RDP107

An Overview of SAP Sybase IQ and What's New in Version 16

Andrew Neugebauer / SAP TIP HANA P&D Product Management

October, 2013

Disclaimer

This presentation outlines our general product direction and should not be relied on in making a purchase decision. This presentation is not subject to your license agreement or any other agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or to develop or release any functionality mentioned in this presentation. This presentation and SAP's strategy and possible future developments are subject to change and may be changed by SAP at any time for any reason without notice. This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP intentionally or grossly negligent.

Agenda

Product Success

Market Drivers

Product Overview

Architecture Details

Summary

Product Success

SAP Sybase IQ server

Mature, industrial-strength, and analytic database-management system (DBMS)

Leadership

- Industry-leading performance and scale benchmarks
- Recognized leader in enterprise data warehouse (EDW) market by Gartner Inc. and Forrester Group
- Pioneering technology with >10 patents

Adoption

- >4,500 installations in >2,150 accounts
- >96% customer satisfaction rates – consistently

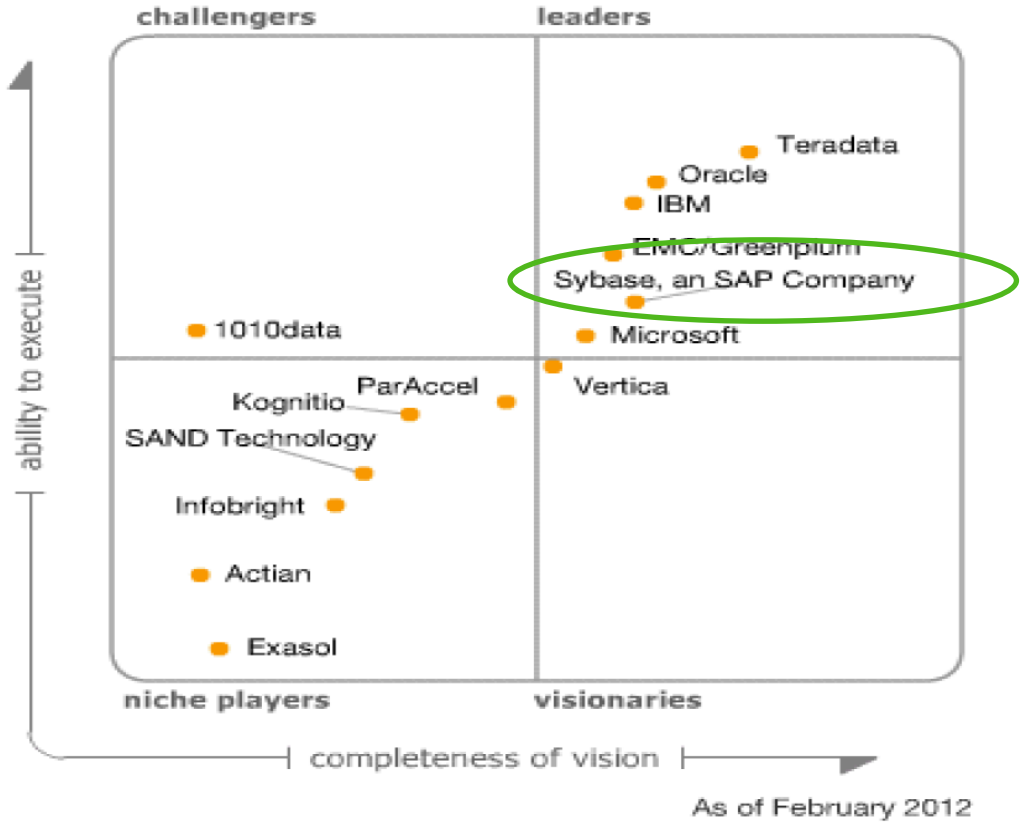
Momentum

- Double the growth rate of the data warehouse market (last four years)
- Fast-paced product releases
- Versions 15 and 15.1 in 2009, version 15.2 in 2010, and version 15.3 and 15.4 in 2011
- Version 16 in March 2013

Ericsson • Sungard • Nielsen • BNP Paribas • Telefonica • hmv.com • comScore • Agricultural Bank of China

Sybase positioned in leaders quadrant for data warehouse DBMS

Gartner 2012 Data Warehouse Database Management Systems: Magic Quadrant



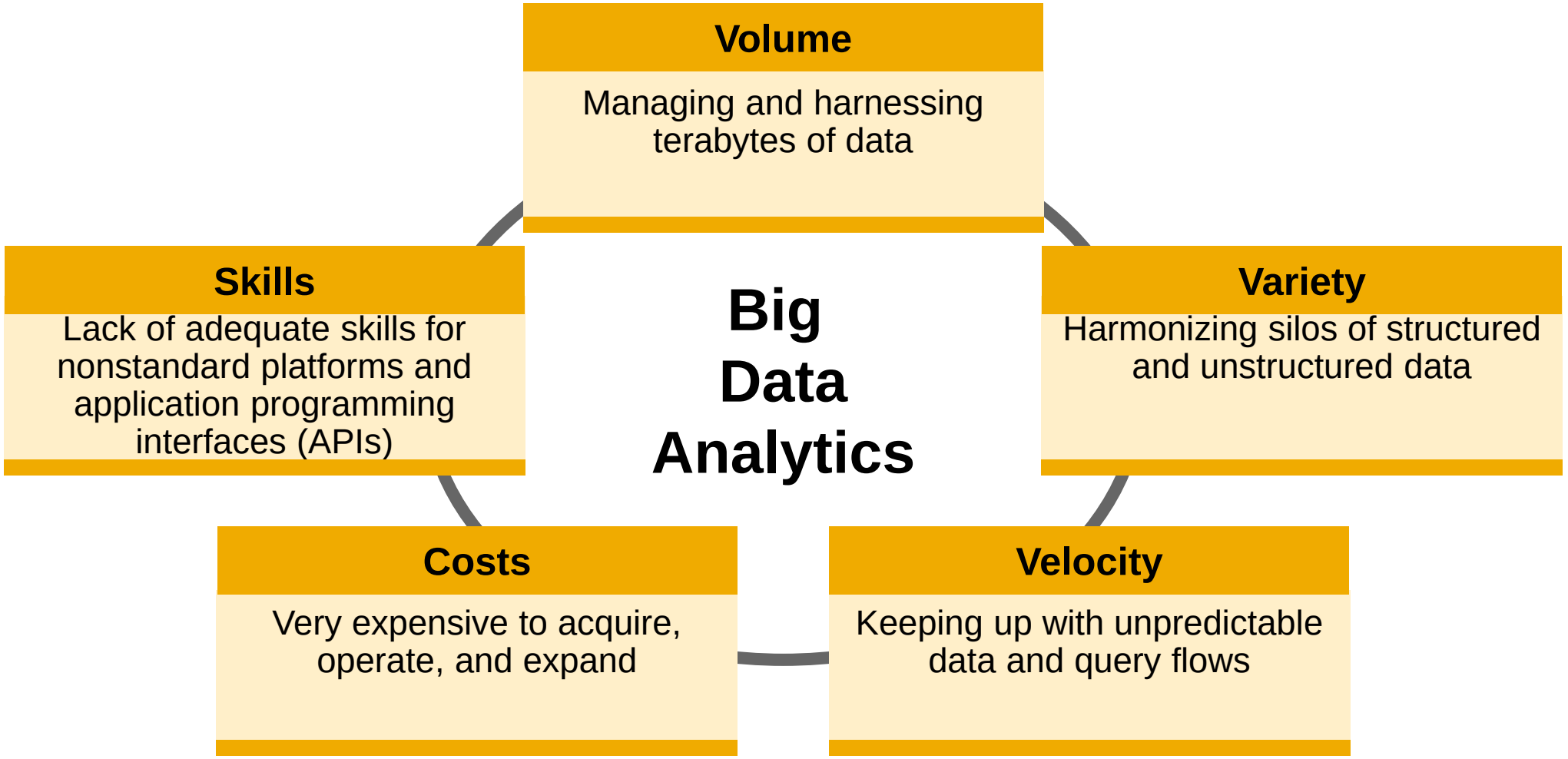
This Magic Quadrant graphic was published by Gartner Inc. as part of a larger research note and should be evaluated in the context of the entire report. The Gartner report is available upon request from Sybase.

Source: Gartner
Published: February 6, 2012

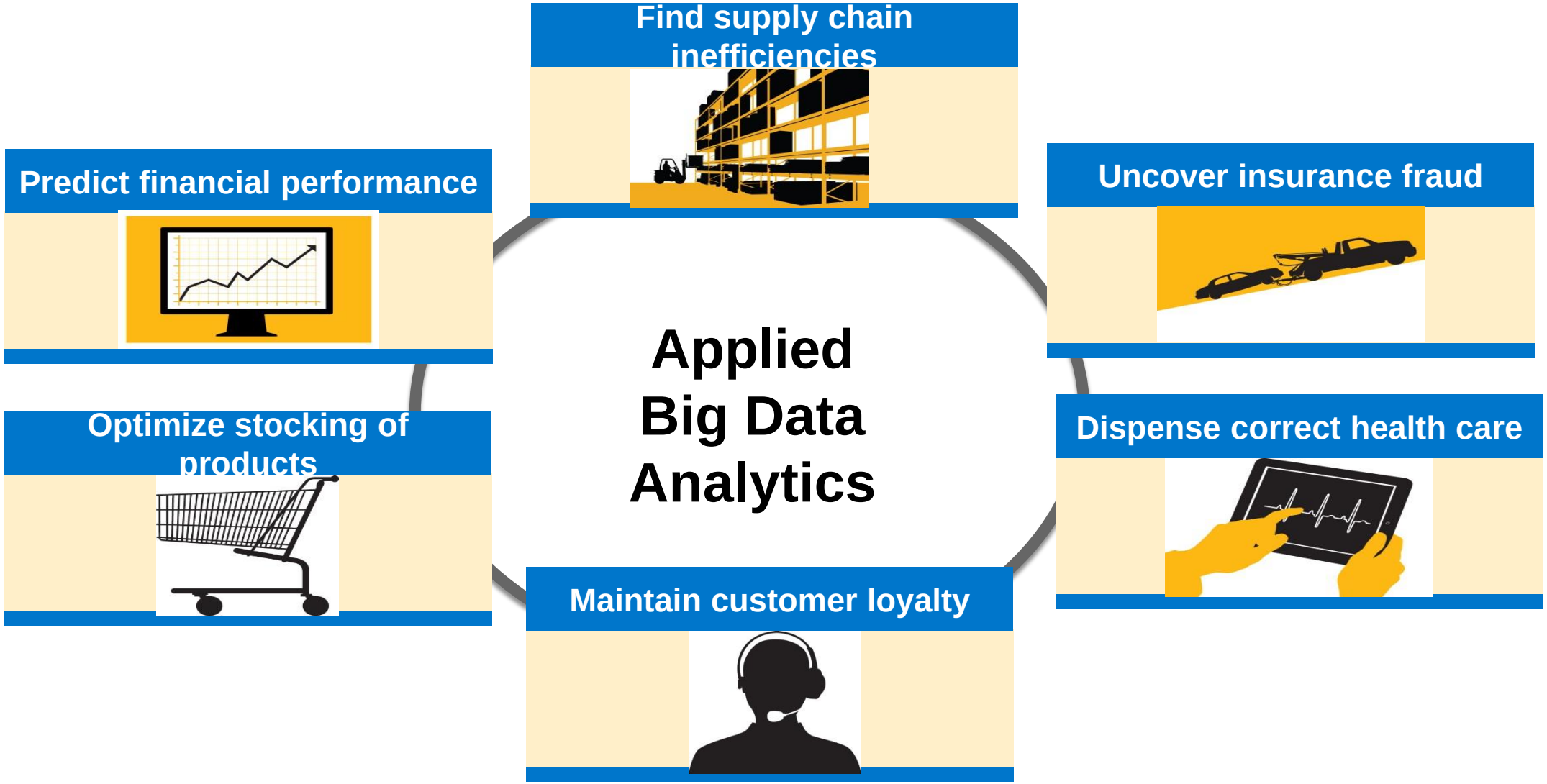
Market Drivers

Big-data analytics: issues

Dealing with volume, variety, velocity, costs, and skills



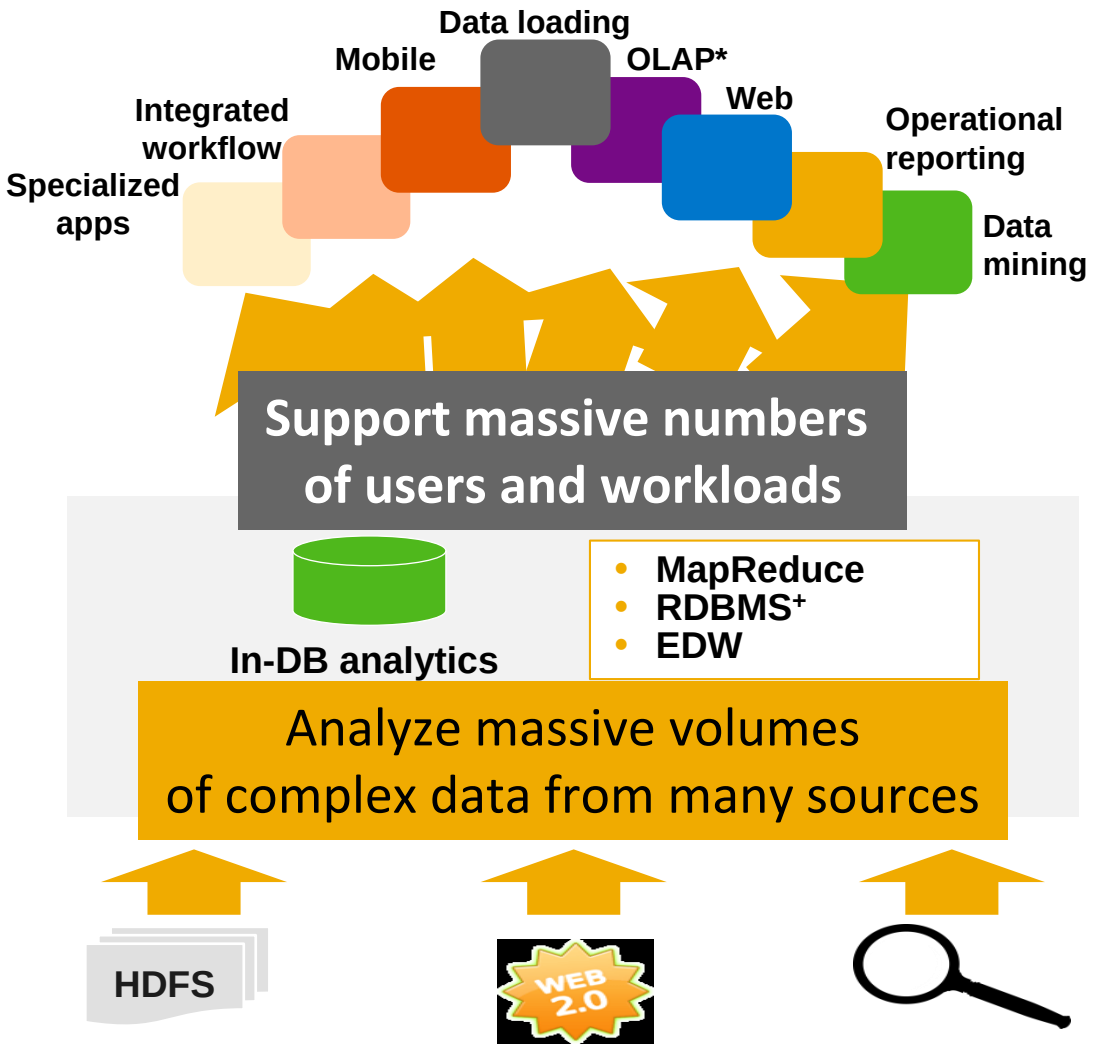
Getting Value from Big Data



Big-data analytics plus data warehousing

Deserves a new platform

- Volume
- Velocity
- Variety
- Costs
- Skills

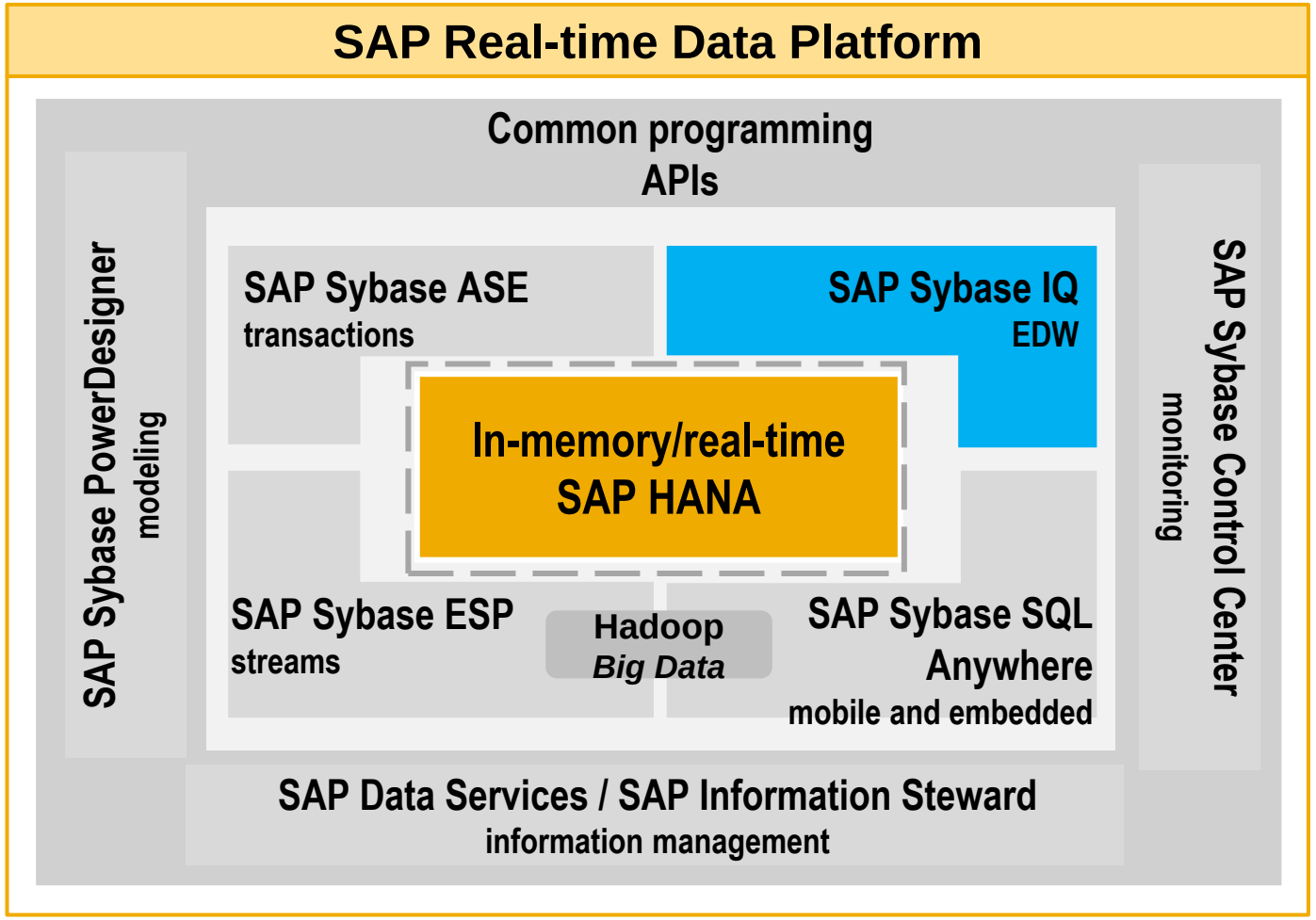


- Platform accessible to all business processes and all business users
- Requirement for data and algorithms together in the platform
- Ability to distribute interactions throughout the enterprise

