IBM 2010 System z 高峰论坛
暨新品发布

客聚天下 • 智绘高远

Infrastructure Architecture for Transactional, Operational & Analytical Data Process

David Heap
IBM大中华区系统与科技事业部高级IT咨询师

Topics

- **Overview and challenges of current data infrastructure in large banks**
- Guidelines of data integration architecture design in large banks
- Optimized transactional, operational and analytic data flow infrastructure
- Business and technical value
- Suggestion

Transactional, operational and analytic data feature comparison

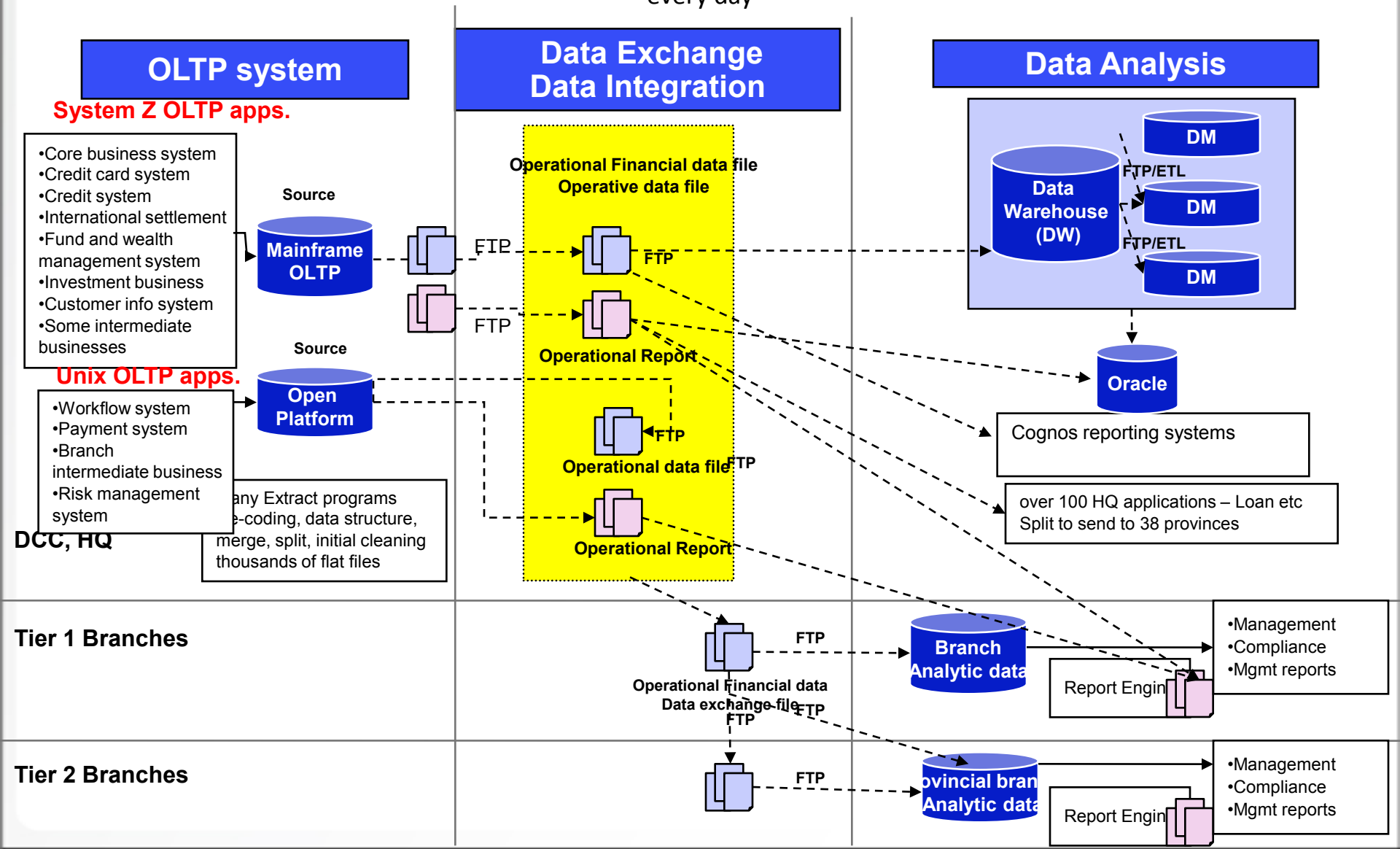
80% core system (z)

