60.4	37.4		61.62		
(-) Cap Ex	(5.4)	(6.4)	(14.0)	(23.7)	(19.3)		(32.05)		
FCF, Pre SBC	$152.3	$217.4	$230.9	$257.2	$214.4		$202.4	$260.5	$226.6
FCF Margin	*25%*	*29%*	*25%*	*25%*	*20%*		*19%*	23%	20%
FCF, Post SBC	$128.8	$185.1	$182.0	$196.0	$160.8		$156.1	$211.4	$178.5
	21%	*25%*	*20%*	*19%*	*15%*		*15%*	19%	16%
Taxes									
GAAP EBIT	$88.60	$115.16	$154.37	$177.72	$144.42		$129.82		
(-) Taxes @ 25% ETR	$22.15	$28.79	$38.59	$44.43	$36.10		$32.46		
	Nov-09	Nov-10	Nov-11	Nov-12	Nov-13			Sponsor	No Sponsor
Shares	167	164.522	166.287	164	163.169				
(x) Share Price	$9.6	$19.7	$23.9	$22.0	$22.5				
Equity Value	-1 $1,608	$3,243	$3,976	$3,598	$3,668			$4,200	
(+/-) Net Debt	$250.3	$207.4	$242.7	$237.3	$205.9			$100	
Total Enterprise Value	$1,358	$3,035	$3,733	$3,360	$3,462			$4,300	
TEV / NTM PF Revenue	1.80	3.30	3.64	3.14				3.77	3.84
TEV / NTM PF FCF, pre-SBC	6.25	13.15	14.52	15.67				16.51	18.98
TEV / NTM PF FCF, post-SBC	7.34	16.67	19.04	20.90				20.34	24.09
								24.76	

Take-out	
Avg. NTM FCF (Pre-SBC)	12.40
% Premium	33%
Avg. NTM FCF (Pre-SBC)	15.99
% Premium	27%

[12] The major PF assumption we make is a one-time price hike that increases growth ~2-3%

Risks:

- Competitive Sideswipes: While mechanically the business has high-covariance due to the connection between new logos and expansion, qualitatively there is a wide fan around competitive positioning.
    - MongoDB: While we think the risk is limited to Mongo specific customers, and the rise of postgres should help combat any share loss to Mongo, the launch of stream processing poses a huge risk given Mongo's importance as source / target for Kafka data. Ultimately, if Mongo continues to invest in a lightweight version of Flink + Connect, this could significantly impair our platform adoption
    - Snowflake + RedPanda: We have not modeled share gain increases in Snowflake connector adoption given the materiality of these rumors. However, if RedPanda continues to burn in order to reach a scale that Snowflake feels justify their price tag, Confluent's Snowflake connectors could be significantly at risk, as Snowflake invests behind RedPanda's platform build out and core streaming engine. Even absent a RedPanda acquisition, Snowflake may take the Databricks relationship as a threatening move and decide to start building out lightweight streaming functionality. On a related note, RedPanda could continue burning if they can access cheap Silicon Valley capital, which could skew CACs and materially hurt contribution margins for Confluent.
    - AWS: Has largely been a moderate competitor, allowing customers to switch to Confluent freely. Moreover, AWS & Confluent have regularly collaborated on applications, with AWS attending Confluent's 2023 investor day. However, if AWS decided they wanted to simultaneously squeeze Confluent on hosting costs, and then discount significantly to win real-time workloads – there would be very little that Confluent could do. We think this outcome is unlikely given that a) this would damage future hosting relationships, with several companies migrating / switching more of their workloads to alternative CSPs & b) AWS can win the streaming workload (from a storage / compute perspective) without the middleware piece
    - Databricks: Seems to be a key partner for Confluent going forward, with co-marketing & co-selling motions both via direct & system integrators. However, they do offer adjacent products like Apache spark processing, which is another method for running stateful processing on top of streaming data.
- Continued Optimizations / AI pressure: In our base scenario we argue that optimizations have been limited to larger customers given they are closer to in-house spend band & now have more ammunition to optimize their workloads, with cheaper hybrid clusters. While smaller customers are less likely to be at capacity, and thus net downsell, there is the added risk that this behavior travels downwards, especially if new use cases are slower to deploy. Moreover, if AI productivity gains outstrip the demand / requirements for data engineers, you could see companies attack their infrastructure budgets for cost-savings, putting a large % of Confluent's customers at risk (2x productivity increase would almost 2x the number of customers who become at risk for self-managing).
- New Cohorts: We have modeled the cohorts with a slight expansion to our reference cohort. Furthermore, the pricing & packaging changes we have made should open the business up to more network focused deployments. However, if Amazon and or RedPanda decide that these customers are critical to their existence / upsell plan, Confluent may struggle to drive new logo expansion – which will further impact the embedded expansion uplift / SoW that Confluent can reach.
- 1-year Volatility: There is a risk that the stock has a weak Q3 print or fails to see the type of re-acceleration that Mongo saw in their last earnings, which investors may be baking into Q3. Balance sheet indicators signal weaker net terms on payment but encouraging long-term contracts. We think the reality for 2025 is that customers are still planning on how to use Confluent and they are early in their agent deployment maturity cycle.

	Q1 '23A	Q2 '23A	Q3 '23A	Q4 '23A	Q1 '24A	Q2 '24A	Q3 '24E	Q4 '24E	Q1 '25A	Q2 '25A
Deferred Revenue	$324.59	$327.97	$323.77	$353.01	$347.66	$348.45	$374.38	$409.20	$405.38	$455.65
Current Deferred Revenue	$295.7	$300.8	$300.6	$330.6	$330.6	$330.5	$328.8	$349.8	372.87	424.93
% Current	*91%*	*92%*	*93%*	*94%*	*95%*	*95%*	*88%*	*85%*	*92%*	*93%*
Subscription Revenue	$161	$176	$189	$203	$207	$225	$240	$251	$261	$271
Current Billings		$181.6	$189.1	$232.7	$206.9	$224.6	$238.2	$271.6	$284.0	$322.9
Total Billings		$179.9	$185.1	$232.0	$201.6	$225.5	$265.8	$285.5	$257.1	$321.1
LTM Revenue	$582	$631	$682	$729	$775	$824	$874	$922	$976	$1,022
Accounts Receivables	$169.7	$188.7	$183.2	$230.0	$199.8	$257.5	$278.7	$314.3	297.862	357.607
DSO	*106.44*	*109.12*	*98.09*	*115.13*	*94.07*	*114.11*	*116.35*	*124.41*	*111.38*	*127.69*
Days Deferred Outstanding Current	*185.51*	*173.92*	*160.96*	*165.49*	*155.60*	*146.47*	*137.30*	*138.46*	*139.43*	*151.73*
Days Deferred Outstanding Total	*203.64*	*189.65*	*173.36*	*176.72*	*163.65*	*154.42*	*156.31*	*161.98*	*151.59*	*162.70*