Power Week 2025

18 - 19 - 20 novembre 2025
IBM Innovation Studio Paris

S103 – L'IA simplifiée et rapide pour IBM Power

20 novembre 09:45 -10:45

Ronnie Moodley
Global Principal Sales Leader
ronniem@ibm.com

Marc Bouzigues
solution architect
marc_bouzigues@ibm.com

IBM

common
FRANCE

Generative AI is poised to unleash the next wave of productivity

$4T

Allowing businesses to unlock value & growth in ways that couldn't be imagined a few decades ago.

annual GDP productivity value expected from AI by 2030.[1]

1. The Economic Potential of Generative AI: The next productivity frontier

98% of enterprises have experimented with Gen AI

But only…

26%

…have scaled experiments into production.[1]

Nearly all enterprise pilots are stuck at the starting line

with only…

5%

...see ROI from AI.

The time for scaling AI is now
but enterprises face obstacles.

TOP 3 OBSTACLES

Data & integration complexity

Skills

Security & compliance

Deploy enterprise AI across critical processes

Scale AI across the enterprise

- Transform enterprise processes with simplified AI integration

- Boost productivity with assistants and agents

- Unlock enterprise data with data fabric and platforms

5x Increase in business process rate with AI integration into existing enterprise workflows[a]
Uwe Remppel, head of ZSI department, Geis Group

18x Increase in developer productivity with AI-powered code assistant[b]
Jasmine Kaczmarek, VP Technology, MR Williams

23% lower TCO while meeting GenAI performance requirement vs competition[c]
Martin Rydén, Head of Architecture & Projects, Shibuya

Ecosystem
Equitus.ai | Rocketgraph | Finacle|Wallaroo.AI

Assistants and Agents
Project Bob

Data Fabric
watsonx.data[2] | IBM DataStage | IBM Knowledge Catalog

AI Foundation
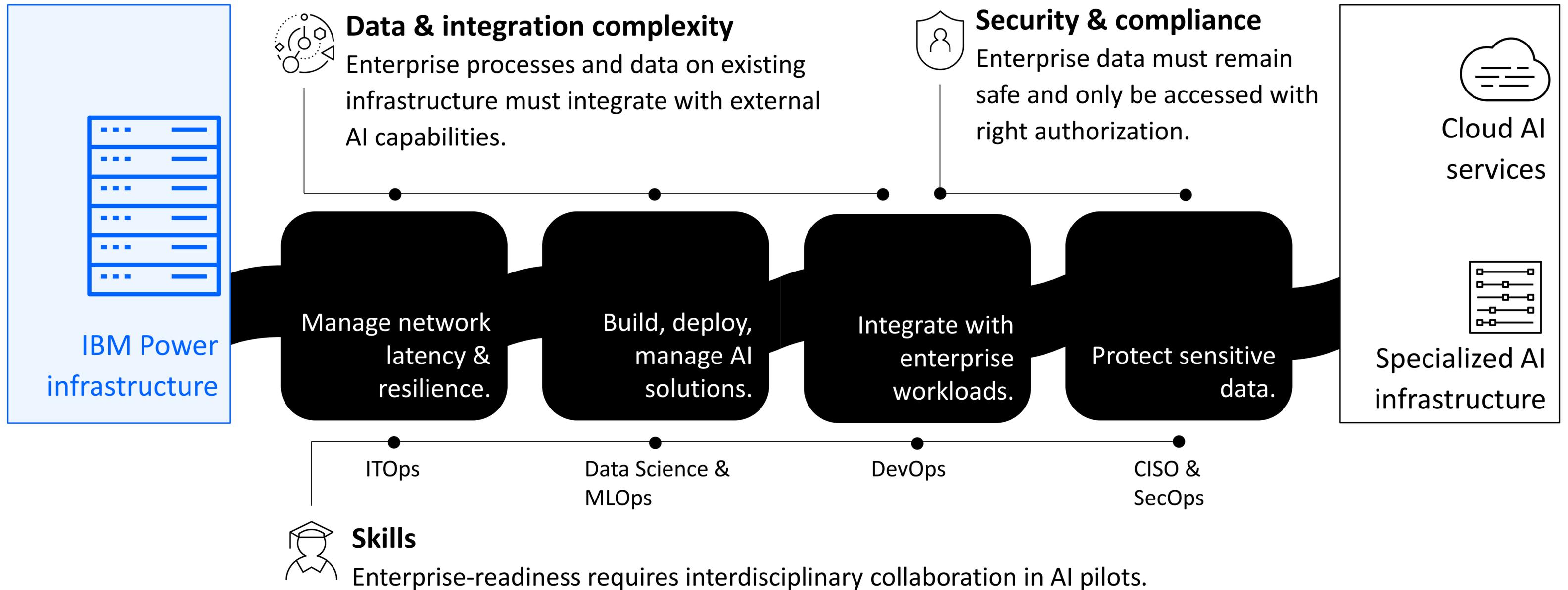Red Hat OpenShift AI[3] | Python Ecosystem for IBM Power