Data integration platform or data exchange platform

Data warehouse

	Transactional transaction syste	Operational report and	Analytic report and
Audience	•First line staff and customers	•First line staff and managers	•Professional analyzer and mid & high-level management
System function	•Recording the transaction flow •Recording real-time of critical information such as the accounting, customer and product •Providing real-time business facts	•Data integration, providing the data sharing for the production system; •Customer data presentation: e.g.: query, reporting (for production), dynamic analysis (for production analysis); •Data quality audit and monitor management.	•Supporting multi-dimensional topic analysis; •Trend analysis and issue discovery; •Analytic customer behavior analysis; risk; profitability analysis; assets and liabilities management.
Data source	•Real-time business change over service channels	•Production system	•Maybe from operational platform •From source system if not available from operational platform.
Data structure	•Organizing the data as per the transaction flow and function requirement •Providing efficient random access based on random change	•Organizing the data as per topic domain from the production perspective; •Providing fine granular operation data, also with some coarse granular summary data •Providing operative data reporting and query.	•Organizing the data as per topic domain from analytic perspective; •Providing coarse granular summary data, with more and complex analytic dimensions •Providing analytic data presentation and analysis
Data storage time	•A large number of dynamic updates everyday •Transaction logs are kept for one day	•Summary data is kept for 1 year; •Common data is kept permanently; •Detailed data is kept for 1-3 months; •Other integration layer data is kept for 13 months.	•Summary data is kept for 5 years; •Common data is kept permanently; •Detailed data is kept for at least 2 years online; •Other integration layer data is kept for at least 2 years.
Technical characteristics	•Database technology subject to efficient random change; •Efficient OLTP data type and data operation.	•Database optimization focuses on both the indexing and partitioning techniques; •Supporting OLTP type and OLAP type database operation.	•Database optimization focuses on the partitioning technique; •Supporting OLAP type database operation.

There are a large number of data conversions among transactional, operational and analytic platform every day
There is 300-400GB transaction data (>1TB in special day), thousands of operational files and tens of thousands of reports every day



