

Data Mart/Warehouse: Progress and Vision

Institutional Research and Planning
University Information Systems

What is data warehousing?

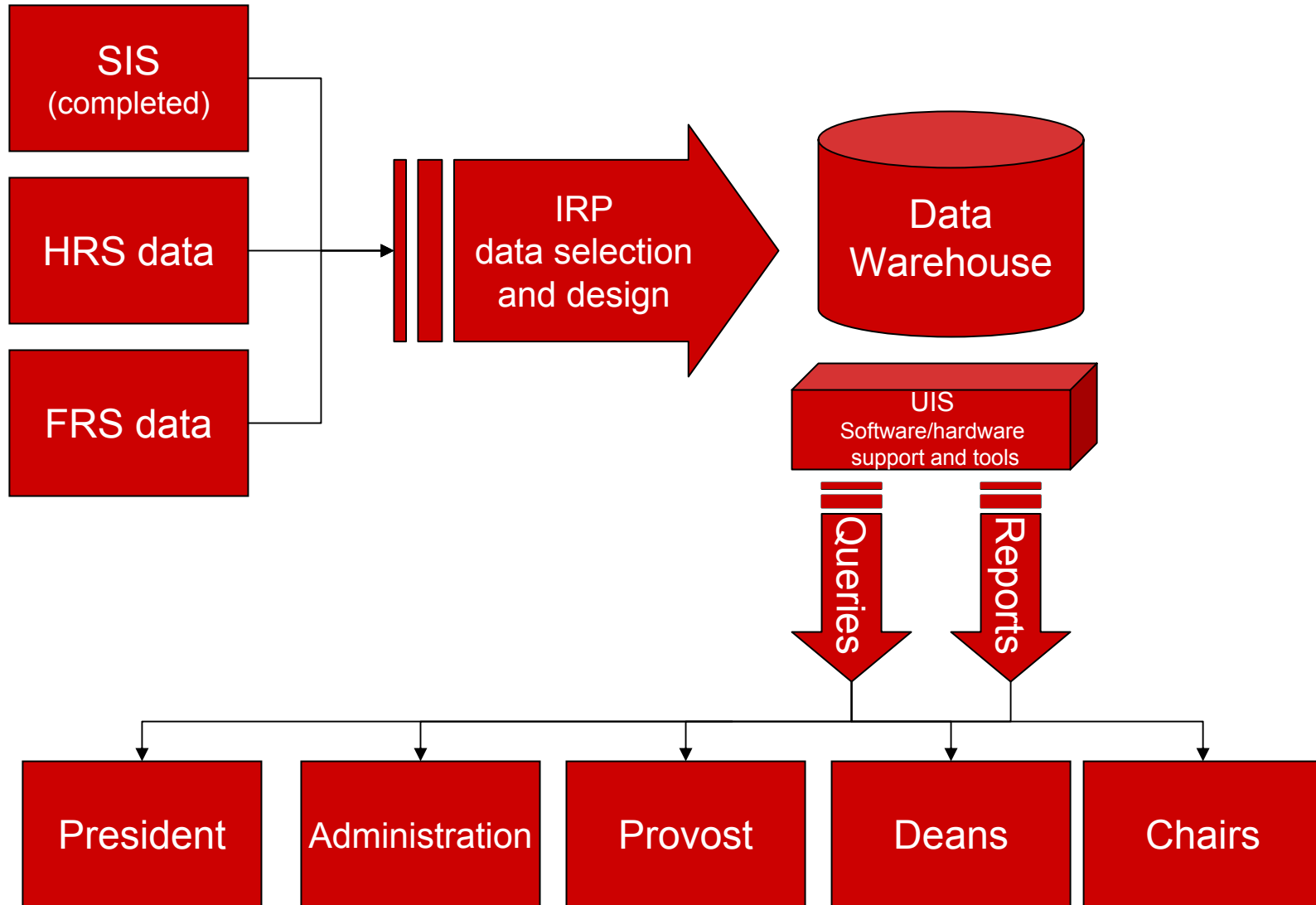
A data warehouse:

- is a single place that contains *complete, accurate* data from multiple sources, making data analysis easier.
- can streamline complex administrative functions and support the decision-making process.
- is specifically designed for querying and reporting.
- is separate from operational/legacy systems (e.g. SIS).

