WHITE STAR CAPITAL

Data-driven Transformation

Through the eyes of an international investor

2023

Executive Summary

$22.1tn

Potential use-case driven impact of generative AI technologies[1]

$1.2tn

Projected data infrastructure and management market size by 2032[2]

$8tn

The global cost of cybercrime in 2023[3]

95%

Of businesses struggle with unstructured data[4]

Looking at Data-driven Transformation from a VC perspective

An unprecedented surge in data generation, enhanced computational capabilities, increased cybercrime, and advancements in data science and generative AI related technologies are all factors that have made data leaders scrutinise previous tools and infrastructure frameworks previously renowned in the modern data stack. The time is right for a new data-driven transformation where companies and tools are designed to help organisations extract the most value out of their data while providing high cost-effectiveness, return on investment, and low implementation time.

Current conditions are particularly conducive for a data-driven transformation. The rapid maturation of technologies, coupled with an increase in skilled data and engineering talent, lays a solid foundation for growth. Historical data suggests that companies focusing on innovation during economic disruptions tend to outperform their counterparts significantly, with up to 240% higher shareholder returns.[2] Furthermore, the substantial economic impact of generative AI, and its potential value creation across multiple use cases, underscores the critical role of data as a key driver of growth. This environment not only encourages but necessitates a shift towards data-centric business models, making it an optimal time for VCs to invest in data infrastructure, management, security, and tooling sectors. We believe the best opportunities within these sectors should address (i) evolving organisational needs amidst the data surge, consolidation of tools and data categories, and (ii) how companies must adapt their tech stacks for effective AI integration and leverage AI-centric technologies.

With this in mind, this report details the significance and potential influence of data and AI, explores the core aspects of the data value chain, including both its upstream and downstream processes, and outlines key investment opportunities.

Sector Lead

Alberto Toledo
Partner, CTO

Sanjay Zimmermann
Partner

Authors

Usamah Khan
Vice President

Neel Patel
Associate

Source: [1] McKinsey (2023), [2] Custom Market Insights, [3] Cybercrime Magazine, [4] Fingent, WSC Analysis (2023)

WHITE STAR CAPITAL

Contents

WHITE STAR CAPITAL

I

Data-driven Transformation: Overview

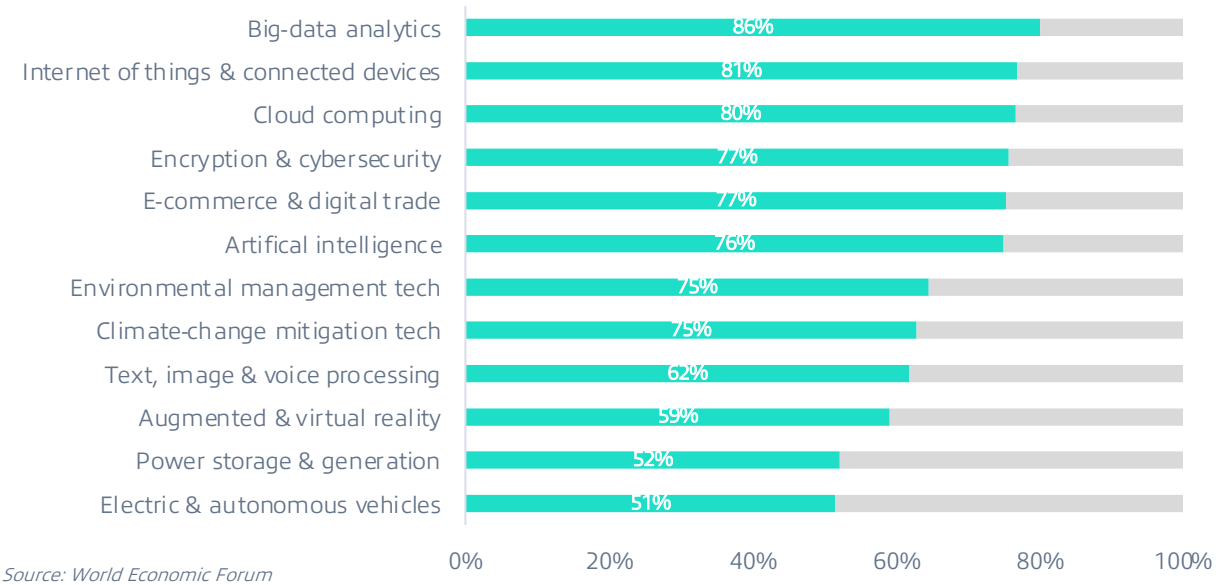
Data-driven Transformation: maximising asset value through data

The surge in data generation requires companies to responsibly manage, secure, and provision data, going beyond usage and analysis, to harness its full value

Coupled with the vast increase in data generation each year, data also now exhibits remarkable variety and complexity. We are in an era where the physical and digital worlds are increasingly intertwined. Everything from cameras and traffic sensors to heart rate monitors is producing data, offering deeper insights into the behaviours of humans and objects.

Consequently, data has emerged as a new class of corporate asset. For companies to truly capitalise on this asset, fully embracing digital transformation is essential. By digitising both their internal and external operations, companies can gather critical data. The current era offers an unprecedented opportunity to leverage this data. This is due to an ever-growing abundance of data, coupled with rapidly advancing computational power that makes managing this data more feasible, and breakthroughs in mathematical techniques and data science which enable us to extract more value from this data than ever before.

Technologies ranked by the share of organisations surveyed who are likely or highly likely to adopt certain technologies from 2023-2027



Source: World Economic Forum

Innovative solutions, related to data infrastructure, management, governance, and tooling, continue to emerge to help companies leverage their data. As a result, tech stacks and strategies are constantly and rapidly being augmented to better support enterprise businesses to operate, serve customers, and scale. Ultimately, an organisation's success hinges on its agility—the speed at which it can make and adjust strategic decisions to enhance results. A strong technological infrastructure is crucial for making informed, data-driven decisions, executing strategies efficiently, and collecting vital insights for continuous improvement.

Because of the above factors, "digital" and "data" have become buzzwords in contemporary discussions. While analogies likening data to 'new oil' or 'new gold' are prevalent, there are still technical and business complexities in unlocking its full value which present significant opportunities. These challenges in data-driven transformation offer fertile ground for venture capitalists to invest in solutions that address these issues.

Source: WSC Analysis (2023)

Economic boom driven by Data and AI

Convergence of rapidly maturing technologies, historically high returns on innovation, and the substantial economic potential of generative AI is reshaping the data landscape

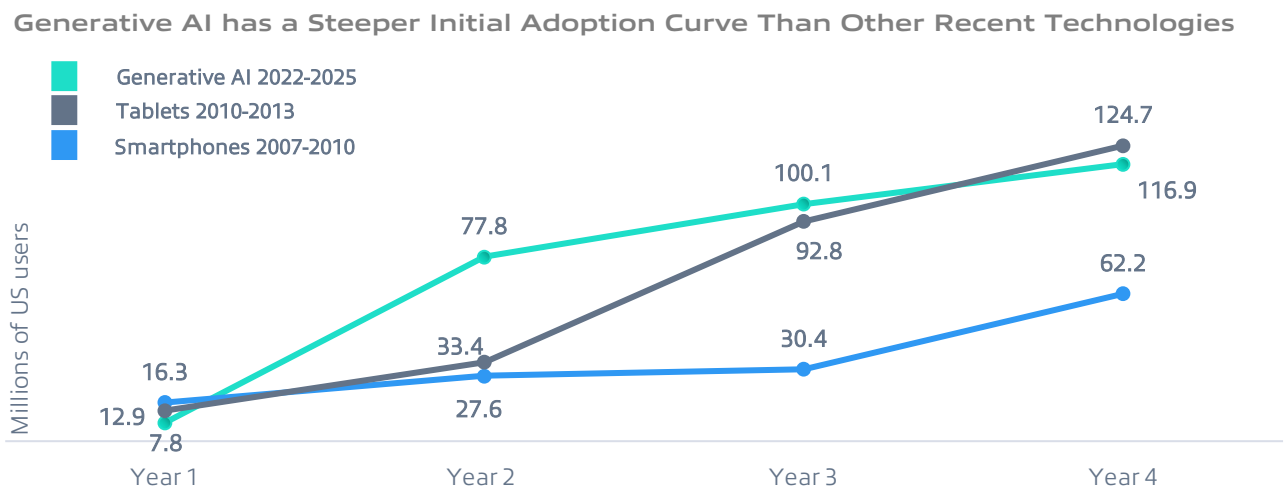
Why now?

Several factors are creating a favourable environment for technology and data-centric businesses. First, there's a rapid maturation of technologies and the increased availability of skilled data and engineering talent. Second, historical data from past recessions and economic disruptions indicate that companies that continuously invested in innovation during these times achieved 240% higher shareholder returns than their counterparts[1]. Third, the potential economic impact of the generative AI, as an emerging technology, is substantial, with the possibility of **adding $2.6tn to $4.4tn annually across 63 use cases and a total potential impact of $13.6tn to $22.1tn** as analysed by McKinsey[1]. These developments are not only exciting but also highlight the critical role of data as the driving force behind this new era of growth.

Impact of Generative AI

While traditional tools that comprise the modern data stack are unlikely to become obsolete, the rapid rise of new technologies in AI is creating a need for developers and engineers to upskill and re-tool. Data teams and vendors must adapt to this shift, where AI applications with compute and data processing needs that were once unthinkable are now a new standard across a plethora of industries, and as such incumbent infrastructures need to be scaled to compete in this new environment.

Data leaders and management teams are now recognising AI's immense capabilities, its potential applications, and the ripple effects it can create. As a result, there's an uptick in investment to develop AI capabilities while managing associated risks.



Source: Insider Intelligence

What does this mean for WSC?

We are at the very beginning of a new data-driven transformation, driven by macro and micro tailwinds, which should see the development of entirely new industries, jobs, and businesses. This provides us an exciting opportunity to invest in multiple avenues related to data and AI that leaves us excited for the future. At White Star Capital, we believe we can play here, without having to redesign our investment strategy, by identifying the sectors and businesses that match our investment criteria.

Source: [1] McKinsey, WSC Analysis (2023)

10-year horizon predictions for Data-driven Transformation

As we look to invest in the top startups from across the globe, we believe the following key trends will result in a growing number of opportunities within the Data-driven Transformation theme

Chief Data Officer roles will be prevalent

Chief Data Officers (CDOs) and their data teams will operate as a distinct business division with accountability for profits and losses.

This division will collaborate with other business groups to set governance and process guidelines, brainstorm innovative data usage methods, create comprehensive strategies to manage company data that integrates with the overall business strategy, and generate new revenue streams by capitalising on data services, data sharing, and data exchange.

This will be driven by (i) prevalent data literacy within unit leaders and internal teams to enable collaboration with CDOs, (ii) recruitment of top data experts and talent that will focus on innovation and creative data use cases, and (iii) continuous experimentation to extract value from data.

Data-ecosystem collaboration will become necessary

Large, intricate organisations will increasingly adopt data-sharing platforms to foster collaboration on data-centric projects, both internally and with other entities.

In the data economy, companies will actively participate in pooling data, yielding richer insights for all involved. Data teams will also work together to establish policies and governance for standardisation, consistencies, and continuous improvement across organisations; akin to collaboration seen in open banking.

Data marketplaces will facilitate data exchange, sharing, and enhancement, enabling businesses to create unique data products and extract valuable insights. This trend, already evident with open-source models, is expected to grow with the demands of AI model training.

Cybersecurity: table stakes for businesses of all sizes

Cybersecurity will become a necessity across all levels and sizes of business due to the escalating frequency and severity of cyberattacks, which are already causing severe financial damage. Small and medium businesses, midmarket enterprises, and large organisations are all targets due to less sophisticated security measures in one or more areas.

Shortage of cybersecurity talent and increasing regulatory pressures to safeguard data will make cybersecurity solutions and service providers extremely vital.

Commoditisation of AI

As AI and ML commoditise, paralleling payment processing's evolution, the future will favor those with proprietary cutting-edge models, datasets, established brand and distribution, and those providing innovative infrastructure and tools for developers. This shift diminishes the strategic significance of AI itself, placing greater emphasis on modular, best-in-class tools that enable efficient and customised management of AI workflows.

Winners will be those who invest in creating flexible, robust infrastructure, catering to the specific needs of diverse organisations, thereby providing a competitive advantage in an increasingly AI-driven tech stack and world.

Source: WSC Analysis (2023)

Succeeding against incumbents

Emerging companies with data and/or AI offerings must leverage all their advantages to successfully compete with incumbents

Incumbents	Emerging Data + AI Companies
Vast troves of data	Innovative and creative; not limited to legacy tech stack
Significant capital	Attracts top-tier tech talent
Established infrastructure and resources	Ability to move and adapt quickly
Costly and slow to adapt to regulations	Can integrate regulatory requirements by design
Potentially slow to adapt to major trends	Requires substantial funding

Key traits required to succeed against incumbents

- **Repeatable enterprise sales**
  This age-old formula will continue to be prevalent and important
- **Deploying product led growth**
  This is a key GTM for businesses in the post-modern data stack era and tooling companies
- **Solving problems in proven categories**
  Data categories will continue to converge into one or get bundled, so emerging companies must stay ahead of this to remain differentiation

In thriving markets, companies with a singular focus can prosper. However, past market downturns have shown that businesses offering a suite of interconnected enterprise services often prevail. This trend is likely to persist.