Infrastructure
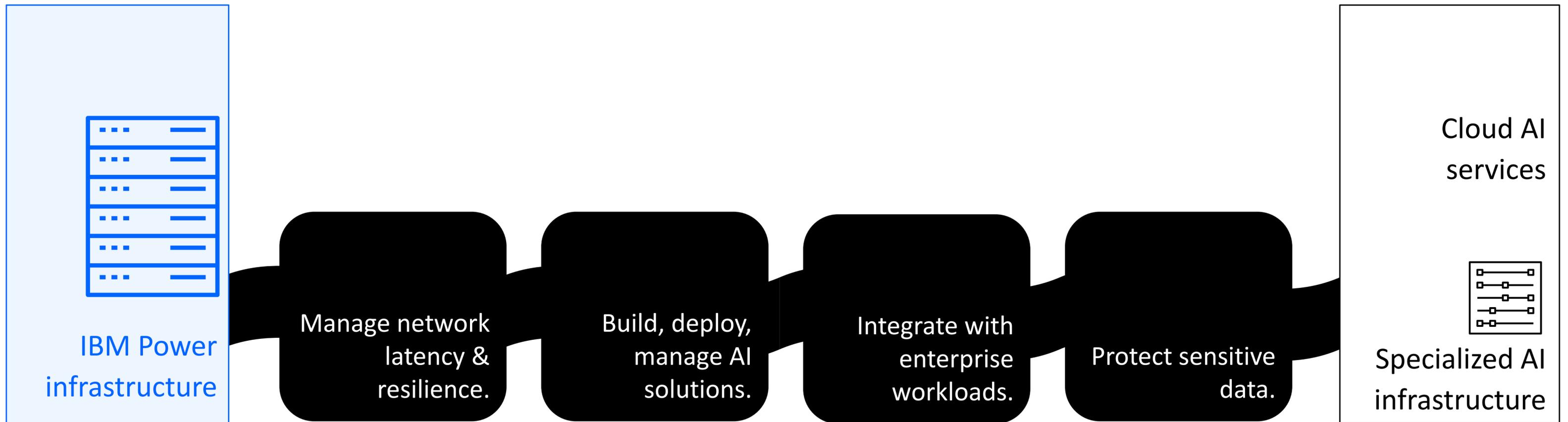IBM Spyre Accelerator[4.] | IBM PowerVS Toolkits

1. Public Preview
2. GA 4Q '25

3.Tech Preview, SoD GA (MMA) 4Q '25, GA (Spyre) 1Q'26
4. Tech Preview

Core enterprise workflows get the largest ROI from AI, if obstacles are addressed.
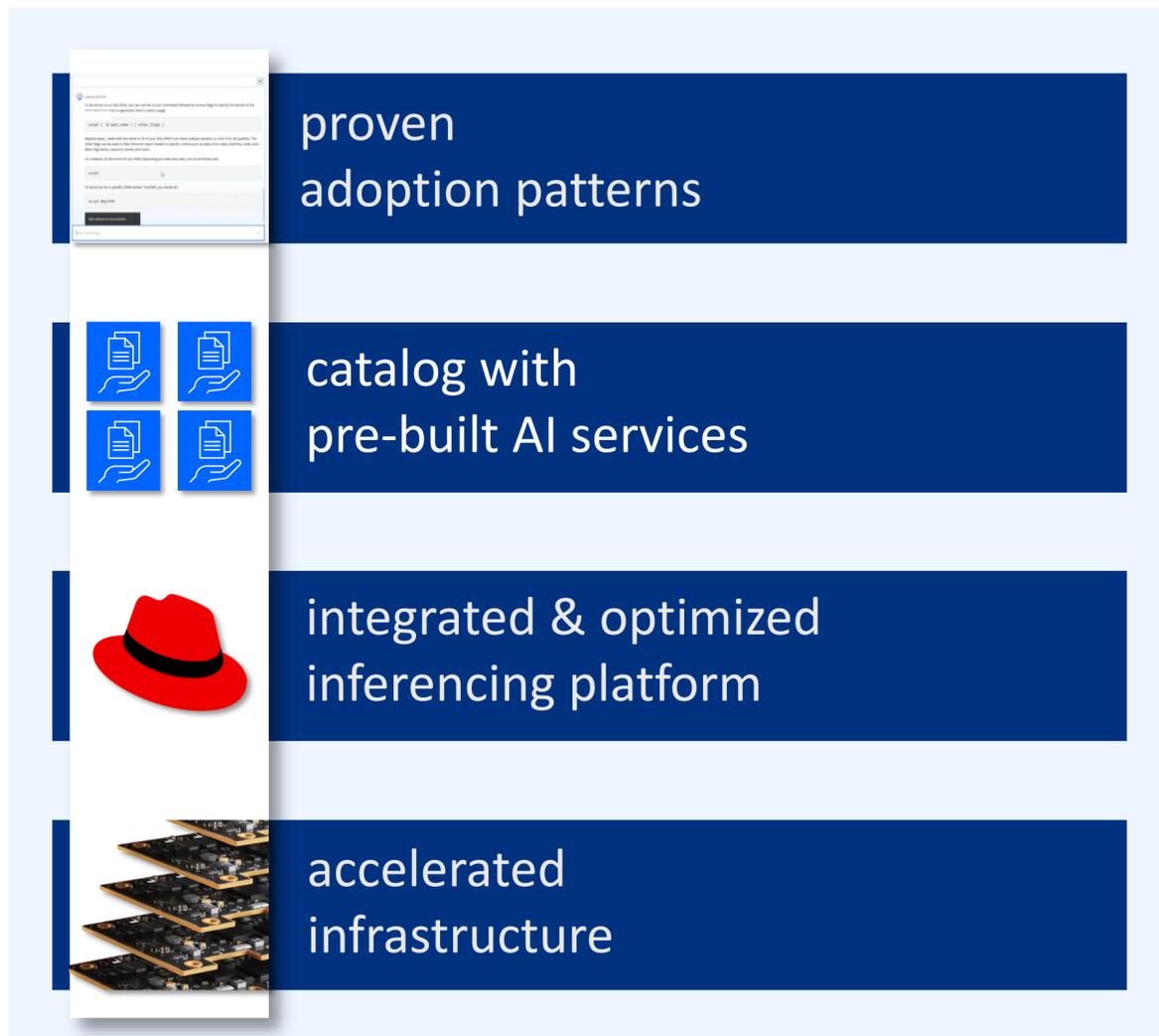
Data & integration complexity
Enterprise processes and data on existing infrastructure must integrate with external AI capabilities.