How data warehousing works

- Information is collected from a variety of different sources and formats.
- The data is “cleaned” – transformed into a common format.
- The cleaned data is added to the warehouse, and made available through data marts.
- Data is frozen in time for comparative and forecasting purposes.
- Users can then write queries or make reports using the data warehouse, instead of linking separate databases

IRP Data Mart/Warehouse



- **Reporting**

Consistency and standardization of data. Users only have to learn how to use one database to write queries and reports and display data in graphical form.

- **Internal** – Enrollment, retention & graduation rates, etc.

- **External** – IPEDS, US News & World Report, Princeton Review

Data Driven Decision Making

- Monitor data
 - Data warehousing helps monitor important university processes.
- Rapid queries
 - Users can generate ad-hoc reports “on the fly” without needing to spend time writing and re-writing the code needed to extract data from multiple sources
- Trend Analysis
 - Administrators can easily analyze admissions and enrollment trends using historical data.
- Daily Decision Making
 - Increase the efficiency of the daily functioning of departments and services

Benefits of Data Warehouse

- **Competitive advantage**
 - Administrators can respond to changing conditions based on data.
 - Consistent public image
- **Reputation for quality and excellence**
 - Enhance public image
- **Facilitates accreditation**
 - By using the data warehouse to “drill down” into the college level and department level, more detailed reports can be generated for accreditation