Therefore, new companies should either aim to develop such integrated solutions or ensure behavioral shift in consumers or enterprises. This approach is essential for them to effectively challenge incumbents and stand out.

WHITE STAR CAPITAL

2

Data-driven Transformation Ecosystem

Ecosystem Highlights

$129bn

Raised by Data-driven Tech Seed – Series B startups in the last 3 years

$38bn

Raised by Data-driven Tech startups in 2023YTD[2]

1,857

Mega rounds[1], out of which 281 in 2022 and 141 in 2023YTD[2]

20.4%

Increase in number of exits from 2020-2022

$20.1bn

In seed funding from 2016-2023YTD[2]

333

VC-backed Data-driven Tech unicorns[2]

1,076

Data-driven Tech exits in 2022

4.0x

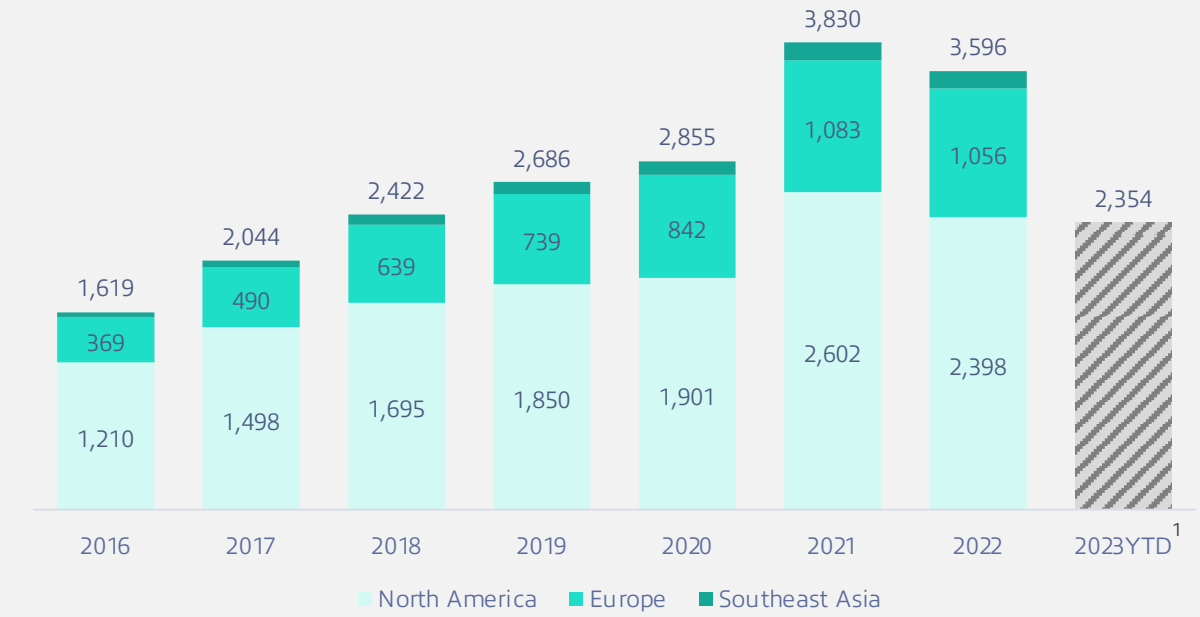
Increase in total capital invested from 2016 to 2022

Source: Pitchbook (2023)
Notes: 1) Rounds >$100m, 2) As of November 2023, inclusive of OpenAI and Anthropic 2023 rounds

Deal activity has seen strong growth over the last couple years

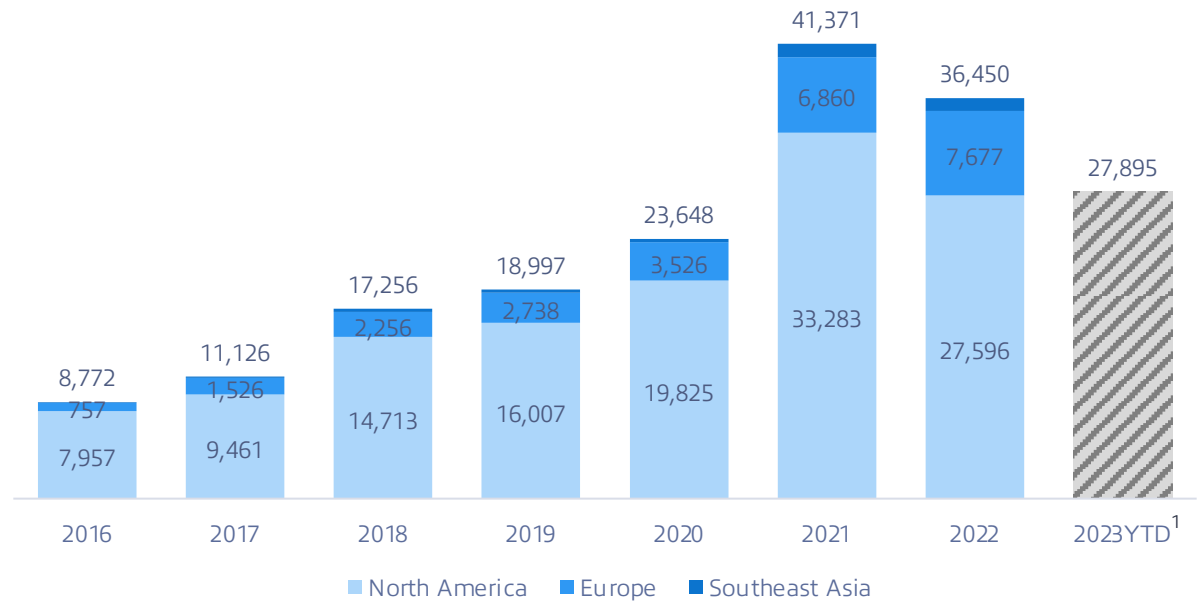
Deal volumes grow 20% on average in Europe and SEA, while North America sees steady 10% growth YoY

Data-driven Tech Deal Count across Seed, Series A and Series B by Region



While deal values have grown 2x on average from 2019-2022, with 2021 being an exception

Total Capital Invested in Data-driven Tech across Seed, Series A and Series B by Region ($m)



Source: Pitchbook (2023)
Notes: 1) As of November 2023

Most of the active funds investing in data-driven tech are based in North America

Part 1

Top generalist VC funds in order of assets under management

Investor	Founded	HQ	AUM [2]	Stage	Deal Count [1]	Notable Deals
Post-Modern Data Stack						
a16z	2009		$35.0bn	Seed / Series A+ / Growth	31	databricks, Tabular, Starburst
NEA	1977		$25.0bn	Seed / Series A+ / Growth	34	databricks, CATALOG, DATARIUM
Accel	1983		$19.1bn	Seed / Series A+ / Growth	39	AIRBYTE, Lightdash, MONTE CARLO
Data Management						
INSIGHT PARTNERS	1995		$90.0bn	Seed / Series A+ / Growth	27	TONIC, Acryl Data, honeycomb.io
ALUMNI VENTURES GROUP	2014		$1.0bn	Pre-Seed to Growth	46	OCIENT, ApertureData, data.world
GREYCROFT	2006		$5.7bn	Early to Growth Stage	24	Amperity, ClearGraph, MMUTA
Data Security						
Lightspeed	2000		$25.0bn	Early to Growth	49	ARCTIC WOLF, confluera, noname
Bessemer Venture Partners	1911		$20.0bn	Early to Growth	48	virtru, AXONIUS, BigID
Index Ventures	1996		$4.0bn	Seed / Series A+	51	castle, expel, persona

[1] Represents number of deals in Data-driven Transformation sectors since 2016, [2] As of November 2023
Source: WSC Analysis (2023)

Most of the active funds investing in data-driven tech are based in North America

Part 2

Top generalist VC funds in order of assets under management

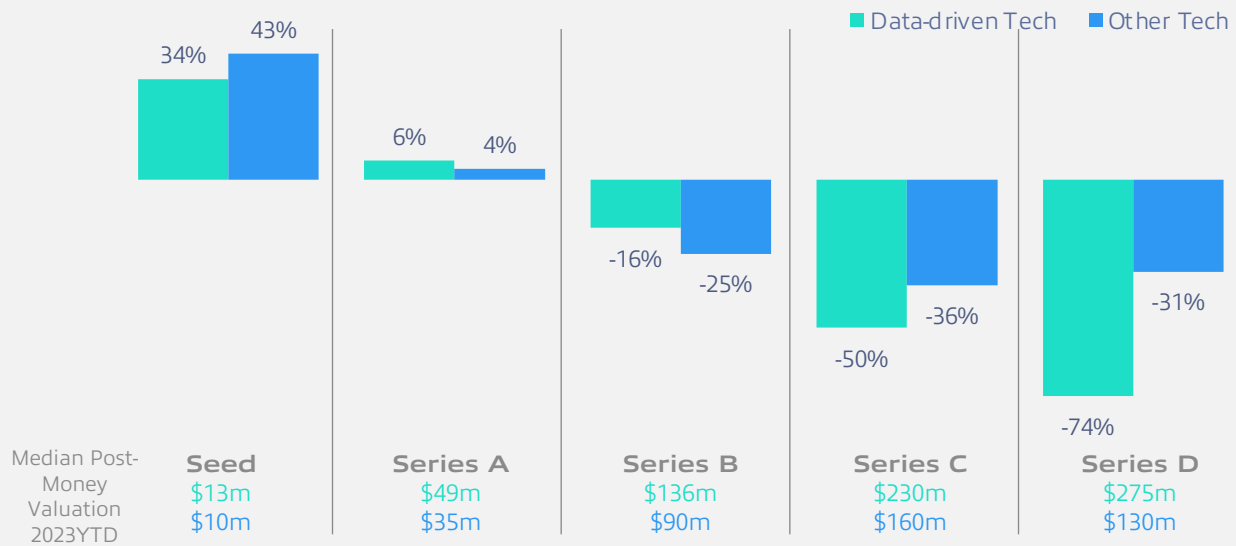
Investor	Founded	HQ	AUM [2]	Stage	Deal Count [1]	Notable Deals
AI & Analytics Tools						
SEQUOIA	1972		$85.0bn	Seed / Series A+ / Growth	86	Amplitude, bigpanda, Hugging Face
a16z	2009		$35.0bn	Seed / Series A+ / Growth	100	anyscale, COACTIVE, = Equals
GV	2009		$8.0bn	Seed / Series A+	39	Mighty Ai, neo4j, Snorkel
Productivity & Developer Tools						
KLEINER PERKINS	1972		$20.0bn	Seed / Series A+ / Growth	10	Figma, netlify, productboard
Accel	1983		$19.1bn	Seed / Series A+ / Growth	15	ATLASSIAN, slack, webflow
Index Ventures	1996		$4.0bn	Seed / Series A+	9	Notion, evervault, elastic
salesforce ventures	2009		$5.0n	Seed / Series A+	20	monday.com, MuleSoft, Airtable
true Ventures	2005		$3.8bn	Seed / Series A	15	syncplicity, tray.io, iterative

[1] Represents number of deals in Data-driven Transformation sectors since 2016, [2] As of November 2023
Source: WSC Analysis (2023)

Data-driven sector has experienced significant valuation decline in the last two years

Since 2021 data-drive companies have seen larger corrections in valuation from Series B onward, influenced by rising interest rates and global uncertainties (COVID-19, Russia-Ukraine conflict). This has led to a market that values profitability over growth, resulting in lower exit multiples and repricing of valuations

% Change in Median Post-Money Valuation by Series Globally 2021 – 2023YTD[1]



Sub-sectors seeing varying valuation impacts from 2022-2023YTD[1]

Median Post-Money Valuation per Sub-Sector

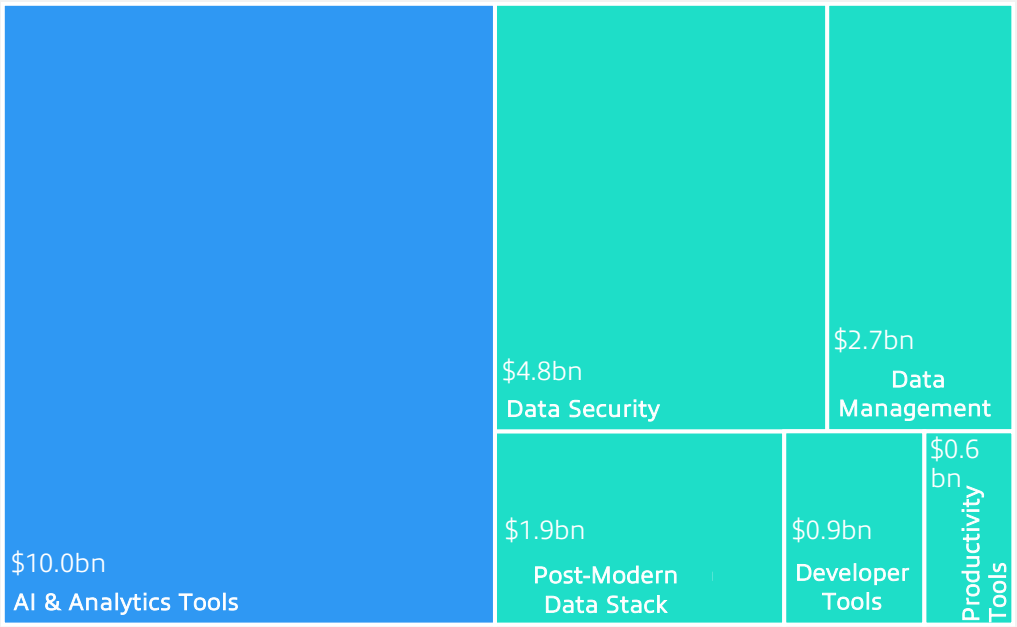
Sub-sector	2022	2023YTD[1]	% Change
Post-Moden Data Stack	40.1	32.0	*-20.0%*
Data Management	24.9	29.5	*+18.9%*
Data Security	38.0	29.8	*-21.6%*
AI & Analytics Tools	25.0	28.0	*+12.0%*
Productivity Tools	30.4	27.9	*-8.3%*
Developer Tools	15.6	20.4	*+31.2%*