Security & compliance
Enterprise data must remain safe and only be accessed with right authorization.

IBM Power infrastructure

Cloud AI services

Specialized AI infrastructure

Manage network latency & resilience.

Build, deploy, manage AI solutions.

Integrate with enterprise workloads.

Protect sensitive data.

ITOps

Data Science & MLOps

DevOps

CISO & SecOps

Skills
Enterprise-readiness requires interdisciplinary collaboration in AI pilots.

What if all of this now comes out-of-the-box?

IBM Power infrastructure

Manage network latency & resilience.

Build, deploy, manage AI solutions.

Integrate with enterprise workloads.

Protect sensitive data.

Cloud AI services

Specialized AI infrastructure

IBM Spyre™ for Power
Turnkey AI for enterprise workloads.

proven adoption patterns

catalog with pre-built AI services

integrated & optimized inferencing platform

accelerated infrastructure

1 click

...to install AI services from the IBM-supported catalog.[1]

1 configuration

...to move AI services of the IBM-supported catalog between IBM Power & IBM Power Virtual Server.[2]

>8 million/hour

...document embeddings for knowledge base integration using IBM Spyre™ Accelerator for Power with batch and prompt sizes of 128.[3]

Disclaimer: 1: AI services of the IBM-supported catalog are delivered as one or a set of containers that can be deployed with a single deployment command. The provided UI for the catalog executes such commands in the backend based on a single click within the UI page of the respective AI service. 2: A single configuration is enabled by exposed industry standard APIs to decouple services at the top and the backing inferencing service for all AI services that are part of the IBM-supported catalog. Any service that requires AI inferencing capabilities can connect inferencing services that provide OpenAI API or watsonx.ai API compliant inferencing endpoints (Spyre endpoint, RH AI Inferencing Server, IBM Cloud, OpenAI, Azure, AWS, GCP, ...). Services can run either on IBM Power or on IBM Power Virtual Server. 3: Based upon internal testing running 1M unit data set with prompt size 128, batch size 128 using 1-card container. Individual results may vary based on workload size, use of storage subsystems and other conditions.

Enterprise use cases

IT Ops	Development	Enterprise Resource Planning	Banking and Finance	Healthcare	Insurance	Public	Other

Code Assistant	Detect & Fix Agent	Forecast & Plan Capacity Assistant	IT Service Desk Assistant

Adoption patterns

Data & Content Management	Deep Process Integration	Digital Assistant	Forecasting	Fraud Detection	Image & Video Analytics	Recommender System

Pre-built AI services

Digitize Documents	Extract & Tag Information	Generate Reports	Knowledge Management	NLP to SQL	Q&A	Serve Models	Similarity Search

Transcribe	Translate & Summarize

Install services

Turnkey AI aligns with core workloads.

Cross-industry

ITOps | Development

IT service desk assistant	Code assistants (RPG, Ansible, …)
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Supply chain forecasting
Order processing assistant	Product sales assistant

Industry-specific

Banking & Finance

Analyst assistant (frauds, NPAs, …)	Predict NPAs
Open account agent	Anti-money laundering

Healthcare

Medical assistant	Medical transcription assistant
Claims & EHR matching agent	Medical image analysis assistant

Insurance

Claims & policy management agent
Predict risk & underwrite assistant

Public

Private documents assistant
360-degree view assistant

…and more

Agriculture assistant
Real estate assistant

…

proven
Adoption patterns

Digital assistant (RAG, …)	Data & content management	Recommender system	Deep process integration	Fraud detection	Forecasting	Image & video analytics

…

pre-built
AI services

Manage knowledge (VectorDBs)	Serve models	Digitalize documents (manual, invoice,…)	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information (PII, meta data, …)	Transcribe (meetings, phone calls, …)	NLP to SQL (Db2, Oracle, SAP HANA, …)

…

Digital assistants
being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

System house DACH / **IT service desk** assistant	Code assistants (RPG, Ansible, …)
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Supply chain forecasting
Order processing assistant	Product sales assistant

Industry-specific

Banking & Finance

Analyst assistant (frauds, NPAs, …)	Predict NPAs
Open account agent	Anti-money laundering

Healthcare

Medical assistant	Medical transcription assistant
Claims & EHR matching agent	Medical image analysis assistant

Insurance

Claims & policy management agent
Predict risk & underwrite assistant

Public

Gov. client DACH / **Private documents** assistant
360-degree view assistant

…and more

SEMICON INDIA / Agriculture assistant
BanFast FÖRVALTNING / Real estate assistant

…

proven
Adoption patterns

Digital assistant (RAG, …)	Data & content management	Recommender system	Deep process integration	Fraud detection	Forecasting	Image & video analytics	…

pre-built
AI services

Manage knowledge (VectorDBs)	Serve models	Digitalize documents (manual, invoice,…)	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information (PII, meta data,…)	Transcribe (meetings, phone calls,…)	NLP to SQL (Db2, Oracle, SAP HANA, …)	…

Support Agent AI

watsonx 8:57 PM

Hi, I'm your IT Support Assistant! Ask me anything about IBM Systems.