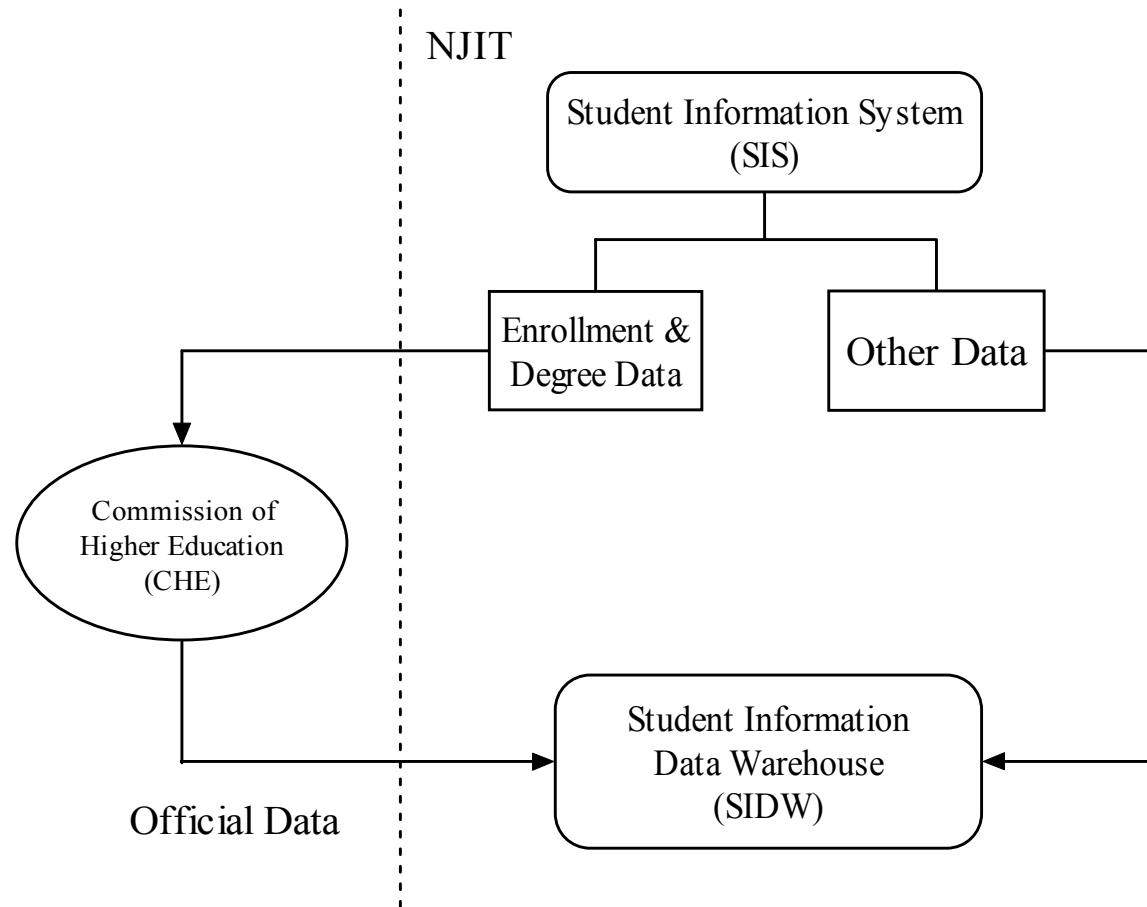
Current data

- SIS
 - Enrollment
 - Degrees
 - Admissions
 - Graduation
 - Retention

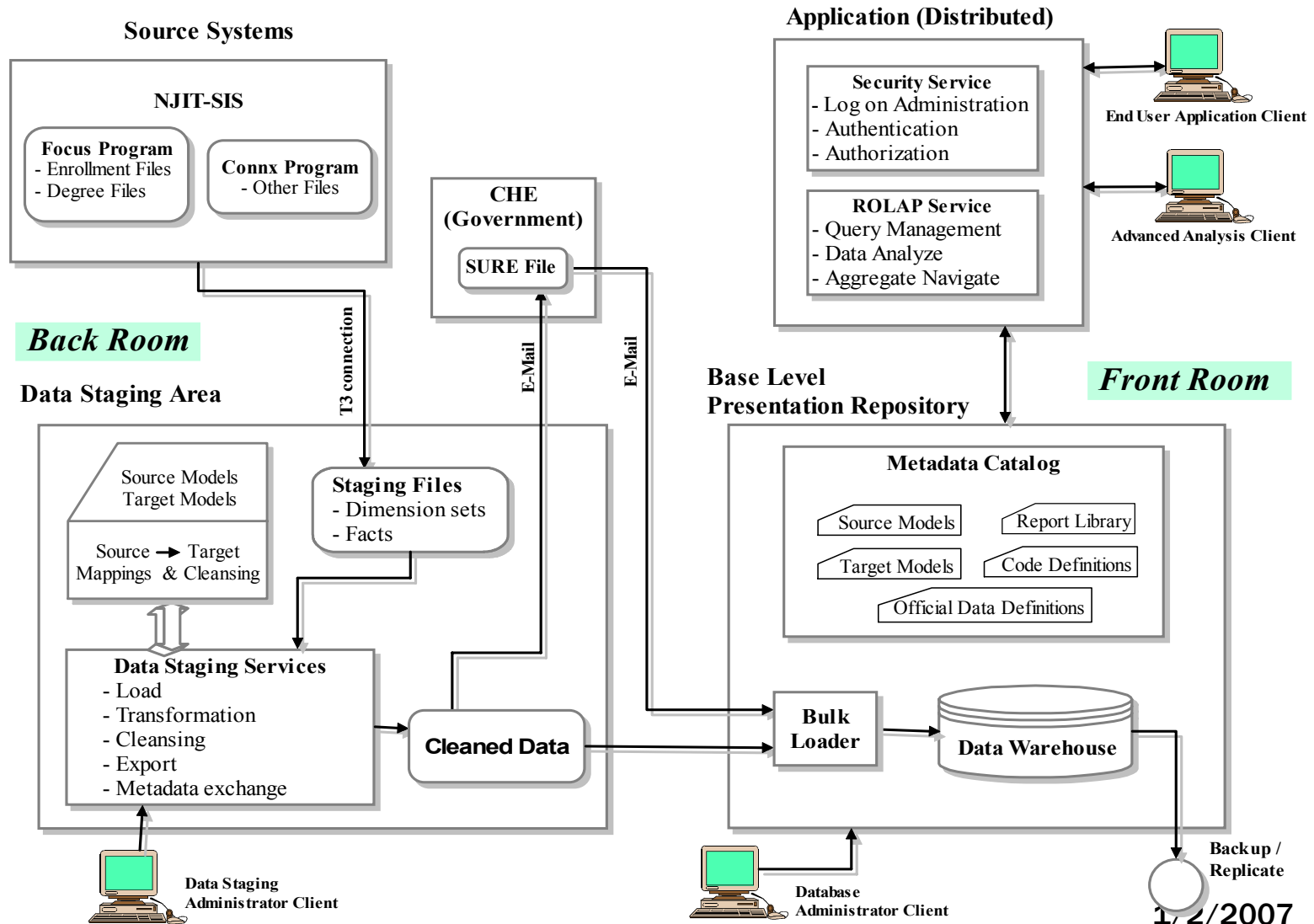
Databases to be added

- Research expenditures (2 years)
- Course distribution (2 semesters w/ ability to go back further)
- Faculty file to 1995 (in progress)
- Staff file (in process of being cleaned)
- Budget file (in conceptual design stage)
- Athletics (in progress)
- Faculty Load (in development)
- Student services (in the future)
- Housing (in the future)