Source: Pitchbook (2023)
Notes: 1) As of November 2023

AI & Analytics Tools led the way for funding while Data Security has seen the most exits

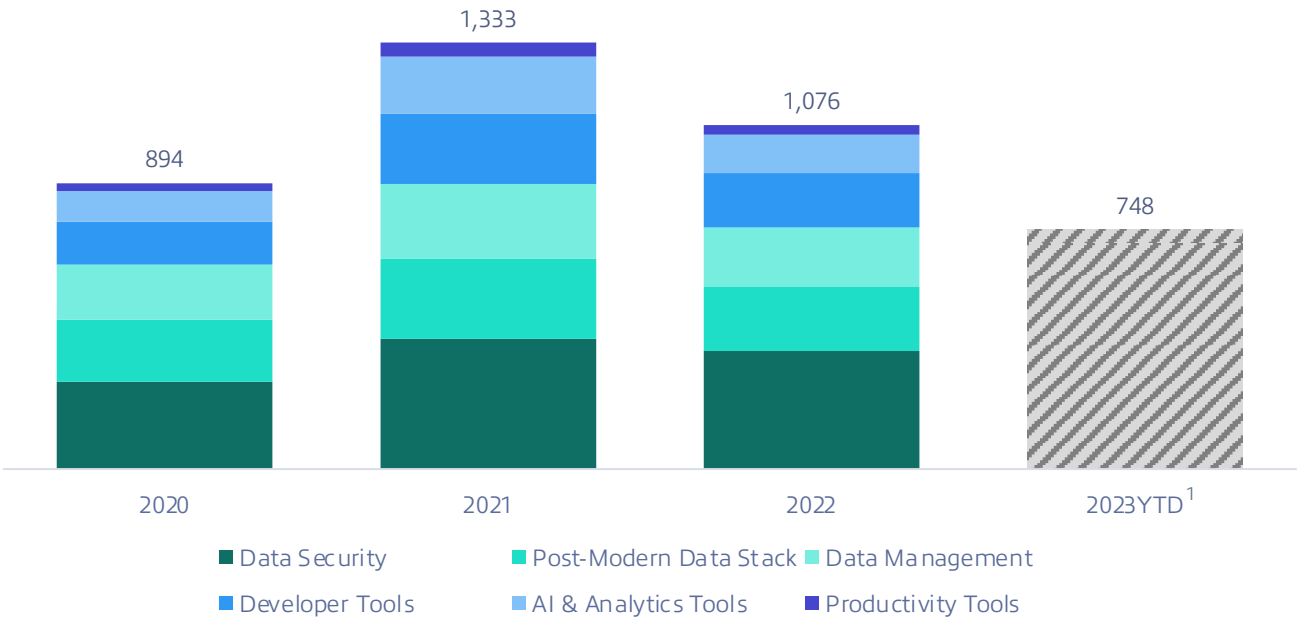
Startups in the AI and Analytics Tools sector attracted the most funding as of 2023YTD[1]

2023YTD[1] Funding per Subsector



Data Security, Post-Modern Data Stack & Data Management led the count of exits from 2020-2023YTD[1]

Count of Data-driven Tech exits across M&A and IPO since 2020 by sub-sector



Source: Pitchbook (2023), funding per sector not inclusive of OpenAI and Anthropic 2023 rounds
Notes: 1) As of November 2023

The exit environment is strong, which is key to the long-term viability of the ecosystem

Exit activity in the Data-driven Transformation sectors saw a 30% increase from 2020 to 2022

93%

Of exits from 2020 to 2023YTD[1] were M&A transactions

45%

Of exits since 2020 came from 2022 to 2023YTD[1] alone, signaling a robust exit environment

4,382

Data-driven Tech companies have exited via M&A, SPAC, and IPO since 2020

$650bn

In total Exit Value from 2020 to 2023YTD[1] disclosed through IPOs and acquisitions

Large Data-driven Tech exits across key verticals since 2020

IPO 2020 — Palantir
(NYS: PLTR) Share price of $9.5 valued at $20.6bn

M&A 2021 — slack
Acquired by Salesforce (NYS: CRM) for $27.1bn

M&A 2021 — proofpoint.
Acquired by Thoma Bravo for $12.3bn

M&A 2020 — interxion
Acquired by Digital Realty (NYS: DLR) for $7.0bn

M&A 2021 — maxim integrated
Acquired by Analog Services (NAS: ADI) for $27.1bn

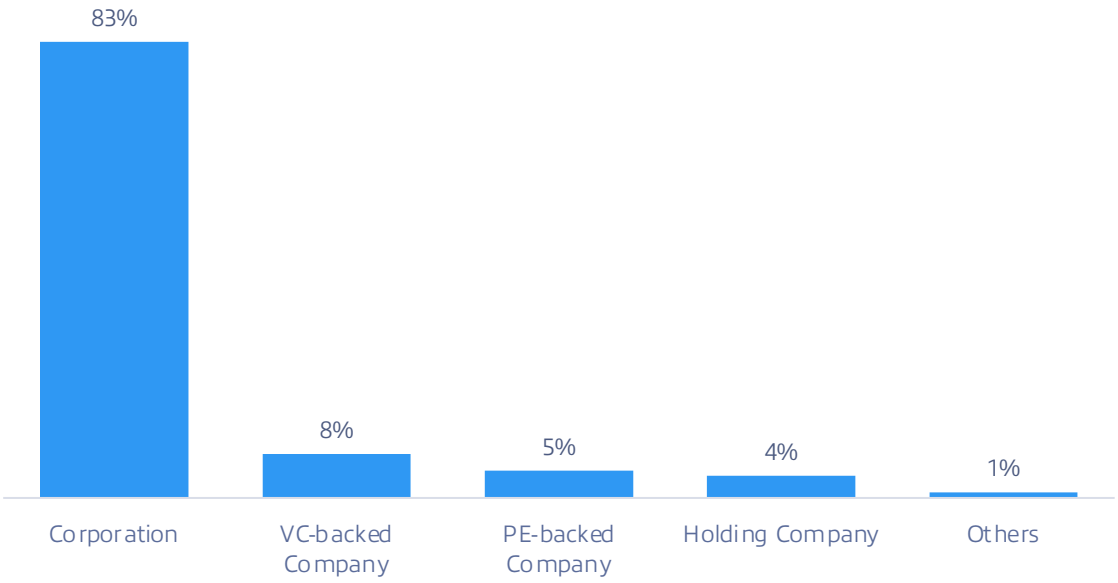
IPO 2022 — snowflake
(NYS: SNOW) Share price $120, valued at $33.2bn

Source: Pitchbook (2023)
Notes: 1) As of November 2023

Corporates and large tech-giants have historically been active acquirers

Corporates were the most active acquirers of Data-driven Tech companies

Top 100 acquirers (acquisition count) by acquirer type from 2020 to 2023YTD[1]



The most active acquirers are larger tech-giants

Most active acquirers of Data-driven Tech companies by count since 2020

Source: Pitchbook (2023)

Select top data-driven unicorns

OpenAI

Research organisation that focuses on creating and advancing AI technologies; known for its creation of ChatGPT

$85bn Valuation

databricks

Data intelligence platform intended to offer an open and unified platform for data and artificial intelligence sectors

$43bn Valuation

ANTHROP\C

AI safety & research company intended to build reliable, interpretable and steerable large-scale artificial intelligence systems

$25bn Valuation

COHESITY

Web-scale platform designed to simplify the way companies protect, manage, and extract value from their data

$10bn Valuation

LACEWORK

Unified cloud security platform designed to automate cloud security at scale

$8.3bn Valuation

Source: WSC Analysis (2023)

WHITE STAR CAPITAL

3

Our Approach to Investing in Data-driven Transformation

Conditions are right for new data-driven transformative companies, funded by VC

The market and macro conditions today are ripe for outsized data investing returns and impact

1- Data leads to revenue generation: Companies are increasingly using data as a fundamental part of their operations. Instead of relying on extensive, multiyear road maps to tackle issues, they're now equipped to consider how cutting-edge data methods can solve problems in a matter of hours, days, or weeks – often opening new avenues for generating revenue

2 - AI is changing the tech stack: AI is reshaping tech stacks, necessitating an understanding of each layer's dynamics and their differences from previous stacks. There are opportunities for innovation ranging from the foundational data infrastructure to effective management and governance to make AI widely accessible.

3 - Talent interest: Talented and creative technical professionals are either drawn to work on groundbreaking data and AI projects or motivated to start their own businesses in this field. This drive often stems from the challenges they encountered while part of internal tech teams in established companies which have been held back by their legacy data architecture and stack.

4 - Increased regulations and adoption of data security: As digital enterprises mature, they are integrating more cybersecurity strategies to address heightened risks from identity management and data privacy, to mitigate escalating costs and cyberattack threat. Increased focus on data sovereignty and the usage of AI across different regions is likely to drive fragmentation.

5 - Lower Barriers to Entry: Digitally mature businesses with MLOps teams and BI investments have been the fastest to adopt and implement AI in their toolkits. Start-ups who are developing tools to streamline development and make integration easier are poised to have first mover advantage in this space.

In keeping with WSC's investment guidelines, it is still possible to invest in Data-driven Transformation if companies can:

Offer solutions that converge across data and security categories

Innovate on business models with repeatable enterprise sales

Create solutions for companies of varying maturity

Attract top-tier technical talent with business acumen

Solve problems in proven categories and avoid the bundling trap (absorption by tech giants)

Source: WSC Analysis (2023)

We've created a WSC VC Investability Index to assess sectors and subsectors of interest

The purpose of this report will be to help shed light on this burgeoning ecosystem, looking at every single industry and identifying opportunities for investment as a generalist venture capital fund

Scoring: Yes = 1 / No = 0

Market Size
Does the potential future market size (up to 2028-2032) at scale correspond to our criteria ($5bn+)?

Market Readiness
Are we seeing pull from the market and C-suite adaptation of these technologies?

Innovation
Are these businesses engaged in a meaningful innovation to stay competitive?

Barriers to Entry
Are there barriers to entry? Will regulation make it difficult to scale?

International Scalability
Is this an internationally scalable model?

Qualified Talent
Are product and engineering talent willing to work within this sector?

Recurring Business Model
Is this a recurring business model? Is there potential for high revenue retention?

Technology Readiness
Is the technology mature and stable? Is there a robust development process in place?

Margins
Are there software-like margins and potential for margin improvement at scale?

Stage
Do these businesses have strong growth trajectory?

Exit Environment
Have we seen large exits in this sector ($1bn+)?

Legend

Out of Scope	Work to be done	Within WSC Scope
Score 0-3	Score 4-6	Score 7-9

Source: WSC Analysis (2023). Scoring is done based on the criteria above on the back-end by WSC team

Data-driven Transformation is broken down into 5 sectors and 34 subsectors

WSC will explore the following subsectors for potential investment opportunities

Post-Modern Data Stack

- DataOps
- Data Streaming
- Data Mesh Enablers
- Data Reliability

Data Management

- Active Metadata Management
- Data Marketplaces
- Synthetic Data Generation
- Vector Databases
- Data Curation Tools
- Data Catalogs

Data Security

- Identity & Access Management
- Threat & Vulnerability Management
- Cloud & Infrastructure Security
- Governance, Risk & Compliance

AI & Analytics Tools

- Predictive Analytics
- Visualisation Dashboarding
- Generative AI UI/UX
- AL/ML-Enhanced DevTools
- AIOps

Productivity and Developer Tools

- Software Delivery Lifecycle
  - Project Management Tools
  - Collaboration Tools
- Developer SaaS
  - Design Tools
  - Low-code/No-code Platforms
- Developer Infrastructure
  - Containerisation & Orchestration Tools
  - Configuration Management Tools
  - Package Managers
  - Continuous Integration/Continuous Deployment (CI/CD) Tools
  - Testing Frameworks
- Build & Development Tools
  - Cloud Services/Platforms
  - Integrated Development Environments (IDEs)
  - Source Control/Version Control Systems
  - Build Tools
  - Code Review Tools
  - Static Analysis/Code Quality Tools

A thematic approach to investing

WSC will focus on these four key sectors undergoing rapid structural change

Data infrastructure and management
Data volume and velocity will reach 465 exabytes/day by 2025, requiring storage and management for unstructured and structured data created

Data protection and governance
Data driving real-time business decisions creates a need for security and compliance tools which overcome risks of cybercrime, threats, and leaks

Productivity and Developer tools
Digital transformation requires tools to streamline application development, enhancing agility and innovation in a tech-driven economy

DevOps and MLOps tools
As AI models grow in complexity, DevOps and MLOps tools are essential to seamlessly develop, deploy, and monitor within intricate tech environments

Using the WSC VC Investability Index, we have identified the following subsectors as the most investable for generalist VCs and thus WSC