You 8:58 PM

How can I list all devices on an AIX system?

watsonx 8:58 PM

To list all devices on an AIX system, use the command `lsdev -C`. This command will display the characteristics of all devices in your system. If a device is not listed, you'll need to define a new one.

Get reference documents

Type something...

"By integrating our digital assistant, Charlie, with IBM Spyre for Power, we'll be delivering a truly plug-and-play AI solution integrated with proprietary data on the IBM Power platform that maintains strict data compliance and, ultimately, accelerates our business forward."

– Alexander Edsvärd, CEO, Edsvärd Hållbarhet AB

EDSVÄRD

"With IBM Spyre for Power, we can host our digital AI assistant locally in Sweden. This enables us to serve a wide range of clients, including the ones who demand the highest data protection standards and compliance with European and national regulations."

– Lars Nylund, CEO, BanFast Förvaltning AB

BanFast
FÖRVALTNING

Disclaimer: BanFast is a real estates company and a client of Edsvärd, using their digital AI assistant Charlie. Edsvärd and BanFast have tested IBM Spyre™ for Power as part of a Tech Preview program. Edsvärd uses the OpenAI API standard to connect their digital assistant Charlie to AI inferencing services. The IBM Spyre™ for Power endpoint also provides OpenAI APIs (via Red Hat AI Inferencing Server). The IBM Spyre™ for Power stack allowed Edsvärd to move to IBM Spyre™ for Power in a plug-and-play fashion. Strict data compliance covers protecting sensitive information, such as personal data and credit card information, from unauthorized access or data breaches according to the most common data compliance regulations and standards such as HIPAA, GDPR, and CCPA and is achieved through an on-premises deployment in a protected IBM Power environment. The IBM Spyre™ for Power stack does not require additional networks to be setup that potentially increase attack vectors because all IBM Spyre™ for Power Accelerators are connected through PCIe cables (and not via network), ensuring protection against network attacks.

Digital assistants
being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

System house DACH	Code assistants (RPG, Ansible, …)
IT service desk assistant	
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Supply chain forecasting
Order processing assistant	Product sales assistant

Industry-specific

Banking & Finance

Analyst assistant (frauds, NPAs, …)	Predict NPAs
Open account agent	Anti-money laundering

Healthcare

Medical assistant	Medical transcription assistant
Claims & EHR matching agent	Medical image analysis assistant

Insurance

Claims & policy management agent
Predict risk & underwrite assistant

Public

Gov. client DACH
Private documents assistant
360-degree view assistant

…and more

SEMICON INDIA	
Agriculture assistant	
BanFast FÖRVALTNING	
Real estate assistant	

…

proven
Adoption patterns

Digital assistant (RAG, …)	Data & content management	Recommender system	Deep process integration	Fraud detection	Forecasting	Image & video analytics	…

pre-built
AI services

Manage knowledge (VectorDBs)	Serve models	Digitalize documents (manual, invoice,…)	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information (PII, meta data,…)	Transcribe (meetings, phone calls,…)	NLP to SQL (Db2, Oracle, SAP HANA, …)	…

Deep process integrations

being validating in Spyre™ Tech Preview.

Cross-industry

Industry-specific

ITOps | Development

IT service desk assistant	Code assistants (RPG, Ansible, …)
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Supply chain forecasting
Order processing assistant	Product sales assistant

Banking & Finance

Analyst assistant (frauds, NPAs, …)	Predict NPAs
Open account agent	Anti-money laundering

Healthcare

Medical assistant	Medical transcription assistant
Claims & EHR matching agent	Medical image analysis assistant

Insurance

Claims & policy management agent
Predict risk & underwrite assistant

Public

Private documents assistant
360-degree view assistant

…and more

Agriculture assistant
Company Sweden
Real estate assistant

…

proven
Adoption patterns

| Digital assistant (RAG, …) | Data & content management | Recommender system | Deep process integration | Fraud detection | Forecasting | Image & video analytics | … |

pre-built
AI services

| Manage knowledge (VectorDBs) | Serve models | Digitalize documents (manual, invoice,…) | Find similar items | Q&A | Translate & summarize | Generate reports | Extract & tag information (PII, meta data, …) | Transcribe (meetings, phone calls,…) | NLP to SQL (Db2, Oracle, SAP HANA, …) | … |

Real-life example:
Extract ordering details in the order process automatically

Hans Geis, a logistics provider running their core ERP system on IBM i, integrates AI for a...

"*5x increase in business process rate* with AI integration into existing enterprise workflows."
– Uwe Remppel, head of ZSI department, Geis Group[1]

Disclaimer: 1: "We achieve a 5x increase in business process rate with AI integration into our existing processes on IBM Power leveraging on-chip accelerator (MMA and SIMD) technology." - Uwe Remppel, head of ZSI department, Geis Group. Measured in staging system of Hans Geis. Without AI it takes a senior employee 2 minutes / mail = 120 seconds / mail; with AI, they can completely automate 80% with 0 minutes processing time: (80%*0 minutes+20%* 2 minutes) / mail = 24 seconds / mail; leading to a factor of 5x: 120/24 = 5.

Simplify AI on-boarding

by moving between cloud & on-premises seamlessly.

Explore.

minutes
change one configuration

**Explore cloud-based
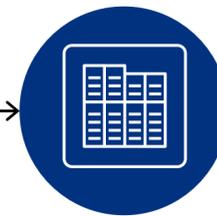AI services**
IBM Cloud, AWS, Azure,
GCP, OpenAI, …

Test.
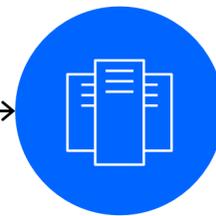
minutes
change one configuration
(when Spyre™ infrastructure is available)

**Co-locate AI services with
system of records**
IBM Power on-chip acceleration

Launch.