Data source analysis diagram



Architecture model



Current IRP Data Mart Screens

Screenshot: Predefined query

Predefined Query

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Pre-Defined Query -

<input type="radio"/> Common Data Set	<input type="radio"/> Graduation Rate
<input checked="" type="radio"/> Enrollment Data	<input type="radio"/> Retention Rate
<input type="radio"/> Degree Data	<input type="radio"/> Student Life
<input type="radio"/> Admission Data	<input type="radio"/> Miscellaneous

Note: This data is provided for internal purposes. In some cases, official data may vary slightly.

Submit

Screenshot: Enrollment data query

Enrollment Data

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Enrollment Data -

Variable to Choose:	Variable Chosen:	Year:
Gender	Gender	2004
Ethnicity / Citizenship	Ethnicity / Citizenship	
Program / School	Class Level	
Class Level		Special Group:
AttStatus / MatStatus		All Student
		All Student
		AMP Student
		EOP Student
		FTFTF
		Honor Student
		International Student
		Transfer Student

Screenshot: Enrollment cross-tab

		Citizenship		Ethnicity								
		Citizen							Not Citizen		Grand Total	
		Afr.	Asn.	Hisp.	Nat.	Unk.	Wht.	Total				
		+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -			+ -
U/G	Gender	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Graduate	Female	75	176	43	1	70	166	531	345		876	
	Male	125	318	107	2	161	527	1240	747		1987	
	Unknown		2			14		16	4		20	
	Total	200	496	150	3	245	693	1787	1096		2883	
Under	Female	165	282	149	3	191	218	1008	71		1079	
	Male	391	853	521	9	606	1606	3986	243		4229	
	Unknown	7	9	6		28	4	54	4		58	
	Total	563	1144	676	12	825	1828	5048	318		5366	
Grand Total		763	1640	826	15	1070	2521	6835	1414		8249	

Screenshot: Admissions data query

Admission Data

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Admission Data -

Variable to Choose: Variable Chosen:

Gender
Ethnicity / Citizenship
Program / School

Gender
Program / School

Year: 2004

Special Group:
Graduate Student - PhD
Freshmen
Graduate Student - Master
Graduate Student - PhD
Transfer Student

Clear Selection

Submit

Screenshot: Admissions cross-tab

				Gender ▾			
				Female	Male	Unknown	Grand Total
				+ -	+ -	+ -	+ -
School ▾	NJITMajor ▾	AppResult ▾	+ -	Count	Count	Count	Count
ARC	USYS	Admitted	+	7	2		9
		Not Admitted	+	7	3		10
		Total	+	14	5		19
	Total	+	14	5		19	
CCS	CS	Admitted	+	22	38	2	62
		Not Admitted	+	13	20		33
		Total	+	35	58	2	95
	IS	Admitted	+	6	6	1	13
		Not Admitted	+	2	12		14
		Total	+	8	18	1	27
Total	+	43	76	3	122		
NCE	BMED	Admitted	+	2	1		3
		Not Admitted	+	5	19		24
		Total	+	7	20		27
	CE	Admitted	+	3	14		17
		Not Admitted	+		2		2
		Total	+	3	16		19
	CHE	Admitted	+		8		8
		Not Admitted	+	17	22		39
		Total	+	17	30		47
	COE	Admitted	+	4	8		12
Not Admitted		+	1	7		8	
Total		+	5	15		20	
EE	Admitted	+	25	70	2	97	

Screenshot: Graduation rate query

Graduation Rate

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Graduation Rate -

Variable to Choose: **Variable Chosen:**

Gender
Ethnicity / Citizenship
Program
School

Gender
School

Year: 1997

Period: 6-year Graduation Rate

Special Group: All Student

All Student
AMP Student
EOP Student
Honor Student

Note: Overall Graduation Rate =
Graduation Rate of "Graduated in School/Program" +
Graduation Rate of "Graduated in Other School/Program"