*Online analytical processing
+Relational database management system

SAP Vision and Emerging Reality: SAP Real-Time Data Platform

Unified open software platform for real-time business



SAP Real-Time Data Platform foundations

- **Cross-paradigm data access** for new models of value discovery.
- **Hyper-performance** on all classes of application and usage scenarios
- **Price-Performance** value across all use cases

Benefits

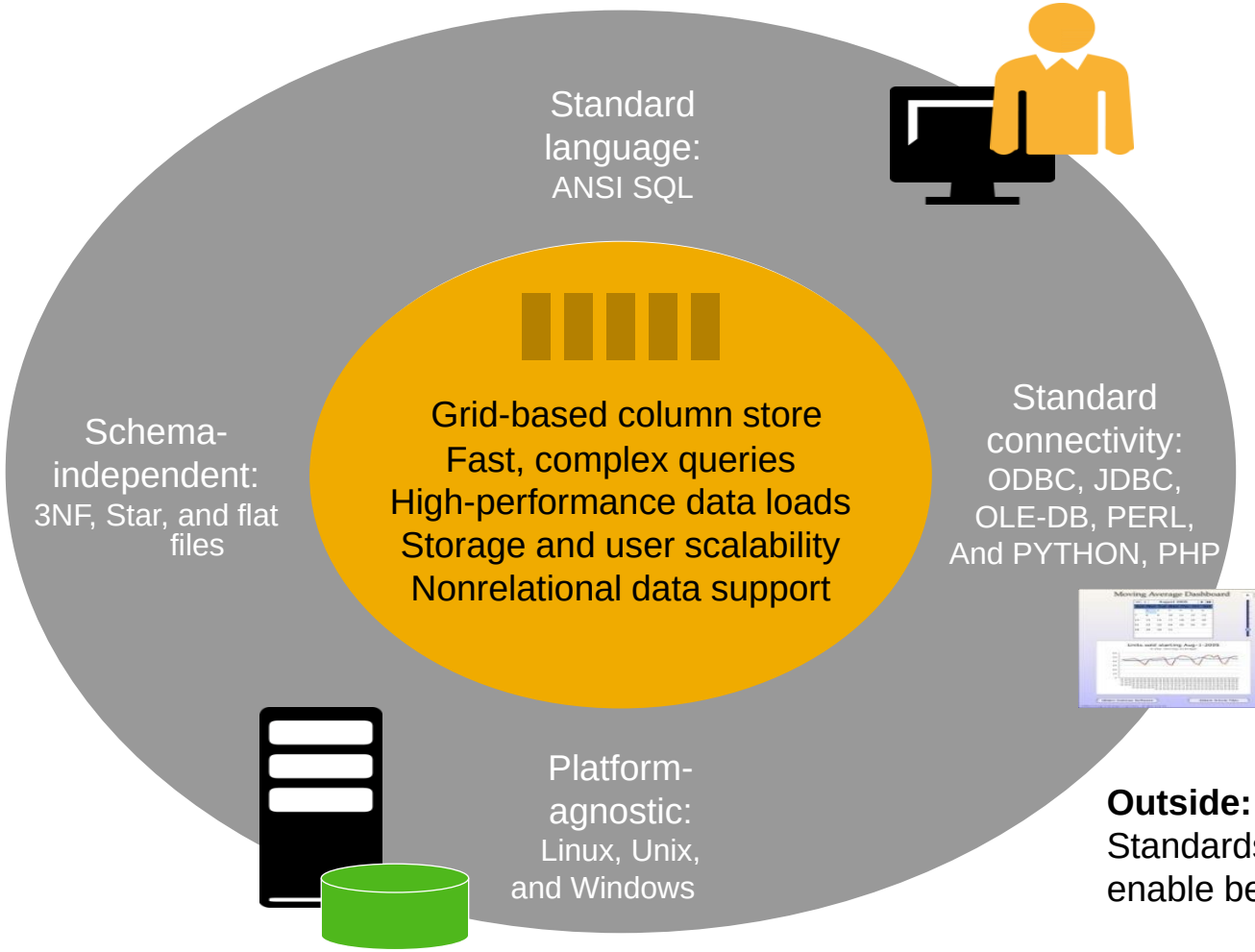
- Execute, record, analyze, and optimize without system limitations
- Embrace and extend across variations of data forms and processing models
- Common modeling, integrated development environment, shared systems management infrastructure, and deployment-independent solutions

Product Overview

What is the SAP Sybase IQ server?

High-performance analytics platform

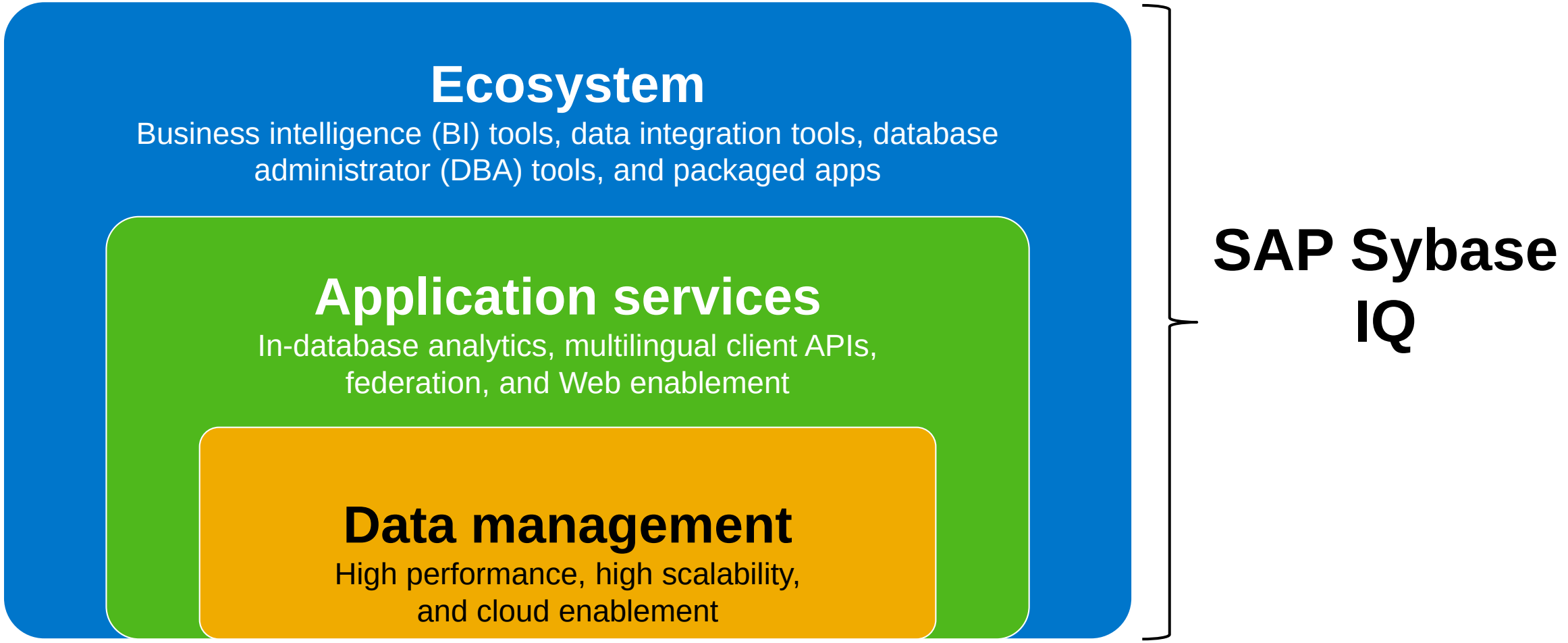
Inside:
Architected from the ground up for high-performance analytics



Outside:
Standards-based, open interfaces enable best-of-breed ecosystem

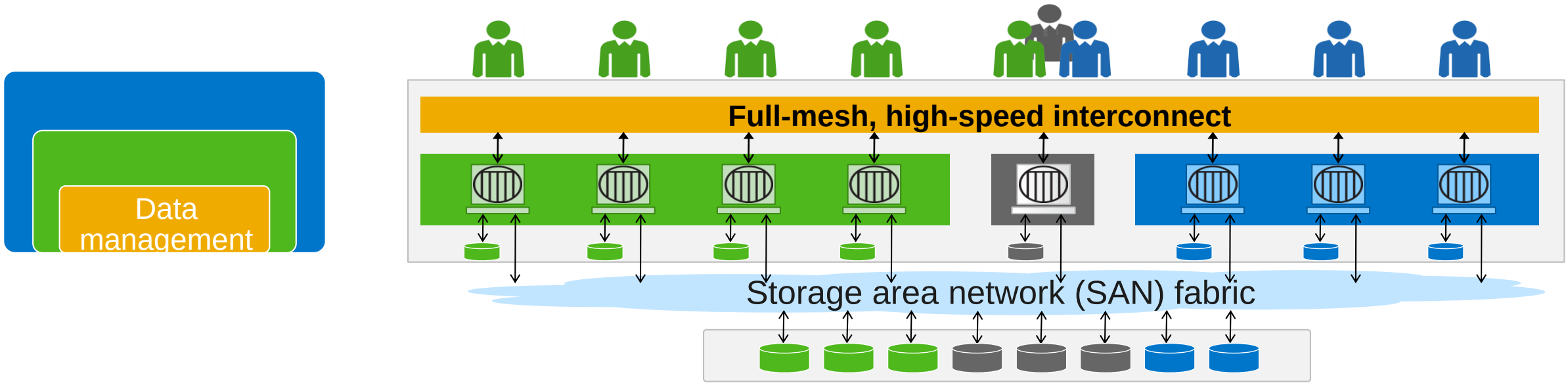
SAP Sybase IQ 16

A comprehensive, three-tier, XLDB analytics platform



SAP Sybase IQ 16

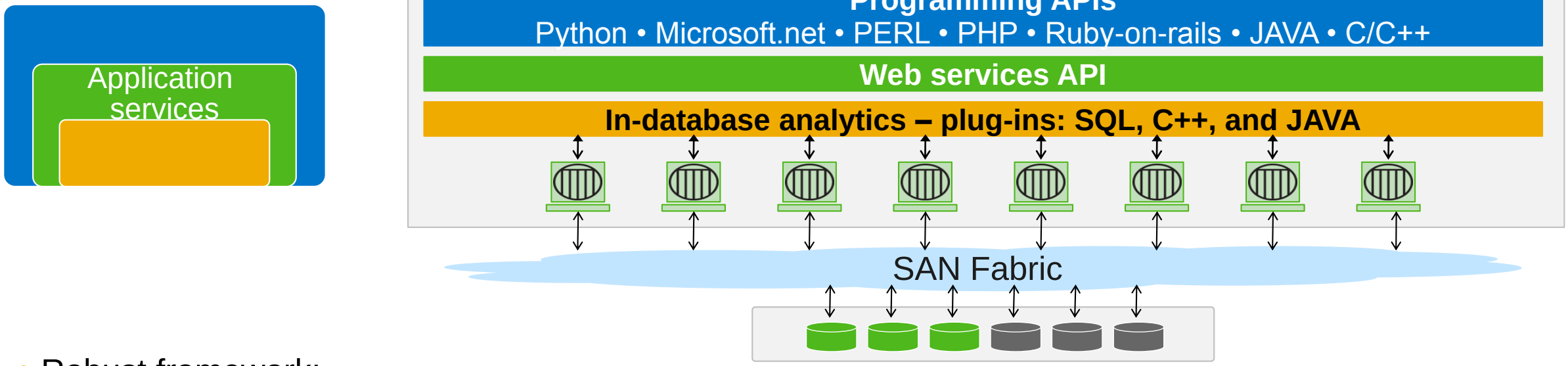
Reliable data management for XLDB analytics



- Industry-leading column store technology for speed, compression, and ad hoc analysis
- Grid framework:
  - Massively parallel processing of complex queries
  - Cloud-enabled, elastic, and virtual data marts for user communities
- Logical and physical partitioning for information lifecycle management
- Productivity-focused administration and monitoring

SAP Sybase IQ 16

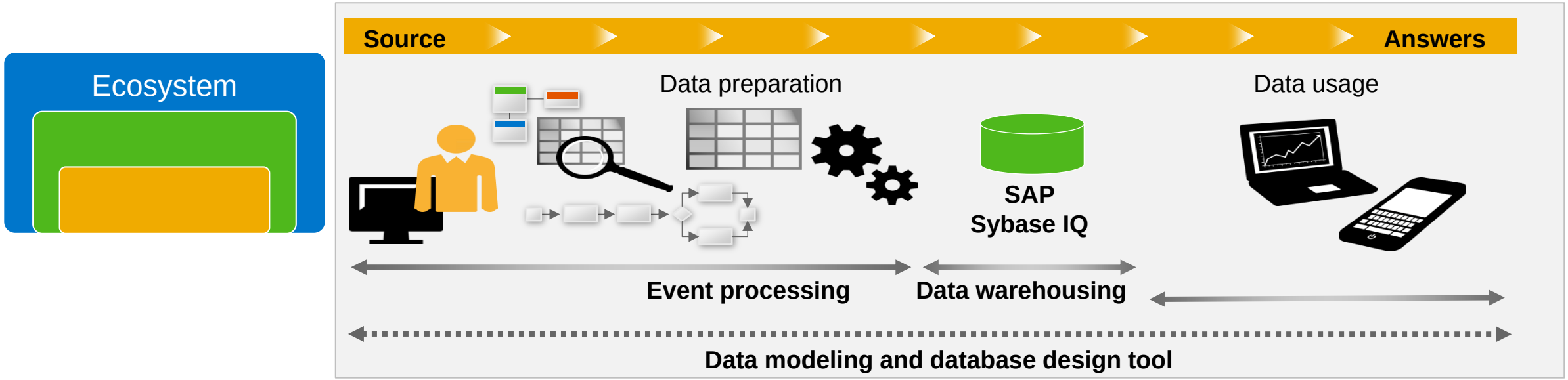
Versatile application services for XLDB analytics



- Robust framework:
  - Comprehensive ANSI SQL including ANSI 2008 specifications
  - In-database analytics – plug-ins: SQL, C++, and JAVA for data mining and statistical analysis
  - In-database Web services with SOAP API
  - Query and data federation via SQL queries
- Multilingual client APIs: C, JAVA, PHP, PERL, Python, Ruby-on-rails, and ADO.NET

SAP Sybase IQ 16

Rich ecosystem for XLDB analytics



- **Certified BI tools:** SAP BusinessObjects BI platform 4.0, Cognos, Microstrategy, and so on
- **Certified data integration tools:** SAP Data Services 4.0, Informatica, Syncsort, and so on
- **Certified data mining tools:** KXEN, SAS, SPSS, Qyte, and so on
- **Certified application tools:** ZEND, Quest, Alteryx, BMMSoft, and so on
- **Certified DBA tools:** Sybase, Bradmark, Whitesands, Symantec, EMC, and so on

Architecture Details

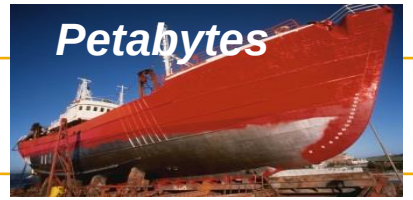
What's New in SAP Sybase IQ 16

SAP Sybase IQ 16

Innovations for extremely large databases (XLDB)

Storage Architecture

- New generation column store
- New partitioning and compression



SAP Sybase IQ XLDB Analytics

Loading Engine

- Fully parallel bulk loading
- Real-time loading into delta store

System Reliability

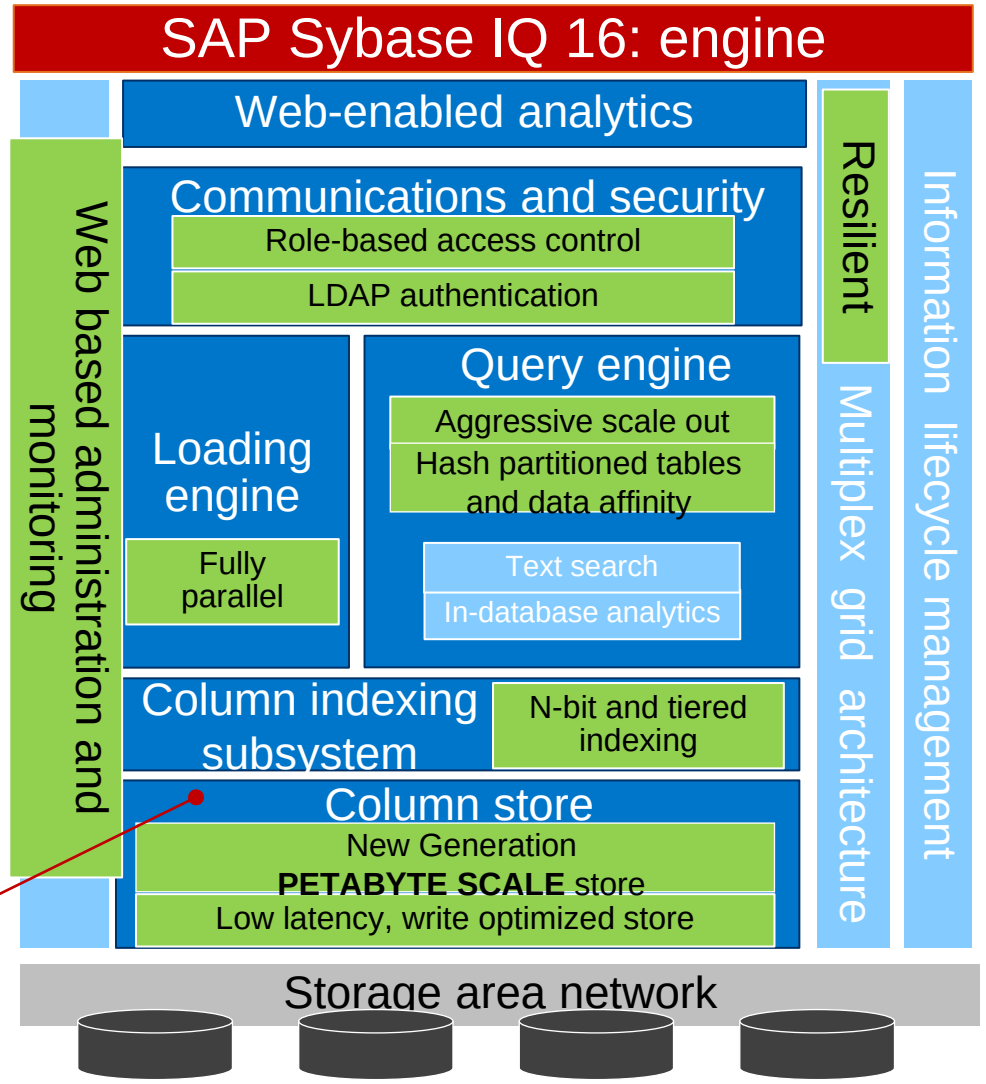
- Grid resiliency
- LDAP and role-based security

Query Processing

- Data affinity
- Aggressively parallel and distributed

SAP Sybase IQ 16

Key building blocks – **new features** highlighted in green boxes

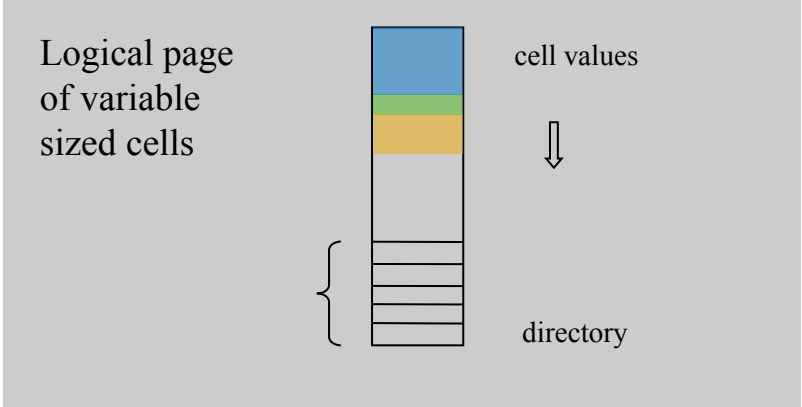


Industry-leading petabyte scale and performance with new write-optimized delta store