	Data Infrastructure & Management	Data Protection & Governance	Productivity & Developer Tools	DevOps & MLOps Tools
Post Modern Data Stack				
Data Reliability	✓			
DataOps	✓			
Data Streaming	✓			
Data Mesh Enablers	✓			
Data Management				
Active Metadata Management	✓			
Data Marketplaces	✓	✓		
Synthetic Data Generation	✓		✓	
Vector Databases	✓			
Data Curation Tools	✓	✓		
Data Catalogs	✓			
Data Security				
Identity & Access Management	✓	✓		
Governance, Risk & Compliance	✓	✓	✓	✓
Threat & Vulnerability Management	✓	✓		
Cloud & Infrastructure Security	✓	✓		
AI & Analytics Tools				
Predictive Analytics			✓	
Visualisation Dashboarding			✓	
Generative AI UI/UX			✓	✓
AL/ML-Enhanced DevTools			✓	
AIOps				✓
Productivity & Developer Tools				
Software Delivery Lifecycle				
Project Management Tools	✓		✓	
Collaboration Tools			✓	
Developer SaaS				
Design Tools			✓	
Low-code/No-code Platforms	✓		✓	✓
Developer Infrastructure				
Containerisation & Orchestration Tools			✓	
Configuration Management Tools			✓	
Package Managers			✓	
Continuous Integration/Deployment Tools			✓	
Testing Frameworks			✓	
Build & Development Tools				
Cloud Services/Platforms	✓		✓	
Integrated Development Environments			✓	
Source Control/Version Control Systems			✓	
Build Tools			✓	
Code Review Tools			✓	
Static Analysis/Code Quality Tools			✓	✓

Source: WSC Analysis (2023)

4

Sector Focus

Post-Modern Data Stack

Post-Modern Data Stack

Signifies a pivotal shift from the resource-intensive modern data stack to an architecture that champions intelligent automation, streamlined processes, and a focus on value-driven data products

$1tn

Expected global data infrastructure market size by 2032[1]

95%

Of businesses struggle with unstructured data[2]

120 zettabytes[3]

Amount of data created in 2023[4]

$29.7bn

Invested in the sector at Seed-Series D stages from 2020-2023YTD[5]

As data leaders scrutinise tools and infrastructure frameworks, previously renowned in the modern data stack, for cost-effectiveness, return on investment, and implementation time, the post-modern data stack is emerging as a solution. It responds to the modern demands of handling vast quantities of high-velocity data and the need to democratize data access for complex use cases in the face of tightening IT budgets, evolving enterprise needs, and AI-centric tech stack.

By reducing the operational overhead associated with the modern data stack, this new approach allows teams to future proof their business to allow AI integration and develop strategic solutions to unlock the full potential of their data. Key features of this evolution involves reworking the data infrastructure for faster data delivery, real-time insights, decentralised architectures, and convergence of previous tools for enhanced data quality.

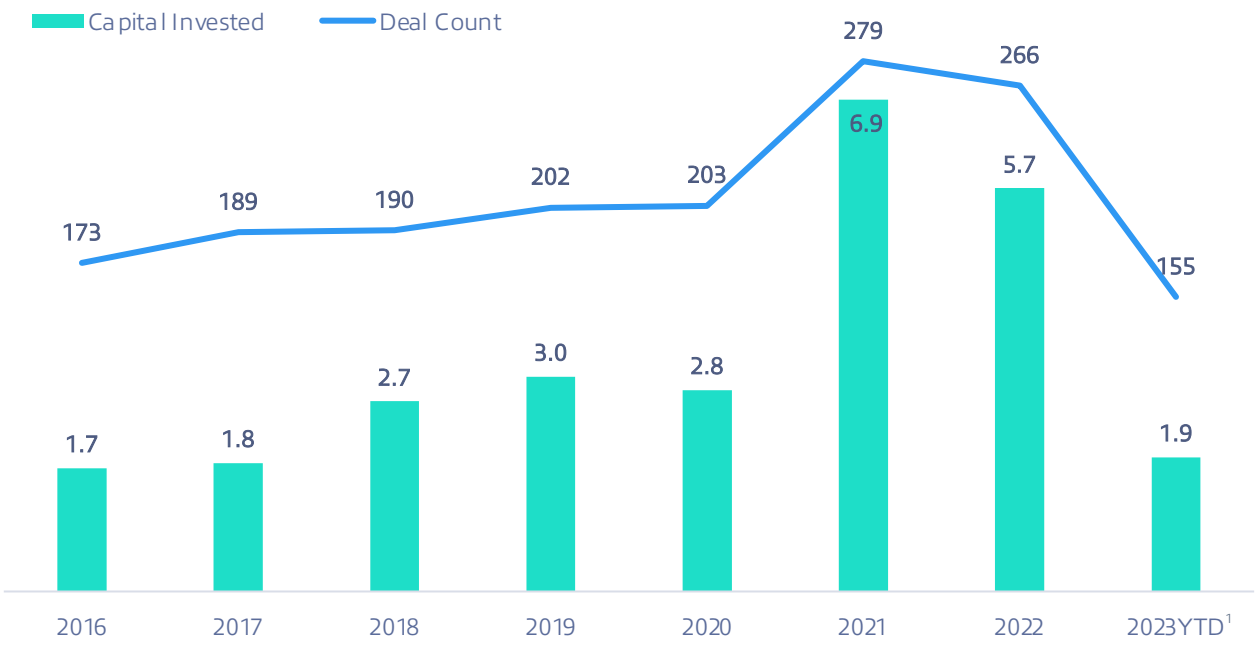
Technology	WSC Score	Notes
Data Reliability	9	Convergence of categories
DataOps	8	Agile data architecture
Data Streaming	8	Real-time data delivery
Data Mesh Enablers	8	Treating data as a product

[1] Custom Market Insight, [2] Fingent, [4] Statista, [5] As of November 2023; Pitchbook, WSC Analysis (2023)
[3] One zettabyte conceptually refers to storing 250 billion DVDs or streaming 36 million years of HD videos

Market drivers

Postmodern data stack era is characterised by a shift from a proliferation of isolated SaaS tools to treat data as a dynamic product, streamlining their architectures to improve data integration, quality, and delivery through agile, cloud-based systems and flexible practices

Capital Invested and Deal Count in Post-Modern Data Stack from 2016-2023YTD[1] ($bn)



Efficiency and Agile Integration

Modern organisations, faced with vast data volumes and rapid data generation, are adopting agile, scalable, cloud-first, microservices-driven systems instead of rigid, monolithic architectures. This is being done to enable better integration and management of diverse data sources. Traditional ETL processes are being replaced by more streamlined workflows that require less manual intervention, enhancing efficiency in data warehouses.

This transition is further propelled by the proliferation of SaaS tools, which, while offering tailored solutions, also lead to tool overload and data fragmentation. The post-modern data stack addresses these challenges by converging data categories and bundling tools, aiming to improve cost containment, data quality, and delivery.

Shift towards Treating Data as a Product

Currently, data is often redundantly stored in expansive, compartmentalised systems, posing challenges for internal users, like data scientists, who need to efficiently locate, retrieve, and amalgamate data for analytical purposes.

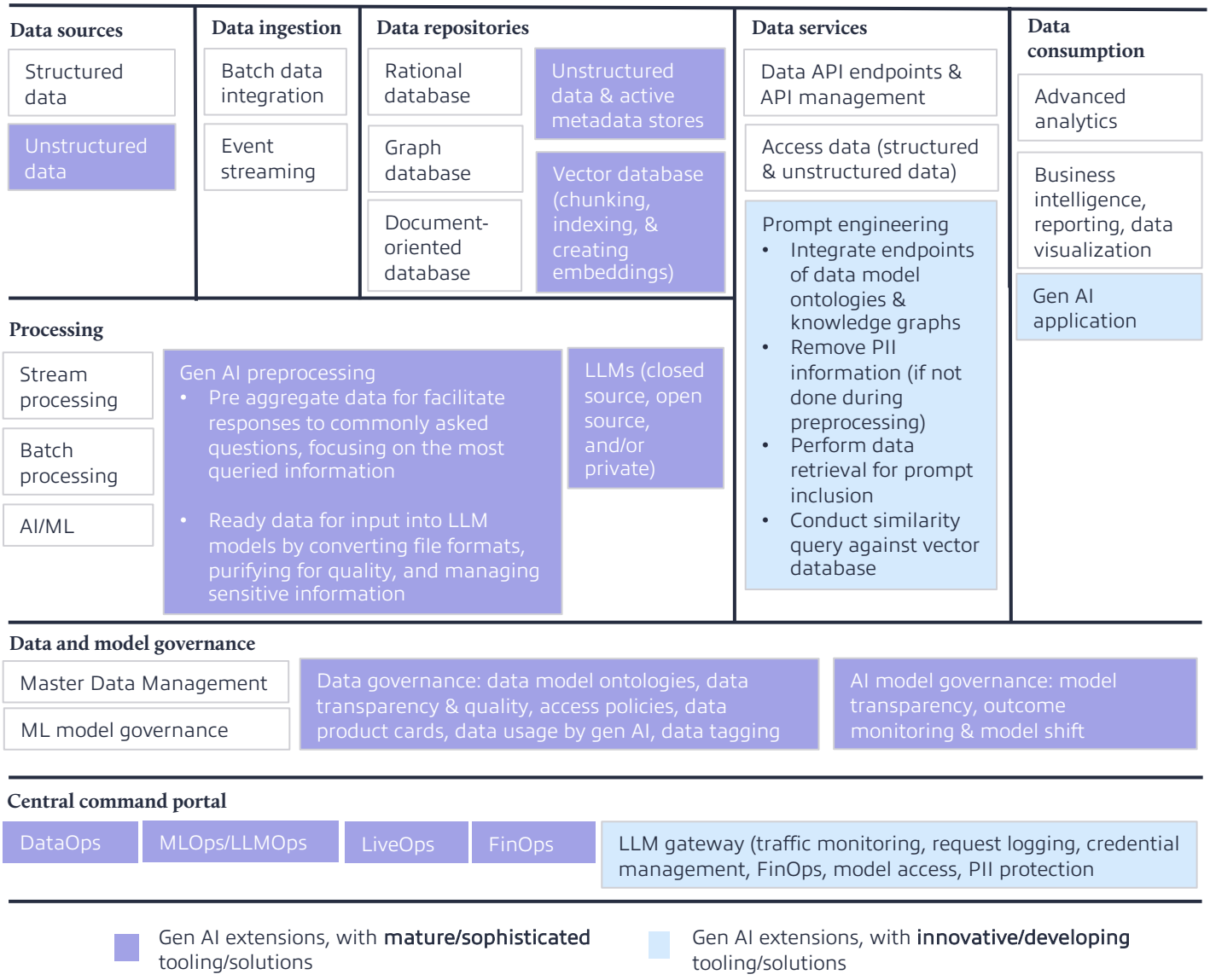
However, there is a shift where data assets are now managed as products with dedicated teams ensuring their enhancement, serving both internal and external users. These teams work on transforming and updating data, as well as deploying self-service and analytical tools, ensuring the data products adapt flexibly to user needs through DataOps and ongoing integration and delivery practices. This will be necessary to leverage generative AI and extract high value uses-cases using the technology.

[1] As of November 2023, Source: Pitchbook

Necessary upgrades to existing data infrastructure to enable Generative AI

The post-modern data stack emphasises the need to overhaul the existing data architecture to facilitate high-value use cases

Potential Revamped Data Architecture



The above upgrades, outlined as sophisticated or developing tooling, to existing data stacks underscore potential opportunities to invest in from a White Star Capital / VC investor standpoint

Source: WSC Analysis, McKinsey

Technologies & subsectors

Part 1

Data Reliability

WSC Score: 9

While data observability and quality have been around for a long time, there is now a convergence of these subsectors into a larger "data reliability" category. This is centered around enabling companies with data hygiene to provide high-quality data and simplification of observability processes.

Tailwinds

+ Companies in the space continue to redefine themselves and attract funding, creating strong attention for the category
+ Significant open-source growth with existing data observability libraries

Challenges

+ Challenges remain in simplifying observability processes while maintaining high data quality standards

DataOps

WSC Score: 8

Agile approach to designing, implementing, and maintaining a data architecture and analytics infrastructure. DataOps tools aim to reduce the cycle time of data analytics, from the gathering of data to deriving insights, while ensuring data integrity.

Tailwinds

+ Market size to reach $14.6bn by 2030 (21.2% CAGR from 2023)
+ Broad understanding and adoption across organisations

Challenges

+ Building a data-first mindset amongst teams
+ Necessitates data science, data engineering, software development, and operations experts
+ Wide range of technology manufacturers, service providers, and consulting companies that offer solutions

Data Streaming

WSC Score: 8

Need for real-time data and operational analytics is driving the growth of data streaming technologies, which are becoming essential for enterprises aiming to transition from data-driven to data-led decision-making. This shift emphasises queries that demand to know "What is happening now?"

Tailwinds

+ Market size to reach $50.1bn in 2026 (26.5% CAGR from 2021)
+ Useful for applications that require immediate action based on live data, such as financial trading platforms, live monitoring systems, etc.
+ High ROI with investment

Challenges

+ Can contribute to data isolation without a common model
+ Fragmented projects and uncoordinated teams/budgets
+ Lack of standalone private company winners as of now

Source: WSC Analysis (2023), Yahoo Finance, MarketsandMarkets

Technologies & subsectors

Part 2

Data Mesh Enablers

WSC Score: 8

Decentralised approach to treat data as a product, allowing organizations domain-specific control over data, which is a shift from monolithic and centralised data lake or warehouse architecture. Technologies enabling this would involve analytics, integration, observability functionalities to allow for scalable and agile decentralised data as a service.