Screenshot: Graduation rate cross-tab

pvtGraduationRateBy_Gender_School_6Yr										
		GradSchool6Yr ▾								
		Grad in School		Grad in Other School		Not Graduate		Grand Total		
		+/-		+/-		+/-		+/-		
School ▾	Gender ▾	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage	
CCS	Female +	8	38.10%	5	23.81%	8	38.10%	21	100.00%	
	Male +	55	38.19%	12	8.33%	77	53.47%	144	100.00%	
	Total +	63	38.18%	17	10.30%	85	51.52%	165	100.00%	
NCE	Female +	14	38.89%	8	22.22%	14	38.89%	36	100.00%	
	Male +	79	39.70%	23	11.56%	97	48.74%	199	100.00%	
	Total +	93	39.57%	31	13.19%	111	47.23%	235	100.00%	
SLA	Female +			1	33.33%	2	66.67%	3	100.00%	
	Male +	4	40.00%	3	30.00%	3	30.00%	10	100.00%	
	Total +	4	30.77%	4	30.77%	5	38.46%	13	100.00%	
SOA	Female +	11	64.71%	1	5.88%	5	29.41%	17	100.00%	
	Male +	5	13.89%	9	25.00%	22	61.11%	36	100.00%	
	Total +	16	30.19%	10	18.87%	27	50.94%	53	100.00%	
SOM	Female +	3	27.27%	2	18.18%	6	54.55%	11	100.00%	
	Male +	3	12.50%	2	8.33%	19	79.17%	24	100.00%	
	Total +	6	17.14%	4	11.43%	25	71.43%	35	100.00%	
UNK	Female +			8	47.06%	9	52.94%	17	100.00%	
	Male +			21	42.00%	29	58.00%	50	100.00%	
	Total +			29	43.28%	38	56.72%	67	100.00%	
Grand Total	+/-	182	32.04%	95	16.73%	291	51.23%	568	100.00%	

Screenshot: Retention rate query

Retention Rate

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Retention Rate -

Variable to Choose:	Variable Chosen:	Year:
Gender	Ethnicity / Citizenship	2002
Ethnicity / Citizenship	School	
Program		Period: 2nd Year Retention Rate
School		
		Special Group: All Student
		All Student
		AMP Student
		EOP Student
		Honor Student

Note: Overall Retention Rate =
Retention Rate of "Retained in School/Program" +
Retention Rate of "Retained in Other School/Program"



Screenshot: Retention rate cross-tab

pvtRetentionRateBy_Ethnic_School_2ndYr										
			RetSchoolIn2ndYr							
			In School at 2nd Yr		In Other School at 2nd Yr		Not Retain at 2nd Yr		Grand Total	
School	Citizenship	Ethnicity	Count	Ret Rate	Count	Ret Rate	Count	Ret Rate	Count	Ret Rate
CCS	Citizen	Afr.	6	60.00%			4	40.00%	10	100.00%
		Asn.	25	59.52%	9	21.43%	8	19.05%	42	100.00%
		Hisp.	4	44.44%	3	33.33%	2	22.22%	9	100.00%
		Unk.	8	50.00%	2	12.50%	6	37.50%	16	100.00%
		Wht.	52	72.22%	4	5.56%	16	22.22%	72	100.00%
	Total	95	63.76%	18	12.08%	36	24.16%	149	100.00%	
	Not Citizen	3	60.00%	1	20.00%	1	20.00%	5	100.00%	
Total	98	63.64%	19	12.34%	37	24.03%	154	100.00%		
NCE	Citizen	Afr.	23	67.65%	3	8.82%	8	23.53%	34	100.00%
		Asn.	54	85.71%			9	14.29%	63	100.00%
		Hisp.	30	78.95%	3	7.89%	5	13.16%	38	100.00%
		Nat.	1	100.00%					1	100.00%
		Unk.	33	82.50%			7	17.50%	40	100.00%
		Wht.	81	81.00%	5	5.00%	14	14.00%	100	100.00%
	Total	222	80.43%	11	3.99%	43	15.58%	276	100.00%	
Not Citizen	13	100.00%					13	100.00%		
Total	235	81.31%	11	3.81%	43	14.88%	289	100.00%		
SLA	Citizen	Afr.	5	100.00%					5	100.00%
		Asn.	3	75.00%	1	25.00%			4	100.00%
		Hisp.	1	100.00%					1	100.00%
		Unk.	8	88.89%			1	11.11%	9	100.00%
		Wht.	4	80.00%			1	20.00%	5	100.00%
		Total	21	87.50%	1	4.17%	2	8.33%	24	100.00%