Promote to production
IBM Power11
+ IBM Spyre™ for Power

Deep process integrations

being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

| IT service desk assistant | Code assistants (RPG, Ansible, …) |
| Detect & fix incidents agent | Forecast & plan capacity assistant |

Enterprise Resource Planning

| BI & HR assistant | Supply chain forecasting |
| Order processing assistant | Product sales assistant |

Industry-specific

Banking & Finance

| Analyst assistant (frauds, NPAs, …) | Predict NPAs |
| Open account agent | Anti-money laundering |

Healthcare

| Medical assistant | Medical transcription assistant |
| Claims & EHR matching agent | Medical image analysis assistant |

Insurance

| Claims & policy management agent |
| Predict risk & underwrite assistant |

Public

| Private documents assistant |
| 360-degree view assistant |

…and more

| Agriculture assistant |
| Company Sweden |
| Real estate assistant |

…

---

proven
Adoption patterns

| Digital assistant (RAG, …) | Data & content management | Recommender system | Deep process integration | Fraud detection | Forecasting | Image & video analytics | … |

---

pre-built
AI services

| Manage knowledge (VectorDBs) | Serve models | Digitalize documents (manual, invoice,…) | Find similar items | Q&A | Translate & summarize | Generate reports | Extract & tag information (PII, meta data, …) | Transcribe (meetings, phone calls,…) | NLP to SQL (Db2, Oracle, SAP HANA, …) | … |

Data & content management

being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

| IT service desk assistant | Code assistants (RPG, Ansible, …) |
| Detect & fix incidents agent | Forecast & plan capacity assistant |

Enterprise Resource Planning

| BI & HR assistant | Supply chain forecasting |
| Order processing assistant | Product sales assistant |

Industry-specific

Banking & Finance

| Analyst assistant (frauds, NPAs, …) | Predict NPAs |
| CREST Open account agent | Anti-money laundering |

Healthcare

| Medical assistant | Medical transcription assistant |
| Claims & EHR matching agent | Medical image analysis assistant |

Insurance

| Claims & policy management agent |
| Predict risk & underwrite assistant |

Public

| Private documents assistant |
| 360-degree view assistant |

…and more

| Agriculture assistant |
| Real estate assistant |

…

proven Adoption patterns

| Digital assistant (RAG, …) | Data & content management | Recommender system | Deep process integration | Fraud detection | Forecasting | Image & video analytics |

…

pre-built AI services

| Manage knowledge (VectorDBs) | Serve models | Digitalize documents (manual, invoice,…) | Find similar items | Q&A | Translate & summarize | Generate reports | Extract & tag information (PII, meta data, …) | Transcribe (meetings, phone calls,…) | NLP to SQL (Db2, Oracle, SAP HANA, …) |

…

IBM i

"Personally Identifiable Information (PII) redaction helps to protect against data breach & ransom scenarios, which can cause fines on banks, insurance and other customer facing organizations, leading to loss of trust and reputation.

We are planning to leverage Spyre for PII redaction, which we believe will decrease the amount of manual work by an organization, improving their ability to onboard customers faster and more efficiently."

– Hemant Prasad, Chief Executive Officer, Crest Infosolutions

Data & content management

being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

IT service desk *assistant*	Code assistants *(RPG, Ansible, …)*
Detect & fix incidents *agent*	Forecast & plan capacity *assistant*

Enterprise Resource Planning

BI & HR *assistant*	Supply chain forecasting
Order processing *assistant*	Product sales *assistant*

Industry-specific

Banking & Finance

Analyst *assistant (frauds, NPAs, …)*	Predict NPAs
CREST — Open account *agent*	Anti-money laundering

Healthcare

Medical *assistant*	Medical transcription *assistant*
Claims & EHR matching *agent*	Medical image analysis *assistant*

Insurance

Claims & policy management *agent*
Predict risk & underwrite *assistant*

Public

Private documents *assistant*
360-degree view *assistant*

…and more

Agriculture *assistant*
…
Real estate *assistant*

proven
Adoption patterns

Digital assistant *(RAG, …)*	Data & content management	Recommender system	Deep process integration	Fraud detection	Forecasting	Image & video analytics

…

pre-built
AI services

Manage knowledge *(VectorDBs)*	Serve models	Digitalize documents *(manual, invoice,…)*	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information *(PII, meta data, …)*	Transcribe *(meetings, phone calls,…)*	NLP to SQL *(Db2, Oracle, SAP HANA, …)*

…

Recommender systems

being validating in Spyre™ Tech Preview.

Cross-industry

ITOps | Development

IT service desk assistant	Code assistants (RPG, Ansible, …)
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Supply chain forecasting
Order processing assistant	**Large retailer** US — Product sales assistant

Industry-specific

Banking & Finance

Analyst assistant (frauds, NPAs, …)	Predict NPAs
Open account agent	Anti-money laundering

Healthcare

Medical assistant	Medical transcription assistant
Claims & EHR matching agent	Medical image analysis assistant

Insurance

Claims & policy management agent
Predict risk & underwrite assistant

Public

Private documents assistant
360-degree view assistant

…and more

Agriculture assistant
Real estate assistant

...

proven
Adoption patterns

Digital assistant (RAG, …)	Data & content management	**Recommender system**	Deep process integration	Fraud detection	Forecasting	Image & video analytics

...

pre-built
AI services

Manage knowledge (VectorDBs)	Serve models	Digitalize documents (manual, invoice,…)	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information (PII, meta data,…)	Transcribe (meetings, phone calls,…)	NLP to SQL (Db2, Oracle, SAP HANA, …)

...