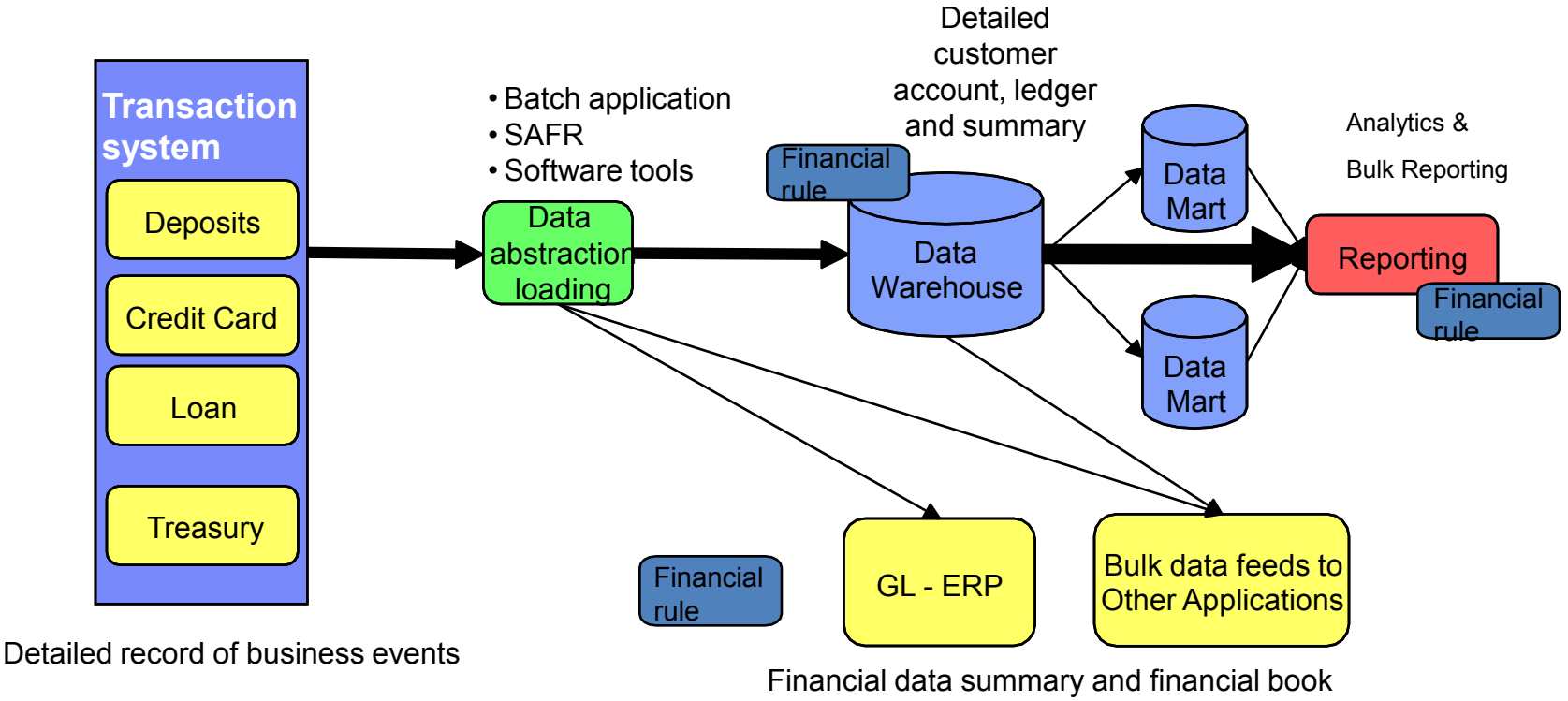
Business and technical drivers

- Increasing demand for correct detail data
  - Large demands for analytic report dynamic
  - Require GL summary and detailed ledger and customer account information to provide flexible business analysis elements for analyzing the customer, product and region
  - Complex account balancing and reconciliation
- Large-scale bank - Existing z architecture us unable to bear analytic application of large database, such as the scanning summary
  - Over 100 million transactions, and over 400 million accounts
  - Must support the architecture of two-level data for serving 30+ provinces
- With the delay of the data availability, it is unable to reach T+1 service level commitment
  - Need to collect source data across multiple heterogeneous platforms
  - No standard data abstraction flow and platform
  - No shared integration data platform, highly redundant data storage and access
- Need to transport a large amount of data over WAN and LAN, resulting in
  - Repetitive abstraction of the same data source
  - The data security can not be guaranteed, especially over LAN
  - High cost-time, bandwidth, system resource

What type of data and business service benefit most from z platform

- Business level - Business-oriented operation control data
  - A most important business area is financial and management accounting of the bank. GL and related operative information system and data. Based on the accounting rule, the data model of this financial information is quite mature and stable
  - The volume of data for detailed customer account and ledger is large. This data is relative clean because they are abstracted from daily business transaction data.
- IT level - Operative data architecture
  - 80% of most critical operative data in the bank operation comes from transactional data of core business system (such as deposit, loan, credit card, etc.), and these system are currently running on z platform
  - Require batch data abstraction, cleaning and consolidation, as well as a period of storage to reflect “only version of truth of financial data” in the bank
  - For this kind of data, the business function should meet dynamic real-time query and a large number of T+1 batch reports, while the reporting demand may be changed easily
  - And this operative data provides basic data for BI to create views, cubes and more required by analytic applications
- Hence, it is a favourable point for building operative BI application for z platform starting from financial and management accounting data

Let's analyze a simplified typical operative data flow



Where to save this detailed financial event data?
How long? - 1 month, 1 year, 5 years...
What is the amount of detailed financial event data?
Which keeping rule? - Accounting rule, business and reporting rule
What happens when the rule changes?
Report serving time requirement? - T+1, T+3, ME+8 ?