Screenshot: Student life query

Student Life

NJIT [Main](#) | [Back](#) | [Help](#) | [Exit](#)

- Student Life -

Residence Hall

State of Residence

Age

Screenshot: Miscellaneous queries

Miscellaneous

NJIT Main | Back | [Help](#) | Exit

- Miscellaneous -

SAT Score

GRE Score

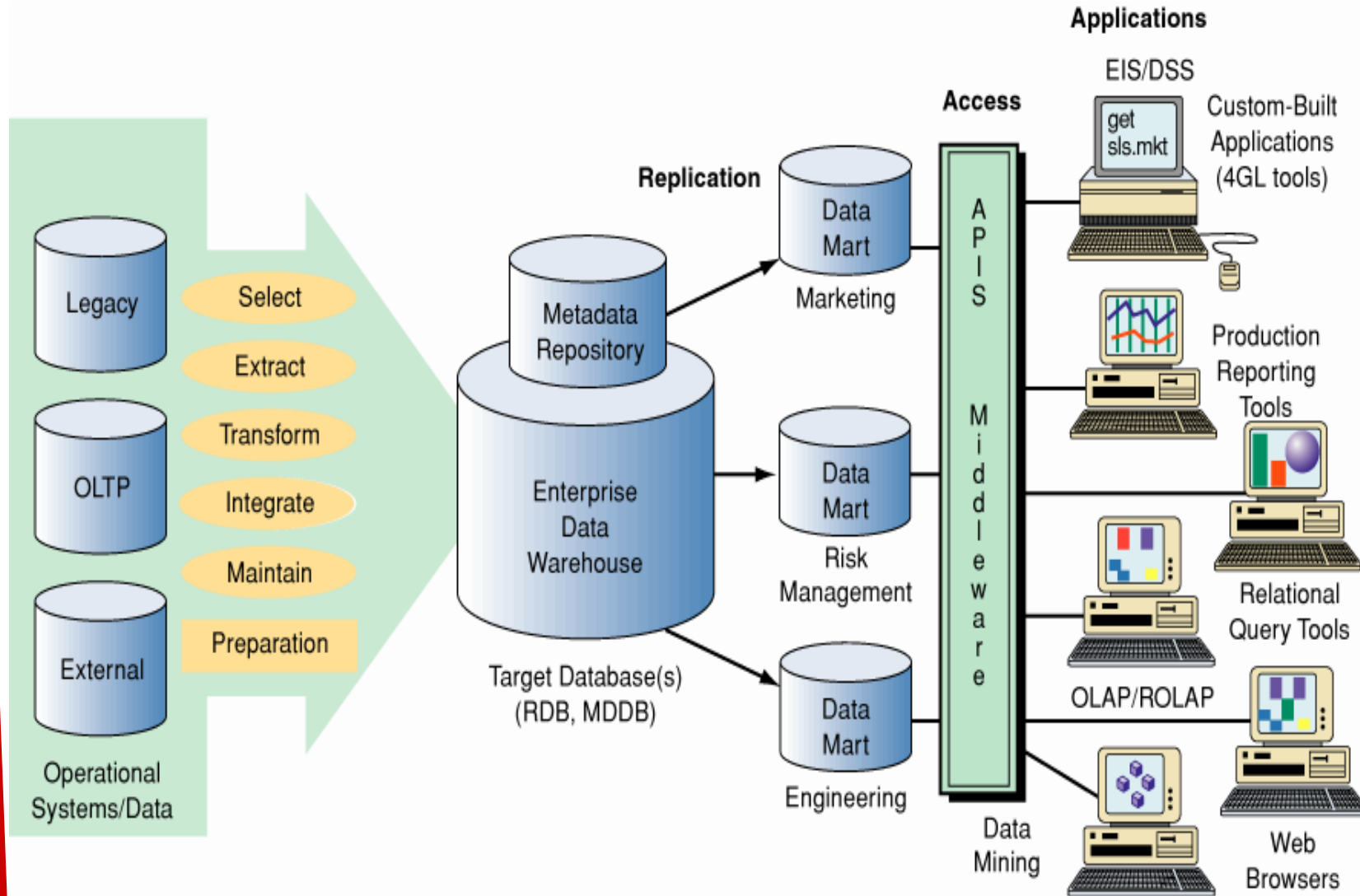
High School Rank Percentile

Current Accumulated GPA

Full-time Equivalent Enrollments (FTE)