Column store processor

Leading data compression technology

- **New generation column store architecture:**
  - Supports a variable number of cells per page
  - Supports various page formats within a column
  - High performance access paths with richer metadata
  - Insert/update/delete of variable length data processed efficiently
  - LZW compression with implicit dictionary

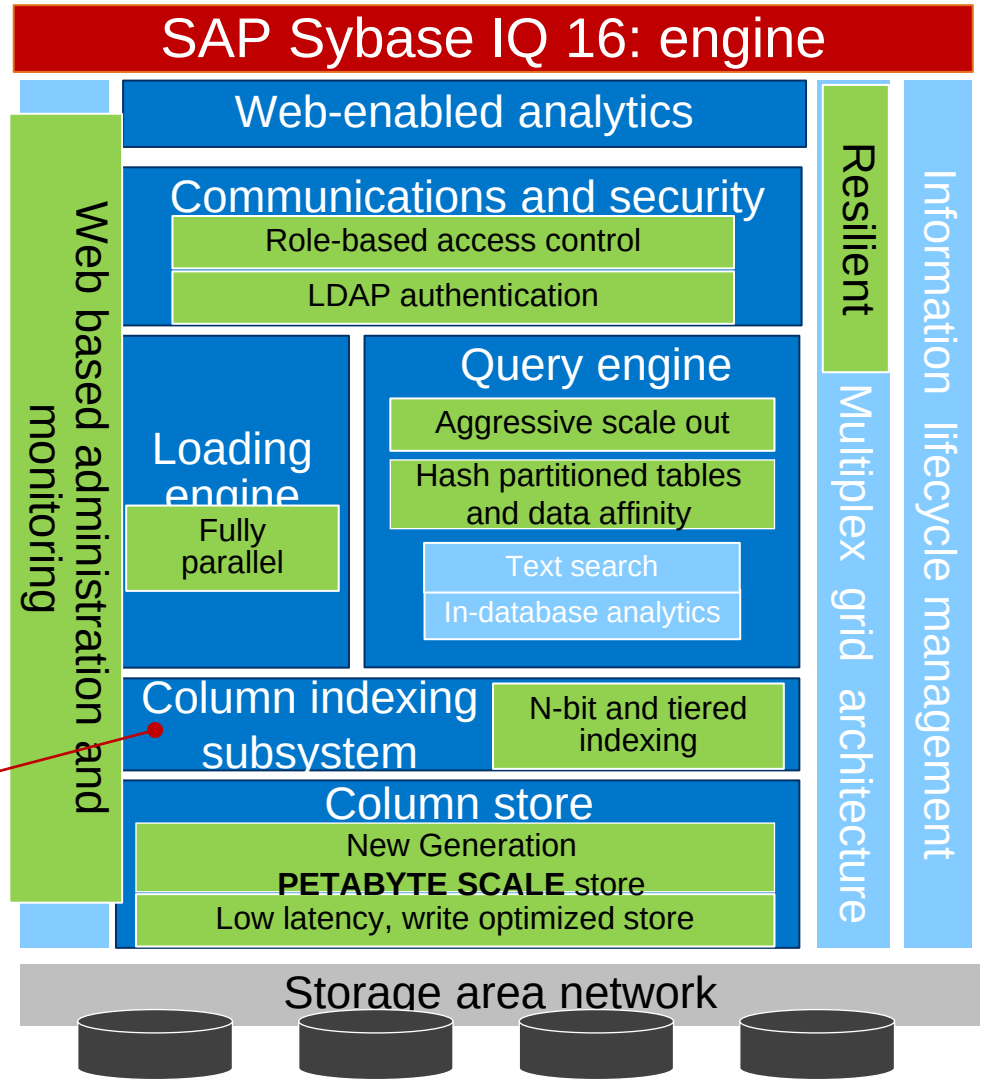


Superior compression achieved for customers!

SAP Sybase IQ: data compression in real world	Raw data loaded	SAP Sybase IQ: compressed	Data explosion in traditional RDBMS
Telefonica	70 TB	15 TB	210 TB to 490 TB
Health insurance review agency	27 TB	12 TB	81 TB to 189 TB
Samsung card	15 TB	7 TB	45 TB to 105 TB

SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes



Patented column indexing technology

Column indexing subsystem

Powerful performance booster

Type	Usage
Fast projection	N-bit compressed raw data – data is stored as an index
Low fast	Low cardinality data (up to 1,000 unique values)
High nongroup	Aggregation on the fly and range searches
High group	Key fields and groupings for high cardinality data
Date, time, and DT	Date ranges and date-part operations
Multicolumn	Indexes for multi-column keys
Word and text	Used for sophisticated key word or phrase string searches with Boolean, ranking, proximity, and fuzzy features
Compare	Column comparisons

Key characteristics:

- Data immediately indexed as Fast Projection (FP) when it is loaded
- Most columns will have at least one index
- Index selection decisions are based on column cardinality (number of unique values)
- Indexes and columns are stored separately
- Multiple indexes used to resolve a query
- Indexes are self-maintaining
- No optimizer statistics to update
- Indexes are compressed
- Index advisor demystifies index selection

Column indexing subsystem

N-bit dictionary compression and bit-mapped secondary indexes

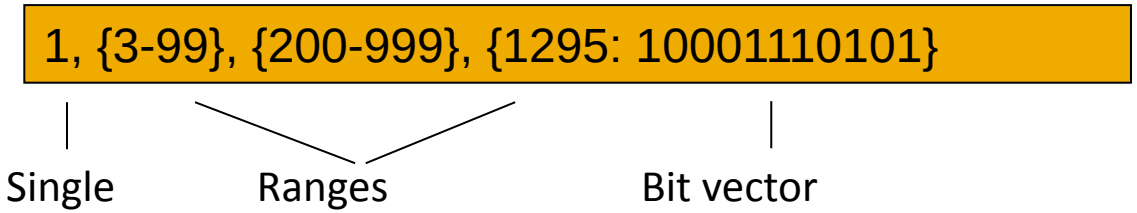
- **N-bit fast projection indexes:**
  - Bit-level instead of 1, 2, and 3-byte FPs
  - Different data pages for same column can have different values of “N” for N-bit
  - No more requirement to rollover FP format for all column data
  - Options provided to set threshold for rollover to flat (to prevent large dictionaries)
  - Options provided to prevent rollover to flat (to prevent long rollover time)

Raw Data = 400 MB; 1 Billion 4-byte integer values fn (N)

N	Token Size	Savings
2	(1B * 2) / 8 = 250MB	93.75%
3	(1B * 3) / 8 = 375MB	90.6%
4	(1B * 4) / 8 = 500MB	87.5%
	…….	
24	(1B * 24) / 8 = 3000MB	25%

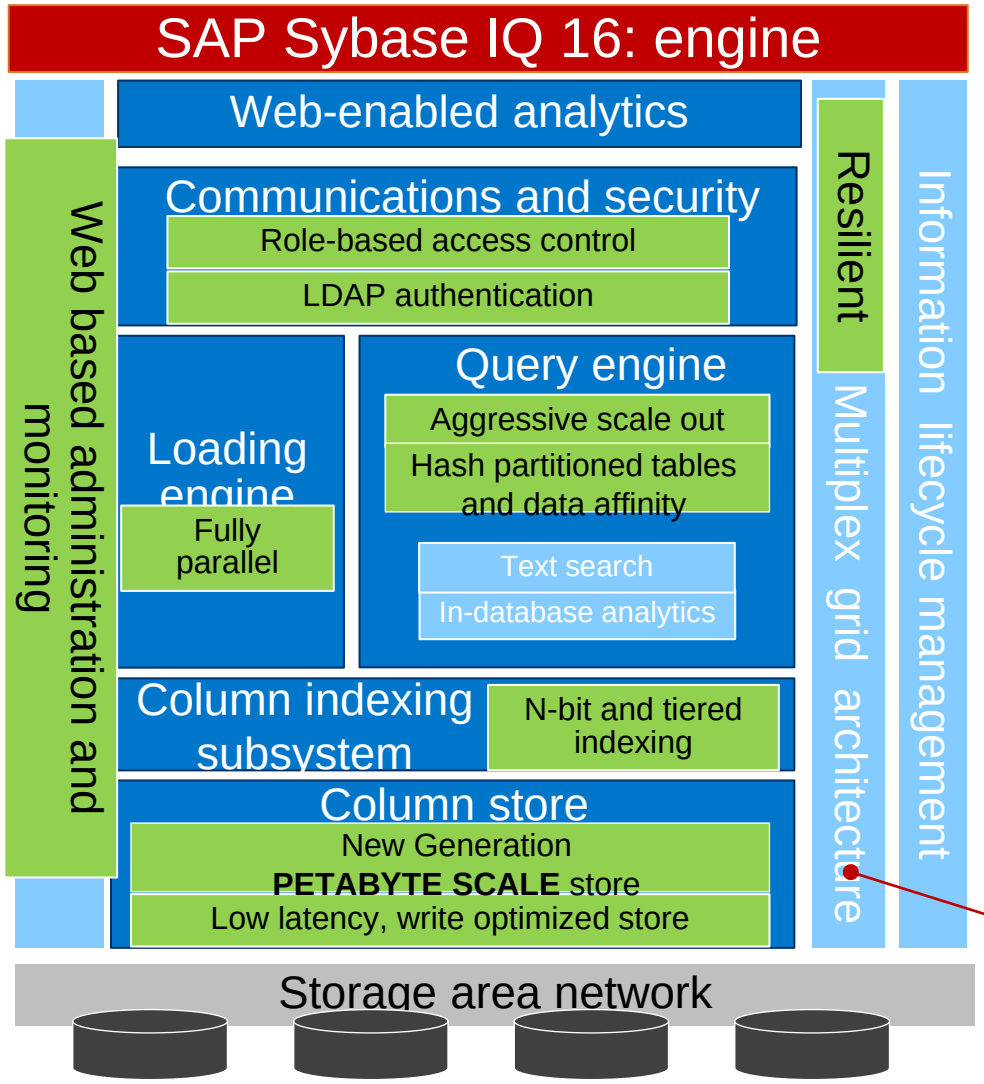
```
2->3
3->4
4->5
5->6
6->8
8->10
10->12
12->16
16->21
21->24
```

- **Bitmapped secondary indexes:**
  - ROWIDs stored in multiple formats
  - Format based on locality of ROWIDs



SAP Sybase IQ 16

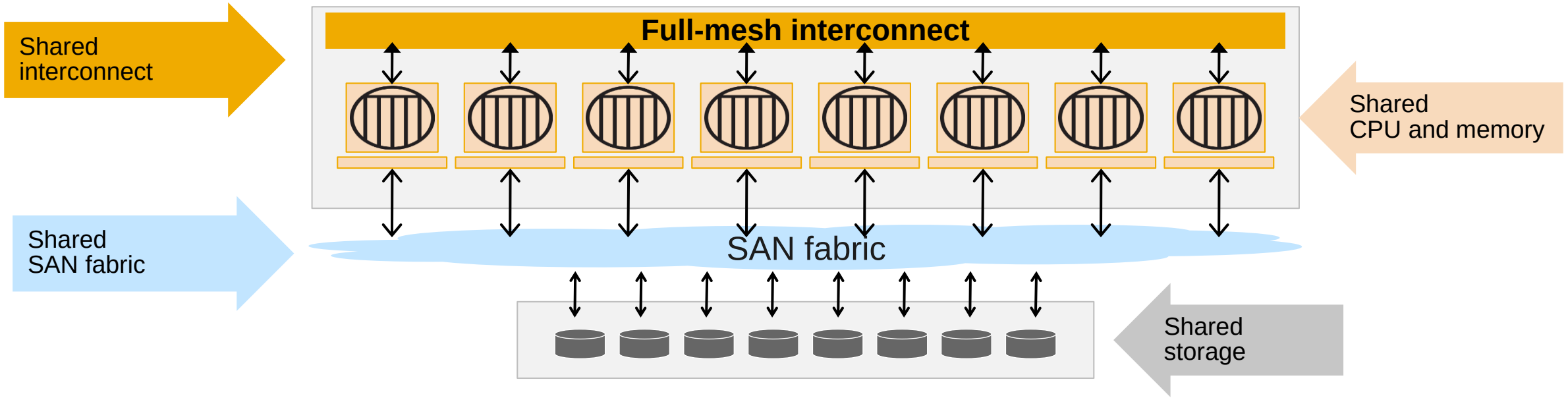
Key building blocks – new features highlighted in green boxes



Industry-leading , scalable Multiplex grid for concurrent and massively parallel processing (MPP) analytics

Multiplex grid

Architecture overview



Compute nodes:

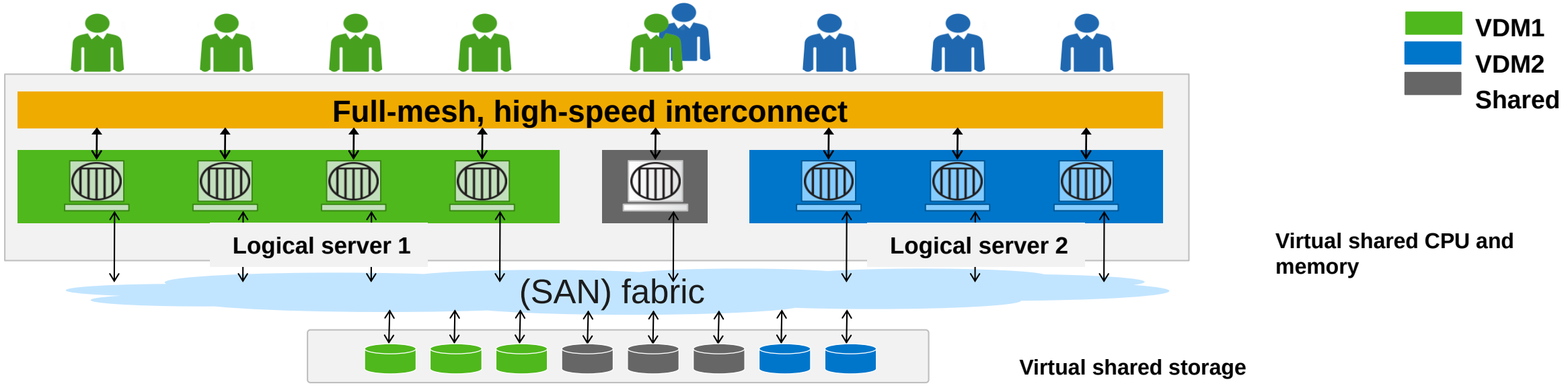
- Can be x86, IBM Power, HP Itanium, and Oracle Solaris Sparc servers
- Each node can have a different CPU core count, memory size
- One node is designated as the coordinator for meta data management
- Any node can be a write node
- Any node can be a read node

Shared storage:

- Can be SSD, FC, SATA, or a combination of all
- Shared storage can be physical SAN with FC or iSCSI
- Shared storage can be virtual SAN over DAS
- Storage can be logically grouped with compute
- Storage can be tiered for data aging

Multiplex grid

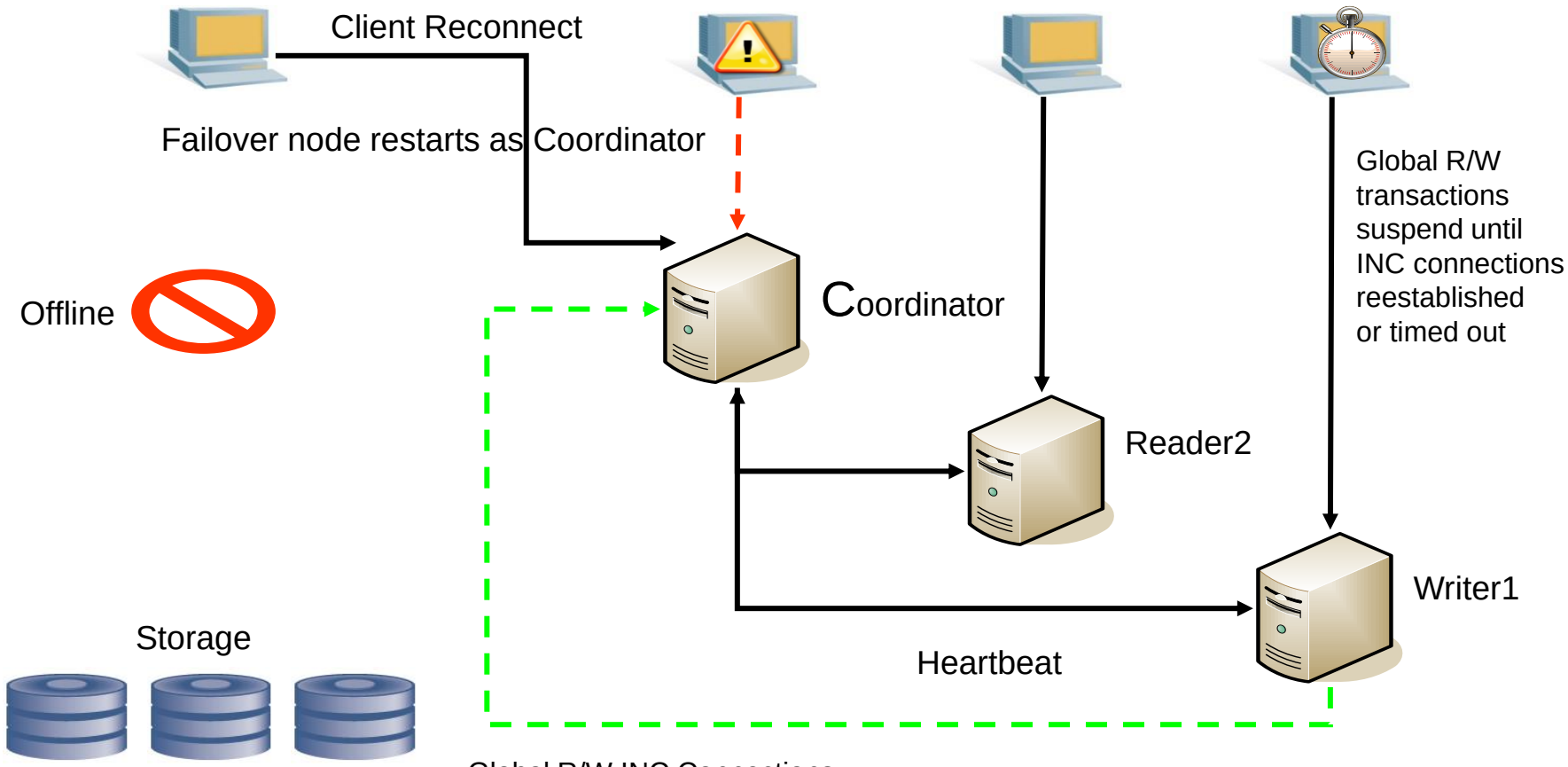
Architecture usage: elastic virtual data marts



- **Virtual data marts (VDM):**
- VDM is logical binding of mutually exclusive nodes, memory, and storage:
  - Logical server (LS) is a mutually exclusive logical binding of nodes and memory
  - Logical server (LS) is a subset of VDM
  - Bindings are elastic – they can dynamically grow and shrink
  - Automatic login redirection and load balancing within a logical server
- VDM works via login permission control
- VDM can isolate applications, workload, user communities
- Distributed query processing (DQP) within VDM boundaries only

Multiplex grid

Global transaction resiliency

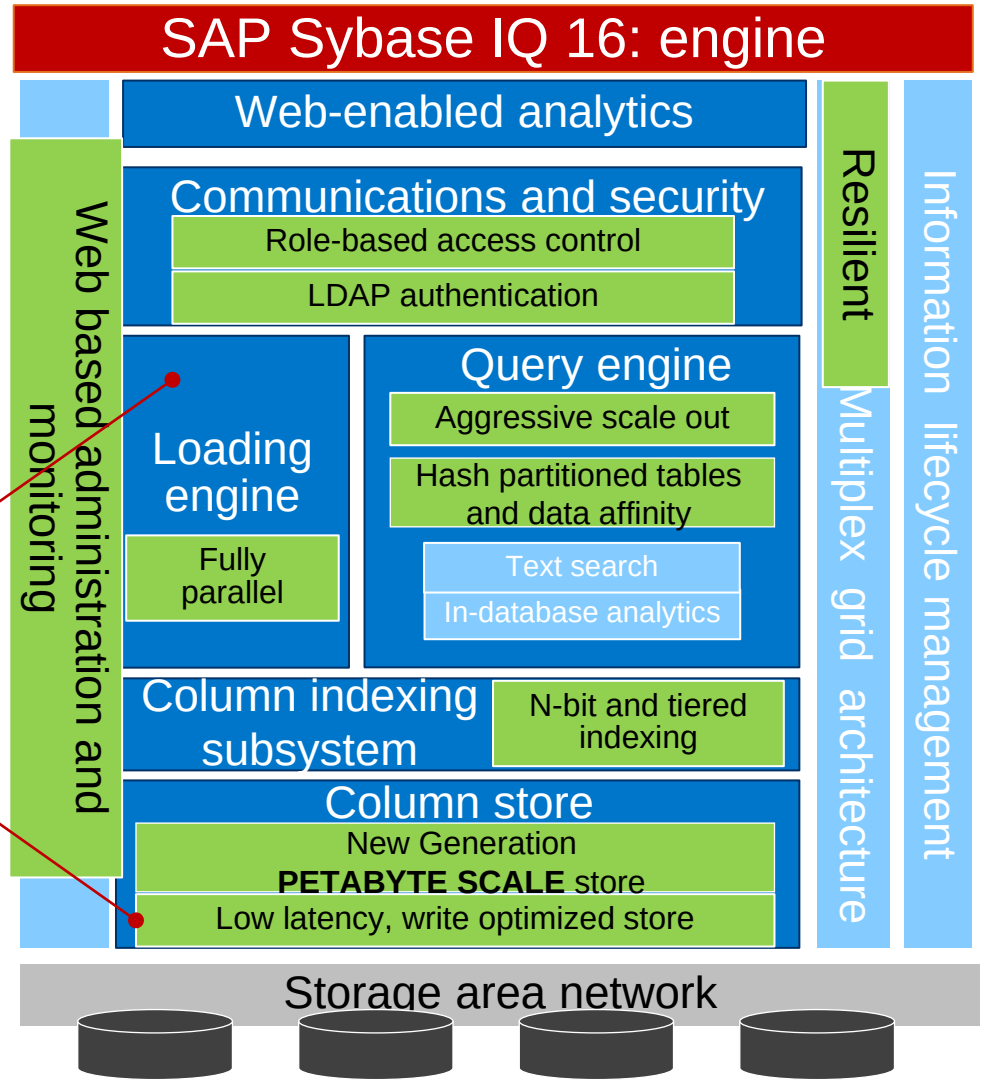


Global R/W INC Connections
automatically reconnect and resume transaction
after a coordinator failover.
In most cases long-running loads will transparently resume.

SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes

Fully parallel bulk, incremental batch, and low-latency concurrent loading



Robust bulk load engine

Loading can be from multiple nodes:

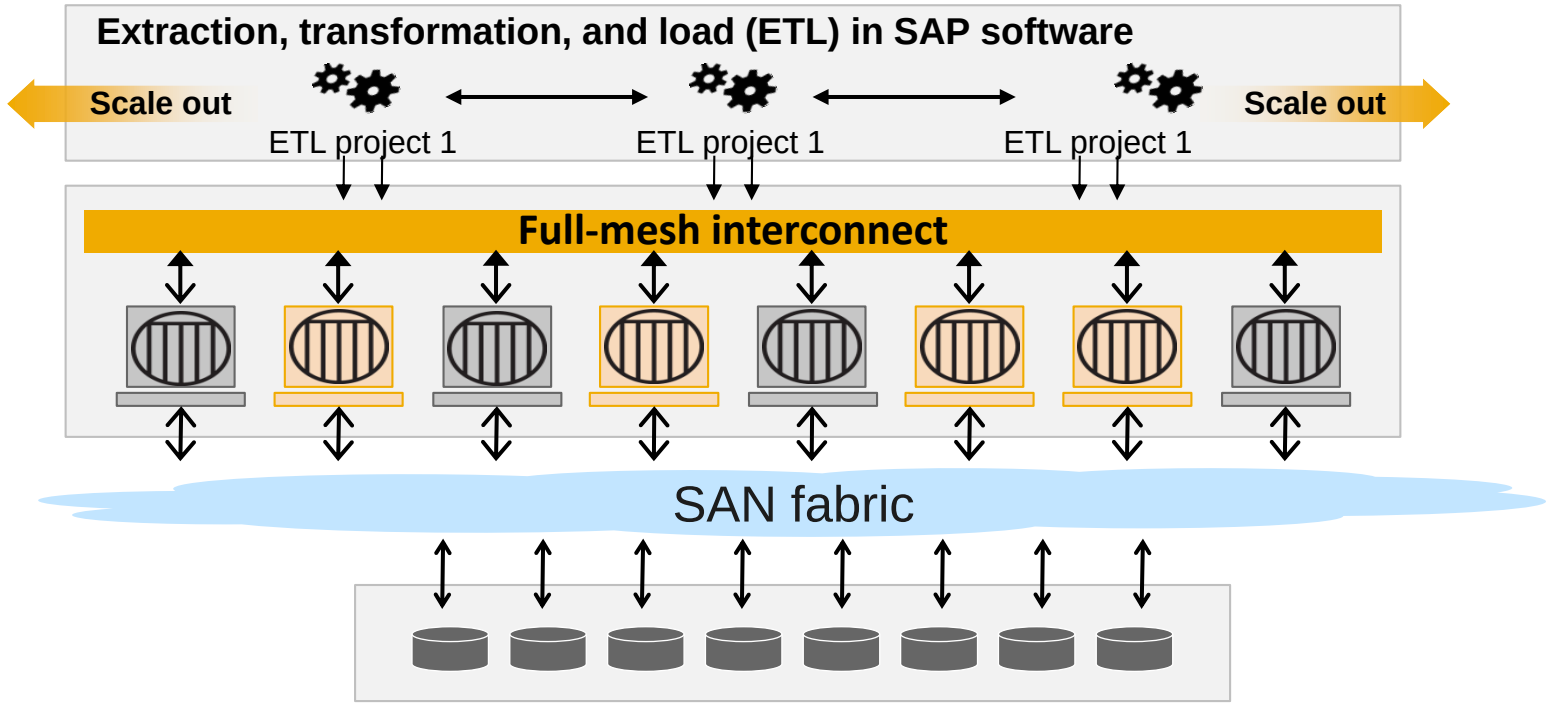
- Load rates in excess of 250 GB/hour are common even with modest-sized hardware nodes
- Incremental loads via microbatching (change data capture)

Page-level snapshot versioning:

- No locking (only table-lock in-memory catalogs) – allows nonblocking concurrent loads and queries

Load from client machines:

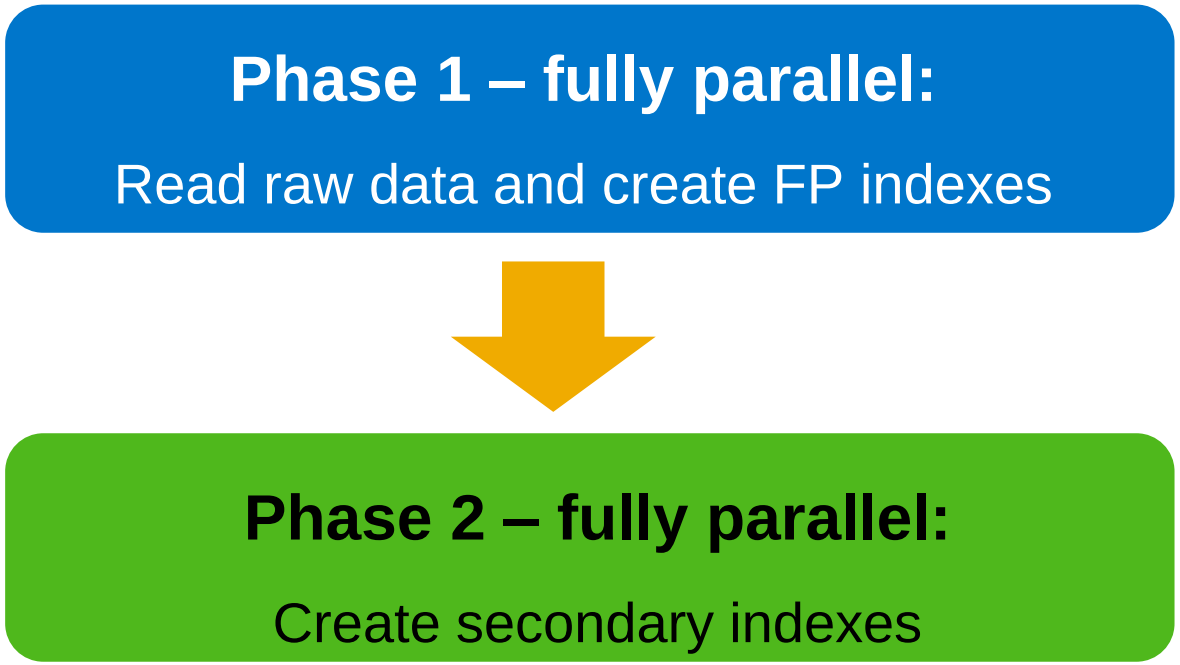
- Eliminates requirement for load files to reside on database server machine



Bulk loading

Fully parallel to maximize use of available processing cores

- Bulk loading commands are fully parallel:
  - LOAD TABLE
  - INSERT…LOCATION
  - INSERT…SELECT
- Load an index/column concurrently with multiple threads
- Remove all bottlenecks which contribute to inefficient use of CPU and storage
- Dynamically scale up and down degree of parallelism based on workload
- New graphical loading plans for diagnostics



Incremental batch loads

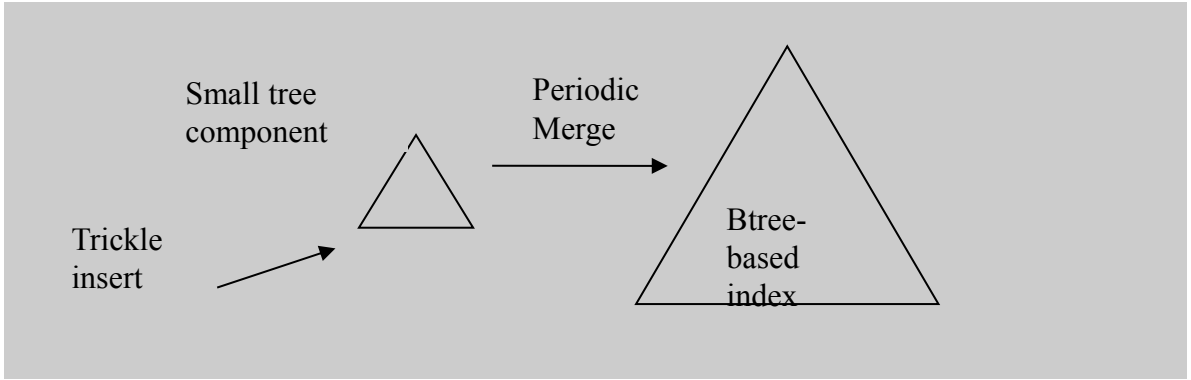
Fast, predictable performance with Tiered High Group (THG) indexes

Value proposition

- Improved performance of frequent, small batch loads
- Predictable performance of small batch loads:
- performance is proportional to the size of the data being loaded, not the table being updated

Architectural considerations

- Inserting into a large High Group (HG) b-tree index is costly
- HG index may have a tiered structure with a small tree component and a large tree component
- Small, batch loads into the HG index are written to the small tree component quickly and synchronously
- The small tree component is periodically merged into the large tree component as a background task



High velocity data loading

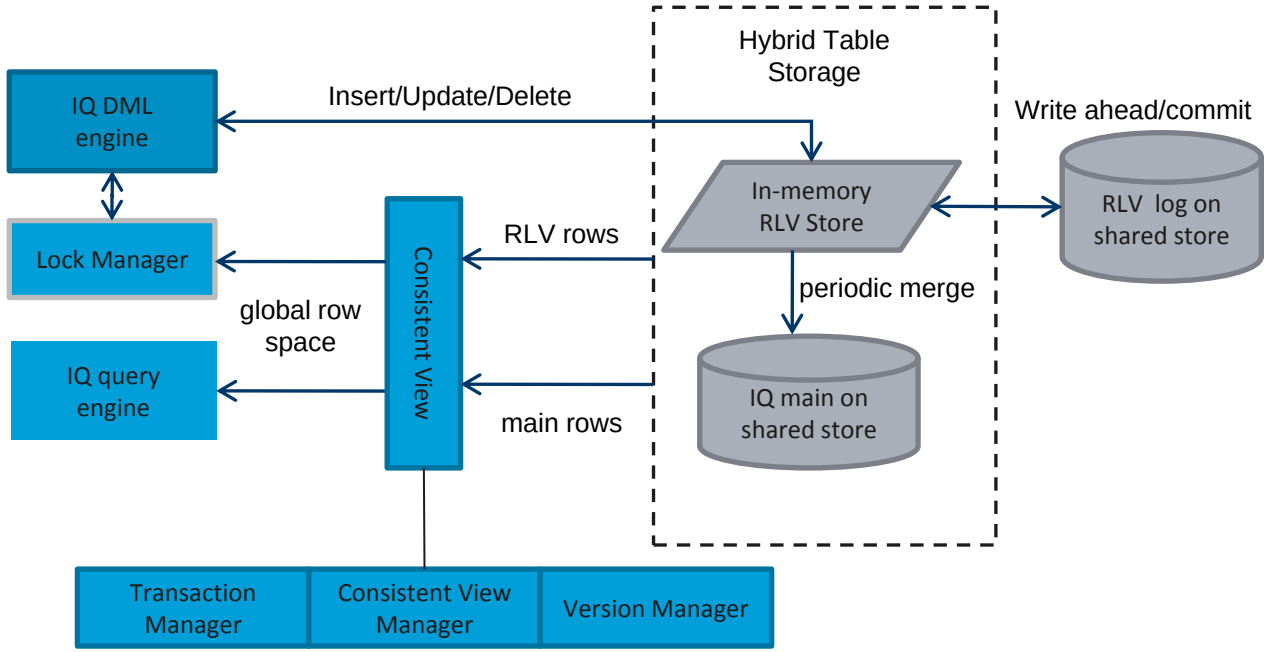
Row-level versioned (RLV) store

Value proposition

- Continuous analytics over operational data
- High velocity, concurrent data modifications
- Exploit large memory and core footprints

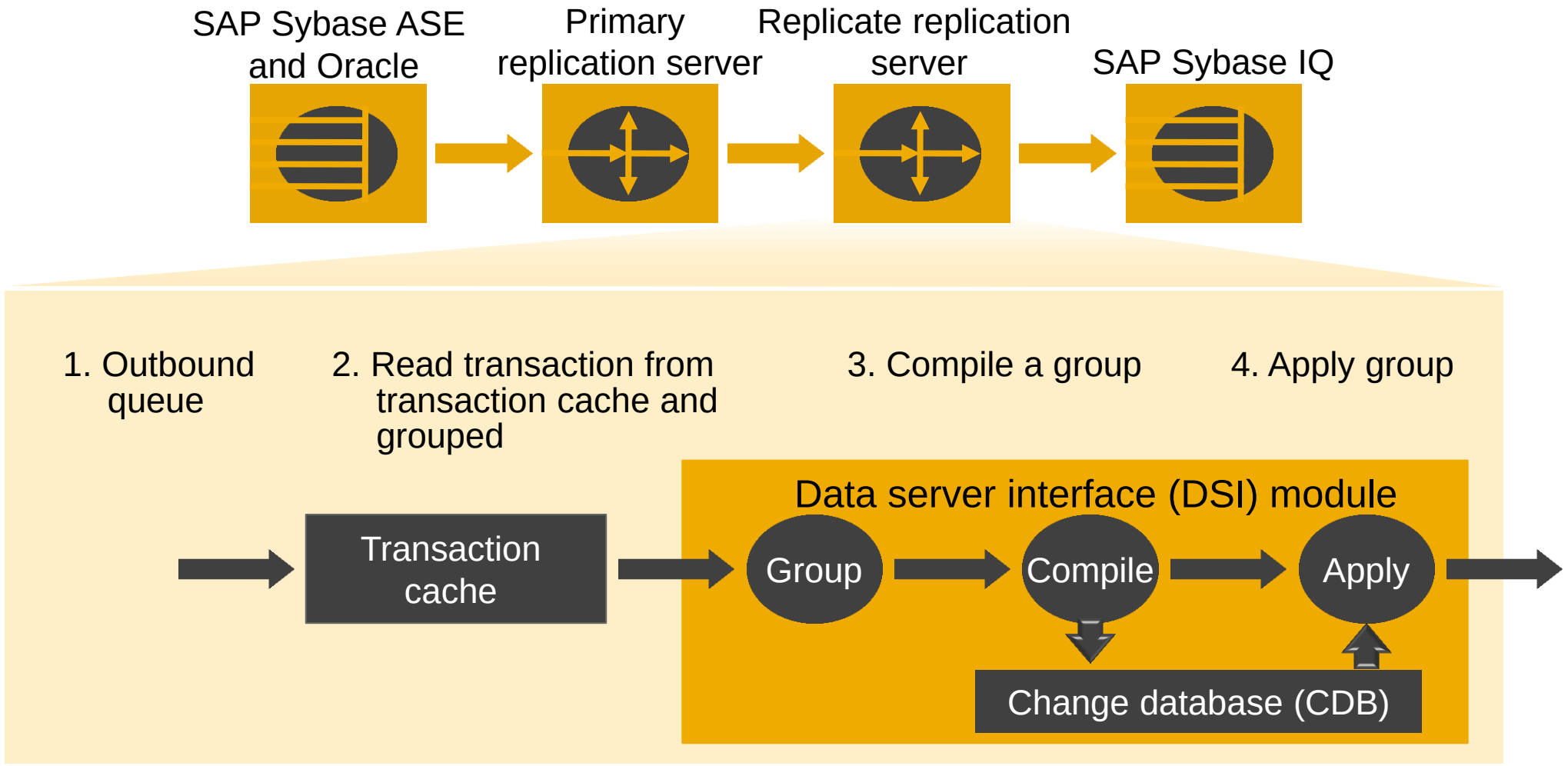
Architectural considerations

- Write optimized in-memory In-memory RLV (Row-level versioned) store
- Row level locking, and statement snapshot isolation
- Low latency micro operations
- In-memory RLV store has reduced compression, no sorting, no indexing
- Fully recoverable with dedicated transaction log
- Asynchronous data transfer from In-memory RLV store to IQ main store
- Users choose which tables are In-memory RLV tables
- Simplex only (not available on Multiplex)