Topics

- Overview and challenges of current data infrastructure in large banks
- **Guidelines of data integration architecture design in large banks**
- Optimized transactional, operational and analytic data flow infrastructure
- Business and technical value
- Suggestion

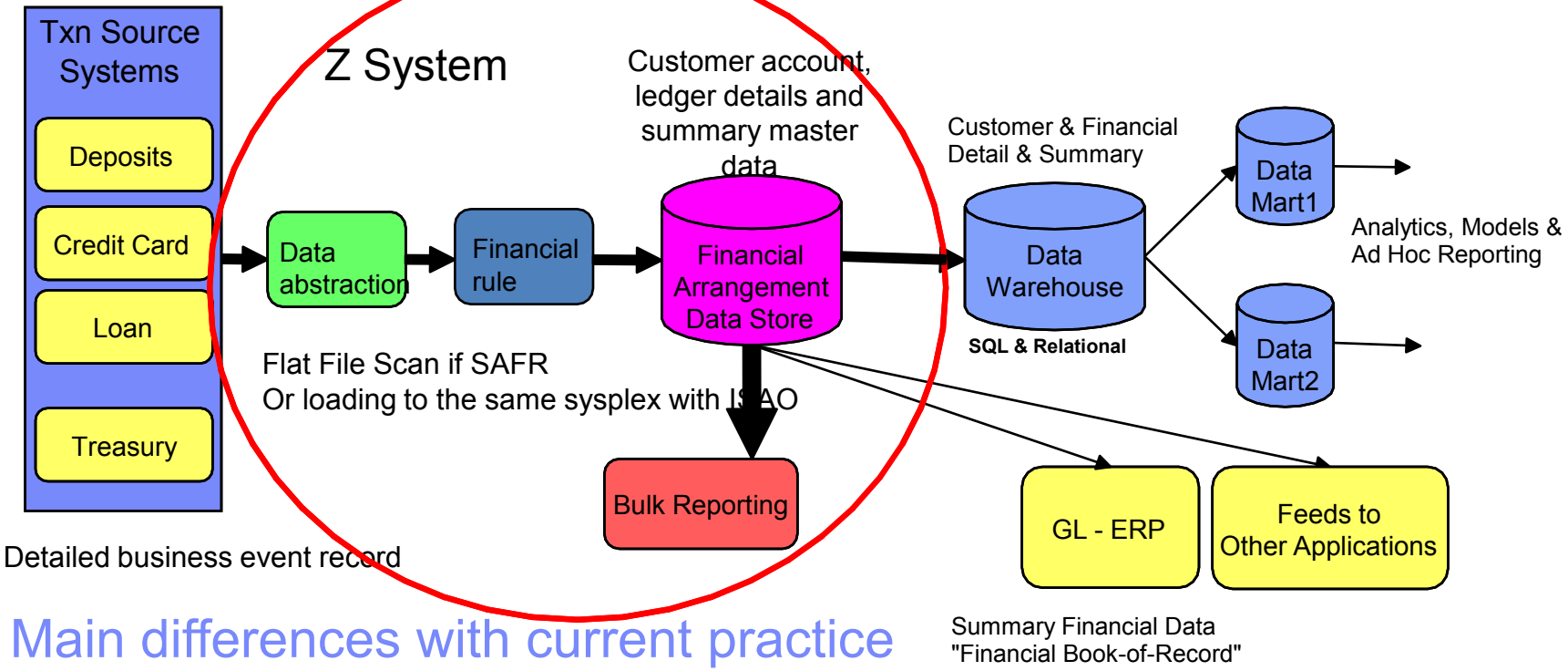
Optimizing current transactional data, operative data and analytic data infrastructure needs to comply with basic principle of data integration of large banks

- Abstract financial data from the source only once
  - Ensure all data is from same time-slot
  - Consolidate processing against production systems
- Centralized data cleansing and conversion, centralized rule management, rule-based data processing and reporting
  - Ensure same technology and business rules are implemented for all data
  - Centralized rule to simplify the business rule maintenance
  - Rule-based report generation, but not bundled with the program, to provide the reporting flexibility
- Keep lowest level of detailed data, and get close to the source data
  - Ensure the data transparency - drill-through from summary data
  - Ensure the summary data consistence (all data is summed from the same source data)
  - Reduce the data movement