Tailwinds

+ Part of the expected $122.9bn advanced data management solutions market size by 2025 (9.5% CAGR from 2020)
+ Enables self-service and governed data access and removes data silos for business users

Challenges

+ Organisational structures need to adapt to enable a steady flow of data products

Source: WSC Analysis (2023)

Market map

Data Reliability

unravel, MONTE CARLO, MANTA, Collibra, Datafold, Databand, acceldata, metaplane, talend, SODA, Bigeye, great_expectations, Anomalo, precisely, lightup

DataOps

dbt Labs, databricks, unravel, Datafold, castordoc, Census, Y42, ASTRONOMER, MOZART DATA, OpenLineage, MARQUEZ, hightouch, great_expectations, Nessie, lakeFS

Data Streaming

aws, Google Cloud, CONFLUENT, databricks, ORACLE, SAP HANA, decodable, striim, GIGASPACES, GridGain, meroxa, VOLTRON DATA, ESTUARY, Redpanda, HAZELCAST, MotherDuck, conduktor, aiven, upsolver, ably, tinybird, Stream Native

Data Mesh Enablers

aws, Starburst, ORACLE, talend, k2view, atlan, Informatica, denodo, CONFLUENT, MONTE CARLO, databricks, ESTUARY, nextdata, snaplogic, CLOUDERA

Source: WSC Analysis (2023)

Top asset-light startups

snowflake

Data lake, warehousing, and sharing platform

$5.3bn raised to date*

databricks

Cloud-based, big data analytics platform

$4.2bn raised to date

Collibra

Data intelligence software

$640m raised to date

aiven

Open-source cloud data platform for data pipeline creation

$533m raised to date

dbt Labs

Data transformation and workflow software platform

$415m raised to date

* Indicates public company; included to showcase a category-defining company
Source: WSC Analysis (2023)

Data Management

Data Management

Cornerstone for any company looking to generate revenue from Data and AI. Data management is vital in enhancing data's role throughout the entire data stack and AI lifecycle

706

Number of exits in the sector from 2020-2023YTD[1]

80%+

Of Data Analysts spend around half their time cleaning and organising data[2]

$222bn

Global data management market size by 2030[3]

$22bn+

Total funding for generative AI application and model layers since 2019[4]

Data management has and will continue to be vital as the process of collecting, storing, organising, maintaining, and using data effectively and efficiently. The sector is undergoing changes to prioritise fewer, more unified tools and embrace automation powered by centralised metadata to allow engineering leaders to unlock full potential of their data.

Meanwhile, the AI-centric tech stack is transforming, highlighting the need for a deep understanding of each layer and their interplay. Effective data management becomes pivotal for companies looking to extract maximum value out of their data and those implementing AI. Reliance on high-quality, pre-cured datasets, exploring synthetic data, and leveraging vector databases, among other factors will allow companies to fully harness the capabilities of data and AI.

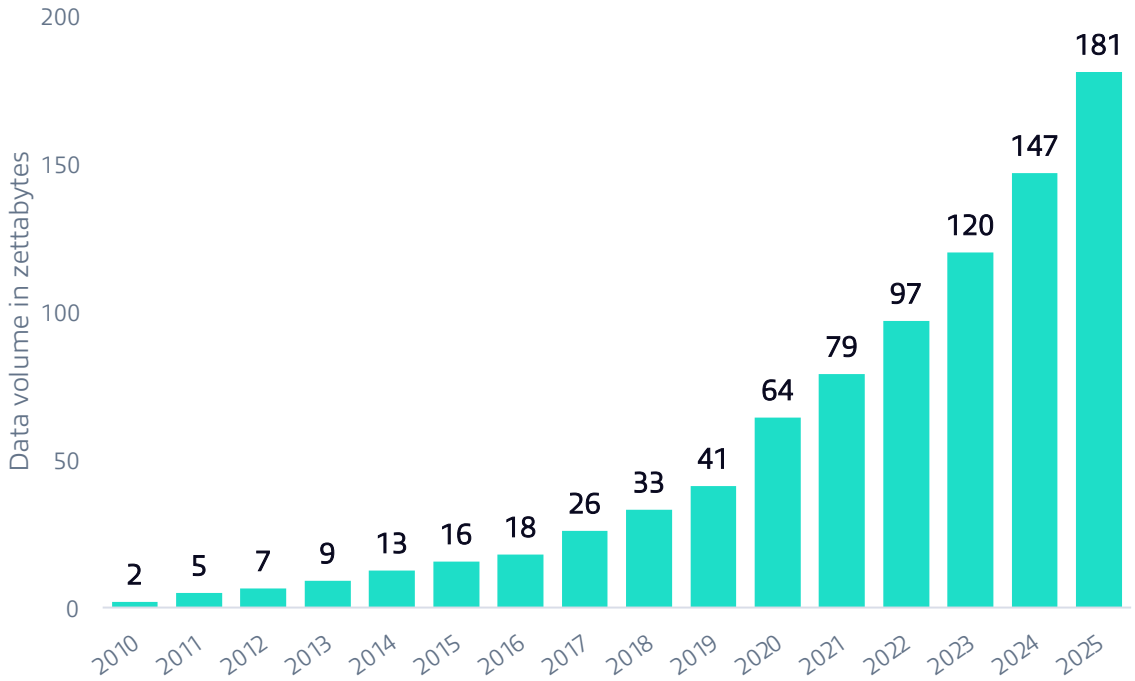
Technology	WSC Score	Notes
Active Metadata Management	10	Key to multiple use cases
Data Marketplaces	8	Will be necessary for ML models
Synthetic Data Generation	8	Effectiveness leading to usage
Vector Databases	8	Suited for GenAI applications
Data Curation Tools	7	Increases quality of ML models
Data Catalogs	6	Replaced by active metadata

[1] As of November 2023; Pitchbook, [2] Forbes, [3] GrandView Research, [4] TheNextWeb

Market drivers

Efficient data management is key to harnessing the deluge of modern data and unlocking competitive advantages, particularly as generative AI enhances the ability to process and analyse diverse, unstructured data forms within an AI-centric data ecosystem

Volume of data created, captured, copied and consumed worldwide (zettabytes)



Source: Statista

Data-driven Decision Making

Data plays a pivotal role in the operation and functionality of businesses. This is especially true for those that are navigating digital transformations and handling immense data volumes. For companies to extract meaningful insights from the myriad of data generated by various systems and technologies, an effective data management model is essential. Without proper management, the abundance of data can become overwhelming and unproductive.

However, with appropriate data management tools, it can unlock revenue generating opportunities and create a competitive advantage.

AI-centric Stack

Data practitioners have started to adopt an array of database types to better organise and manage data. This change is speeding up the process of querying and interpreting unstructured and semi-structured data, leading to rapid advancements in AI-driven capabilities and the discovery of new data relationships.

When volume of data and its potential value are skyrocketing, especially with generative AI's adeptness at handling varied data forms like chats, videos, and code, organizations are compelled to refine their data management strategies. This shift to embrace an AI-centric data stack is essential for maximising value extraction.

Source: McKinsey, WSC Analysis (2023)

Generative AI convergence across the data value chain

Generative AI can enhance business processes with its applications across the data value chain, speeding up and refining tasks. Its effectiveness depends on the management and structure of the underlying data

WSC Area of Interest

1. Data discovery
Performs queries to create interactive data profiles for each major data repository, making them user-friendly and accessible

2. Data ingestion
Develops templates for data ingestion, complete with connection strings and ports, generates/replenishes tokens, and prepares data for deployment, thereby simplifying the system for data engineers

3. Data storage and curation
Applies advanced data compression and archiving algorithms, informed by usage logs, to maximise storage efficiency and minimise costs

4. Data processing and streaming
Identifies the most suitable computing frameworks for processing large data volumes in real-time, tailored to specific use case profiles

5. Data access
Establishes access guidelines by analysing policies and active directory metadata to determine which individuals should or should not have access to specific types of data

6. Data consumption
Connects with business intelligence and reporting tools to enhance data storytelling and facilitate better collaboration within the organisation

7. Data governance
Synchronises with data privacy and security measures, including data masking and encryption, while delivering continuous monitoring and immediate alerts for data breaches and irregular activities

8. Data interpretation
Enhances understanding of data by offering suggestions derived from the analysed insights

There are opportunities to invest in tools and/or products that enable parts of the above value chain from a White Star Capital / VC investor standpoint

Source: McKinsey, WSC Analysis

Technologies & subsectors

Part 1

Active Metadata Management

WSC Score: 10

Dynamic approach to handling metadata, the descriptive information about data that helps in its discovery, understanding, and management, and ensuring it flows effortless across the entire data stack. Unlike passive metadata, active metadata is continuously updated and managed through automated processes to facilitates real-time, bidirectional updates across different tools.

Tailwinds

+ Functions as a layer on top of modern data stack by leveraging open APIs to connect all tools in your data stack and ferry metadata bi-directionally
+ Holds the key to many use cases like observability, cost management, remediation, quality, security, programmatic governance, optimised pipelines

Challenges

+ Existing passive metadata players in the space transitioning to active metadata

Data Marketplaces

WSC Score: 8

Platforms or dynamic hubs where users can buy, sell, exchange or license different types of datasets and data streams. With the rise of generative AI, access to datasets to train models creates tremendous opportunities within this space.

Tailwinds

+ Backbone of foundational generative AI models are datasets
+ Major tech companies with proprietary data cracking down on data access
+ Major tech giants also getting sued for accessing data without adequate licenses

Challenges

+ Potential regulations may create complexities within this subsector and/or ability to licence data assets

Synthetic Data Generation

WSC Score: 8

Synthetic data is artificially created data, generated using models or simulations that can be based on real data, rather than being directly collected from real-life environments. To be usable, this data has statistical properties similar to real data, and synthesis techniques which can be applied to both structured and unstructured data using various algorithms

Tailwinds

+ Mission-critical for many industries, creating prevalent use cases
+ Has become prevalent for ML model training
+ Quality of synthetic data has been deemed reliable, accelerating its usage

Challenges

+ Small market size, expected to reach $6.9bn by 2032 (37.5% CAGR from 2022). May be much larger, however, with ML model training
+ Adhering to privacy laws like GDPR and ensuring high quality of synthetic data does somewhat remain a challenge

Source: WSC Analysis (2023)

Technologies & subsectors

Part 2

Vector Databases

WSC Score: 8

Vector databases, key to interpreting unstructured data and central to search functions, are akin to the "shovels" in the generative AI gold rush. Their widespread use in LLM-powered apps, soon to be integrated into major applications, positions investing in companies providing managed vector databases as a strategic move in this modern gold rush.

Tailwinds

+ Search queries for vector databases having low latency makes it well-suited for generative AI applications
+ Significant VC funding in the space in 2022 and 2023

Challenges

+ Small market size, expected to reach $4.3bn by 2028 (23.2% CAGR from 2023)
+ Uncertainty if capped at around ~$10bn market size or much larger
+ Optimal implementation of vector databases often requires specialised knowledge

Data Curation Tools

WSC Score: 7

In regard to AI/ML model training, these tools are essential for selecting high-quality, relevant data and ensuring efficient labeling to ultimately enhance model accuracy and reduce development time. These tools offer scalability, diversity in data, and advanced features for managing large datasets. Unrelated to AI/ML model training, this category is also part of the large data preparation sector

Tailwinds

+ Market size to reach $12.9bn in 2028 (18.6% CAGR from 2021)
+ Pre-curated datasets can strongly improve model training quality and reduce costs
+ Significant opportunities exist within visual and audio datasets, given they are often incorrectly labelled, broken, missing, or have duplicates

Challenges

+ Data accuracy concerns require curation early in the data lifecycle to eliminate imbalance and errors
+ Security and privacy issues with data leakage and hacking, especially for government or public sector projects

Data Catalogs

WSC Score: 6

Catalogs act as an inventory of data assets, enabling improved data discovery, governance, and usage. Companies are now looking to active metadata management tools to essentially enhance and replace their data catalog tools.