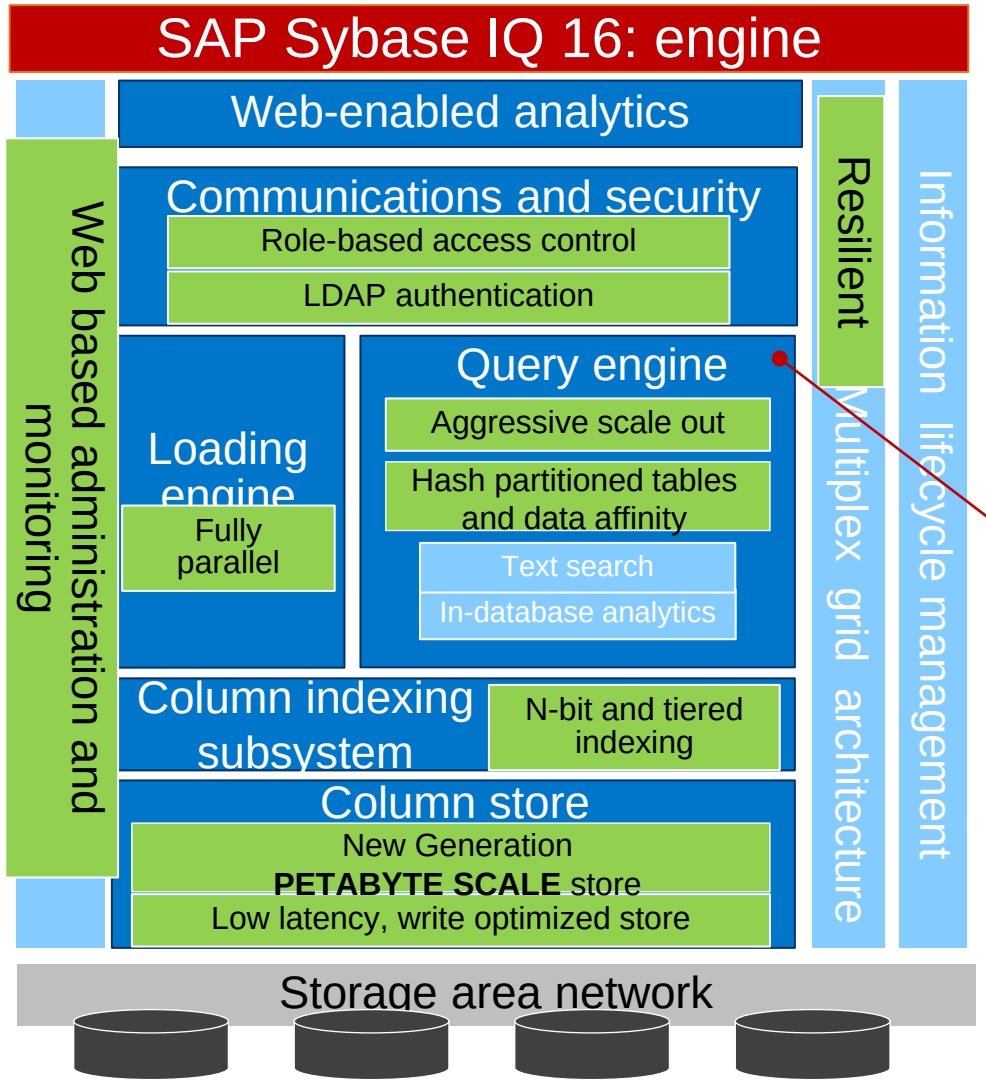
Continuous real-time loads

Using SAP Sybase Replication Server



SAP Sybase IQ 16

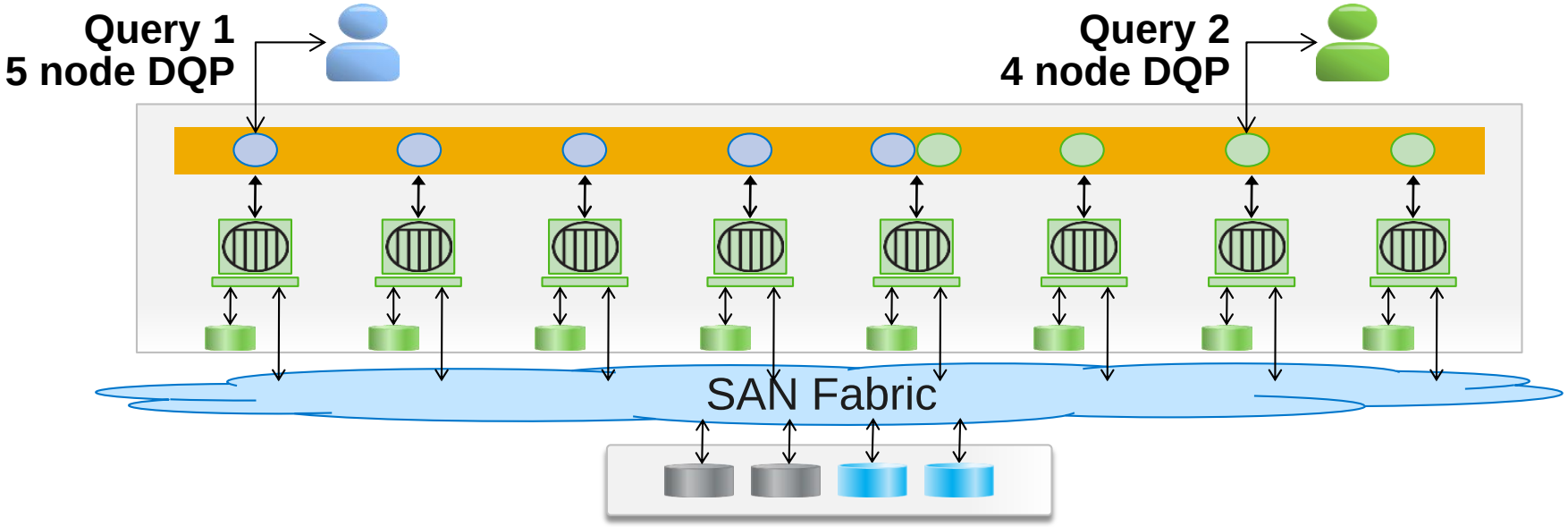
Key building blocks – new features highlighted in green boxes



Intelligent query engine with data affinity for “shared nothing” performance on a flexible “shared everything” architecture

Query engine scale out

Distributed query processing



Massively parallel processing

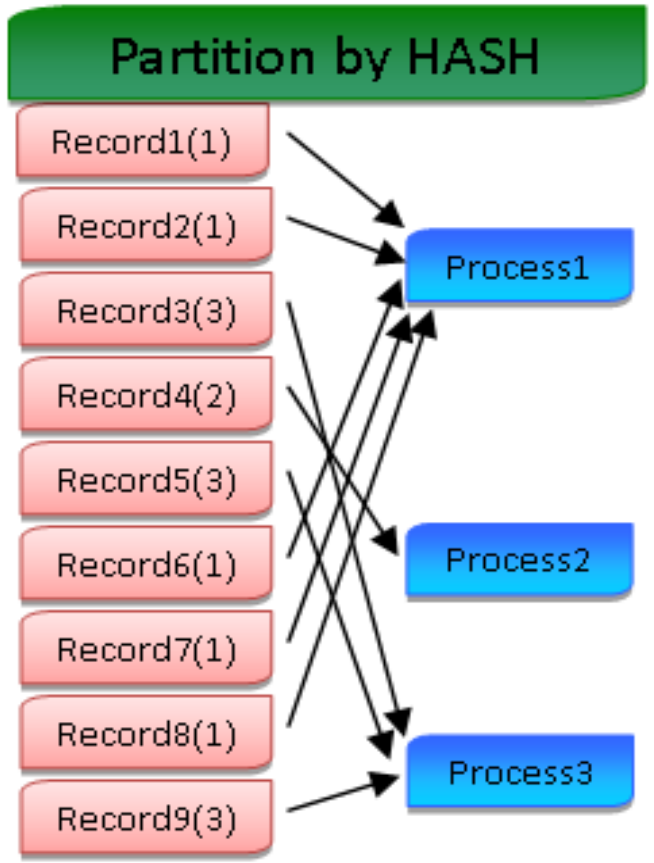
- **Leader node:** Receives and initiates queries, including user-defined functions (UDFs)
  - Any node can be a leader, one leader per query, many concurrent leaders possible
  - Leader node may satisfy query within itself
- **Worker node:** Nodes picking up work units from leader
  - Many worker nodes per query, same worker node can serve multiple queries
  - Worker nodes are enlisted only if leader cannot satisfy query on its own

Query engine scalability

Hash partitioned tables for query scale up and scale out

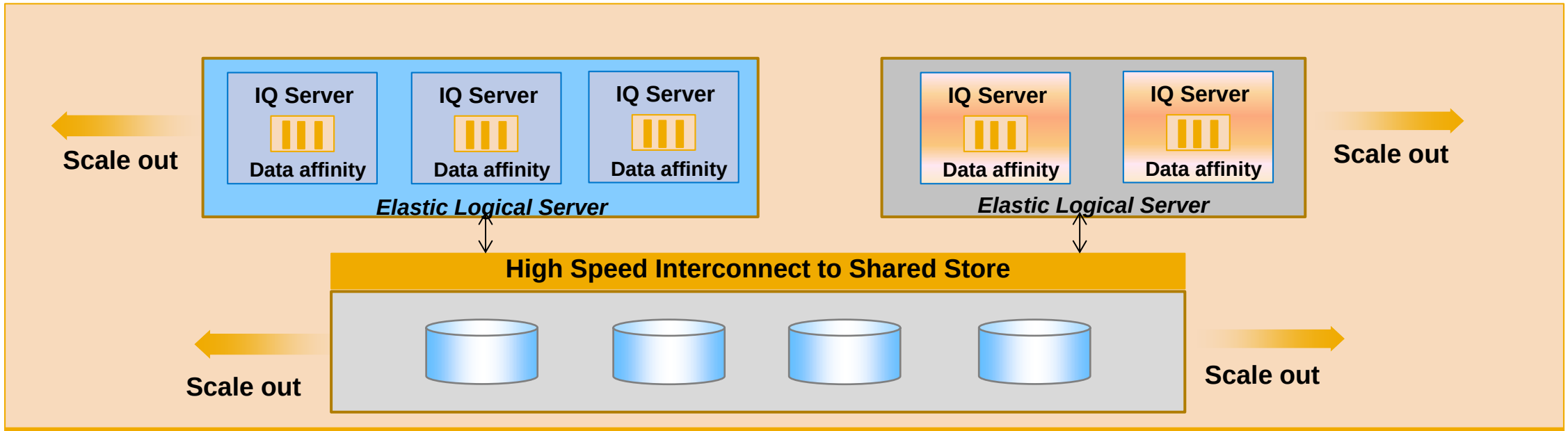
Hash partitioned tables support scalable query performance

- Data is automatically partitioned during loading with built-in hash algorithms
- Data is divided into persistent subsets
- Reduces results sharing
- More efficient CPU usage
- Reduces instantaneous temp usage
- Optimizer will use hash partitions for join and group by when available



Query engine scale out

Data affinity for “shared nothing” performance

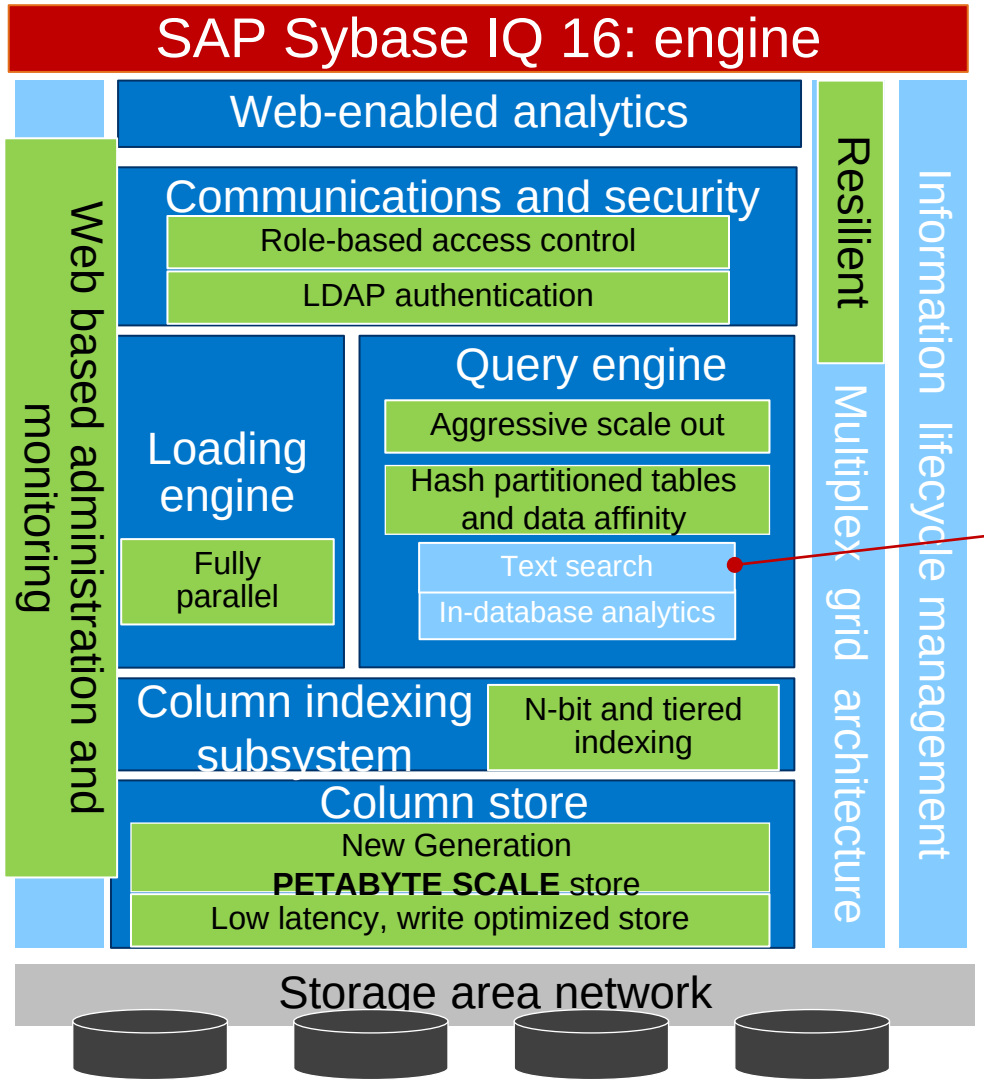


Data affinity

- Query optimizer tracks which data is in which server cache
- Distributed query work assignments are intelligently assigned based on which server has which data set in cache
- Data must be hash partitioned
- Each partition is automatically assigned to a specific node as queries execute
- Data affinity is self managed – no user interaction required
- Caches stay “hot” and I/O is reduced

SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes



Comprehensive, built-in full-text search

Text search and analysis

Enable analytics on textual data and structured relational data

- Text index — SQL based on terms and phrases, prefix, proximity, and scoring
- Interface to plug in third-party document converters or term breakers
- Text analytics library in SAP Data Services
- Third-party text analytics library

Relational data

+

Nonrelational text data

SAP SYBASE IQ

Combined analytics with in-database text indexing

Optional third-party prefilter, text segmentation, or entity extractors

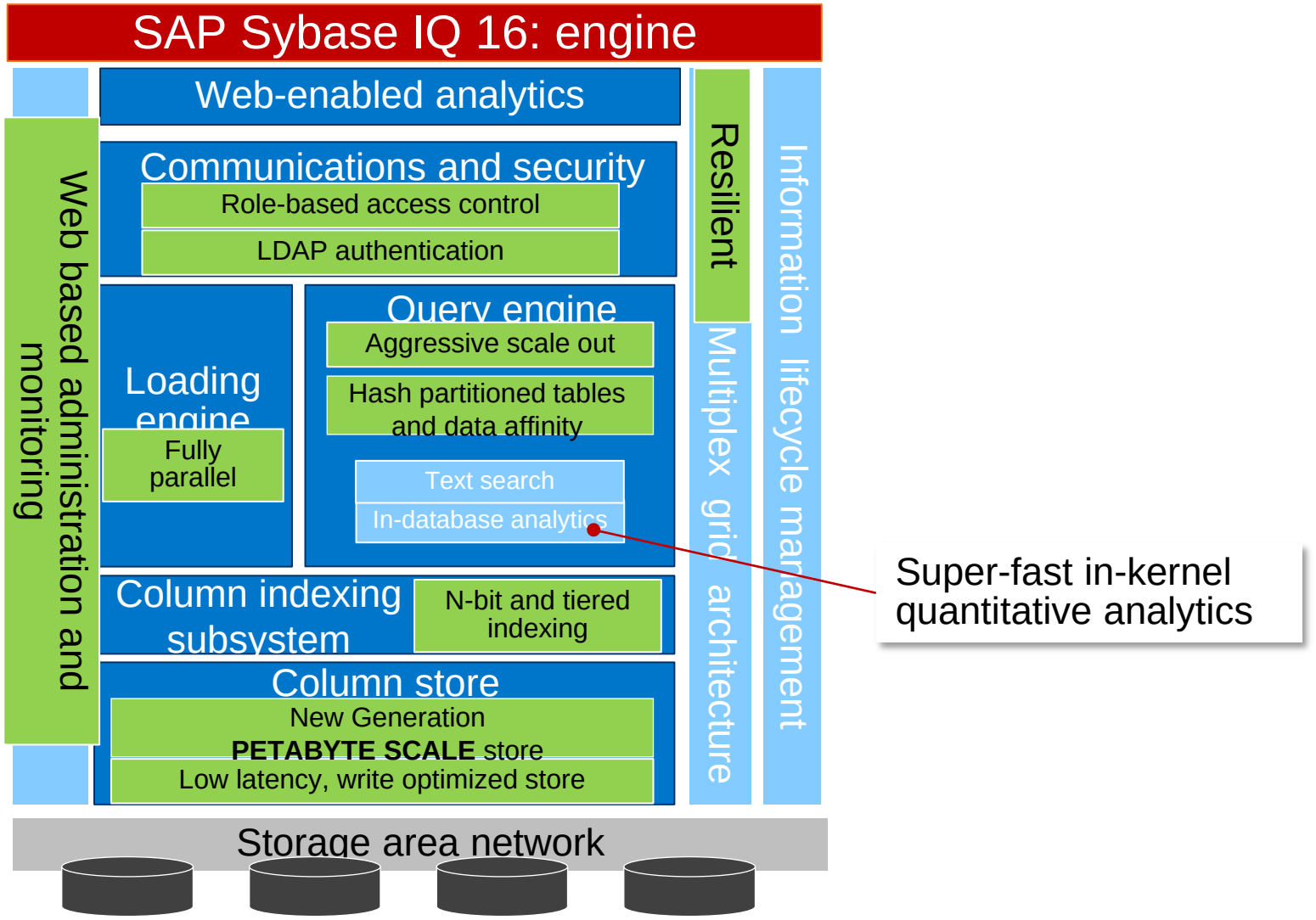
Text search: E-mail archiving, e-discovery, and e-library

Text analytics: Fraud detection, risk analytics, and news feed analysis

Text mining: Clustering, categorization, and sentiment analysis

SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes



Problems with complex analytics

Data to logic

Fetch data from database

Create datasets for analytical packages

Time-consuming process –
could run into memory constraints with large data sets

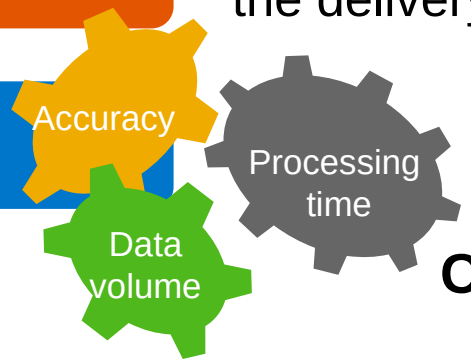
Analyze data using statistical functions on proprietary platforms

Proprietary platforms make it very difficult to embed in applications

Store results from datasets back into database

Another time-consuming process that could slow down the delivery of results to end users

Generate reports

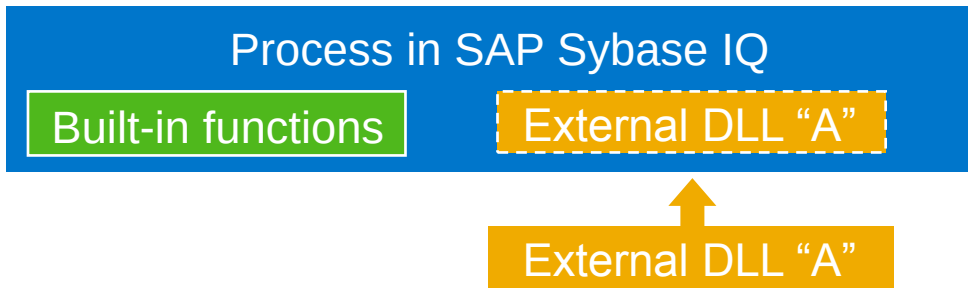


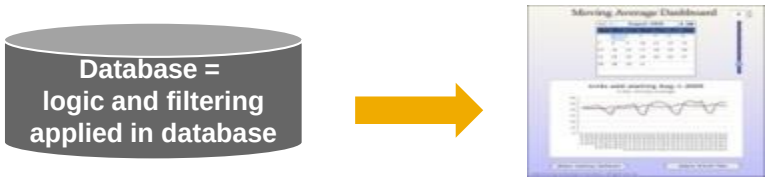
Compromise on at least one key area

In-database analytics in SAP Sybase IQ

No compromise for complex analytics:

- Basic to advanced analytical functions available to SQL directly from engine of SAP Sybase IQ
- Data never leaves the database until results are materialized
- Analytics code and models must be shareable yet must allow ad hoc analysis
- Analytics code and models must be applicable to the latest data set
- Standards-based access; concept extensibility is compulsory
- Performance and scalability is a given
- Average developer must be able to build in database analytical models





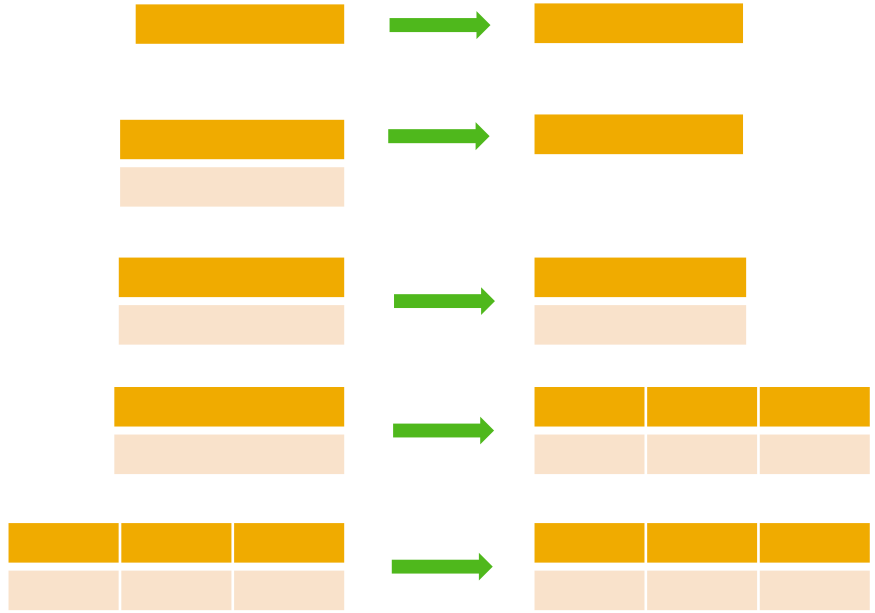
Analytics simplified: Logic to data = fast and efficient

In-database analytics in SAP Sybase IQ

Custom-function APIs

Several different forms of C++ and JAVA UDF APIs for building custom in-database analytics, each valid at different locations within queries:

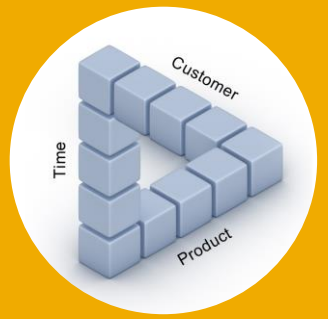
1. {Scalar} to {Scalar functions} like sin, cosine, and so on
2. {Scalar set} to {Scalar functions} like max, min, and so on
3. {Scalar set} to {Scalar set} like OLAP windows and so on
4. {Scalar set} to {Tables} like join result sets and so on
5. {Scalar set, Tables} to {Tables} like MapReduce and so on



All variants are parallelizable, but 5 is also distributable across the grid

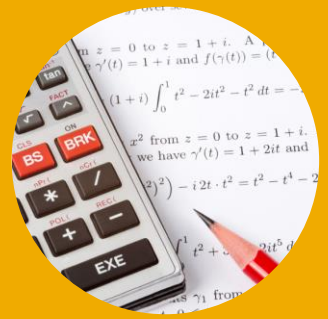
In-database analytics in SAP Sybase IQ

Prebuilt functions available natively and through partners



OLAP

- Windowing
- Ranking
- Cubes
- Roll-ups
- Correlation
- Covariance
- Weighted average



Mathematical

- Basic math
- Matrix algebra
- Gamma and beta functions
- Area under curve
- Interpolation methods

Statistical

- Descriptive statistics
- Distance measures
- Hypothesis testing
- Cross tabulation
- Anova

Univariate distributions

- Monte Carlo simulation
- 30 univariate distributions available

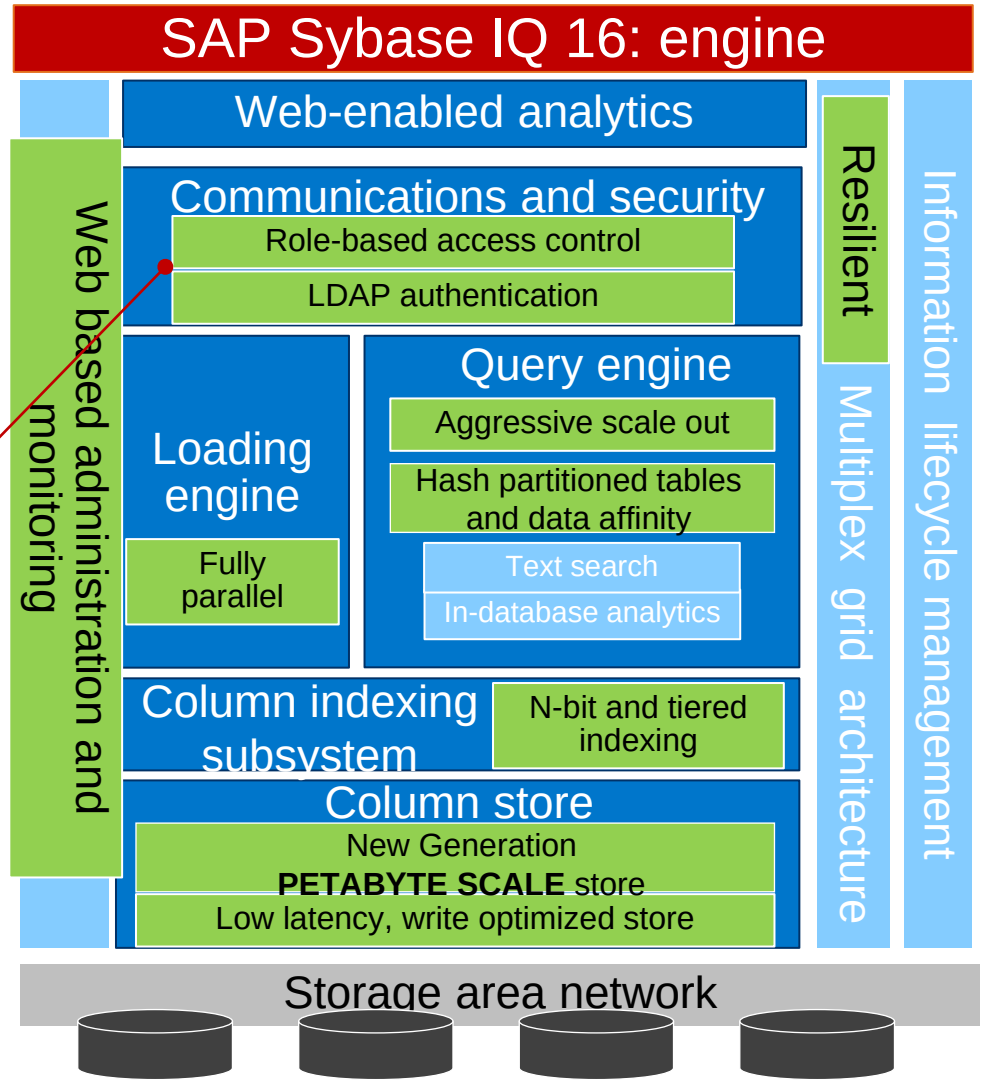
Data mining

- Linear regression
- Logistic regression
- Principal component analysis (PCA)
- Cluster analysis — five models available

SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes

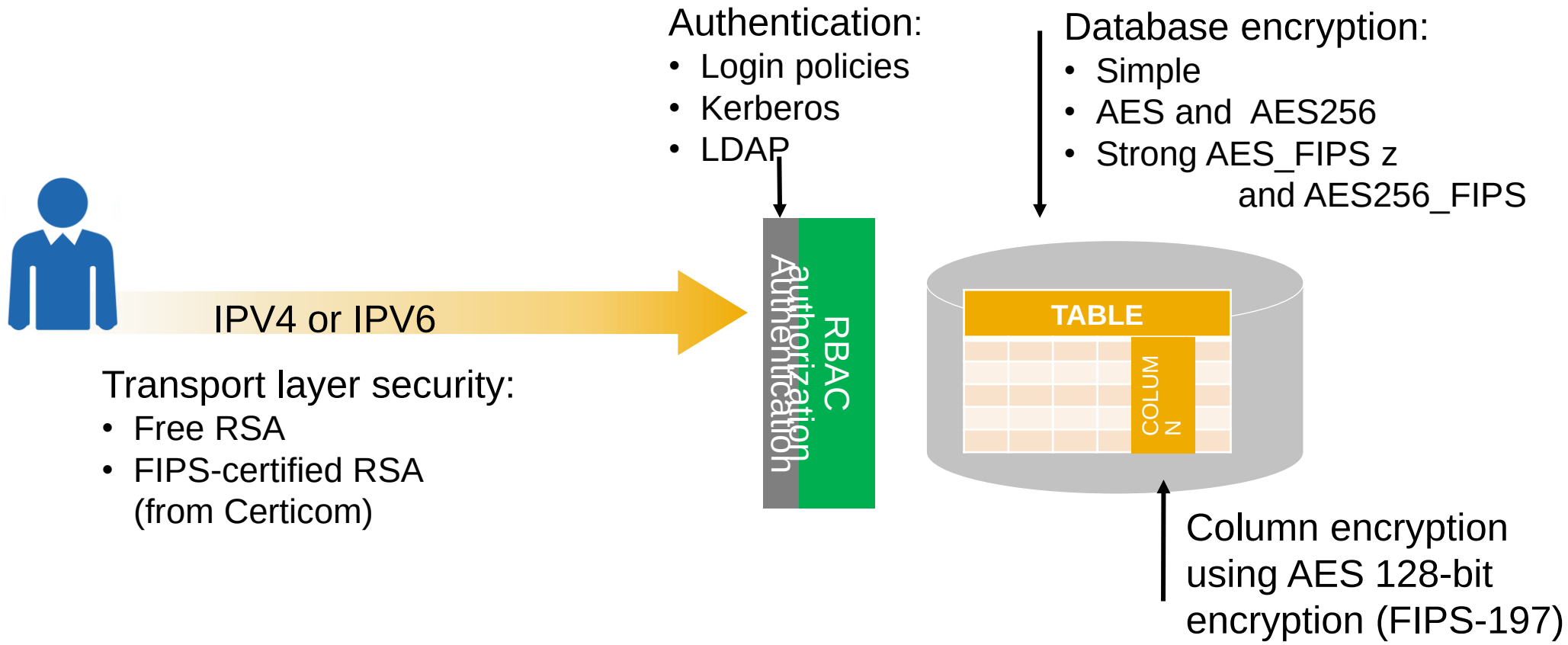
Strong encryption, authentication and role-based authorization



Security

Authentication, authorization, and encryption

Advanced security option:
Strong FIPS encryption, Kerberos and **LDAP authentication**



Security

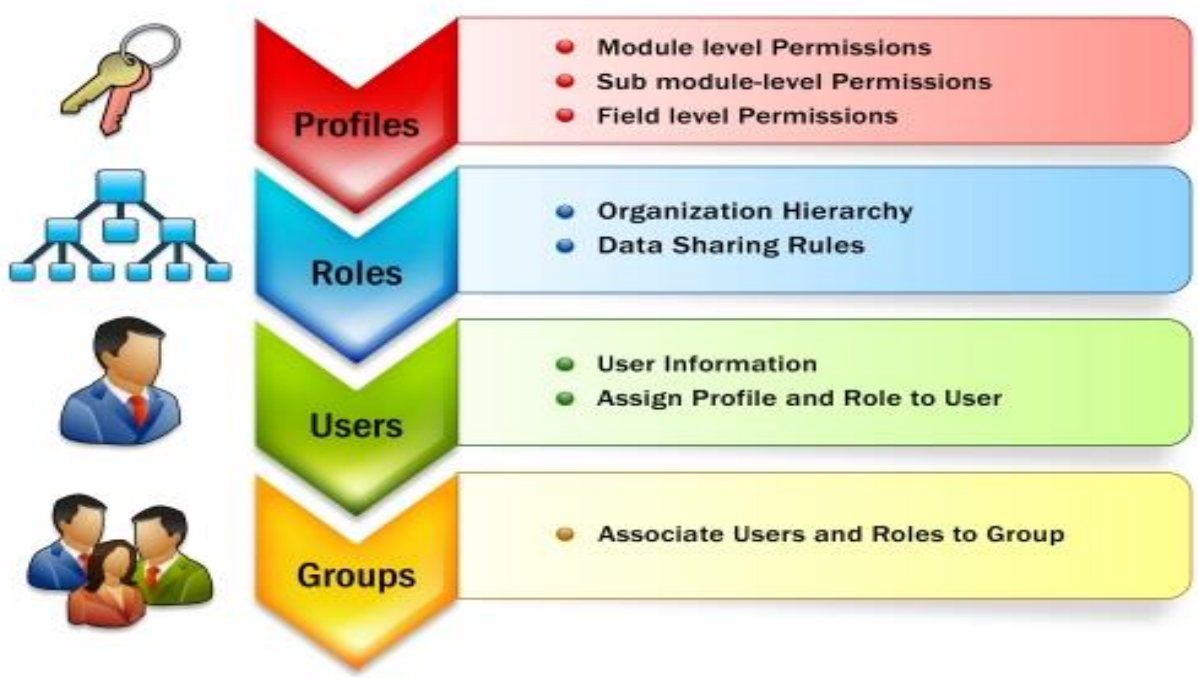
Role-based access control

Value Proposition

- Support separation of duties and principle of least privilege
- Breakdown privileged operations into fine grained sets that can be individually granted
- Control over granting and propagation of privileges
- Complete backwards compatibility and clean migration

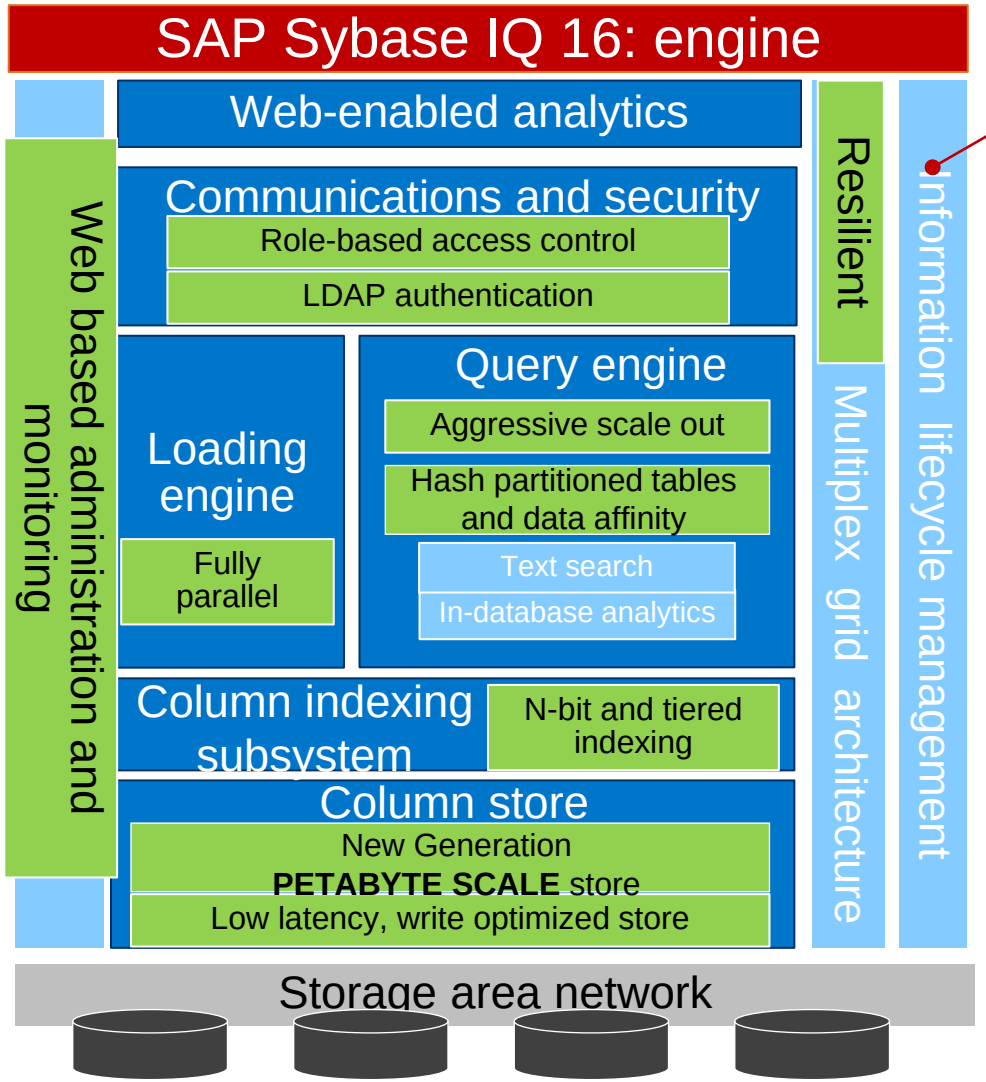
Architectural Considerations

- Support ANSI SQL role semantics, system defined roles and user defined roles
- Minimum number of role administrators
- Grantable privileges for privileged database operations
- Secure system stored procedures with SQL SECURITY INVOKER
- Excellent performance with connection level cache mechanism
- Restrict impersonation through SET USER to adhere to RBAC model



SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes



Manage data through its existence in the data warehouse

Information lifecycle management (ILM)

Building blocks

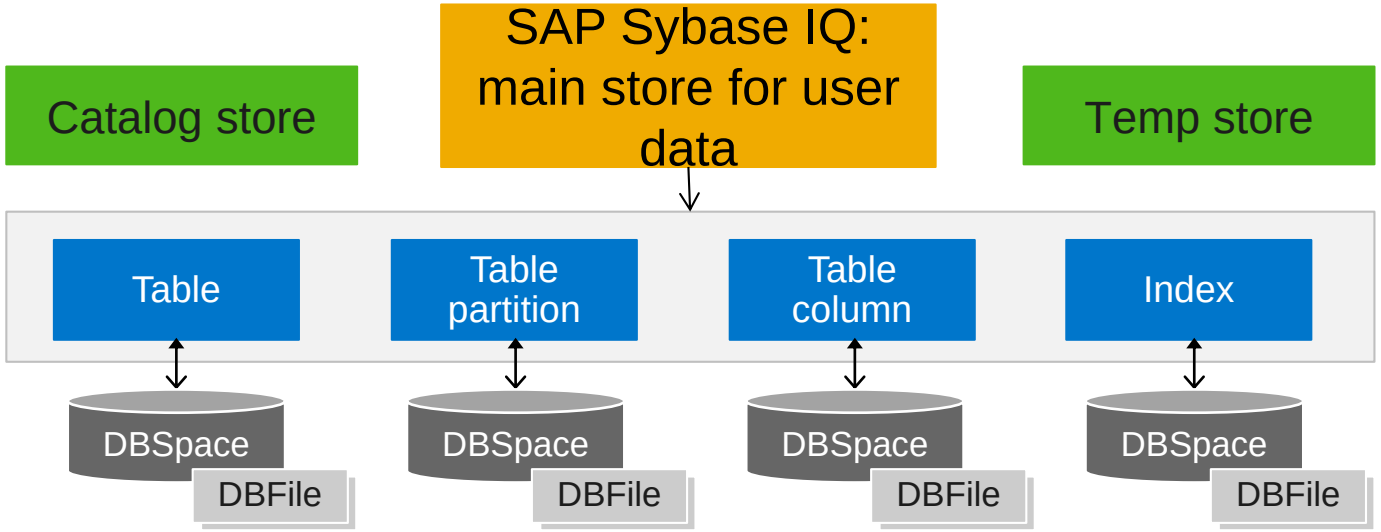
Value-based data partitioning on columns