Topics

- Overview and challenges of current data infrastructure in large banks
- Guidelines of data integration architecture design in large banks
- **Optimized transactional, operational and analytic data flow infrastructure**
- Business and technical value
- Suggestion

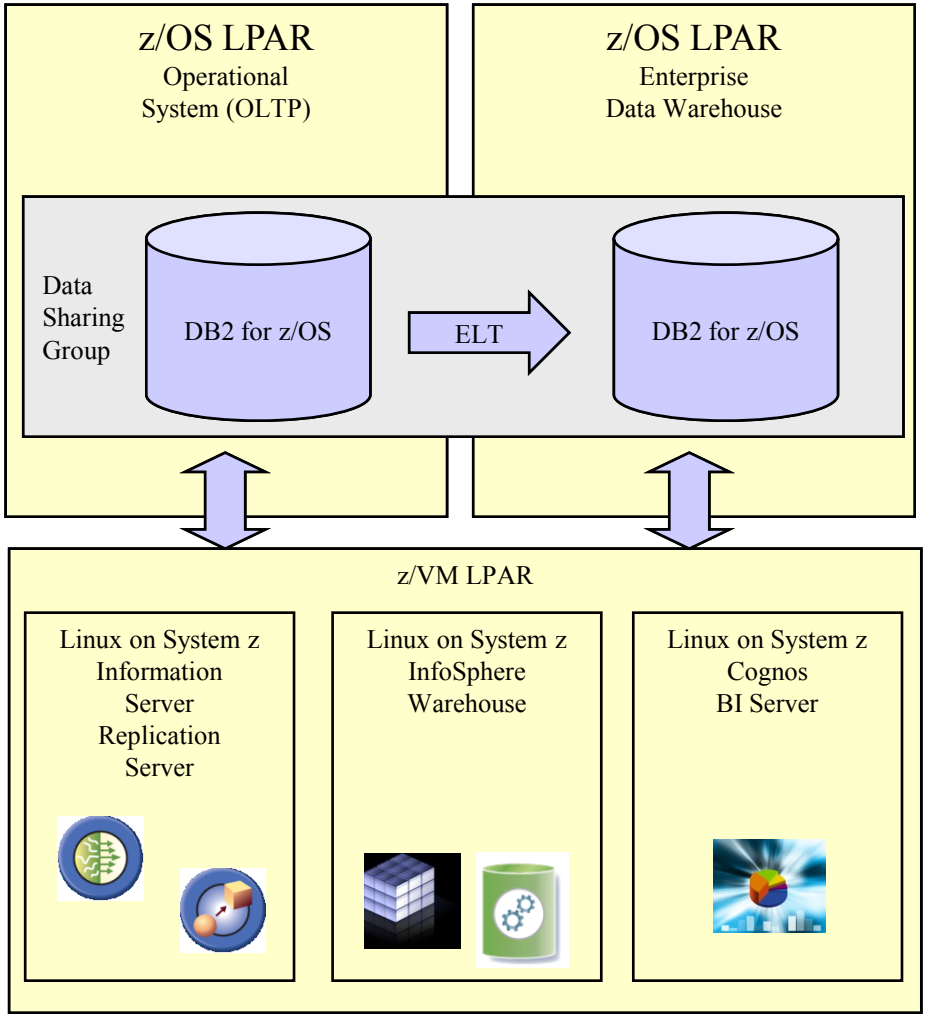
Build operative data store for ledger details
Detailed Financial Arrangement Data Store