Tailwinds

+ Definite need in enterprises with large unstructured and structure data volumes

Challenges

+ Will get merged into a larger data-related category
+ Existing large players in the space
+ Small market size, expected to reach $8.1bn by 2032 (27.1% CAGR from 2021)

Source: WSC Analysis (2023), Yahoo Finance, Grand View Research

Market map

Active Metadata Management

ORACLE, ataccama, Alation, atlan, Acryl Data, Collibra, Alex, MANTA, cinchy, Informatica, erwin

Data Marketplaces

DAWEX data exchange technology, explorium, narrative, Eagle Alpha, snowflake, nomad data, Datarade, DROP, Google Dataset Search Beta

Synthetic Data Generation

k2view, MOSTLY·AI, hazy, sogeti Part of Capgemini, gretel, datagen, TONIC

Vector Databases

supabase, atlas by NOMIC, Pinecone, Weaviate, zilliz, chroma, planetscale, redis, SingleStore, MongoDB., vespa, drant, kinetica, Epsilla, Milvus, marqo

Data Curation Tools

Informatica, talend, ataccama, alteryx, scale Nucleus, LIGHTLY, atlas by NOMIC, Argilla, scale, Weights & Biases, Cleanlab, mlflow, Carbon, Snorkel

Data Catalogs

Acryl Data, Alation, ataccama, atlan, castor, Collibra, data.world, Demyst, Hitachi Vantara, IMMUTA, Informatica, IBM, metaphor, ObservePoint, OKERA, SailPoint, securiti, SELECT STAR, Skyhigh Security, stemma

Source: WSC Analysis (2023)

Top asset-light startups

Fivetran

Automated data integration platform

$853m raised to date

Alation

Enterprise collaborative data platform

$340m raised to date

IMMUTA

Data management platform designed to automate data governance

$266m raised to date

data.world

Cloud-native data catalog platform

$132m raised to date

STRATIO

Hybrid virtual data management platform

$91m raised to date

Source: WSC Analysis (2023)

Data Security

Data Security

Attack surfaces continue to expand while cybercriminals continue to exploit new vulnerabilities, emphasising the need for cybersecurity solutions more than ever

1,318

Number of exits in the sector from 2020-2023YTD[1]

800,000

Number of cyber attacks per year[2]

$8tn

Estimated global cost of cybercrime in 2023[3]

49 days

Average number of days it takes an organisation to identify a cyberattack[4]

As digital enterprises mature, cybersecurity has become a crucial focus area. Emerging organisations are busy implementing tooling and defining core policies and risk management strategies. However, the proliferation of SaaS solutions in enterprises, though cost-effective and productivity-enhancing, introduces heightened exposure to cybersecurity risks. This trend has reached a critical juncture where, if not properly managed, it can escalate costs and amplify cyberattack threats.

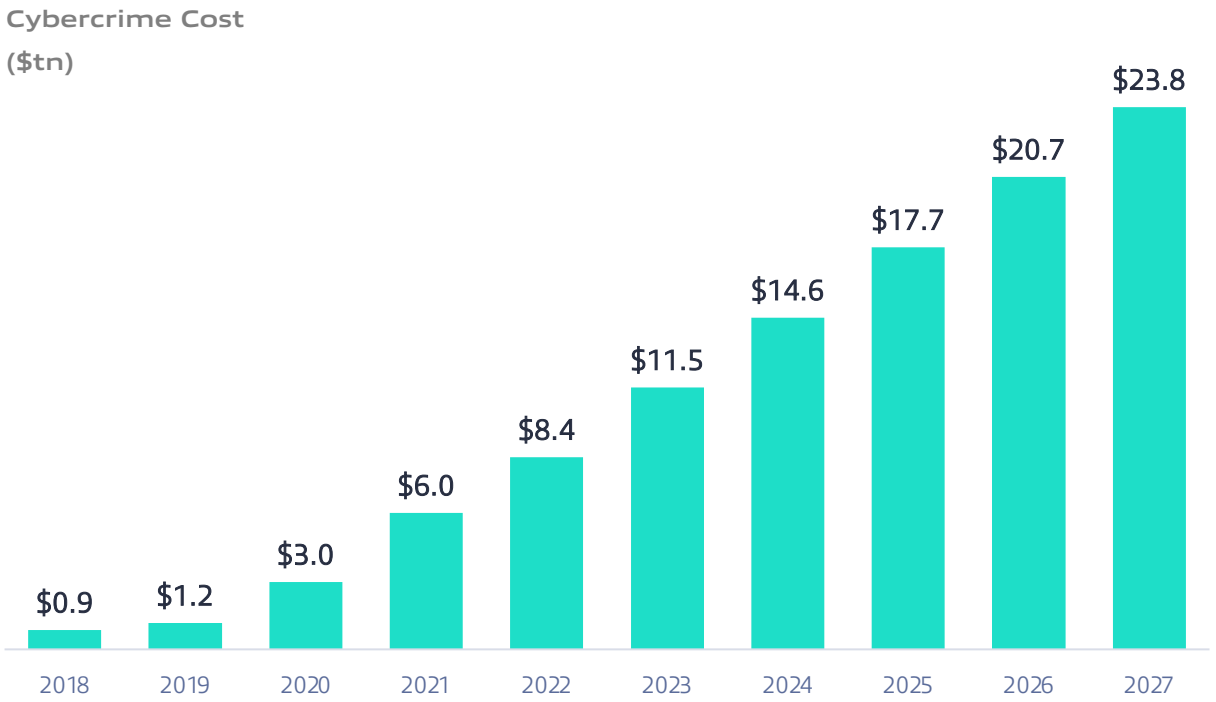
Thus, it's imperative for enterprises to develop comprehensive cybersecurity strategies encompassing all aspects of their digital infrastructure. Key aspects like identity and access management, increased threat velocity, AI/ML applications, and persistent talent shortages in cybersecurity further underscore the urgency for organisations to adapt and strengthen their cybersecurity posture.

Technology	WSC Score	Notes
Identity & Access Management	9	Increasingly important for CISOs
Governance, Risk & Compliance	9	Regulations driving adoption
Threat & Vulnerability Management	8	Proactive approach is key
Cloud & Infrastructure Security	8	Large market size

[1] As of November 2023; Pitchbook, [2] Cobalt Labs, [3] Cybercrime Magazine, [4] Houlihan Lokey

Market drivers

Evolving regulations, increased awareness of data rights, decline in cybersecurity efficacy, and the need for CISOs to adapt strategies to manage the growing challenges in cloud-based and SaaS environments is leading to prioritisation of data privacy and security



Source: Statista

Regulations Leading to Prioritisation of Privacy, Security

Companies now prioritise data privacy, ethics, and security as essential skills, influenced by changing laws like the Virginia Consumer Data Protection Act (VCDPA), General Data Protection Regulation (GDPR), and California Consumer Privacy Act (CCPA). This shift is also due to growing public knowledge of data rights and the significant impact of security breaches.

Erosion of Cybersecurity Efficacy

Additionally, security leaders report a decline in cybersecurity effectiveness, highlighting an expanding technological divide between cyber attackers and defense systems. Only 48% of security leaders indicated their cybersecurity defenses were effective against common security threats. This underscores the continued necessity for innovation in cybersecurity.

CISOs are Becoming Enablers of Innovative and Adaptive Security Strategies

Although many CISOs possess an extensive array of tools for securing data in the cloud, the rapid expansion of cloud providers, SaaS solutions, and emergence of generative AI has pushed data sharing to its limits. Conventional methods, effective in on-premises settings, are struggling to match the surge in users, data sources, and policies that need governance, management, and security in the current landscape.

This situation is compelling CISOs to take proactive measures, implement strategies around the data stack to adapt and keep up with the rapid pace of data evolution.

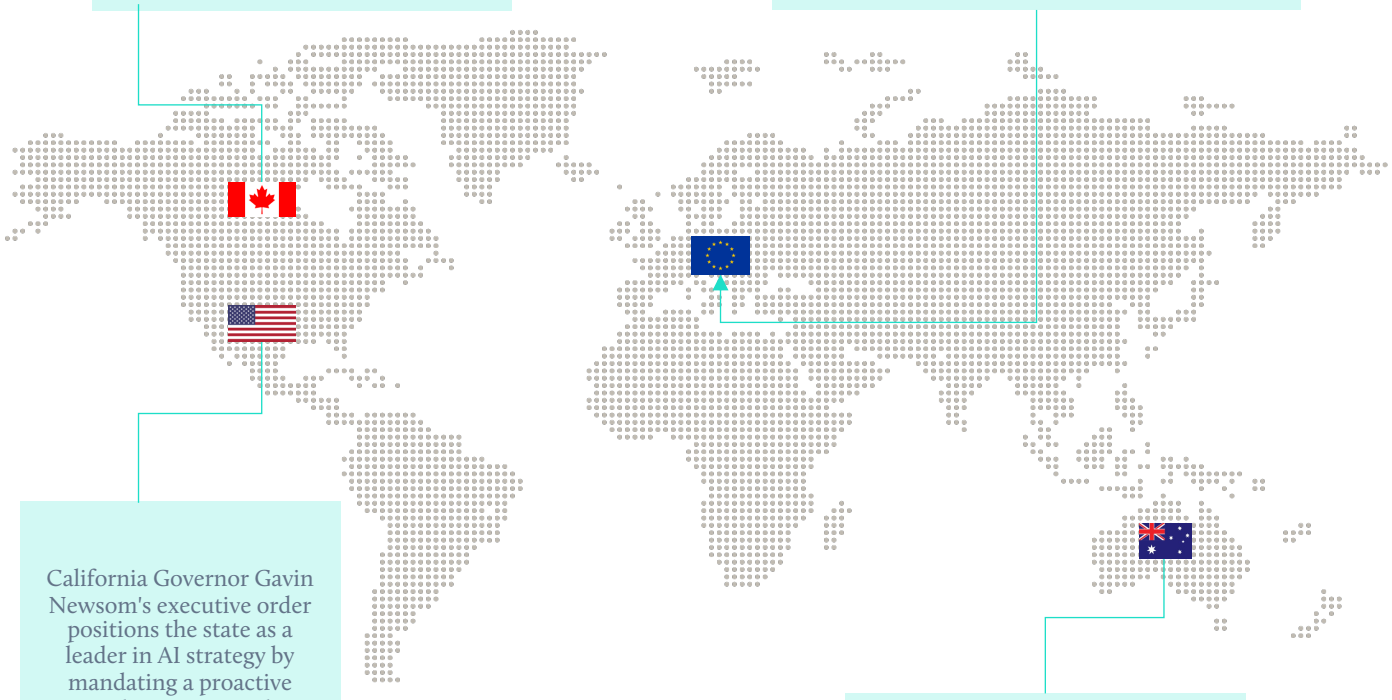
Source: Solutions Review, Statista, Houlihan Lokey

Evolving regulatory landscape in AI

Regulators are on notice with discussions over regulation of generative AI heating up in the EU, U.S., and around the world. This will only lead to increased importance and adoption of data and cybersecurity solutions as it pertains to AI related data and content

Proposed legislation set for 2025 aims to enhance current consumer protection and human rights laws by acknowledging the importance of "high-impact" AI systems adhering to human rights and safety regulations

The EU has provisionally agreed on the Artificial Intelligence Act, which addresses the challenges of AI in terms of safety, fundamental rights, and business development. The act mandates obligations for AI systems based on their risk level, focusing on transparency for general-purpose AI and stricter rules for high-impact AI models. This legislation is set to become EU law pending formal adoption by the European Parliament and policies.



California Governor Gavin Newsom's executive order positions the state as a leader in AI strategy by mandating a proactive regulatory approach, promoting trustworthy AI through state procurement, emphasising fairness and equity in AI products and services, and prioritizing workforce development for AI deployment

2023 Australian budget unveiled the Responsible AI Network, a program aimed at ensuring the ethical implementation of AI technologies

There are ongoing debates around using analytics and biometric data for training models which highlights the need for enhanced privacy safeguards

Source: Industry.Gov, DLA Piper, NY Times, ISED, Data Guidance, LOC.Gov, CNBC, Reuters

Generative AI has high-value usage in cybersecurity

Threat Intelligence
Models can analyse extensive historical data to provide insights for predicting and understanding emerging threats. They identify patterns, assisting security analysts in prioritising defense strategies

Malware Detection & Analysis
Algorithms trained on extensive malware datasets can identify common features and behaviors. This training facilitates the development of advanced malware detection systems capable of identifying and neutralising new malware strains.

Anomaly Detection
LLMs assist in anomaly detection by modeling typical system behavior and spotting deviations. They continuously learn from network traffic, user behaviors, and system logs, enabling them to detect and flag suspicious activities

Detecting GenAI Texts in Attacks
Detecting AI-generated text helps identify phishing emails, polymorphic code, and unusual email senders, and verifies if links in texts lead to known malicious websites

Examples of industry-wide adoptions

CROWDSTRIKE	Introduced Charlotte AI, a generative AI security analyst that uses the world's highest fidelity security data and is continuously improved by a tight feedback loop with CrowdStrike's threat hunters, managed detection and response operators, and incident response experts.
DARKTRACE	Used its advanced AI technology, including Darktrace DETECT™ and RESPOND™, to protect more than 8,400 customers worldwide from security and privacy risks associated with generative AI tools and LLMs, enabling businesses to leverage the power of AI safely and responsibly while guarding against potential threats and data breaches.
SentinelOne	Introduced Purple AI, a generative AI dedicated to threat-hunting, analysis and response. It uses a variety of models both open source and proprietary and aims to increase the organisation's efficiency by arming security analysts with an AI engine that can help identify, analyse and mitigate threats using conversational prompts and interactive dialog.
FLASHPOINT	Announced an expansion of its partnership with Google Cloud to leverage Security Workbench capabilities to revolutionise threat detection and risk reduction for organisations, enabling security professionals to improve decision making processes. Using NLP, the platform will transition from the traditional search-based interaction model to a conversation-based experience.
Recorded Future®	Announced the launch of its AI platform, which incorporates its 100+ terabyte intelligence cloud and 10+ years of threat reporting experience with OpenAI's GPT LLM to provide a host of benefits for both executives and analysts.
flare	Introduced an automated cyber threat intelligence linguist that can take inputs from any language on earth and deliver actionable intelligence at scale. Flare AI Assist also seamlessly explains complex technical exposure at a level that enables senior security professionals to work faster, and junior security professionals to come up to speed.

Source: WSC Analysis (2023), Houlihan Lokey

Technologies & subsectors

Part 1

Identity & Access Management

WSC Score: 9

Managing digital identities and regulating access to organisational resources is a vital priority. This subsector encompasses areas like authentication and access control and is essential for cybersecurity approaches like "zero trust," emphasising strict verification before access is granted.

