

# SuperSTAR Analytical Software Suite

Seth Grimes  
Alta Plana Corporation

SIGSTAT  
January 15, 2003

*Alta Plana*

# Introduction

## Alta Plana Corporation ([altaplana.com](http://altaplana.com))

- Founded 1997, Takoma Park MD, by Seth Grimes
- Consult on data systems architecture & design

## Seth Grimes

- At Census 4/98 to 9/02: designed 2000 tab system
- Contributing editor & decision-support columnist, Intelligent Enterprise magazine ([iemagazine.com](http://iemagazine.com))

## Space-Time Research ([www.str.com.au](http://www.str.com.au))

- HQ Melbourne, Australia; European operations
- CEO Warren Richter; Operations manager John Ellenberger
- US partners also include Westat, Sabre Systems

# SuperSTAR background

## SuperSTAR architecture

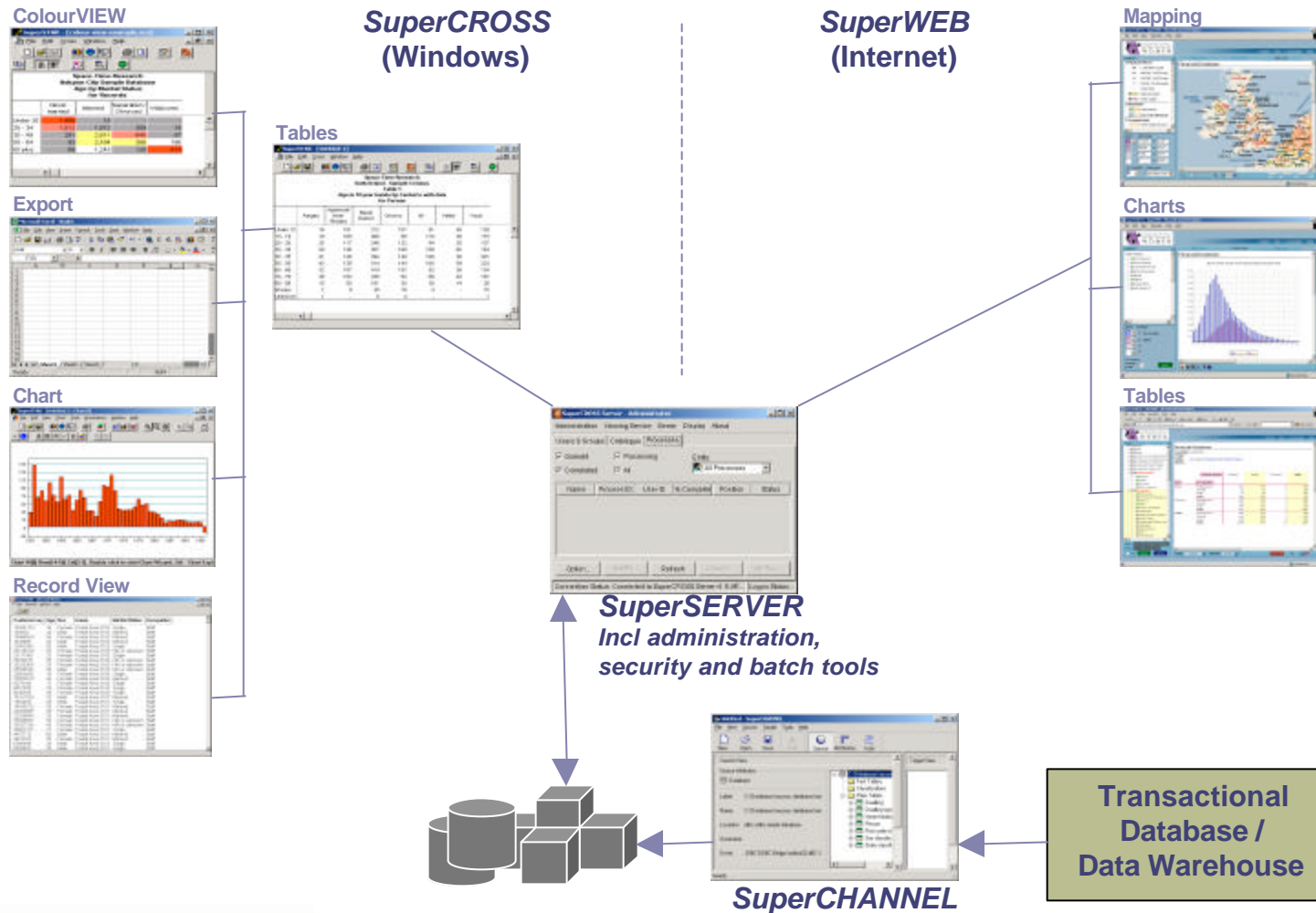
- Dates to 1986 research project
- Graphical (no coding) and scripting interfaces
- High-capacity, shared, multithreaded server
- Similar to OLAP tools, but with special statistical capabilities, higher capacity, larger DBs
- Similar to tabulation functions within statistical tools (SAS, SPSS) but without end-user coding
- Similar to relational DBMSes but optimized for aggregation/tabulation speed
- Similar to tab/dissemination tools (Beyond 20/20, TPL Tables) but more interactivity, higher capacity, larger DBs