Main differences with current practice

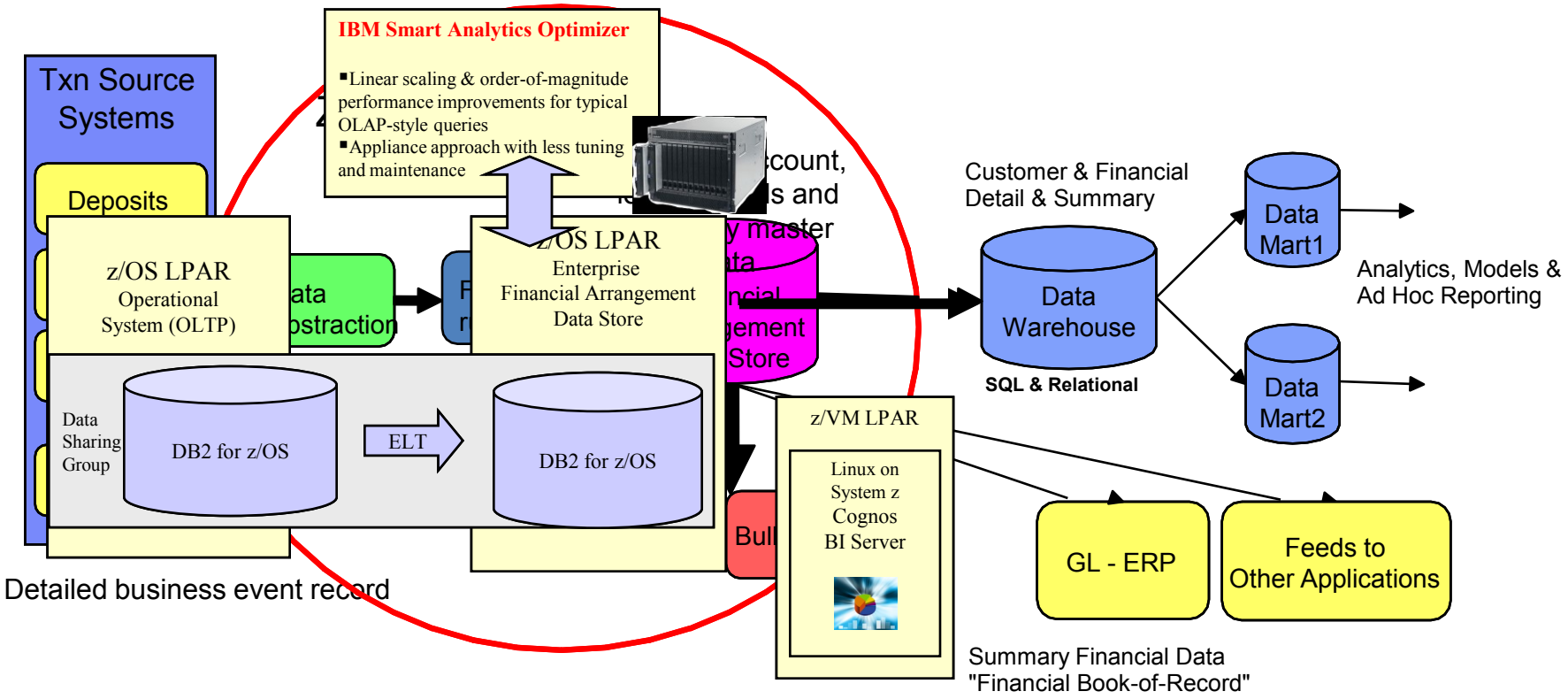
1) Centralize the accounting rule on the same system
2) Implement the business rule at the upstream of the data processing flow
3) Abstract financial data from the source data only once
4) Store the customer account, ledger and summary report in central location
5) Keep fine-granular details for 5 years
6) Provide consistent detailed data for generating the production report

Z BI deployment architecture



- Centralize critical business data on System z
- Backup and govern the workflow environment and skills with high availability of existing system z
- Effective data migration in the same shared environment (not via the network)
- Improve performance and TCO with cubing service (data mart) and DB2 enhancement
- Complete complex data conversion and data quality control with Linux on System z with Information Server