Tailwinds

+ Market size to reach $41.5bn by 2030 (12.6% CAGR from 2023)
+ Growing importance for CISOs and expansion of existing cyber spend on IAM tools
+ ML enhancing authorisation capabilities with granular permissions

Challenges

+ Ongoing lack of skilled cybersecurity talent raises worries around capacity to effectively thwart and tackle cyber threats and attacks
+ Midsized enterprises see a significant drop in budget growth, raising concerns about their ability to address current and emerging cybersecurity threats

Governance, Risk & Compliance

WSC Score: 9

Part of a large cybersecurity framework that helps organisations align their information technology with business objectives, while effectively managing risk and meeting compliance requirements. Companies are continuously seeking solutions that provide real-time insights from governance and risk management to help navigate strategic compliance decisions. This is only becoming more top of mind for CISO in an ever-evolving data and AI landscape.

Tailwinds

+ Market size to reach $134.9n by 2030 (13.8% CAGR from 2023)
+ Increasing regulatory complexities may drive the need for more integrated GRC tech solutions

Challenges

+ Financial constraints, as compliance represents a significant and rising cost to organisations

Threat & Vulnerability Management

WSC Score: 8

Focused on identifying, evaluating, and mitigating cyber vulnerabilities using threat intelligence and insights into information technology and business operations for timely risk management.

Tailwinds

+ Market size to reach $37.1bn by 2032 (10.0% CAGR from 2023)
+ Increasing complexity and frequency of cyber threats demand robust vulnerability management
+ Advancements in automation and AI is improving the efficiency and effectiveness of vulnerability management

Challenges

+ Organisations have historically adopted a reactive rather than proactive approach to security, focusing on recovery after attack
+ Overload of data from vulnerability scans, most of which is irrelevant, leads to difficulty in prioritising threats

Source: WSC Analysis (2023), Grand View Research

Technologies & subsectors

Part 2

Cloud & Infrastructure Management

WSC Score: 8

In the realm of cloud computing, it is crucial to implement protection strategies and solutions that encompass the entire digital infrastructure. These strategies need to dynamically scale alongside the growth of application resources and automatically adjust to new infrastructure demands.

Tailwinds

+ Market size to reach $106.0n by 2029 (18.1% CAGR from 2022)
+ As cloud adoption increases, so does a company's attack surface, given the diverse nature of the cloud infrastructure, with its multiple applications and service model

Challenges

+ Misconfigurations and a lack of user awareness around security features can lead to vulnerabilities

Source: WSC Analysis (2023)

Market map

Identity & Access Management

IBM, Microsoft, Oracle, RSA, SailPoint, One Identity by Quest, BeyondTrust, ForgeRock, okta, Aretec, jumpcloud, auth0, CyberArk, PingIdentity, Trulioo, strongdm, veza, onelogin

Governance, Risk & Compliance

DataGrail, osano, securiti, Cyral, Alation, Fortanix, TrustArc, Relyance AI, Varonis, trust-hub, OKERA, Informatica, MMUTA, Spirion, Transcend, wirewheel

Threat & Vulnerability Management

Cisco, Qualys, Mandiant, SHI, Arctic Wolf, CrowdStrike, paloalto Networks, ThreatQuotient, Recorded Future, ThreatConnect, proofpoint, Rapid7, Flashpoint, Cyware, tenable, Anomali, Analyst1, Red Sift, flare

Cloud & Infrastructure Security

Laminar, Cyral, cyera, Lacework, orca security, Wiz, privacera, Very Good Security, onetrust, MMUTA, Opaque systems, Privitar

Source: WSC Analysis (2023)

Top asset-light startups

LACEWORK

Automated cloud security platform

$1.9bn raised to date

onetrust

Privacy and data governance platform

$1.1bn raised to date

ARCTIC WOLF

Continuous vulnerability and risk management platform

$1.1bn raised to date

WIZ

Cloud security and identity management platform

$801m raised to date

onfido

Verification platform intended to automate the identity verification process

$287m raised to date

Source: WSC Analysis (2023)

AI & Analytics Tools

AI & Analytics Tools

AI is transforming business operations across all sectors. Industry leaders are recognising its vast potential and expansive applications, leading to increased spending to develop and manage AI capabilities in products

496

Number of exits in the sector from 2020-2023YTD[1]

97%

Of surveyed developers are using GenAI and AI tools today[2]

$1.9tn

Estimated market size for AI tools by 2030[3]

$17.8bn

Total funding for Generative AI investments in 2023[4]

Generative AI in particular has emerged as one one of the most important technological developments of 2023. Initially, businesses with advanced digital infrastructure were the first to invest in and implement Generative AI technologies. However, over the past few months, there has been a noticeable shift with a slew of start-ups building tools to support companies at various stages of digital maturity to leverage Generative AI and the associated challenges with implementing and maintaining these tools.

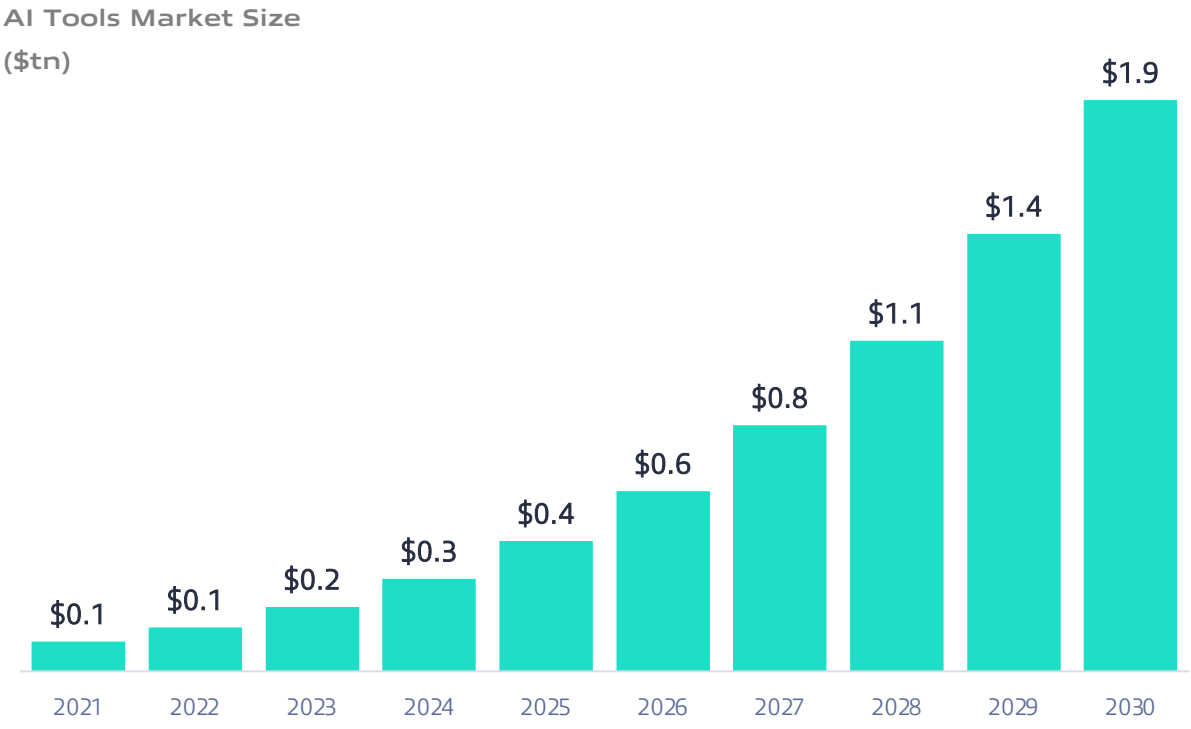
The broad adoption of AI has been to support and enhance Sales, Marketing, and Customer Service. However, AI is seeing the largest adoption among developers. Nearly 100% of engineers in one survey[2] reported using AI in their work. Yet despite this, some of the greatest challenges are faced in models and tools, such as monitoring model performance in production, fine-tuning for better results, and retraining models over time[5] representing an opportunity for companies to build for.

Technology	WSC Score	Notes
Visualisation Dashboarding	10	Priority for data-driven orgs
Predictive Analytics	10	Drives revenue opportunities
Generative AI UI/UX	7	Used to better leverage AI
AL/ML-Enhanced DevTools	7	A new suite of tools for devs
AI/MLOps	7	Critical to manage & deploy AI

[1] As of November 2023; Pitchbook, [2] Sonatype, [3] Statista, [4] Dealroom [5] McKinsey

Market drivers

AI, as a tool available to the public, is still in its early days. For years, it required teams of Data Scientists and Engineers to develop specialised products with narrow focus. Now, individuals from various professional backgrounds are experimenting and using a new suite of AI tools to drive results in their professional and personal lives



Source: Statista

Great Economic Potential

Generative AI is set to revolutionise the global economy, potentially adding $2.6 to $4.4tn annually across various use cases, outstripping the GDP of countries like the United Kingdom. The most significant gains are expected in customer operations, marketing and sales, software engineering, and R&D, with industries like banking, high tech, and life sciences poised to benefit the most.

In banking alone, the impact could be an additional $200 to $340bn each year, and in retail and consumer goods, up to $660bn annually.

Companies are Increasingly Adopting AI Tools in the Workplace

2023 was a year of experimentation for companies. Large tech incumbents such as Google, Meta, Microsoft, and Nvidia saw significant growth due to their involvement in building AI models and infrastructure.

Looking ahead to 2024, companies across all industries will be in a place to integrate AI tools in their processes to reduce costs and enhance productivity. This will be driven driven by learnings from extensive experimentation with AI in 2023, the upcoming launch of more AI products from tech giants and start-ups, and an easing in the market for AI talent making hiring for AI-related roles more feasible.

Source: McKinsey, WSC Analysis (2023)

AI / ML Ops

The development and delivery of AI and ML models is a robust pipeline. Activities such as model training, deployment, and monitoring form the core of AI/MLOps. Picking the right tools and practices that support each stage are crucial to developing and maintaining models that perform exceptionally well

WSC Area of Interest

Data Ingestion

1. Data Collection

The initial stage in developing a ML model is gathering relevant data from various sources. The quality and quantity of data collected significantly impacts the performance of the model being built

2. Data Pre-Processing

This involves cleaning and organising raw data into a suitable format for analysis. Handling missing values, normalising data, encoding categorical variables, and removing duplicates all ensure data is ready for model training.

3. Feature Engineering

This process involves selecting, modifying, or creating new features from the raw data to improve a model's performance. Highlighting patterns and relationships in the data can enhance a model's ability to learn

Model Development

4. Model Training

During model training, a ML algorithm is chosen and applied to learn from the processed data. The model iteratively adjusts its parameters to minimise errors and improve its predictions based on the training data set.

5. Model Selection

Choosing a model is typically based on a model's performance metrics (such as accuracy, precision, and recall on datasets), its complexity, and its suitability for the data and problem.

6. Model Deployment

Once trained & selected, a model is deployed into a production environment where it can process real-time data and provide predictions or insights. Deployment can vary from a simple script to complex cloud-based solutions

ML Operations

7. Performance Monitoring

A common issue is models working well in "lab" environments but not in production. Monitoring is important to assess performance degradation due to changes in data or the environment in which it is deployed

8. Retraining & Tuning

Models need to be periodically retrained. Fine-tuning models is becoming increasingly important for models that are generic by design to bring value to specific verticals. Automating this pipeline is a key aspect of AI/MLOPs

Model Governance

9. Model Governance

A framework of policies, and practices need to be in place to ensure responsible use and management of ML models. This includes addressing ethical considerations, and ensuring compliance with regulations

10. Model Security

Securing the data pipeline, implementing access controls, & safeguarding against threats like adversarial attacks, & data poisoning is crucial to protect the model and its data from unauthorised access and malicious intent

The greatest challenges facing Engineering teams are coincidentally attractive opportunities to invest in from a White Star Capital / VC investor standpoint

Source: KD Nuggets, WSC Analysis

Technologies & subsectors

Part 1

Visualisation Dashboarding

WSC Score: 10

Solutions that can deliver clear and concise insights and visualizations are crucial for businesses looking to become increasingly data-driven.

Tailwinds

+ There is a large pull from C-Suites to implement reporting and SaaS tools and often, this is where budgets are first allocated
+ Organisations are using more tools as part of their tech stacks generating more data than before. Centralising reporting across siloes is becoming increasingly important for organisations

Challenges

+ Unlike other current trends, this is not a new space and there already exist many industry leaders
+ Most products offer analytics tools as a part of their value offering. While these tools are not very effective or robust, the fact that they are integrated and offer free features makes them compelling for teams to keep

Predictive Analytics

WSC Score: 10

A relatively mature space, the future of predictive analytics is shaping up to provide more granular and accurate predictions, allowing for increased personalisation and better behavioral understanding.

Tailwinds

+ New technologies, such as GenAI, are creating new opportunities for platforms to offer more sophisticated tools

Challenges

+ Integrating various tools and ensuring compatibility is complex, especially in environments with a mix of legacy systems and modern applications

Generative AI UX

WSC Score: 7

For the most part, publicly available Generative AI is generic and not tuned to specific verticals. In addition to this, most models are difficult to interface with. While OpenAI has made some strides towards multi-modality, start-ups that are building new ways to interact with GenAI models and make them more accessible to specific verticals are poised to succeed in the next few years.