Client examples & demos
aligned with core workloads; not restricted to Spyre™.

Cross-industry

ITOps | Development

System house DACH	MR WILLIAMS *More Than A Convenience Store Distributor*
IT service desk assistant	Code assistants (RPG, Ansible, …)
Detect & fix incidents agent	Forecast & plan capacity assistant

Enterprise Resource Planning

BI & HR assistant	Large retailer US — Supply chain forecasting
Geis Global Logistics — Order processing assistant	Large retailer US — Product sales assistant

Industry-specific

Banking & Finance

Demo TechXchange 2024 — Analyst assistant (frauds, NPAs, …)	Infosys Finacle — Predict NPAs
CREST — Open account agent	Anti-money laundering

Healthcare

SHIBUYA BY PULSEN — Medical assistant	SiPH — Medical transcription assistant
Claims & EHR matching agent	SiPH — Medical image analysis assistant

Insurance

Demo TechXchange 2025
Claims & policy management agent
Predict risk & underwrite assistant

Public

Gov. client DACH
Private documents assistant
360-degree view assistant

…and more

| SEMICON INDIA	MIDRANGE
Agriculture assistant	…
Company Sweden — Real estate assistant	

proven
Adoption patterns

Digital assistant (RAG, …)	Data & content management	Recommender system	Deep process integration	Fraud detection	Forecasting	Image & video analytics	…

pre-built
AI services

Manage knowledge (VectorDBs)	Serve models	Digitalize documents (manual, invoice,…)	Find similar items	Q&A	Translate & summarize	Generate reports	Extract & tag information (PII, meta data, …)	Transcribe (meetings, phone calls, …)	NLP to SQL (Db2, Oracle, SAP HANA, …)	…