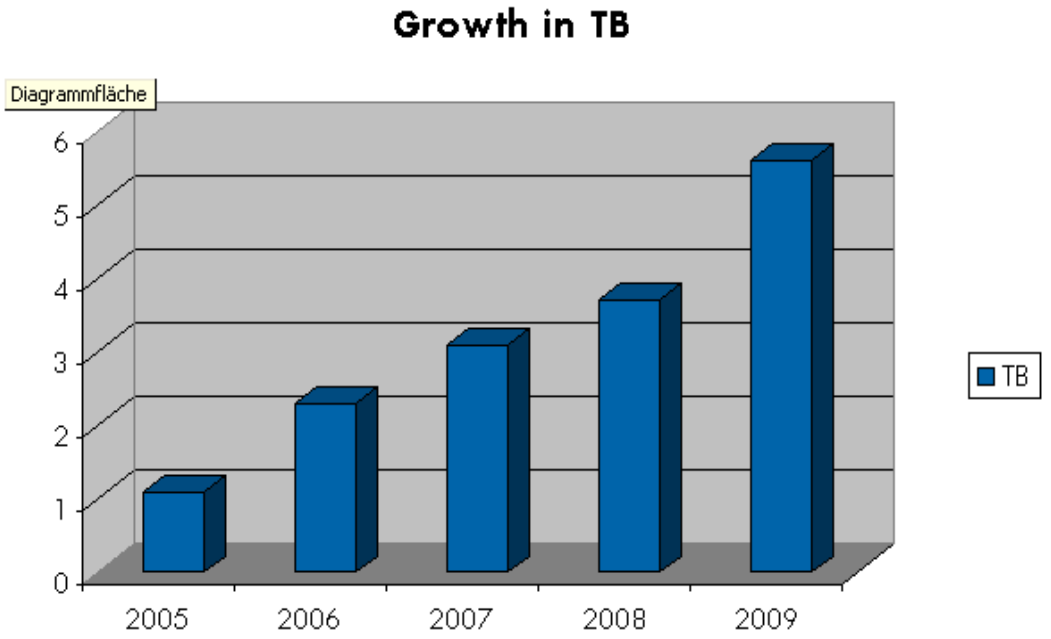
Optimized transactional, operational and analytic data flow infrastructure



- Load financial details report abstracted from OLTP into MQT table. Because in the same database sharing group, so
  - ✓No codec
  - ✓No FTP
- Run the report analysis client over zLinux, and run complex query and summary applications over z hybrid platform
- Enjoy high availability and disaster recovery system in 3 centers of 2 regions
- High timeliness
- Reduce data redundancy

Actual data of pilot customer

- 5 TB data
  - 14 DB, 1568TS, 3517 IX, 1568 Tables, 53118 Columns
- 8000 nights of data warehouse jobs, 4000 daytime analysis jobs
- Data in the data warehouse is refreshed instantly, and the history data is kept for 4 years
- The largest report contains 1 billion records
- Data growth as shown below



Topics

- Overview and challenges of current data infrastructure in large banks
- Guidelines of data integration architecture design in large banks
- Optimized transactional, operational and analytic data flow infrastructure
- **Business and technical value**
- Suggestion

Business and technical value from reconstructing the business data platform to run OLAP on z

- The consolidation of financial reports on z enables the processing of financial reports in standard and modular way to improve the reporting performance and timeliness
- The transaction flow and report generation are centralized on single platform to reduce redundant data mapping, refer and balance the data maintenance overhead
- zNext/ISAO provides the data warehouse query optimization, a mechanism for a large number of concurrent operations and large data memory to reduce z operation cost and improve OLAP capability
- ISAO has powerful linear scalability
- No multifarious tuning of DB2 (MQTs, aggregates, indexes, etc.)
- Maintain QoS of traditional System z and DB2. For the bank management, the production report governance can be incorporated into current host disaster recovery system to extend the business recovery scope

Topics

- Overview and challenges of current data infrastructure in large banks
- Guidelines of data integration architecture design in large banks
- Optimized transactional, operational and analytic data flow infrastructure
- Business and technical value
- **Suggestion**

ISAO POC objective and plan

- Objective:
  - To verify the capability of DB2 on zOS +ISAO with when running analytic applications in terms of the response time, CPU utilization and application transparency, enabling the customer to be fully confident in keeping analytic application OLAP and operative and analytic data on the mainframe
- Plan:
  - Run ISAO workshop to describe the value of OLAP on DB2 for zOS+ISAO
  - Select appropriate existing analytic applications (WAS, Cognos)
  - Analyze to determine appropriate data model which can use ISAP application
  - Design the test scenario
  - Compare the test results to show cost effectiveness of ISAO