Multiple user DBSpaces (containers for database objects)

Separate unstructured data from transactional data

Place for frequently accessed data on fast storage

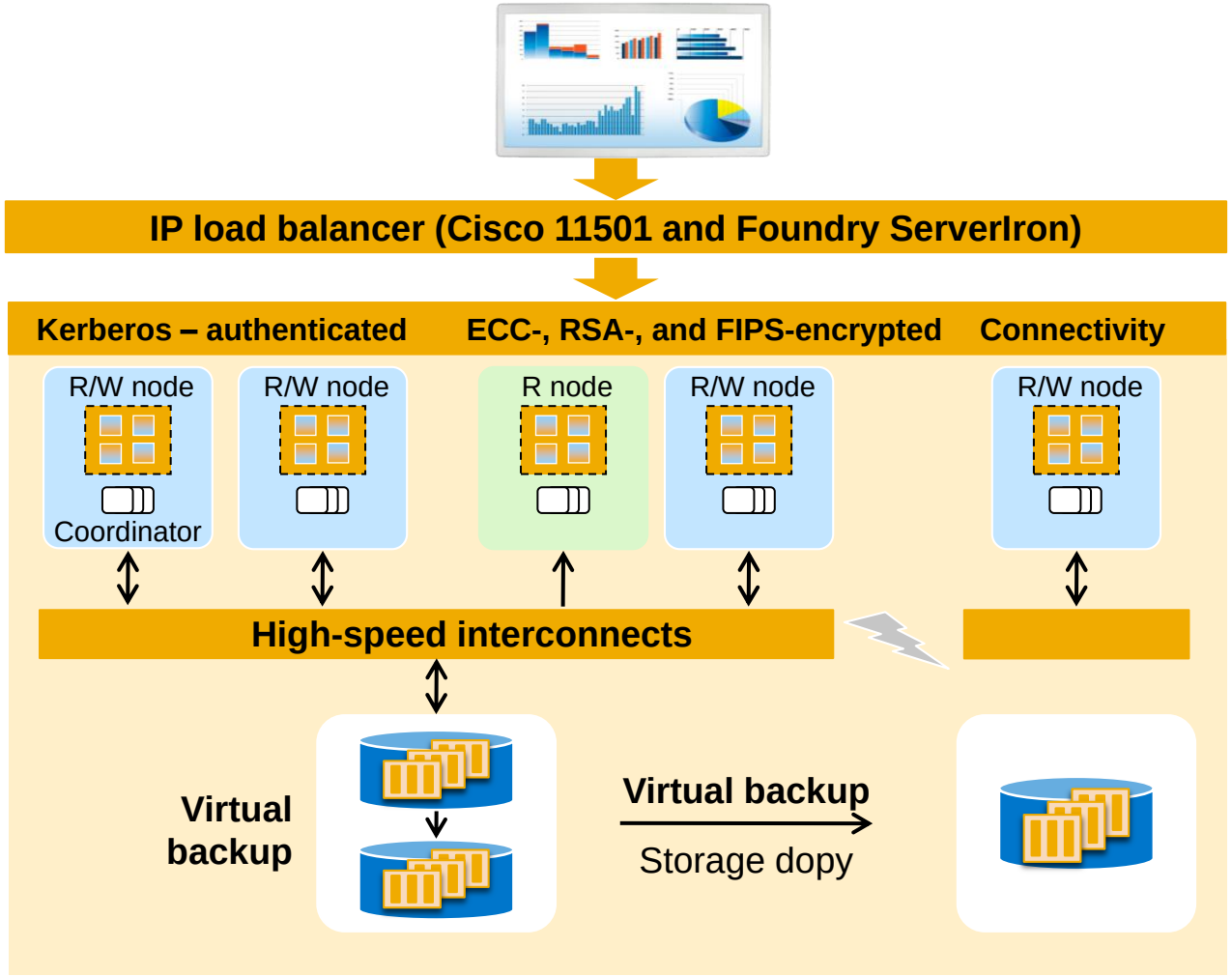
Granular DB admin with read-only, read-write, online, and offline DBSpaces



Information lifecycle management

High availability and disaster recovery with virtual backup

- **Multiplex for scalability and high availability**
- **Virtual backup integration with storage software:**
  Tight loop between virtual backup and storage commands
  ATA-grade storage for database copies
  Examples:
  - EMC CLARiiON
    - Snapview and SAN copy
    - Joint Sybase and EMC whitepaper
    - EMC Symmetrix
      - RecoverPoint and Timefinder
- **Fast database restore of storage copies:**
  Verify backup
- **Disaster recovery:**
  - Storage replication and virtual backup
  - Difficult for shared nothing to implement
  - Requirement for a second system of equal



Virtual backup

Foundation for disaster recovery

Verification of backups:

- Quick restore
- Integrity checking

Testing upgrades:

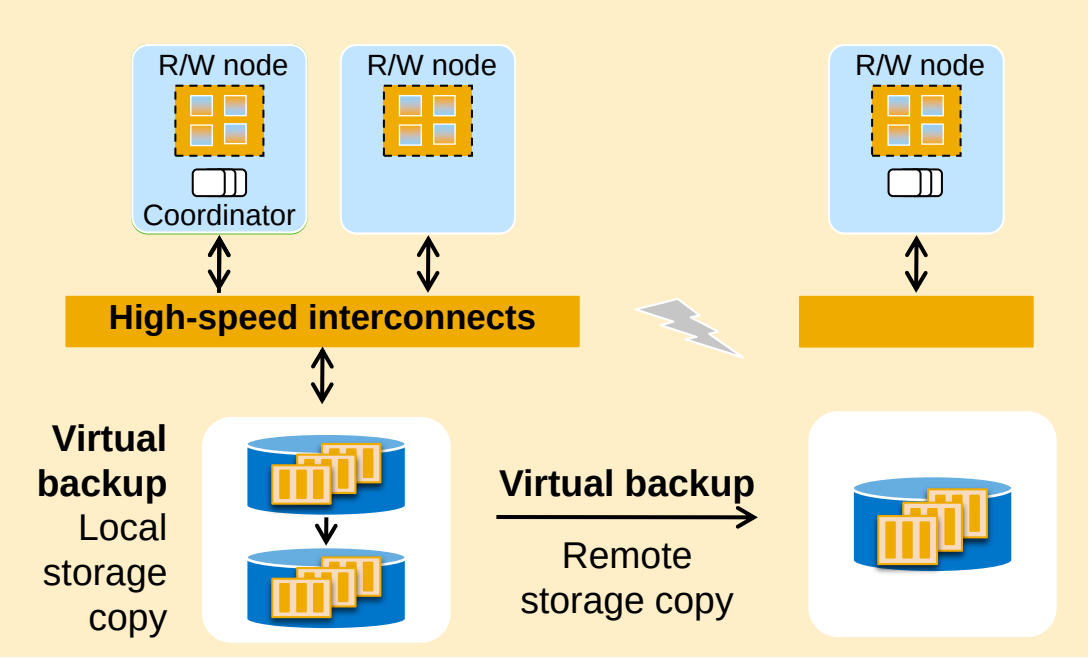
- Application upgrades
- Version upgrades

Development copy:

- Avoid developing against production
- Test against full size data

End-user playground:

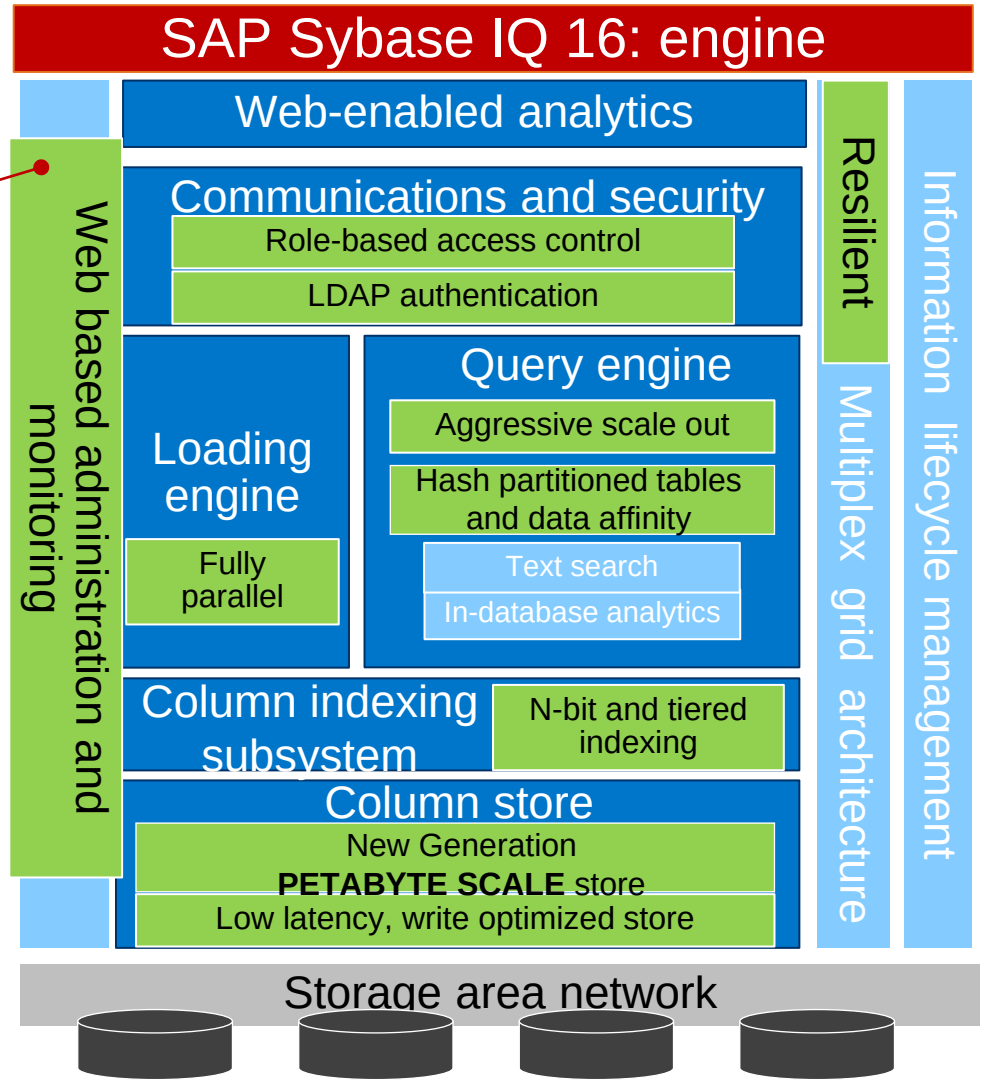
- Run any query



SAP Sybase IQ 16

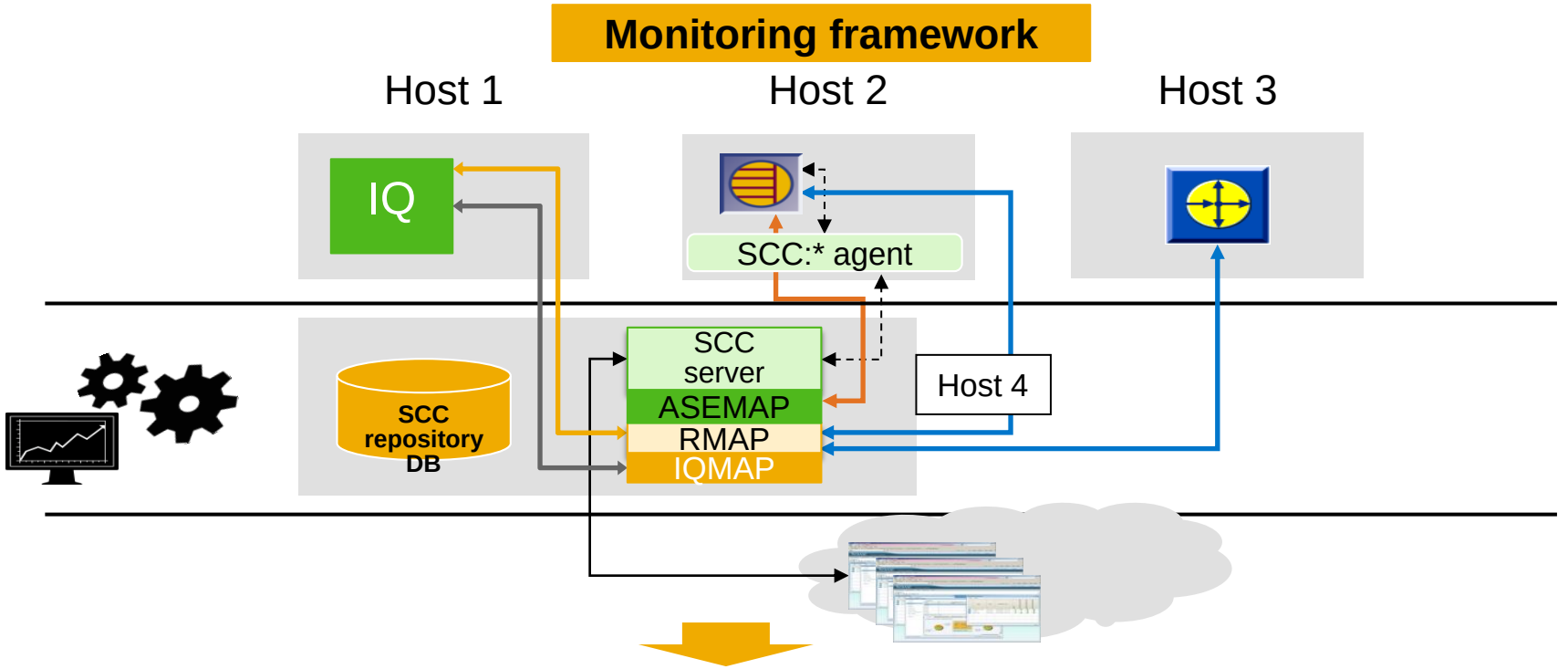
Key building blocks – new features highlighted in green boxes