IBM Spyre™ Accelerator PCIe attached card

- 300+ TOPS

- 75W PCIe gen5 x16
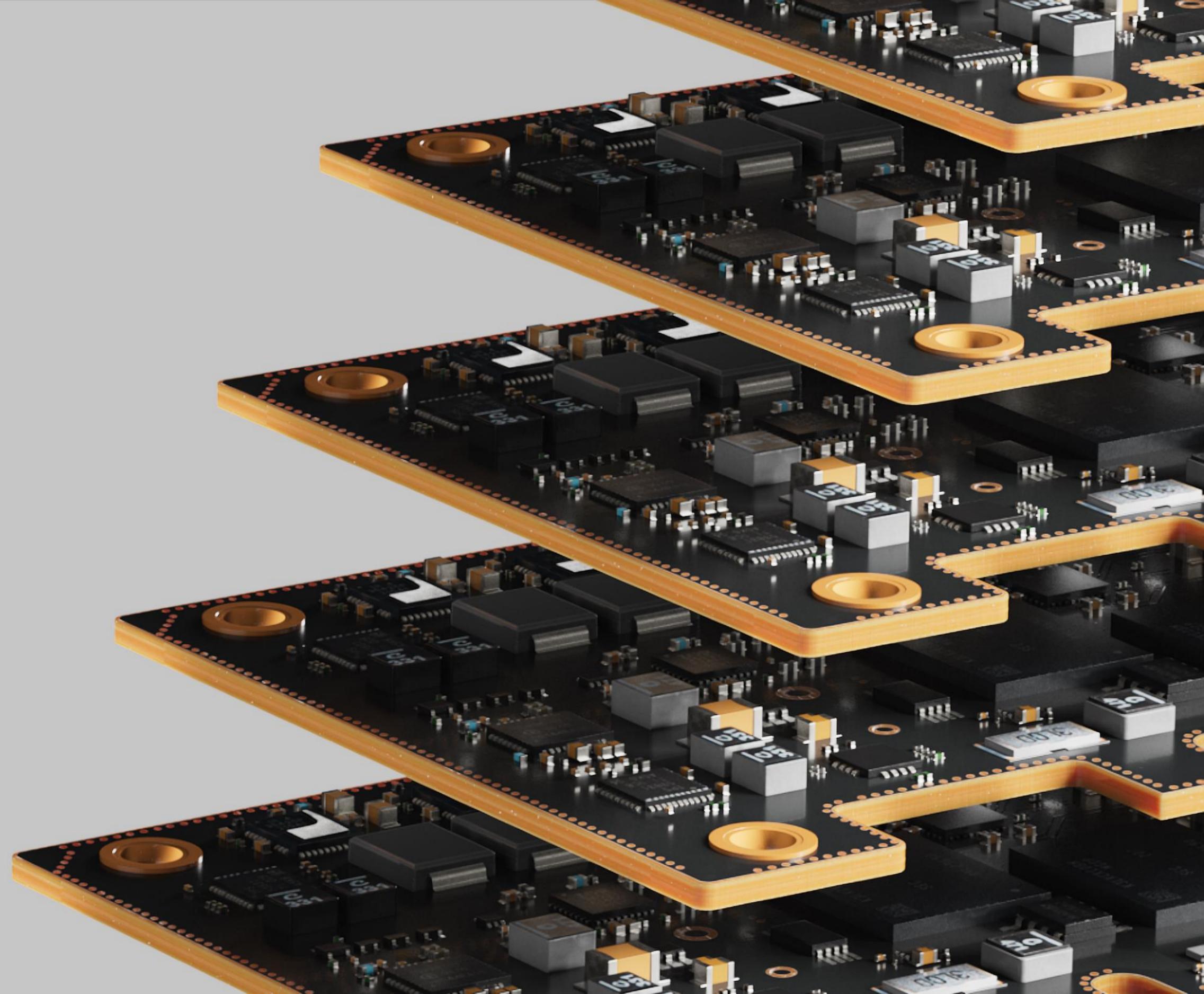 adapter

- 128GB of
  LPDDR5 memory
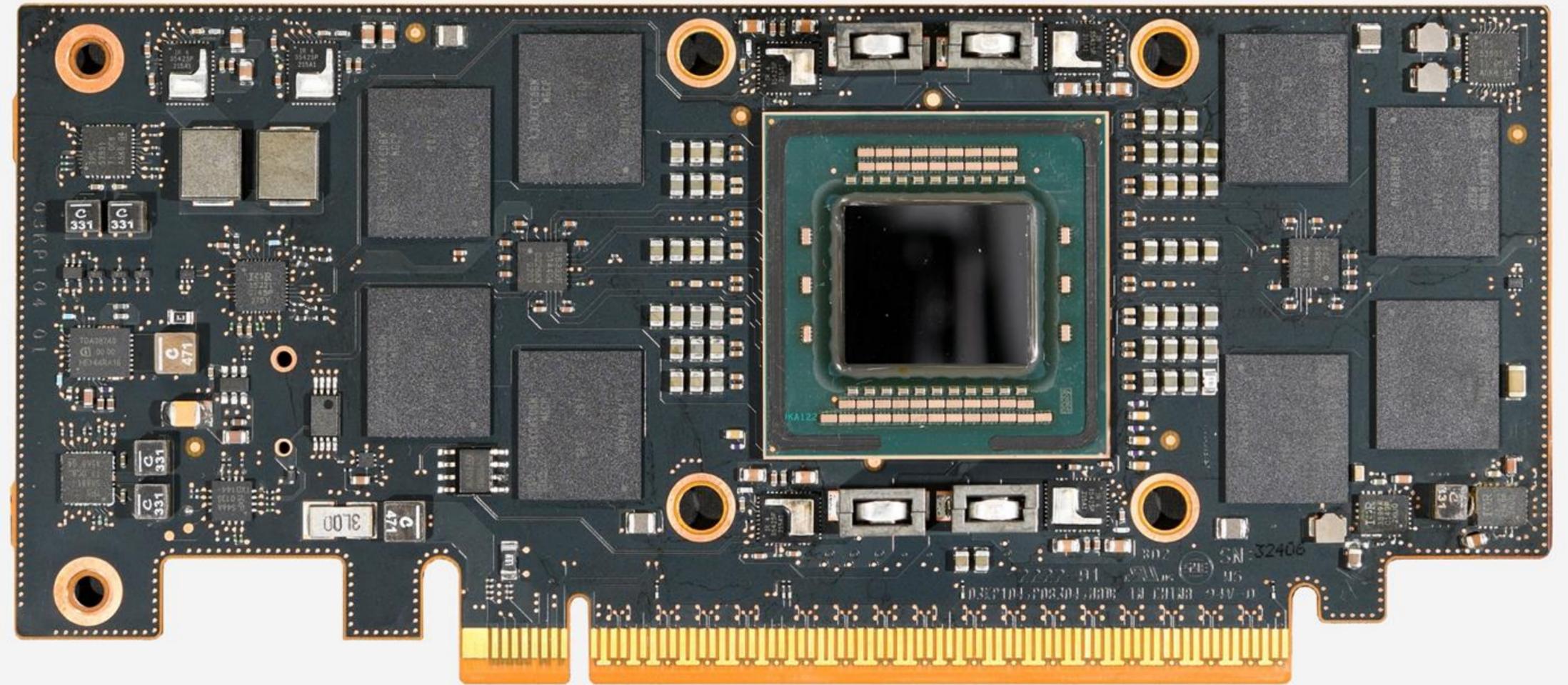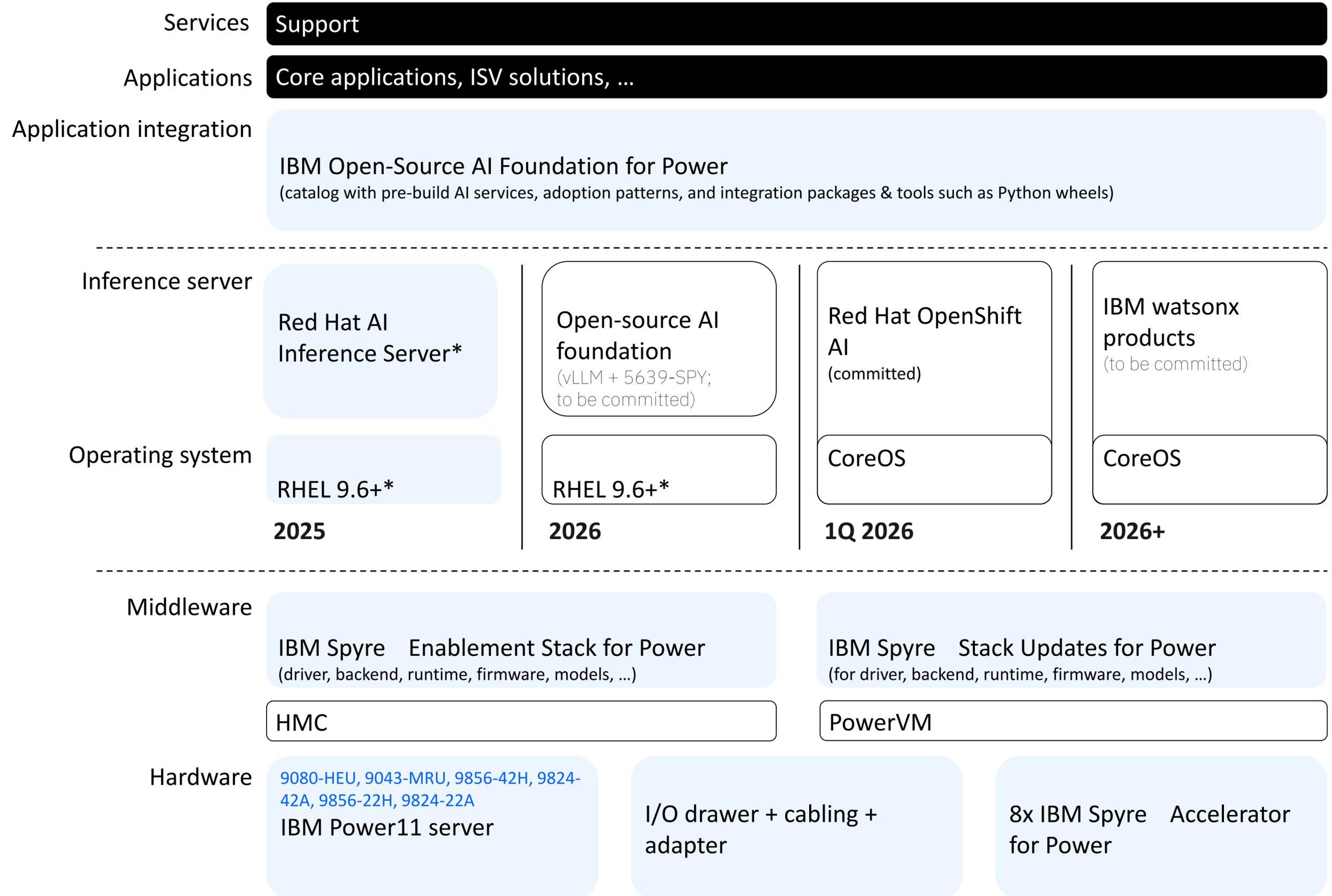
Designed to handle
generative AI use cases