# SuperSTAR background

## SuperSTAR users

- Stat offices in Australia, Denmark, Finland, NZ, Norway, Poland, South Africa, Sweden, Switzerland, UK
- Other government agencies including Australian defense, social security, tax, and transportation
- Private sector includes Telstra (Australian telecomm) and Henry Rak Consulting Partners (strategic market research)

# SuperSTAR background



Alta Plana

©Alta Plana Corporation, 2003

SuperSTAR analytical suite

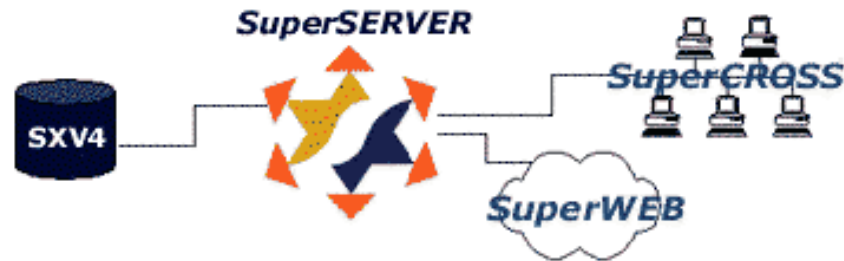
# SuperSTAR background



## SuperCROSS is the tabulation interface

- Drag-and-drop table creation
- Easy to add calculations, create functions, and format
- Integrated charting via SuperCHART
- Relationship analysis via ColourVIEW
- Integrated mapping via SuperMAP

# SuperSTAR background



## SuperSERVER is the tabulation engine

- Manages database catalog and access permissions by users and groups
- Administration interface facilitates managing larger-scale installations
- Runs on Windows, Solaris & AIX and serves networked and local clients, including a command-line client

# SuperSTAR background



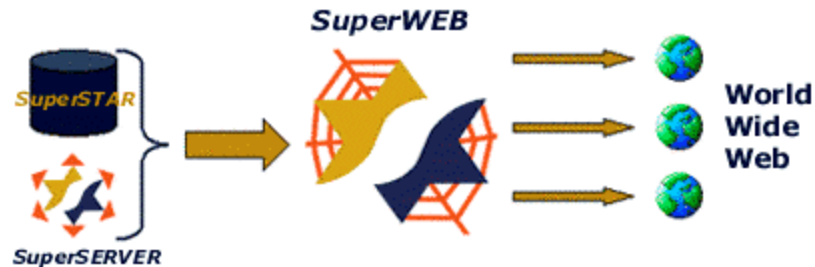
## SuperCHANNEL is the database builder

- Drag-and-drop mapping of tables & fields in input databases to SuperSTAR DB
- Handles hierarchical relationships and joins
- Handles data “cleansing” operations
- Supports database inserts & updates
- Generates Java code for actual DB build

There’s also a command-line DB builder, SNU



# SuperSTAR background



SuperWEB is a “thin client” browser interface

- No Java applets or ActiveX controls
- Customizable user interface
- Implemented with Java Server Pages (JSP)
- Delivers predefined tables and build-your-own table interface
- Integrated charting
- Integrated thematic mapping via ESRI’s ArcIMS
- RecordVIEW and ColourVIEW to come

Demo

then on to

Census 2000 Data Product Production (DPP)  
system and SuperSTAR

*Alta Plana*

©Alta Plana Corporation, 2003

**SuperSTAR analytical suite**

# Census 2000 DPP System

## Summary data products produced by DPP

- Redistricting data (PL 94-171), also adjusted version
- Summary Files 1 & 2: state & national
- Summary Files 3 & 4: state & national
  - SF2 has 250 race/ancestry/ethnicity iterations; SF4 has 336
- 108th congressional districts; school district reimbursable

## Sources and outputs

- Geographic data: up to 10 million tabulation areas
- Detail-file metadata: fields, value sets, file layout
- Detail data: 281m people, 107m housing, 10m geo
- Product specifications: thousands of tables overall