Tailwinds

+ The increasing demand for personalised, dynamic, and multi-modal user experiences with the rapid evolution of AI technologies

Challenges

+ Very challenging to set up. Most interfaces don't work well alongside trainable models and it takes considerable computational effort to implement in live environments

Source: WSC Analysis (2023)

Technologies & subsectors

Part 2

AI & ML-Enhanced DevTools

WSC Score: 7

AI/ML-Enhanced DevTools are development tools augmented with AI and ML capabilities to improve software development processes, including coding, testing, and debugging.

Tailwinds

+ The increasing complexity of software development and the ongoing demand for faster and more efficient development cycles
+ The continuous advancements in AI and ML technologies also contribute to the growth potential

Challenges

+ Managing the quality control of AI models, especially in coding accuracy
+ Data leakage from incorrect use of tools has driven some organisations to restrict the use of AI/ML enhanced DevTools in the workplace

AI/MLOps

WSC Score: 7

AI/MLOps involve best practices, guidelines, and technologies for deploying and maintaining ML models in production.

Tailwinds

+ The rapid adoption of AI/ML in business processes largely due to advancements in GenAI creates a need for scalable and efficient AI/ML model deployment

Challenges

+ Integrating AI/ML workflows with existing IT infrastructure, especially legacy infrastructure, is a challenge for many organisations

Source: WSC Analysis (2023)

Market map

Predictive Analytics

Google Cloud, qualtrics, IBM, Qlik, Pecan, dynatrace, Microsoft Azure, RAPIDMINER, teradata., H2O.ai, data iku, Hewlett Packard Enterprise, alteryx, ALTAIR, DOMO, MAXA, CLOUDERA, TIBCO, Hitachi Vantara

Visualisation Dashboarding

+ableau, Power BI, SAP Lumira, Google Data Studio, Astrato, TOUCAN TOCO, plotly, Observable, GRID, preset, Datawrapper, Infogram, Rill, cumul.io

AI & ML-Enhanced DevTools

kite, GitHub Copilot, replit, Cursor, Sourcegraph, codota, tabnine, intellicode, DEEPCODE, applitools, test.ai, walrus.ai, TabbyML, codium, mutable.ai, STACKHAWK, functionize, Sofy.

AI/ML Ops

new relic, dynatrace, freshworks, PagerDuty, digitate, AISERA, BigPanda, Site24x7, LogicMonitor, meshIQ, splunk>enterprise, APPDYNAMICS, DATADOG, INSTANA

Generative AI UX

Humanloop, glean, nuclia, coveo, algolia, YEXT, SuperAnnotate, ENCORD, elasticsearch

Source: WSC Analysis (2023)

Top asset-light startups

Contentsquare

Digital experience insights optimization platform

$1.4bn raised to date

dataiku

Centralised data platform helping businesses in their data journey

$852m raised to date

BigPanda

Platform of AI for IT operations

$431m raised to date

algolia

Search and discovery application programming interface platform

$334m raised to date

H_2O.ai

Open-source ML automation platform designed to democratise artificial intelligence

$251m raised to date

Source: WSC Analysis (2023)

Productivity & Developer Tools

Productivity & Developer Tools

In today's digital era, nearly every service and product we interact with has been shaped by software. Behind every software, there lies a robust foundation of tools and frameworks designed to streamline the development process

766

Number of exits in the sector from 2020-2023YTD[1]

79%+

Growth in market value for Big-AI focused companies in 2023[2]

$152bn

Global productivity and developer tools market size by 2030[3]

83%+

Of Web Traffic is represented by API calls, the result of digital transformation and cloud-based application deployment[4]

The Software Development Process, comprised of various stages, demands diverse expertise and the need to overcome unique hurdles. These challenges present opportunities for tools and frameworks that facilitate easier, more efficient build processes. The breakneck speed at which software is being developed is mirrored by the speed of growth of tools designed to support developers. Each new advancement propels this sector into an exciting space of innovation.

The nature of Developer tools is that they often solve similar problems but with small or niche differences. As such, tools often face fierce competition amongst each other. Yet once a tool proves its worth, it becomes an invaluable and sticky asset to developers, showing a tendency to rapidly gain traction within organisations and spread across the industry via talent mobility. This stickiness and potential for widespread adoption presents compelling investment opportunities.

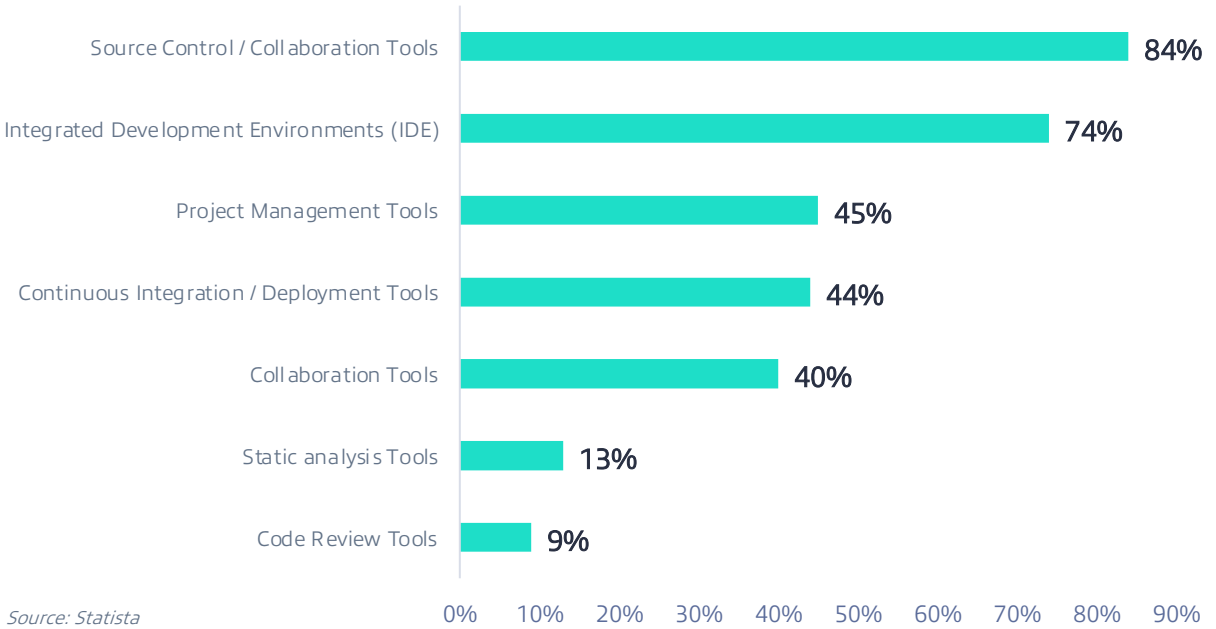
Technology	WSC Score	Notes
Software Delivery Lifecycle	9	Support collaboration
Developer SaaS	8	New opportunities for GenAI
Developer Infrastructure	7	Necessary to deploy software
Build & Development Tools	6	All developers use this

[1] As of November 2023; Pitchbook , [2] The Economist, [3] GrandView Research, [4] Akamai

Market drivers

Building software demands time and people. Enhancing software quality involves superior tools that boost developers' efficiency. These tools streamline time management, enabling developers to concentrate on essential tasks, thereby accelerating the product release process

Programming & Development Tools used by Software Developers Worldwide, 2022



Source: Statista

Increased Complexity in Tech Stacks

Modern applications often require integration with various systems, handling large data sets, and ensuring high performance across different platforms and devices.

This complexity has never demanded more sophisticated development tooling to support this, such as advanced coding assistance tools, debuggers, and integrated development environments (IDEs) that support multiple languages and frameworks.

Tools that offer automation, such as continuous integration and continuous deployment (CI/CD) pipelines, are becoming indispensable for efficient software development. These tools not only speed up the development process but also ensure consistent quality and easier maintenance of codebases.

Agile Methodologies are More Crucial than Ever Before

In a world where products need to be shipped fast and often, DevOps practices and agile methodologies focus on collaboration, flexibility, and speed, facilitate rapid development cycles and seamless communication between developers, testers, and operations teams.

Tools that support version control, real-time collaboration, and project management are crucial in this regard. Cloud-based development environments and tools are also gaining traction due to their scalability and accessibility. They allow teams to work collaboratively, a feature that has become increasingly important in the wake of the global shift towards distributed engineering teams.

Source: McKinsey, WSC Analysis (2023)

The Software Development Lifecycle

The Software Development Lifecycle (SDL) is a structured framework that outlines the stages and processes involved in creating and maintaining software. Developers often use different tools for each stage to support specific requirements and objectives.

Represents WSC Area of Interest

1. Planning
Developers define project goals, scope and requirements. Project planning tools are essential to ensure alignment with stakeholders

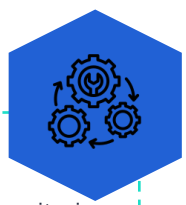
2. Requirements Gathering
Requirement gathering tools include collaborations tools, informational repositories, and other tools to document and detail product builds

3. Design
Design tools like wireframing and prototyping software facilitate collaboration between designers and developers

4. Development
Tools such as integrated development environments (IDEs), code editors, and version control systems help write, test, and manage code efficiently

5. Testing
Developers use testing tools and frameworks to automate testing processes, ensuring that the software performs as expected and meets the defined requirements

6. Deployment
Deployment tools and automation scripts are used to streamline the process of deploying to production environments, making it more efficient and less error-prone



7. Operation & Maintenance
Developers rely on monitoring tools, error tracking software, and customer feedback to continually improve software

8. Analytics Tools
Developers often place analytics on top of data-streams to learn user behaviour and effectively iterate on their products

Innovative tools that address specific pain points in the SDL process are likely to find receptive markets making them attractive opportunities to invest in from a White Star Capital / VC investor standpoint

Source: WSC Analysis

Technologies & subsectors

Part 1

Software Delivery Lifecycle

WSC Score: 9

(Inclusive of Project Management Tools, and Collaboration Tools)

Solutions that coordinate how teams design, develop, and test software. This sector contains further segments of products, services, and solutions that exert an influence on developers throughout the lifecycle of the development process.

Tailwinds

+ Rise in remote working has boosted the demand for tools that facilitate collaboration across distributed teams
+ Tech layoffs have shifted from bottom-up buying motion and putting the purchasing power back in the hands of organisations

Challenges

+ Fragmentation and fierce competition occurs from many different players in the space

Developer SaaS

WSC Score: 8

(Inclusive of Design Tools, and Low-code/No-code Platforms)

These tools and services are accessed via the internet and offer a range of functionalities to assist developers and non developers in creating, deploying, and managing software applications. They often take the form of Low-code, Platform as a Service, or Design Tools.

Tailwinds

+ New technologies, such as GenAI, are creating new opportunities for platforms to offer more sophisticated tools

Challenges

+ Integrating various tools and ensuring compatibility is complex, especially in environments with a mix of legacy systems and modern applications

Developer Infrastructure

WSC Score: 7

(Inclusive of Containerisation & Orchestration Tools, Testing Frameworks, CI/CD Tools, Configuration Management Tools, and Package Managers)

All software that supports the distributed repeatable development of software that engineering teams depend on, such as Version Control, Testing Frameworks, CI/CD Tools, and package managers.

Tailwinds

+ Mission-critical for engineering teams to deploy software at scale efficiently

Challenges

+ Integrating various tools and ensuring compatibility is complex, especially in environments with a mix of legacy systems and modern applications

Source: WSC Analysis (2023)

Technologies & subsectors

Part 2

Build & Development Tools

WSC Score: 6

(Inclusive of Containerisation & Orchestration Tools, Testing Frameworks, CI/CD Tools, Configuration Management Tools, and Package Managers)

Tools that support the day-to-day workflows of an individual engineer. Similar to choosing what pen or notebook to use, developers put together an ecosystem of what works best for them.

Tailwinds

+ All developers of every level have a need for these tools. There will never not be users
+ Developers influence others around them often leading to strong product led growth in this sector

Challenges

+ Many open-source tools making it hard to monetize
+ Individuals can switch, and they switch often resulting in low stickiness
+ Difficult to sell at enterprise level

Source: WSC Analysis (2023)

Market map

Software Delivery Lifecycle

Jira Software, Confluence, Trello, asana, Notion, wrike, Adobe Workfront, monday.com, jupyter, Apache Zeppelin, slack, GitHub, smartsheet, ClickUp, Basecamp

Developer SaaS

Figma, Sketch, webflow, Canva, Adobe Creative Cloud, affinity, Microsoft PowerApps, outsystems, quickbase, mendix, ZOHO, PEGA, servicenow, Retool, kintone

Developer Infrastructure

docker, kubernetes, MESOSPHERE, ANSIBLE, puppet, CHEF, Maven, npm, Jenkins, Travis CI, GitLab, Selenium, JUnit 5, appium

Build and Development Tools

aws, Microsoft Azure, Google Cloud, Visual Studio, PyCharm, eclipse, PVS studios, sonarqube, Crucible, Gradle, Maven, Jenkins, GitHub, GitLab, beanstalk, Review Board, Crucible, Axolo

Source: WSC Analysis (2023)

Top asset-light startups

sonarsource

Coding software designed to manage the code quality of applications

$4.7bn raised to date

outsystems

Application and delivery platform designed to integrate custom code

$802m raised to date

docker

Software container platform

$443m raised to date

asana

Platform for work management that helps teams orchestrate work

$414m raised to date

GitHub

Software development platform for open-source and private projects in organisations

$351m raised to date

Source: WSC Analysis (2023)