User-friendly, web-based administration and monitoring



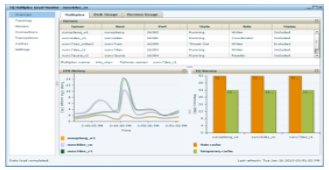
Sybase Control Center

Web-based administration and monitoring

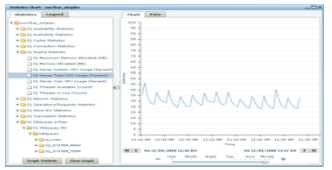


Node-level monitoring

Engine monitoring



Multiplex monitoring

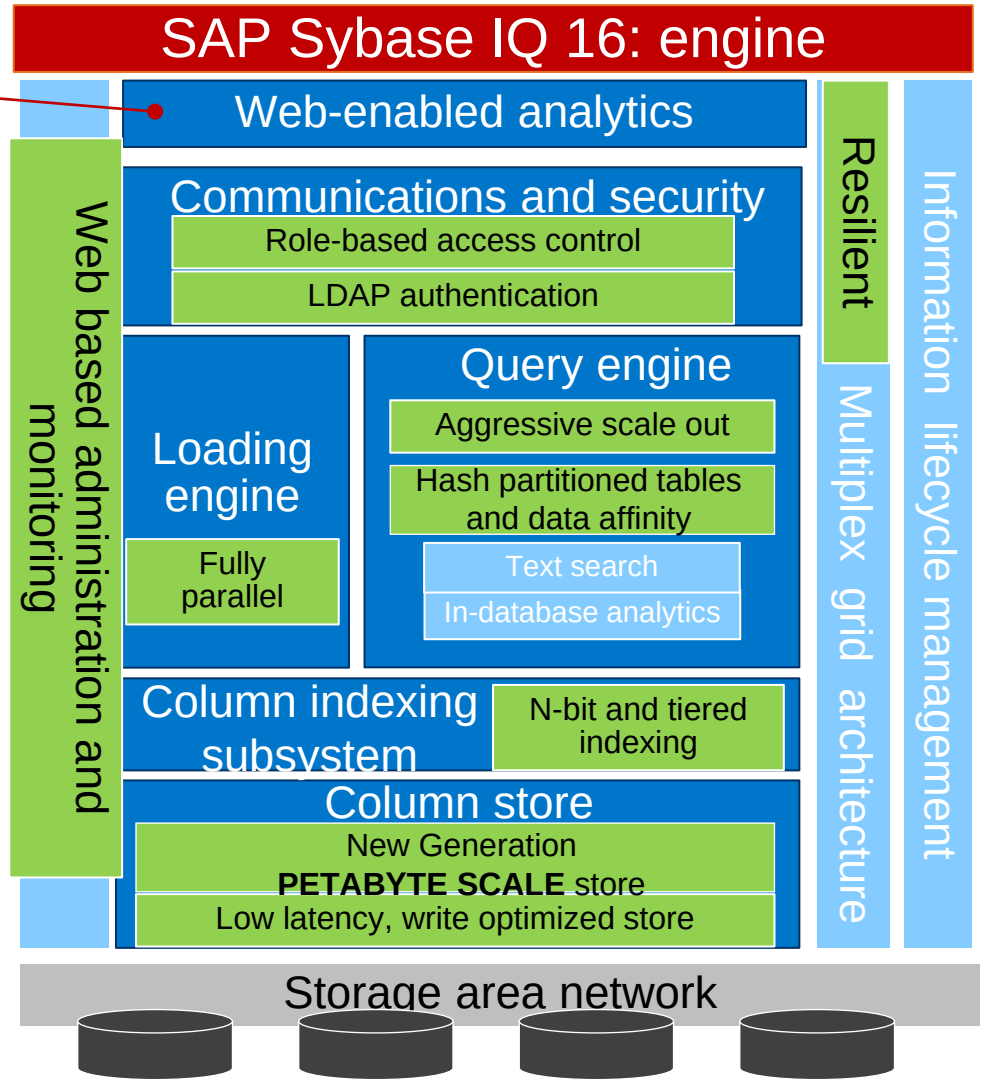


History replay

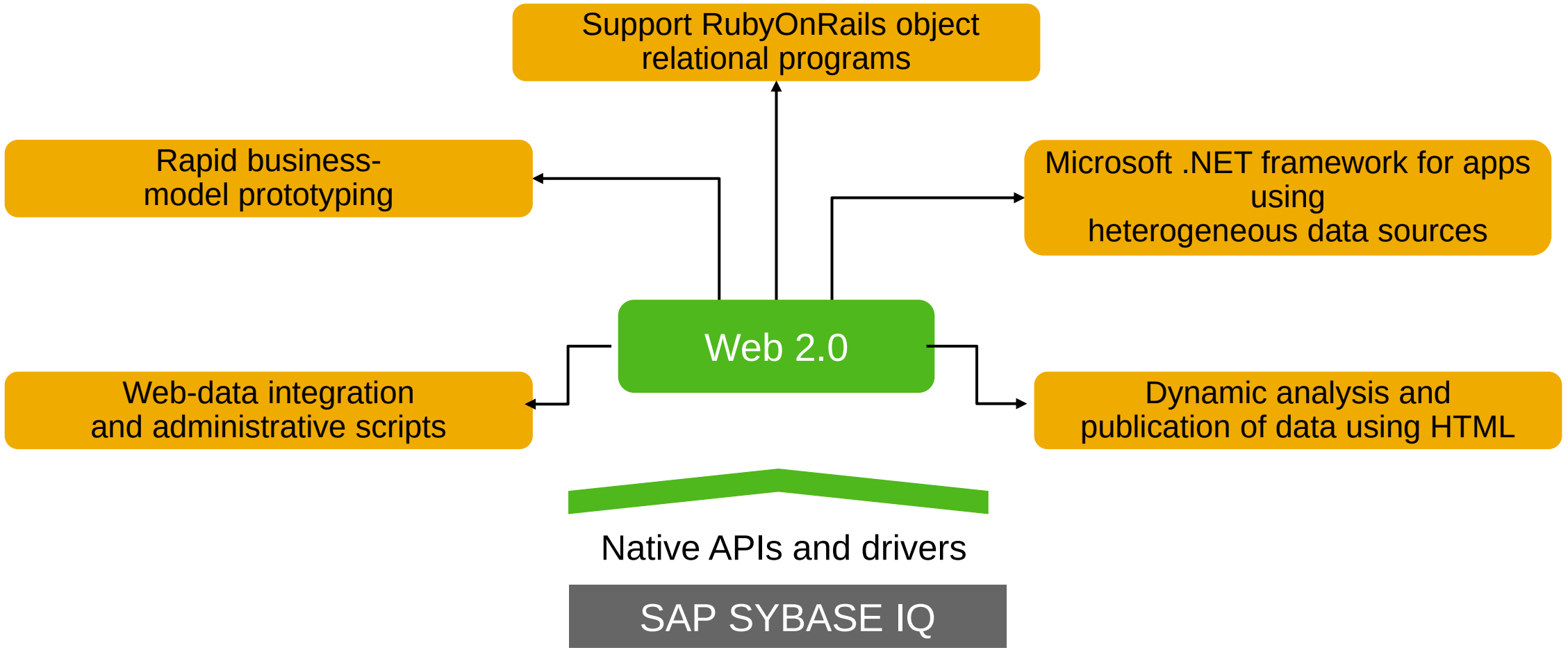
SAP Sybase IQ 16

Key building blocks – new features highlighted in green boxes

Web-enabled application language drivers



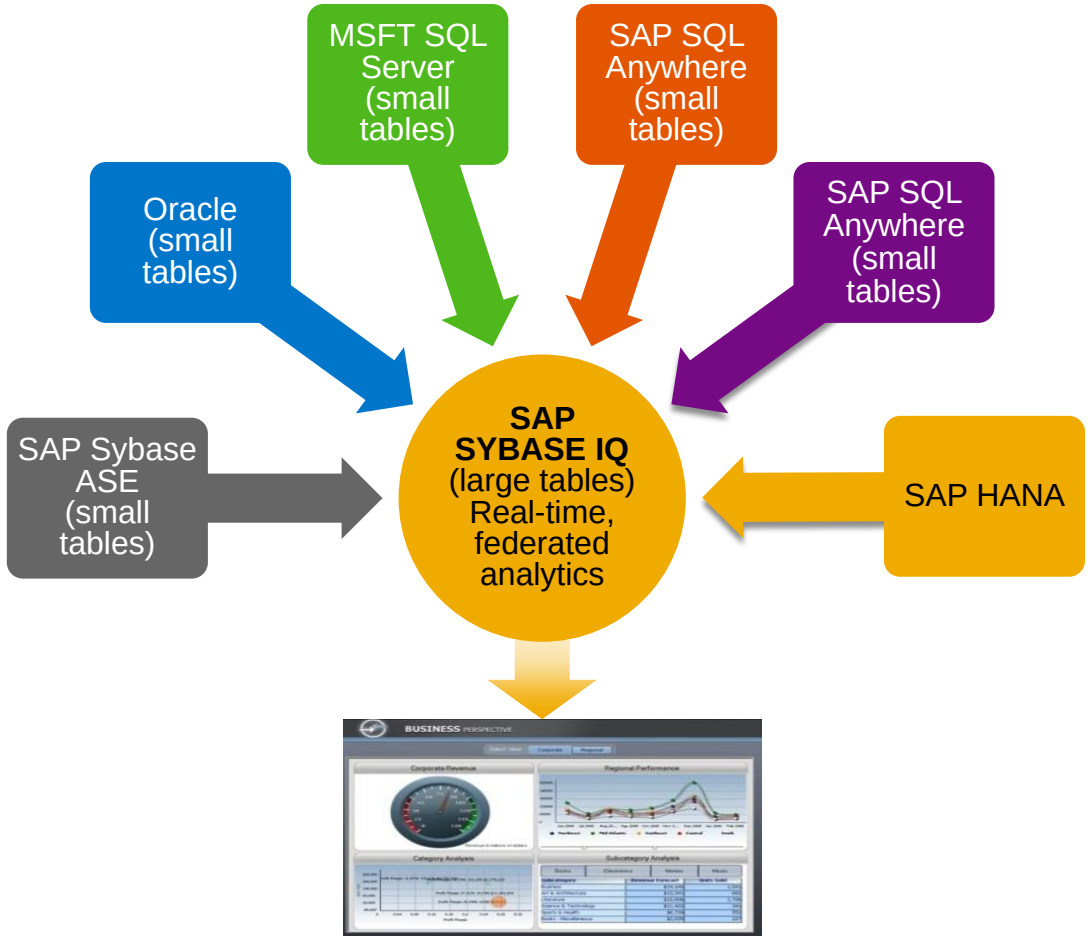
Web 2.0: language drivers



Query federation
With other databases

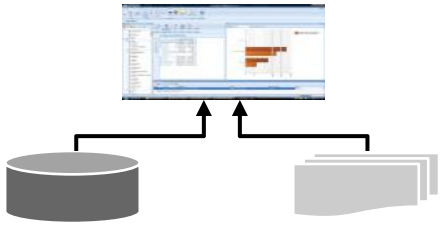
Ability to analyze data in SAP Sybase IQ as well as in remote RDBMS (majority of the data in SAP Sybase IQ and a small set of data in remote systems)

Remote servers could expose materialized views (cubes) that is queried to provide a federated view of the data marts



Federation

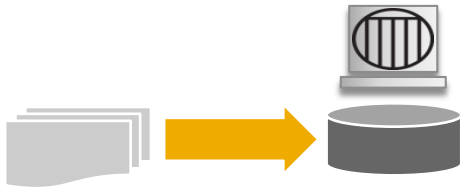
With external file systems (Hadoop distributed file system)

1. 

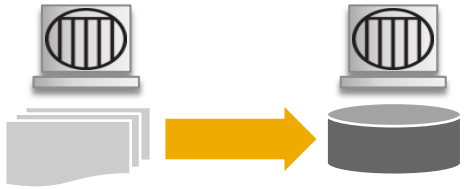
Client-side federation: Join data from SAP Sybase IQ and Hadoop at a client-application level

2. 

Load Hadoop data into column store of SAP Sybase IQ: Extract, transform, and load data from Hadoop distributed file system (HDFS) into schemas of SAP Sybase IQ

3. 

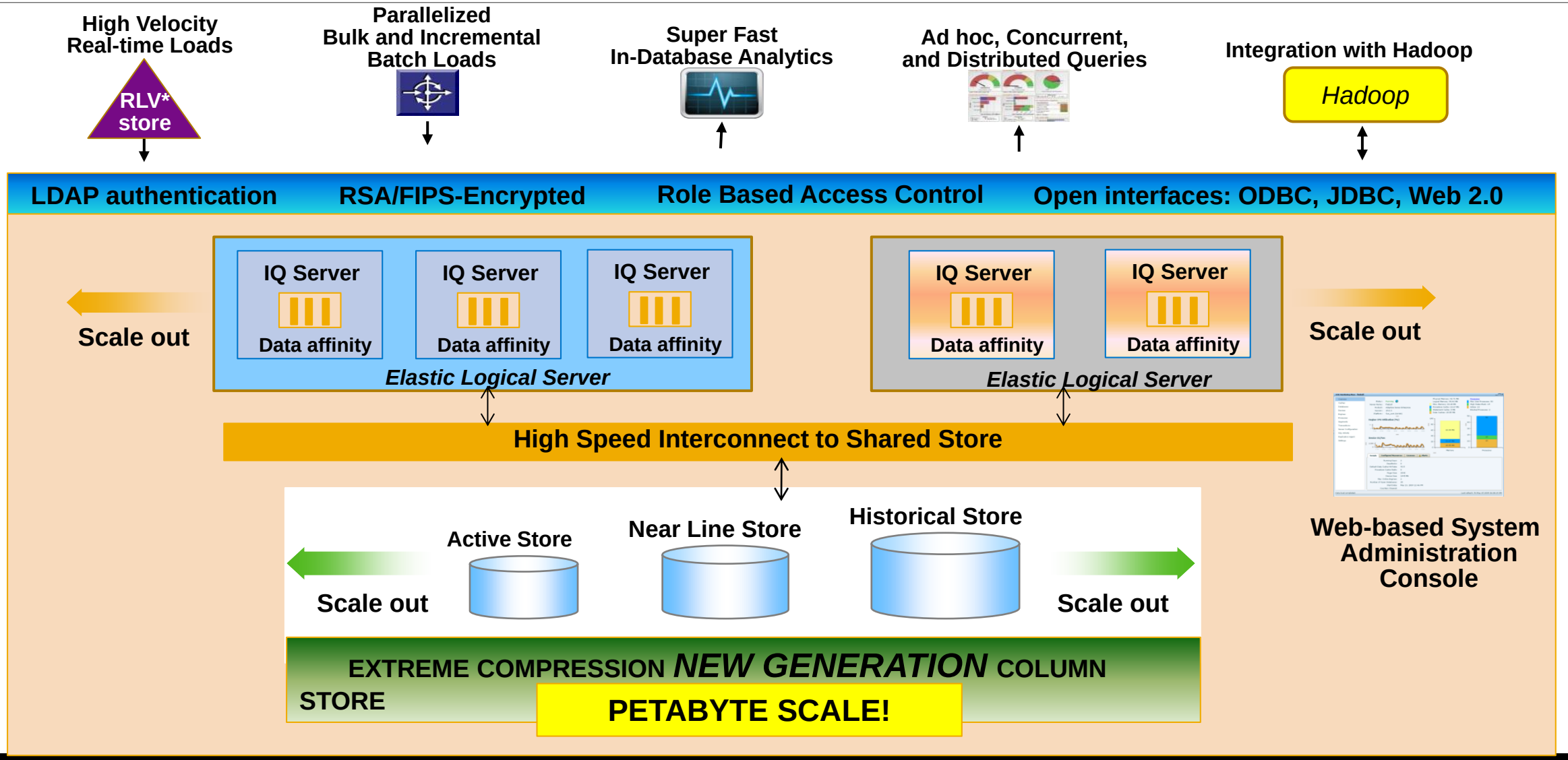
Join HDFS data with data of SAP Sybase IQ on the fly: Fetch and join subsets of HDFS data on demand, using SQL queries from SAP Sybase IQ (***data federation*** technique)



Combine results of Hadoop MapReduce (MR) jobs with SAP Sybase IQ data on the fly: Initiate and join results of MR jobs on demand using SQL queries from data in SAP Sybase IQ (***query federation*** technique)

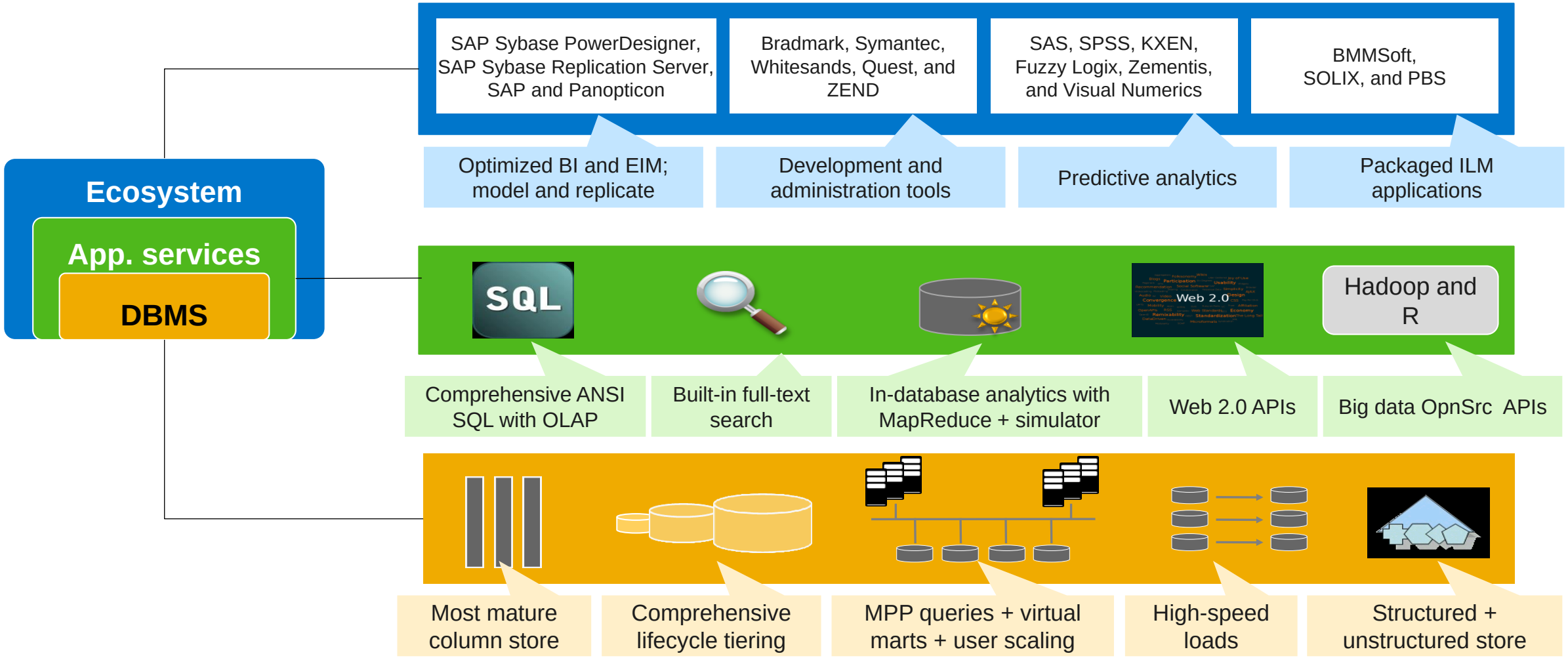
Summary

SAP Sybase IQ 16 Architecture



SAP Sybase IQ 16

A complete platform for XLDB analytics



Summary

SAP Sybase IQ is the #1 columnar, disk-based analytic database with intelligent caching. IQ centralizes "Big Data" analysis of massive volumes of structured and unstructured data for report acceleration, data mining and predictive analytics

- **Lower TCO** – Extreme data compression, data partitioning, commodity hardware, better price-performance
- **Administration** – IQ uses standard RDBMS (ANSI SQL), and is easy to administer, as attested to by many customer testimonials.
- **Parallel computing** - IQ employs both SMP parallel and MPP distributed query execution. IQ automatically reduces/increases parallelization as user query workload changes.
- **User scalability** - IQ scales out easily with its flexible "shared everything" MPP architecture. Adding nodes is trivial and IQ automatically balances resource use, allowing many users to share node resources.
- **Node dedication** - IQ can dedicate nodes for parallel loading with no interference with reporting operations.
- **Disaster recovery** – IQ features a proven Non-Stop architecture, which can utilize a secondary storage system for immediate failover

This is the current state of planning and may be changed by SAP at any time.

SAP Sybase IQ Resources

Where to get information on SAP Sybase IQ

- **SAP Sybase IQ SCN Developer Site:**
  http://scn.sap.com/community/developer-center/analytic-server
- **Free Express and Enterprise Edition Downloads**
  https://www.sap.com/iq-downloads
- **"How to" Tutorials**
  http://scn.sap.com/docs/DOC-43300
- **SAP Sybase IQ 16 (video):**
  https://www.youtube.com/watch?v=Bbcl8uY3tj4
- **SAP Sybase IQ for Big Data(video):**
  https://www.youtube.com/watch?v=XRNgVPxiyVw
- **SAP Sybase IQ as Near-line Storage (video):**
  https://www.youtube.com/watch?v=FdtdAhTQYSg
- **SAP Sybase IQ for BusinessObjects (video):**
  http://www.youtube.com/watch?v=A4hR0n2y8zQ

Further Information

SAP Public Web

scn.sap.com
www.sap.com

SAP Education and Certification Opportunities

www.sap.com/education

Watch SAP TechEd Online

www.sapteched.com/online

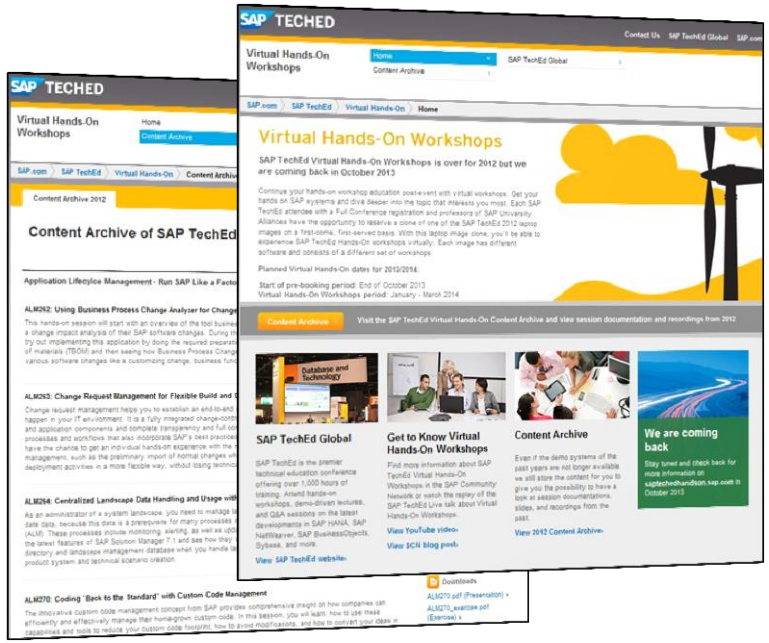
SAP TechEd Virtual Hands-on Workshops and SAP TechEd Online

Continue your SAP TechEd education after the event!

SAP TechEd Virtual Hands-on Workshops

- Access hands-on workshops post-event
- Available January – March 2014
- Complementary with your SAP TechEd registration

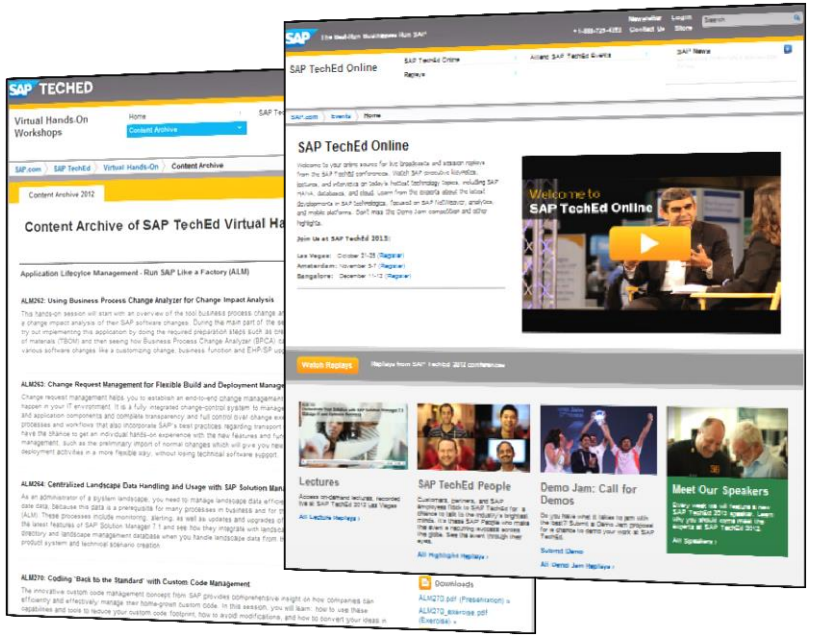
http://saptechedhandson.sap.com/



SAP TechEd Online

- Access replays of keynotes, Demo Jam, SAP TechEd LIVE interviews, select lecture sessions, and more!
- View content only available online

http://saptecheed.com/online



Feedback

Please complete your session evaluation for RDP107

Thanks for attending this SAP TechEd session.

© 2013 SAP AG or an SAP affiliate company. All rights reserved.