# Census 2000 DPP System

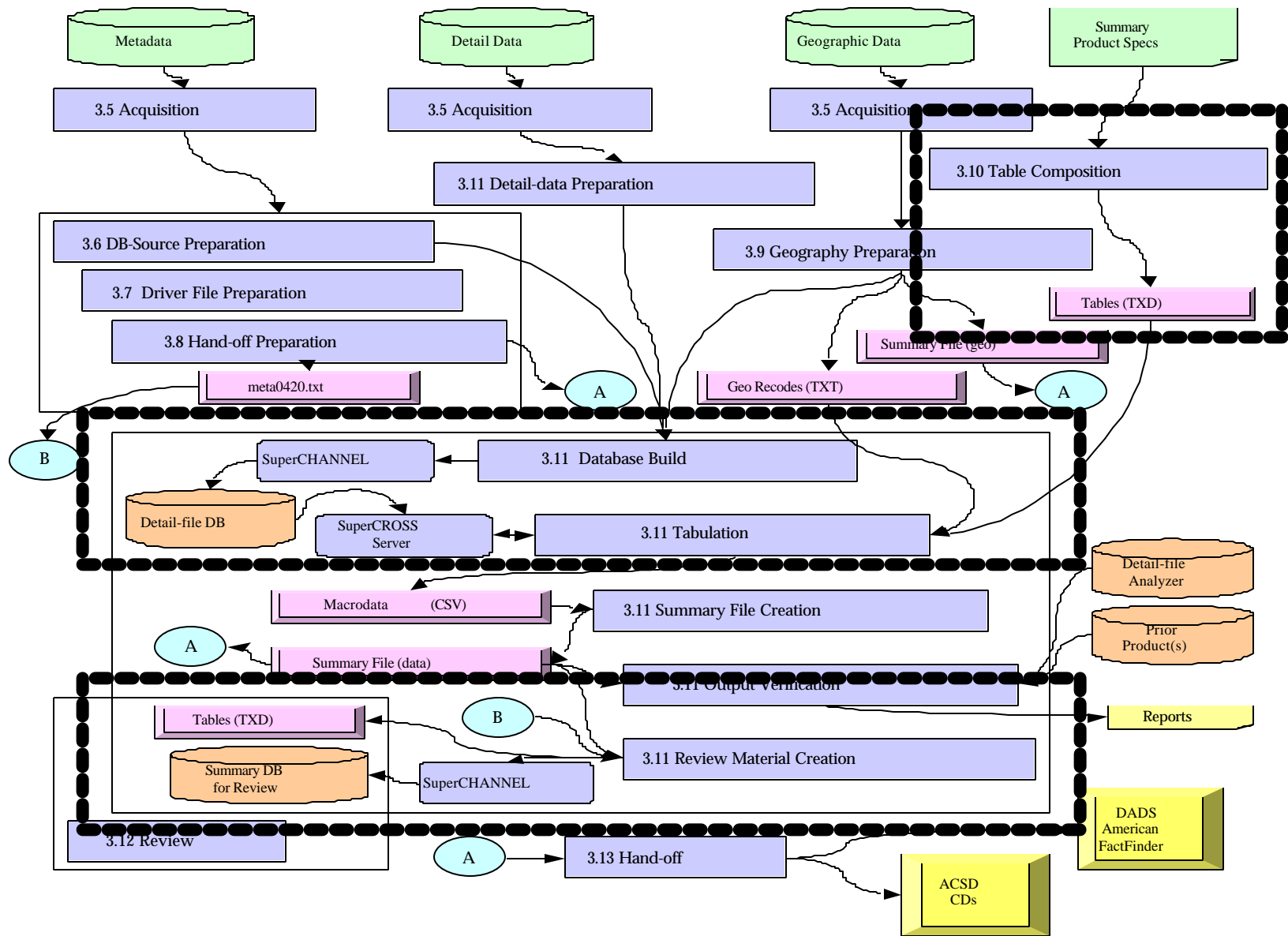
## Data Access and Dissemination System

- IBM Global Services is lead contractor
- American FactFinder for Web dissemination

## DPP platform

- Windows SuperCROSS clients
- IBM RS/6000, AIX operating system server
  - Extensive use of parameter files to create a generic system
  - SAS data preparation and output processing, driven by parameter files and using lots of macro processing
  - SuperSTAR database builder, invoked via scripts
  - SuperSERVER tabulation engine, invoked via scripts
  - Scripting: Korn shell with Perl & Python for utility functions

# DPP System Components



# Contact

Questions? Discussion?

Seth Grimes

Alta Plana, <http://altaplana.com>

301-270-0795 (office); 301-873-8225 (cell)

[grimes@altaplana.com](mailto:grimes@altaplana.com)

Space-Time Research

<http://www.str.com.au>

*Alta Plana*

©Alta Plana Corporation, 2003

**SuperSTAR analytical suite**

# SuperSTAR Analytical Software Suite

Seth Grimes  
Alta Plana Corporation

SIGSTAT  
January 15, 2003

*Alta Plana*