Submit

Components of Data Warehouse

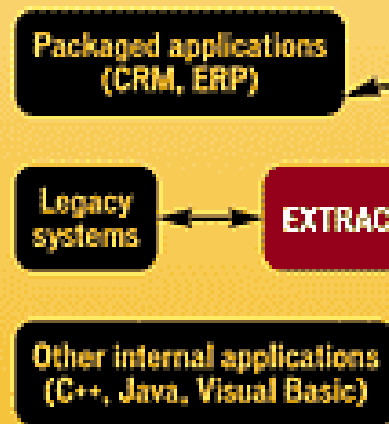


Extract, Transform and Load (ETL)

Moving and Improving Data

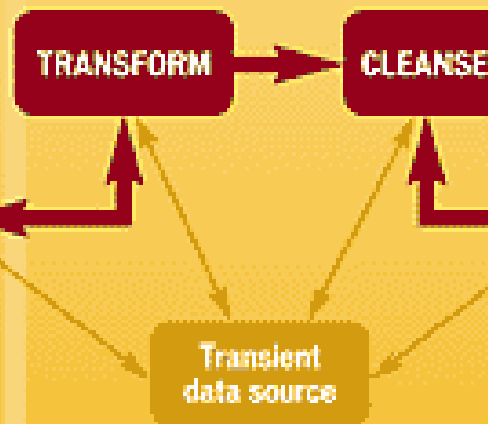
EXTRACT

The process of reading data from a database.



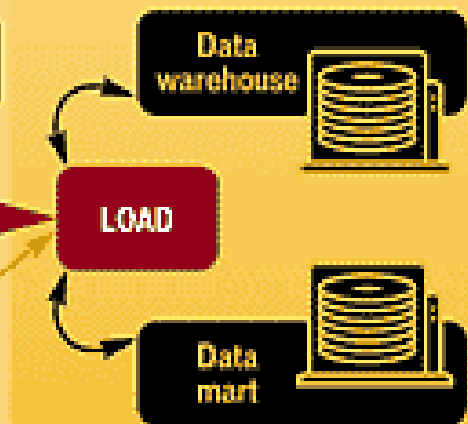
TRANSFORM

The process of converting the extracted data from its original state into the form it needs to be in so it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.



LOAD

The process of writing the data into the target database.