8 cards in I/O drawer
form a logical cluster

1TB of memory

1.6TB per second
aggregate memory
bandwidth

Spyre Stack

Services
Support

Applications
Core applications, ISV solutions, …

Application integration

IBM Open-Source AI Foundation for Power
(catalog with pre-build AI services, adoption patterns, and integration packages & tools such as Python wheels)

---

external
mandatory
optional

*removal allowed if purchased via Red Hat. Support for those components will then come via Red Hat.

Inference server

| Red Hat AI Inference Server* | Open-source AI foundation (vLLM + 5639-SPY; to be committed) | Red Hat OpenShift AI (committed) | IBM watsonx products (to be committed) |

Operating system

| RHEL 9.6+* | RHEL 9.6+* | CoreOS | CoreOS |
| **2025** | **2026** | **1Q 2026** | **2026+** |

Middleware

| IBM Spyre Enablement Stack for Power (driver, backend, runtime, firmware, models, …) | IBM Spyre Stack Updates for Power (for driver, backend, runtime, firmware, models, …) |
| HMC | PowerVM |

Hardware

| 9080-HEU, 9043-MRU, 9856-42H, 9824-42A, 9856-22H, 9824-22A IBM Power11 server | I/O drawer + cabling + adapter | 8x IBM Spyre Accelerator for Power |

Strategic roadmap ensures continued outcome driven AI innovations

Expanding Progressively →

	Q1 2025	Q2 2025	Q3 2025	Q4 2025	1Q 2026
AI-Ready Infrastructure		PowerVS & watsonx SaaS integration toolkit V1	IBM Spyre Accelerator – Tech Preview PowerVS & watsonx SaaS integration toolkit V2 with Satellite Connector	IBM Spyre Accelerator – GA	
Optimized AI Foundation and Ecosystem	Finacle AI Platform Equitus Video Analytics Wallaroo.ai Rocketgraph	RH OpenShift AI Tech Preview (MMA)	Power inferencing microservice – Tech Preview Python Ecosystem for IBM Power IBM Orchestration Pipelines	RH OpenShift AI GA (MMA) RH OpenShift AI Tech Preview (Spyre)	RH OpenShift AI GA (Spyre)
Data Fabric		IBM DataStage (Data Integration) IBM Knowledge Catalog (Data Intelligence)	watsonx.data – Tech Preview	watsonx.data – GA1	
Assistants and Agents	WCA for i [Explain] – Private Preview		WCA for i – Public Preview Project Bob IBM Concert Integration – IBM Power Maintenance	~~WCA for i – GA (Explain, Understand)~~ Project Bob	~~WCA for i – next~~ Project Bob IBM Concert Integration – Risks & Resilience

Discover turnkey AI with IBM Spyre™ for Power.
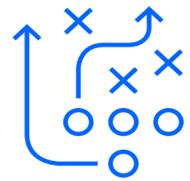
INFORM YOURSELF

Explore AI on IBM Power yourself:
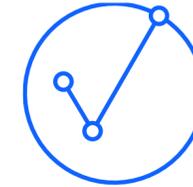ibm.com/products/power/ai

30 minutes or less

GET BRIEFED

Meet with an IBM AI expert for custom demonstrations of IBM Spyre™. Understand where AI can help boost productivity & drive growth without long pilots.

2 – 4 hours

SEE VALUE

Run a use case alignment workshop, try AI services for your use cases at the same day, and plan the enterprise integration.

1 day