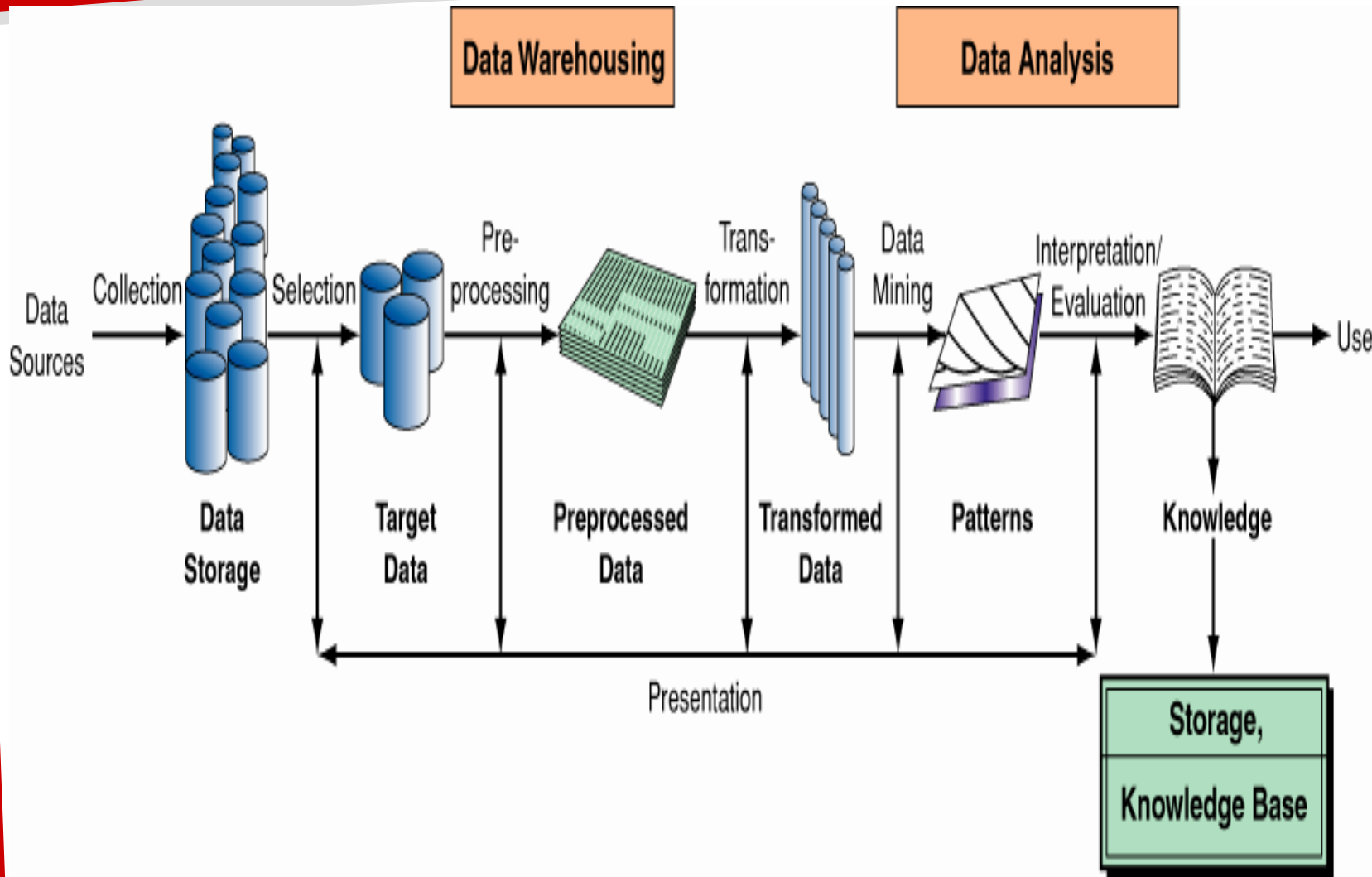


ETL software is used to migrate data from one database to another or to data marts and data warehouses. The central part of the process, the "transform" function, cleanses the data, eliminates duplicates and reformats the data for the target repository.

DW Summary: Characteristics

- Subject-orientation
- integrated
- non-volatile (i.e. not updated)
- time variant (kept for long periods, for forecasting and trend analysis)
- summarized
- large volume
- not normalized
- metadata
- data sources

Data Warehouse Process



- **Intelligent Agents and Agencies - tools work and think for user.**
- **Query Facilities and Managed Query environments.**
- **Statistical Analysis - One of the biggest surprises in the data warehousing marketplace is the resurgence of interest in traditional statistical analysis, and the concomitant resurrection of the popularity to products like SAS and SPSS.**

- **Data Discovery -**
 - **A large class of tools formerly classified as decision support, artificial intelligence and expert systems. They now make use of neural networks, fuzzy logic, decision trees, and other tools from advanced mathematics to allow a user to “sift” through massive amounts of raw data to “discover” new, interesting, insightful, and in many cases useful things about the organization, its operations, and its markets.**
 - **There are many different data discovery tools/products on the market.**

- **OLAP - *On-line Analytical Processing*:**
 - often uses multi-dimensional spreadsheet tools allowing users to look at information from many different angles.
 - Users are able to “*slice and dice*” reports and to look at the same kinds of information at different levels at the same time.
 - Typical OLAP application might allow a product manager to view sales figures for a given product at the national level, see them broken down by division, *drill down* to see territories within a division, check sales numbers for each store within a territory, and then compare them against sales of stores from another territory.

Data Mining

Provides for:

- **Knowledge discovery in databases**
- **Knowledge extraction**
- **Data archeology**
- **Data exploration**
- **Data pattern processing**
- **Data dredging**
- **Information harvesting**

Data Mart Timeline:

	Month1	Month2	Month3	Month4	Month5	Month6
Project Planning/Management/Policy	█	█	█	█	█	█
Development Environment	█					
Software Installation	█					
Data Inventory & Quality Assessment	█					
Database Design	█	█				
Application Design		█				
Data Scrubbing & Loading	█	█	█			
Application Development		█	█	█		
Testing				█	█	
Documentation					█	█
Rollout Complete Project						█